iSCSI Today and Tomorrow

Gene Chesser
Microsoft IOP Program Manager
Hewlett-Packard

Agenda

1. The need for Multi-Protocol Networks
2. Users' Attitudes and Perceptions of iSCSI
3. iSCSI: What is it?
4. iSCSI Products
5. iSCSI Tomorrow
6. Summary

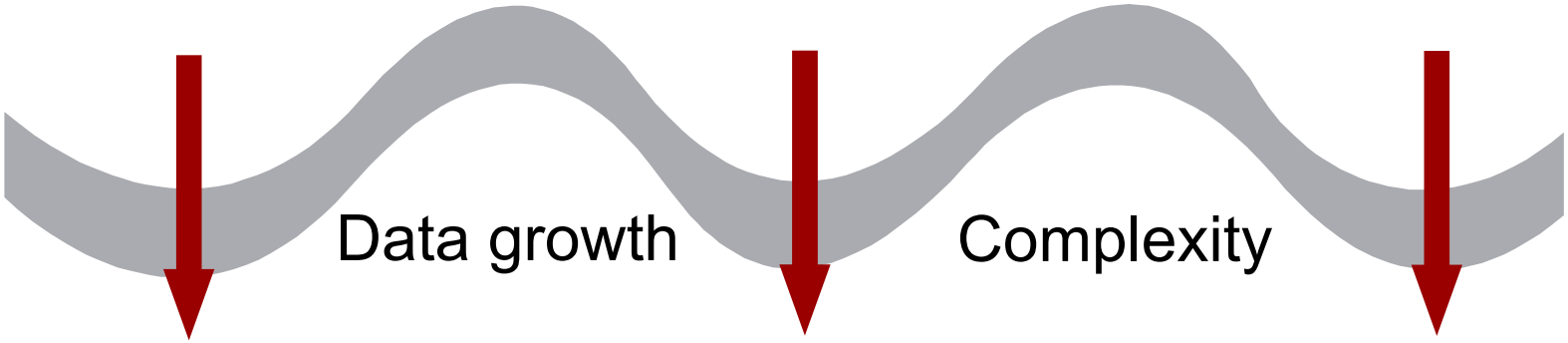
The need for Multi-Protocol Networks

Evolution of Computing – The Next Wave

Disk Scaling

SAN Scaling

Network Scaling



DAS centric

Fibre Channel centric

Multi-Protocol Network centric

standardization waves

Why Multi-Protocol Networks?

Why IP?

- IP backbones span the globe
- Large pool of skilled people
- Security, other attractive features
- *Possible* cost benefits

Why not IP…for storage?

- Need to port storage mgmt applications
- Storage knowledgeable IP people
- Questions around Performance & reliability

A *balanced* argument that dictates a *pragmatic* response

Fibre channel and IP are *complementary*

IP = emerging solution for departmental storage networks and scaling distance

Fibre Channel strengths

- designed for storage needs
- strong market acceptance
- SCSI standards
- adaptive enterprise solutions

IP strengths

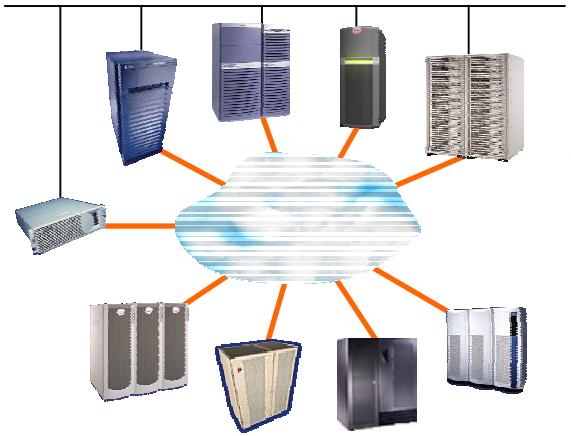
- existing ethernet networks
- standards and compatibility
- shared protocol
- skilled people

iSCSI: evolution not revolution

- fibre channel is a proven storage technology
  - reliability and performance is higher compared to IP
  - rapidly growing install base
  - maturing SAN management tools
  - cost per MB decreasing rapidly

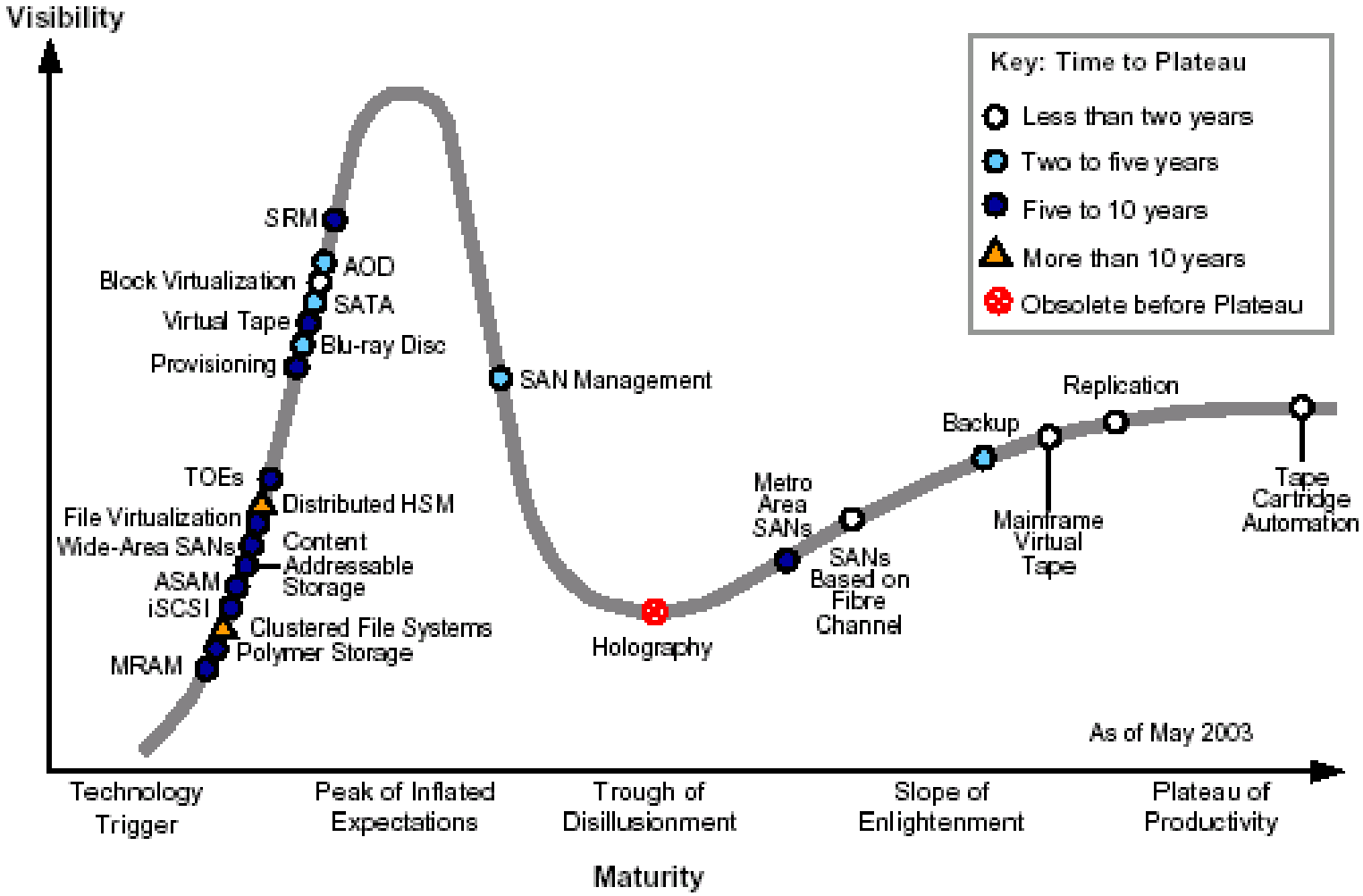
- iSCSI storage has a lot of promise
  - networks are well understood and ubiquitous
  - perception that interoperability issues will be less
  - availability of IP networking expertise compared to FC
  - ability to transport over almost any medium

Internet Protocol Meets SCSI



- FCIP – encapsulate fibre channel in IP packet
  - connect FC SAN island domains through IP fabric
  - connect FC devices directly to IP fabric
- iSCSI – Glue protocol to allow SCSI protocols to run on TCP/IP
  - interconnect iSCSI servers with FC or iSCSI storage
- iSER – RDMA extensions for iSCSI transport

The Technology Hype Cycle



Source: Gartner, May 2003

WW External Disk Storage Systems Revenue by Installation Environment ($M)

Installation Environment	2002	2003	2004	2005	2006	2007	CAGR 03-07
DAS	$ 6,354	$ 5,811	$ 5,349	$ 4,932	$ 4,427	$ 3,708	-10.6%
NAS	$ 1,543	$ 1,772	$ 2,063	$ 2,440	$ 2,863	$ 3,166	15.6%
SAN							
Fibre Channel	$ 5,395	$ 6,042	$ 6,518	$ 7,022	$ 7,391	$ 7,410	5.2%
ESCON/FICON	$ 840	$ 751	$ 718	$ 688	$ 657	$ 630	-4.3%
iSCSI	$ -	$ 15	$ 45	$ 153	$ 568	$ 1,945	238.0%
Subtotal SAN	$ 6,235	$ 6,808	$ 7,281	$ 7,862	$ 8,616	$ 9,985	10.0%
Total	**$ 14,132**	**$ 14,391**	**$ 14,693**	**$ 15,234**	**$ 15,906**	**$ 16,858**	**4.0%**

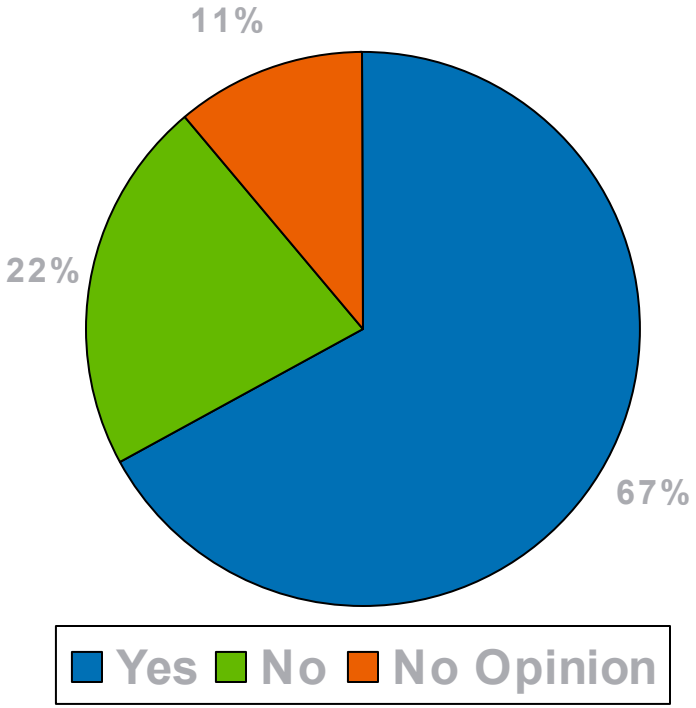
iSCSI grows to 20% of the total SAN market in 2007

Source: IDC, October 2003

Users' Attitudes and Perceptions of iSCSI

General Interest

- IDC asked 300 IT professionals in NA the following question “Knowing what you currently know about iSCSI, **would** you buy iSCSI at a future date”?



Source: IDC, January 2003

Key End User iSCSI Perceptions

- Easy to lean – they already understand TCP/IP
- Leverage standard network equipment for storage and networking
- Should cost less initially and over time compared to FC
- Should be less expensive to manage initially and over time compared to FC
- Will lag FC performance initially but over time have the same or better performance
- Security will not be as good as FC
- Will allow network managers to handle SANs

Factors that would significantly contribute to an iSCSI purchase

1. Products with good management software
2. High availability products available
3. Products with built-in security features
4. Low-cost 10 Gigabit ethernet
5. Large networking vendors heavily involved in iSCSI
6. iSCSI native RAID array available
7. iSCSI on server motherboards for internal storage
8. Solution available from my server supplier
9. Gateways from fiber channel to iSCSI

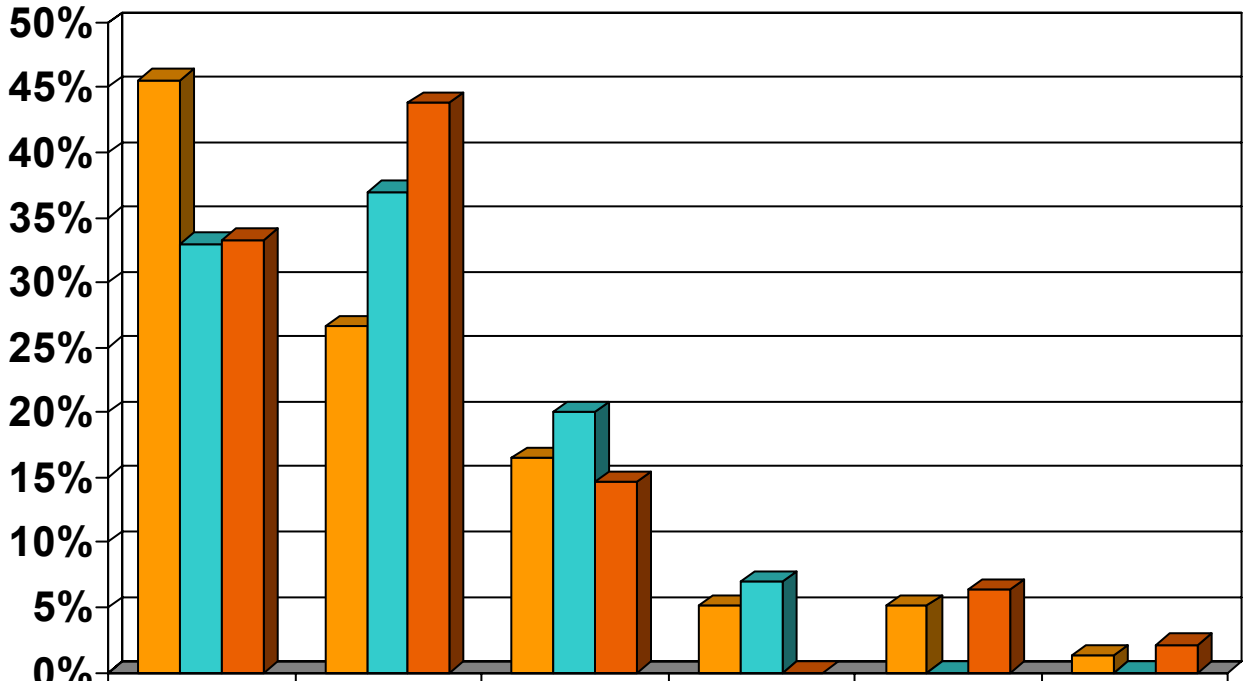
Source: IDC, January 2003

iSCSI Operating Systems Usage

Operating System	Company Size		
	Small 40 - 100	Medium 101 - 500	Large > 500
Microsoft Widows	65.4%	67.6%	56.0%
Other Unix (AIX, HP-UX, etc.)	11.5%	10.3%	16.7%
Sun OS/Solaris	3.8%	8.8%	10.7%
Linux	13.5%	1.5%	7.1%
Novell NetWare	1.9%	10.3%	7.1%
Mainframe (MVS)	1.9%	0.0%	1.2%
All the above	0.0%	0.0%	1.2%
Don’t know	1.9%	1.5%	0.0%

Source: IDC, January 2003

iSCSI Intended Application Uses



	Mixed Storage	Databases	Mainly Backup	Image databases	Mainly e-mail	Application Devel
Large	46%	27%	17%	5%	5%	1%
Medium	33%	37%	20%	7%	0%	0%
Small	33%	44%	15%	0%	6%	2%

Source: IDC, January 2003

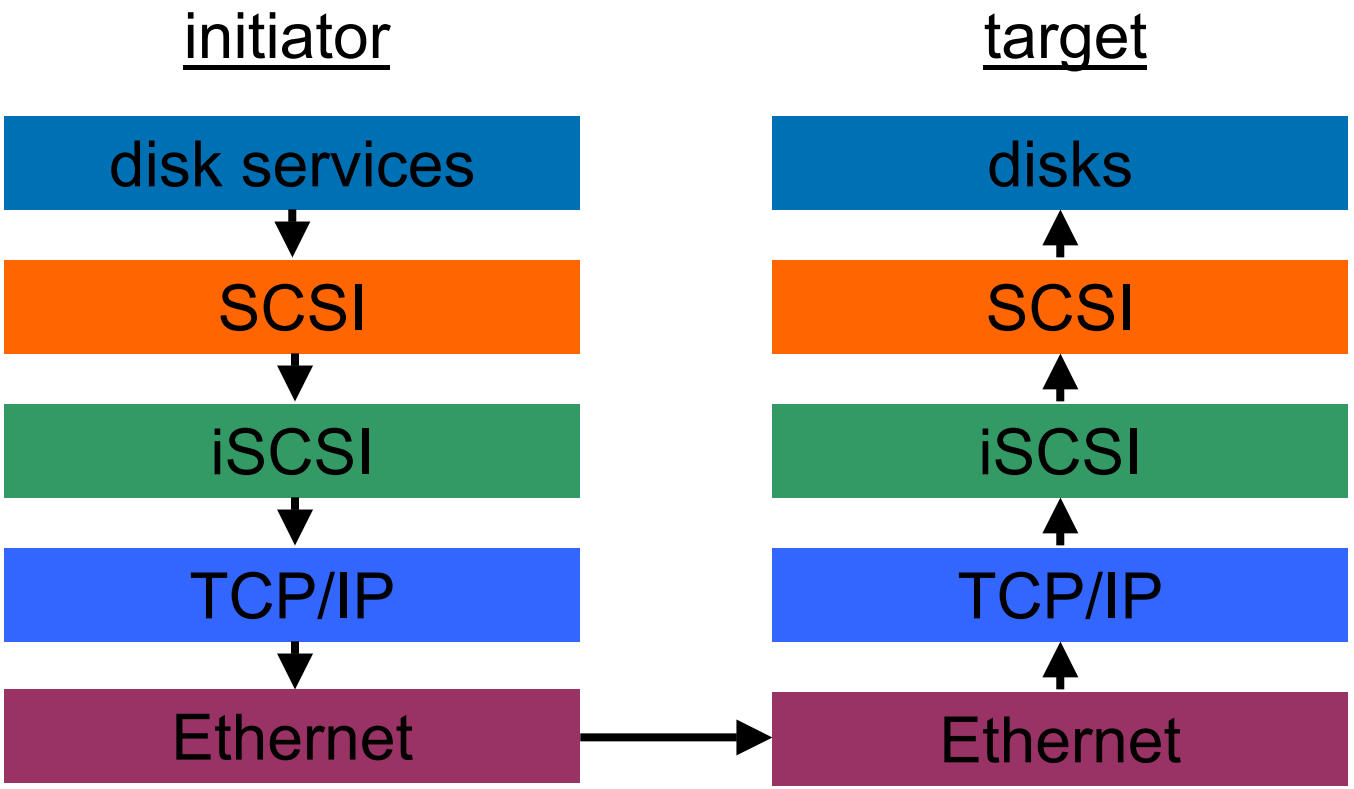
iSCSI: what is it?

iSCSI is internet Small Computer System Interface, a TCP/IP-based storage networking standard for linking data storage facilities, developed by the Internet Engineering Task Force (IETF)

by carrying SCSI commands over IP networks, iSCSI is used to ***facilitate data transfers over intranets and to manage storage over long distances***

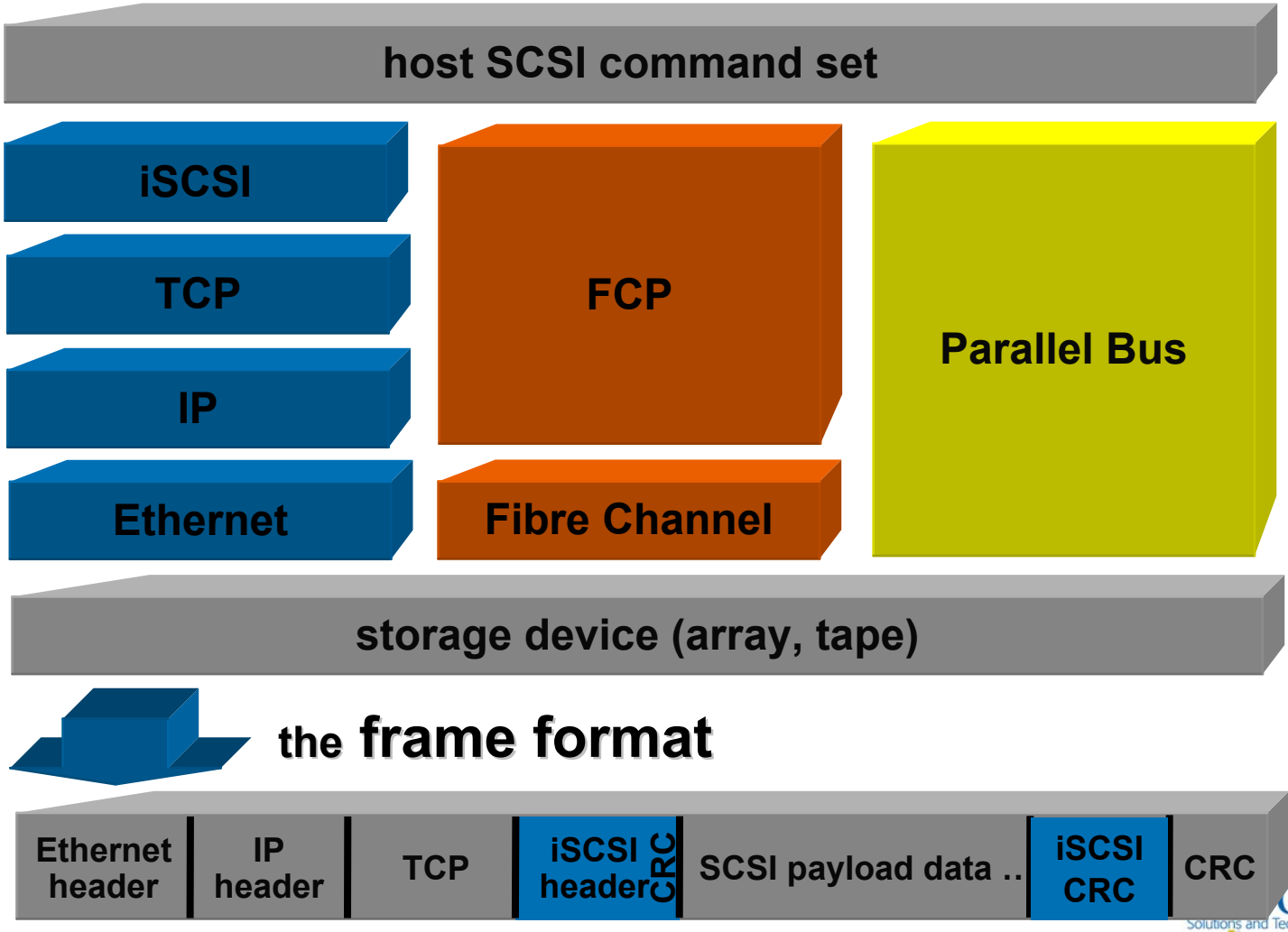
iSCSI: TCP/IP-based protocol for establishing and managing connections between ***IP-based storage devices, hosts, and clients***

conceptual view



iSCSI protocol

a transport protocol alternative for SCSI that operates over TCP/IP



discovery - iSNS

iSNS (Internet Storage Name Server)

- provides registration and discovery of SCSI devices and Fibre Channel-based devices
- in IP-based storage like iSCSI end devices registered with iSNS
- Microsoft is shipping an iSNS Server

discovery - SLP

SLP (Service Location Protocol)

- provides a flexible and scalable framework for providing hosts with access to information about the existence, location, and configuration of networked services, including the iSNS server
- SLP **may** be used by iSNS clients to discover the IP address of the iSNS server
- to implement discovery through SLP
  - a Service Agent (SA) should be co-hosted in the iSNS server
  - a User Agent (UA) should be in each iSNS client
  - eEach client multicasts a discovery message requesting the IP address of the iSNS server(s)
  - the SA responds to this request
  - optionally, the location of the iSNS can be stored in the SLP Directory Agent (DA)

iSCSI security

- authentication
  - performed at connection login
  - initiator and target authenticate each other's identity and authorizes login
  - SRP (Secure Remote Password)
  - may use domain authentication or radius server
- confidentiality
  - IPSec (IP Security)
  - each IP packet is encrypted with ESP (Encapsulating Security Protocol).
  - security associations, key management
- data integrity
  - IPSec
  - iSCSI message digest (CRC)

iSCSI performance

- ***TCP/IP offload***
  - TCP/IP memory copies are very CPU and memory intensive
  - most believe that at 1-Gb Ethernet speeds, TOE (TCP/IP Offload Engine) is required
  - HP believes software-only iSCSI stacks without TOE will be acceptable for some applications when the IO intensity is low, or CPU load is not a worry.
  - TOE will be absolutely required at greater than 1-Gb speeds
- iSCSI will lag Fibre Channel performance for the foreseeable future, at least until 10-Gb TOE is available

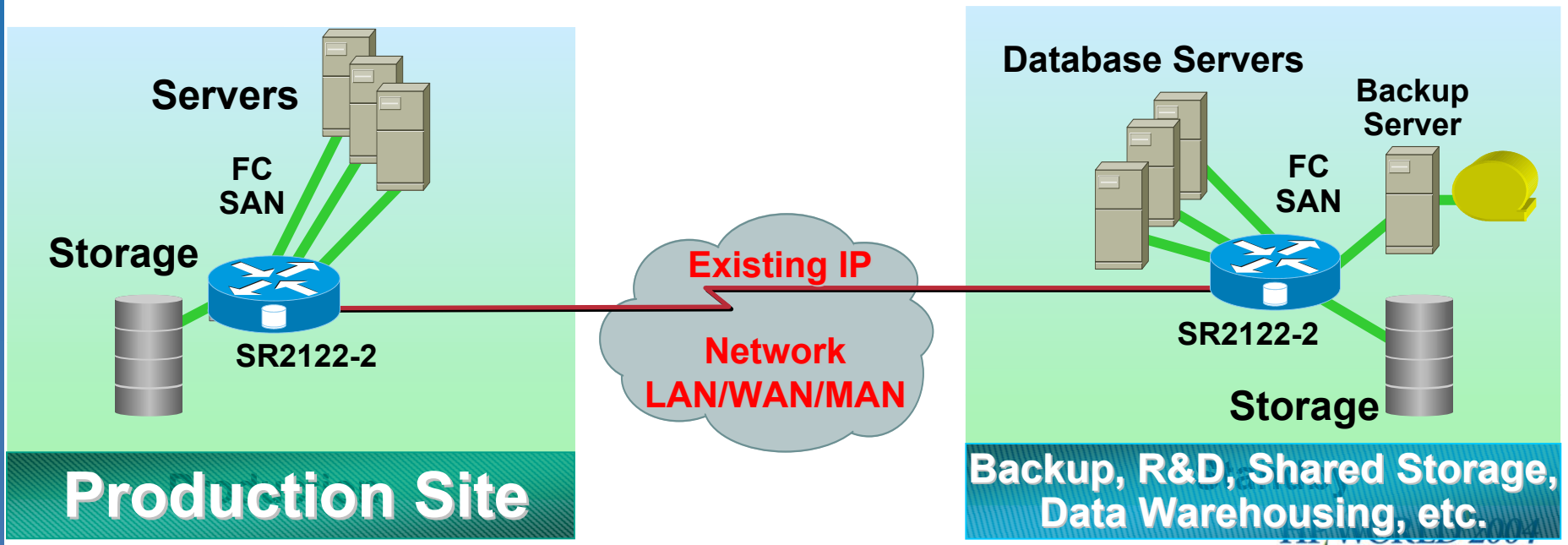
iSCSI boot from network storage

- boot from SAN is here today with Fibre Channel
- iSCSI boot is doable with iSCSI adapter & option ROM
  - iSCSI adapter can establish an iSCSI connection at POST, and keep the connection alive while the OS comes up
- iSCSI Boot is harder with Software stack or TOE NIC
  - how to establish a TCP/IP stack and keep the connection alive while the OS boots up?
  - PXE Boot works in some environments
- expect iSCSI boot to mature in 2005 - Forward

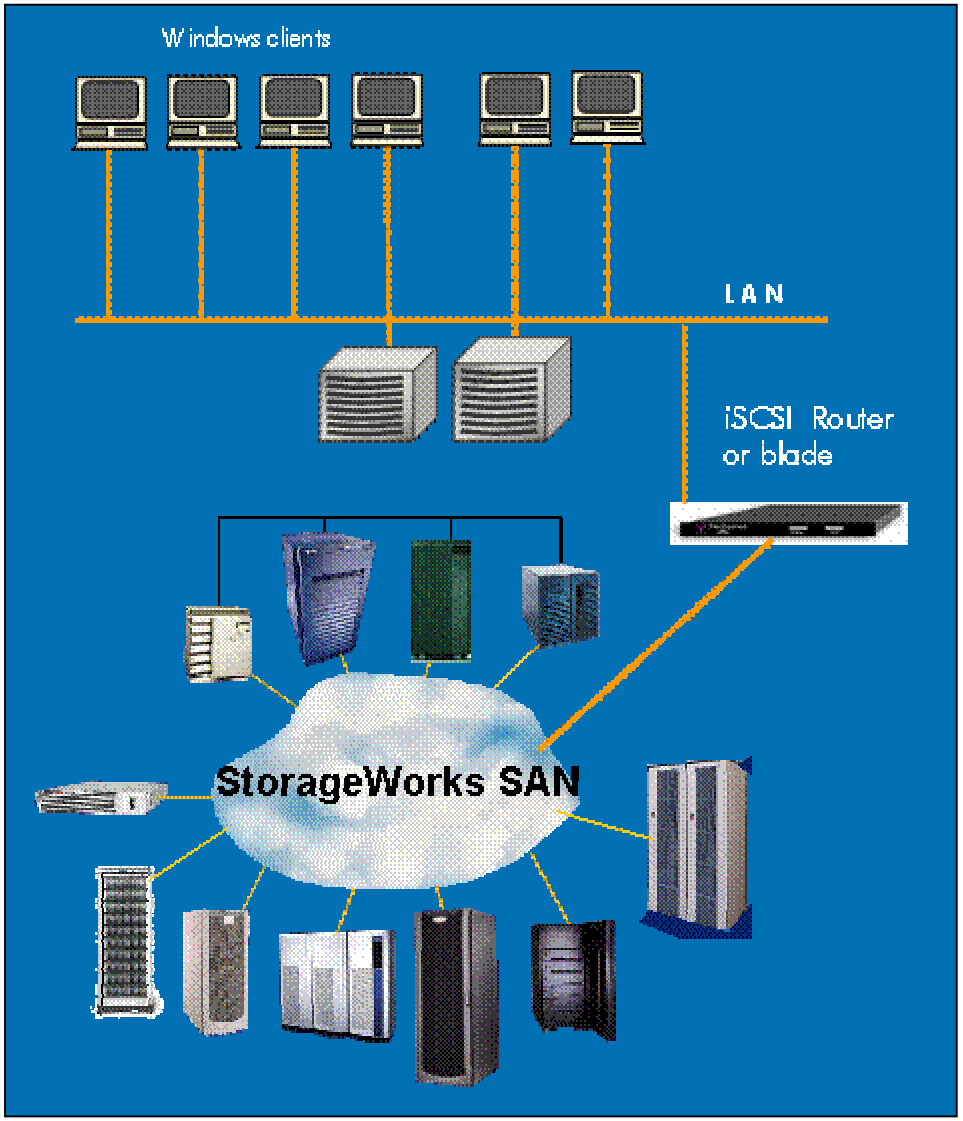
iSCSI Products

SAN Extension via FCIP and the IP Storage Router

- The HP StorageWorks IP Storage Router 2122-2 performs a Fibre Channel encapsulation process into IP Packets and reverses that process at the other end
- A tunnel connection is set up through the existing IP network routers and switches across LAN/WAN/MAN
- No need for dedicated wide area link to connect SAN
- Low cost connectivity for data replication, remote tape backup or data mirroring applications
- Fully supported with HP's Continuous Access solutions in M, and B-series infrastructures



iSCSI bridging for Storage Consolidation



- Provides access to storage resources of the SAN to non-critical or remote servers
- Enables greater flexibility, and asset utilization
- Ties “remote” servers, located across MAN/WAN, to FC SANs
- Connects SAN storage to Blade Servers that do not have FC connectivity
- Integrates the technologies that are important to the user in a single multi-protocol infrastructure

Move Beyond Traditional Boundaries

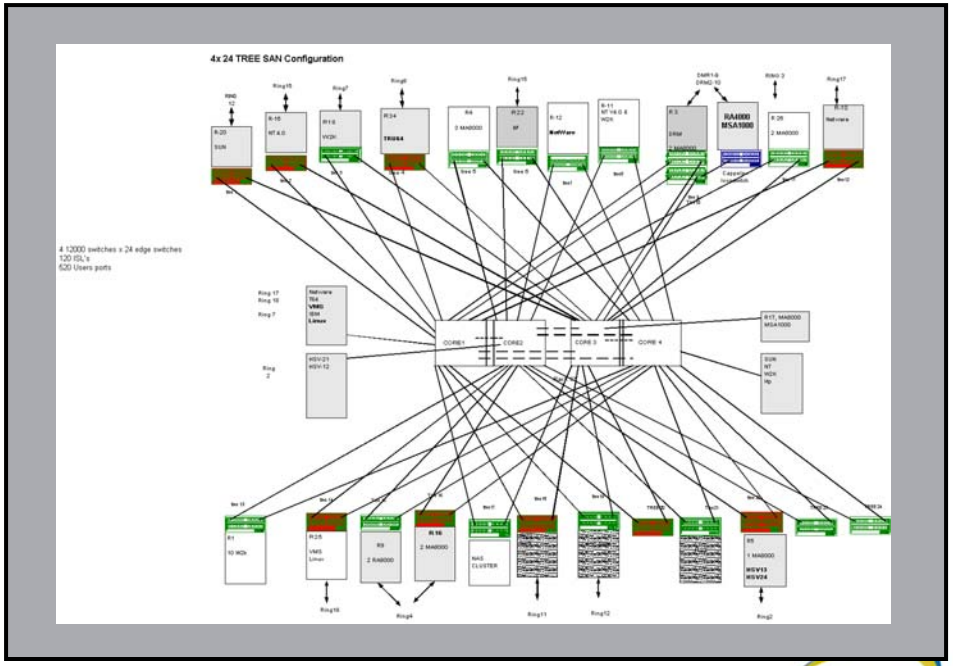
- **StorageWorks SR2122-2 iSCSI storage router**

SR2122-2 iSCSI

storage router

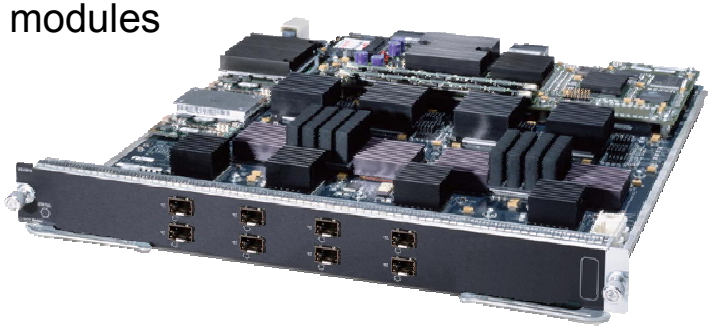
StorageWorks

SAN Design Guide



MDS 9000 IP Storage Services Module for the MDS 9506, 9509 & 9216 Switches

- Interface Feature Highlights
  - 8 Gigabit Ethernet ports with SFP/LC optical interfaces
  - Supports concurrent iSCSI and FCIP on any port
  - Full wire-rate performance for both FCIP and iSCSI
  - Leverages Fibre Channel interfaces on other switch modules
- iSCSI Feature Highlights
  - iSCSI Initiator-Fibre Channel Target
  - Transparent view of all allowed hosts/targets
  - iSCSI to Fibre Channel zone mapping
  - Seamless integration with MDS SAN services
- FCIP Feature Highlights
  - Up to 3 FCIP tunnels per port on all ports (24 tunnels per line card)
  - Optimized for full performance in WAN environments
    - TCP performance options
    - Window Scale, SACK, PAWS, TCP Keep-Alive, RTT
    - VSANs for enhanced stability in WAN environments
- MDS Family Services
  - All MDS SAN services supported on both Fibre Channel and IP interfaces



HP IP Storage Products

- 2003
  - SR-2122-2 Fibre Channel to iSCSI storage router
- Early 2004
  - HP-UX iSCSI software driver was released
- Mid 2004
  - HP qualified iSCSI support on the MDS 9000 IP Storage Services Module
  - Integrated iSCSI blade for the XP Arrays
  - IP device discovery added to OVSAM v.3.2
- End of 2004 to 2005
  - Low-cost NAS product combining file and block level access
  - Enhanced iSCSI/FCIP bridging for Fibre Channel switches
  - Roll out of native iSCSI arrays
  - Accelerated iSCSI NICs for ProLiant servers
  - iSCSI tape libraries

HP Multi-Protocol Strategy Solving the "How" Challenge

- HP offers a growing portfolio of IP storage components that complement the proven HP Fibre Channel SAN
  - New NAS, iSCSI Storage Router, SAN Extension solutions
- Design rules to build StorageWorks SANs supporting iSCSI bridging and NAS capabilities are provided in new edition of the SAN Design Guide
  - Tested, proven IP technologies
- HP SAN integration, planning and support services

iSCSI Tomorrow

- Next Steps for iSCSI
- How does RDMA Work
- Data Center Infrastructure
- Extending the Infrastructure
- Simpler Unified Networks
- RNICs
- Future Data Centers

What are the next steps for iSCSI?

TCP offload engine [TOE]

- use will enable lower CPU utilization and achieve wire speed operations
- presence of TOEs on servers may promote the use of iSCSI

10 Gbps Gigabit Ethernet

- initially expecting IP network bandwidth to exceed demand
- will be available before 10Gbps fibre channel
- may promote iSCSI's use

What are the next steps for iSCSI?

IP Network Traffic Shaping / Quality of Service

- may be possible to partition bandwidth towards storage, management and general IP traffic within a single network via IP QoS mechanisms
- management of network and storage resources can be consolidated to simplify overall management complexity and increase overall system performance

IPv6

- increase in namespace and allowance for mobile devices may increase demand for networked storage

iSCSI in evolution

iSER

- HP was instrumental in defining an overarching architecture to run iSCSI on all RDMA transports.
- iSER allows iSCSI to run on RDMA/TCP/IP.
- HP collaborating with other vendors to enhance iSER to run on IB.

Other work

- HP is an author/co-author of every iSCSI-related specification.
- HP closely worked with UNH on the Linux open-source iSCSI implementation.
- HP collaborating on design for iSCSI CIM (Common Information Model) in SNIA.
- HP working on iSCSI naming extension standardization so both FCP and iSCSI transports can easily co-exist in targets.

iSER and RDMA

iSER – iSCSI Extensions for RDMA

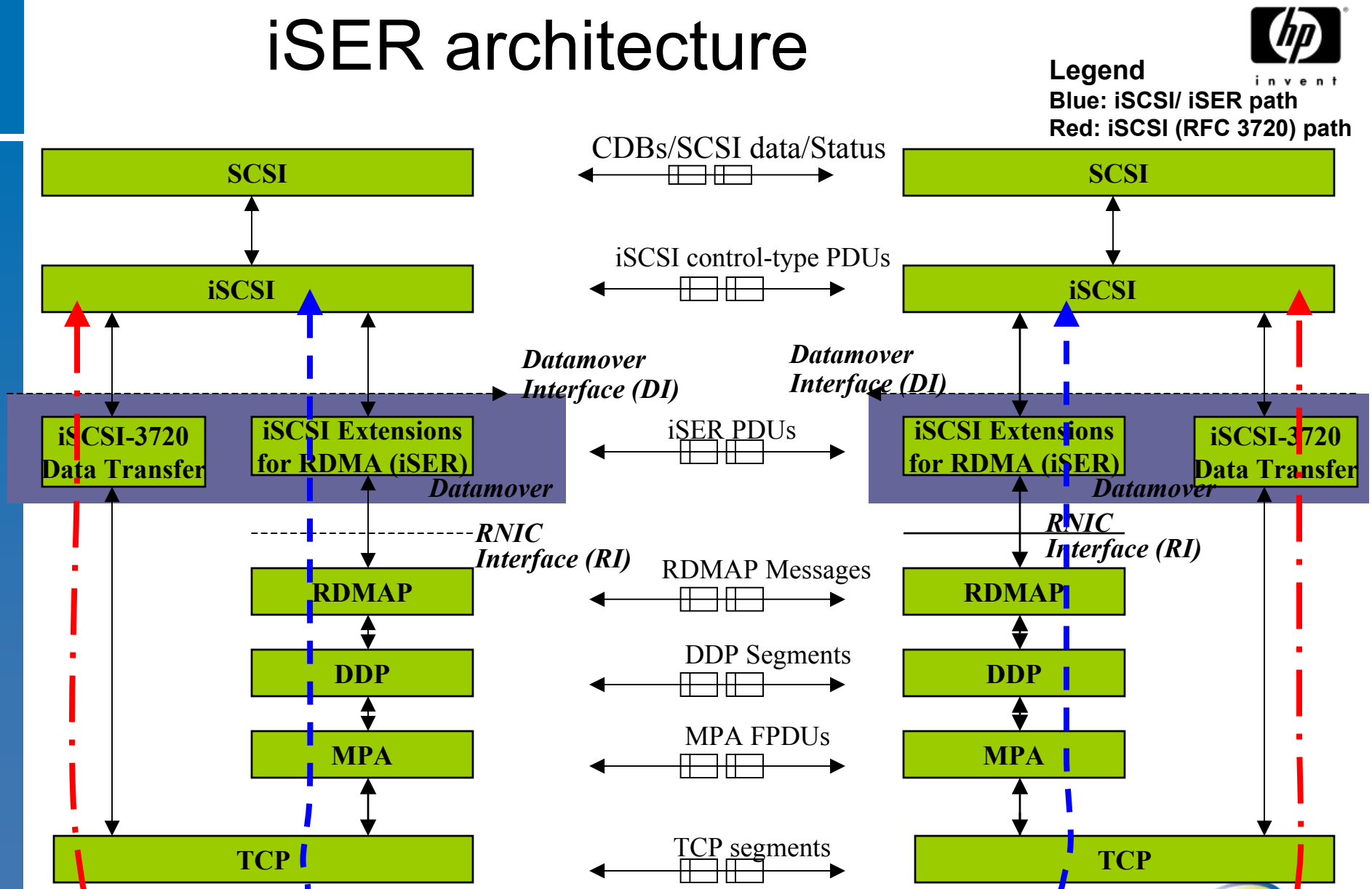
- Direct Memory to Memory semantics using RDMA transport
- Improved Performance through zero-copy data transfers
- Increased system scalability due to RDMA usage
- RDMA offers two things over an iSCSI offload NIC:

✓ *The RNIC is generic--it doesn't have to know iSCSI at all, and it can be used to accelerate other protocols like NFS and Sockets Direct.*

✓ *The RNIC can be cheaper than an iSCSI TOE adapter, because a segment reassembly memory is not required.*

➢ This will bring RNIC simplicity closer to generic Ethernet Cards

iSER architecture



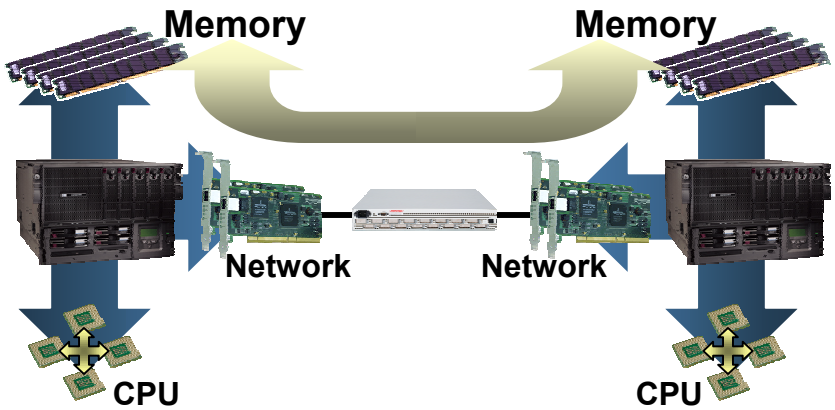
How RDMA Technology Works

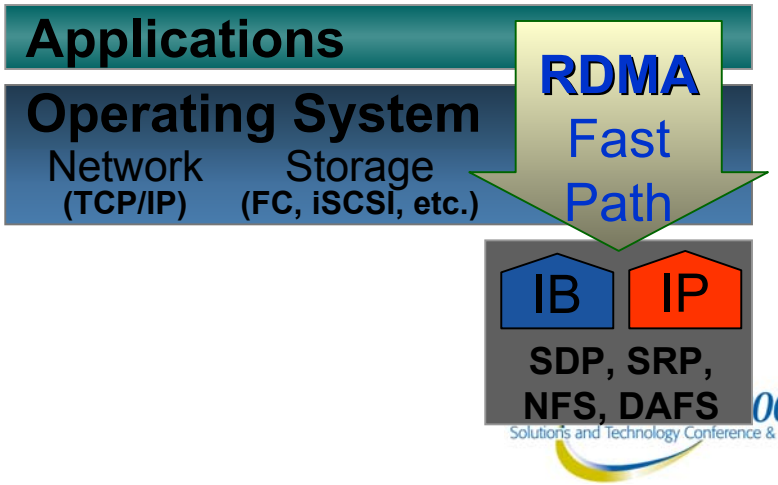
Fast and secure communications

- remote direct memory access (RDMA) provides efficient memory to memory transfers between systems
  - much less CPU intervention needed
  - true “zero copy” between systems, data placed directly in final destination
  - makes CPU available for other tasks
- maintains current, robust memory protection semantics

Protocol overhead & efficiency

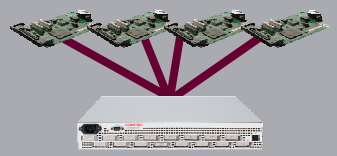
- TCP/IP protocol overhead an increasing percent of CPU & memory workload
  - up to ~1Ghz CPU power per Gb/s
  - memory bandwidth ~3x wire speed
- 3 options to solve the problem:
  - faster CPUs (paced by Moore’s Law)
  - move work to the NIC (TCP/IP offload)
  - create a more efficient, compatible protocol (RDMA/TCP)

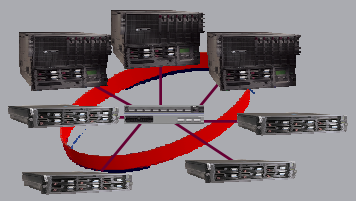


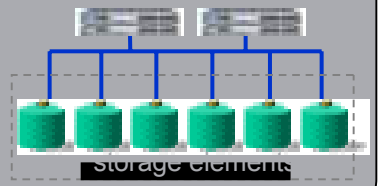


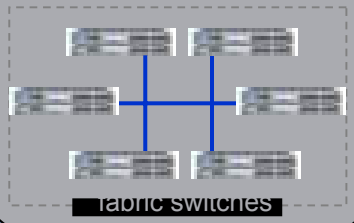
Data Center Infrastructure Evolution

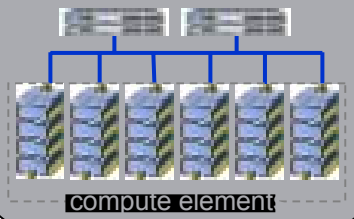
	Today	Tomorrow	
storage	• Fibre Channel • NAS (Storage over IP)	• 10 Gigabit Fibre Channel • iSCSI (Storage over IP)	storage fabric
networking	• 1 Gigabit Ethernet • KVM over IP (Lights-out Management)	• 10 Gigabit Ethernet • IP acceleration (TCP/IP & IP Sec)	data center fabric
clustering	• Proprietary Solutions (ServerNet, Myrinet, etc.)	• IP Fabrics (RDMA/TCP)	compute fabric







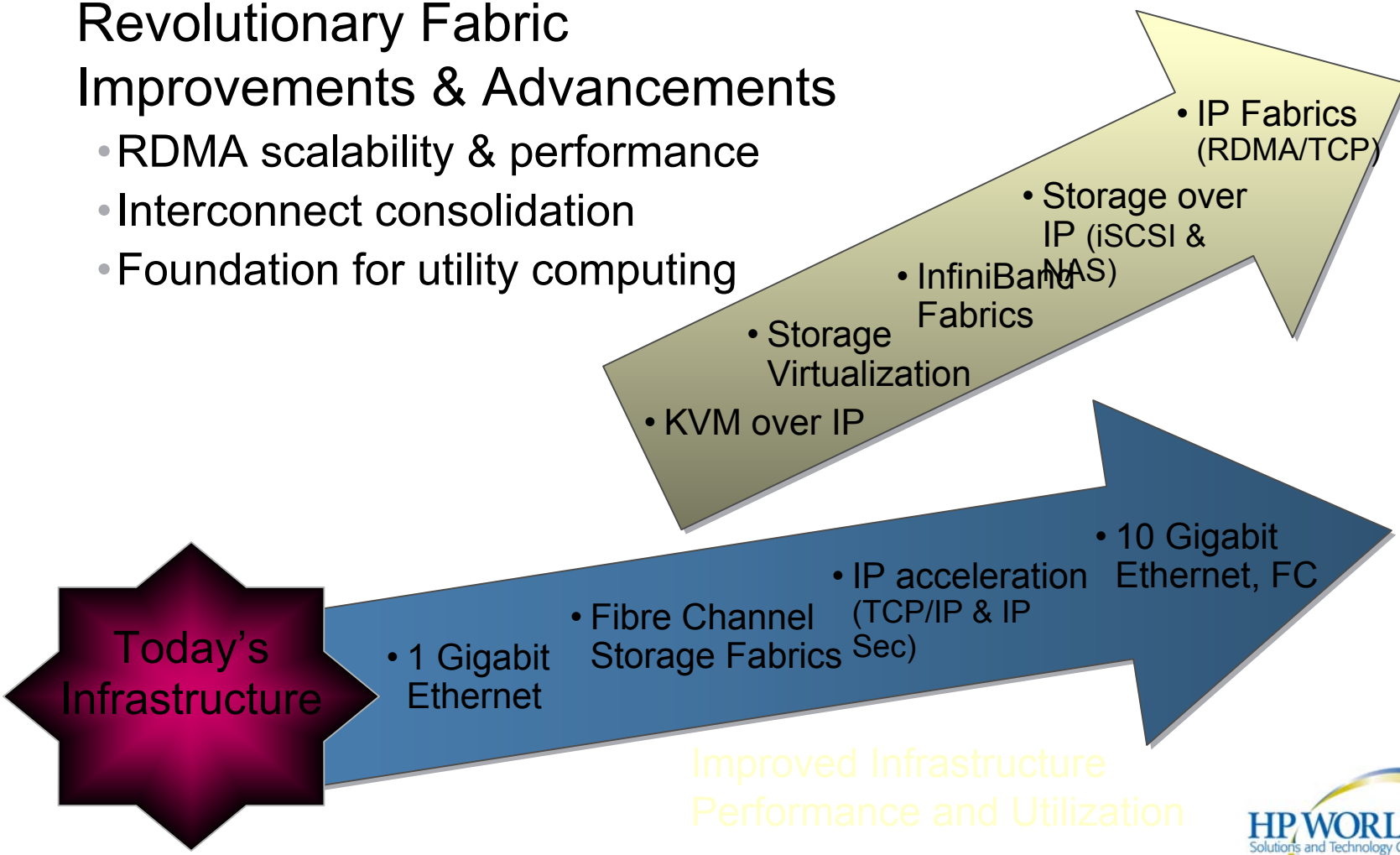




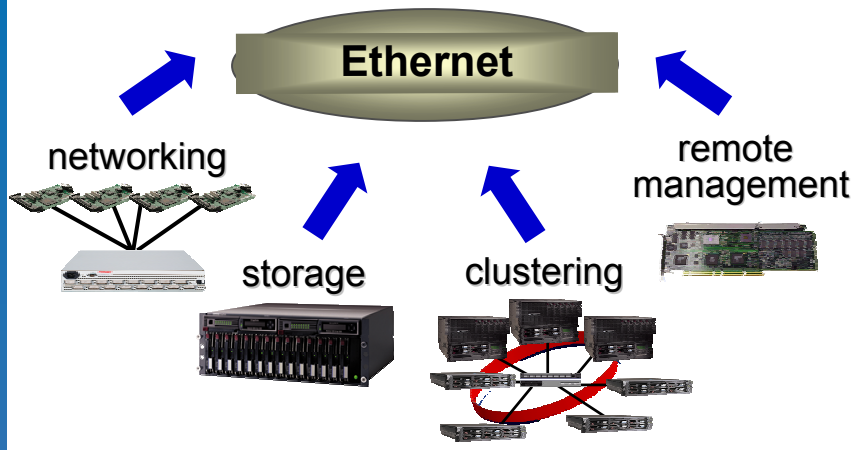
Extending Infrastructure Capabilities

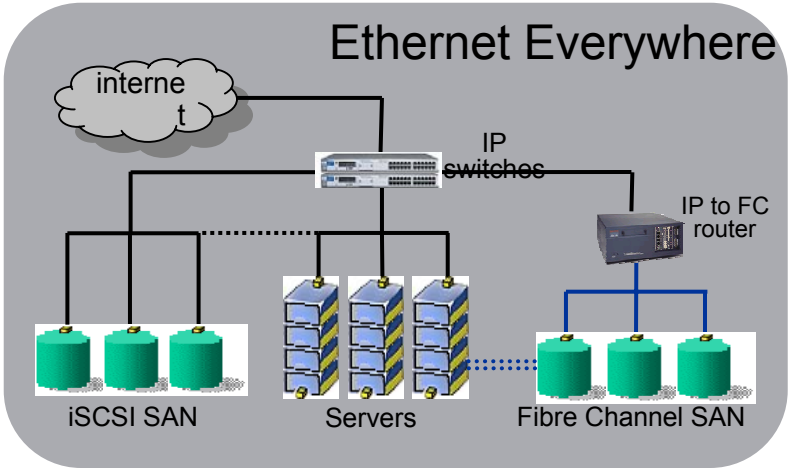
Revolutionary Fabric Improvements & Advancements

- RDMA scalability & performance
- Interconnect consolidation
- Foundation for utility computing



Simpler, Unified Infrastructure





Consolidate ports

- Leverage Ethernet pervasiveness, knowledge, cost leadership and volume
- Consolidate KVM over IP and reduce switch port costs (up to $1000 per port)

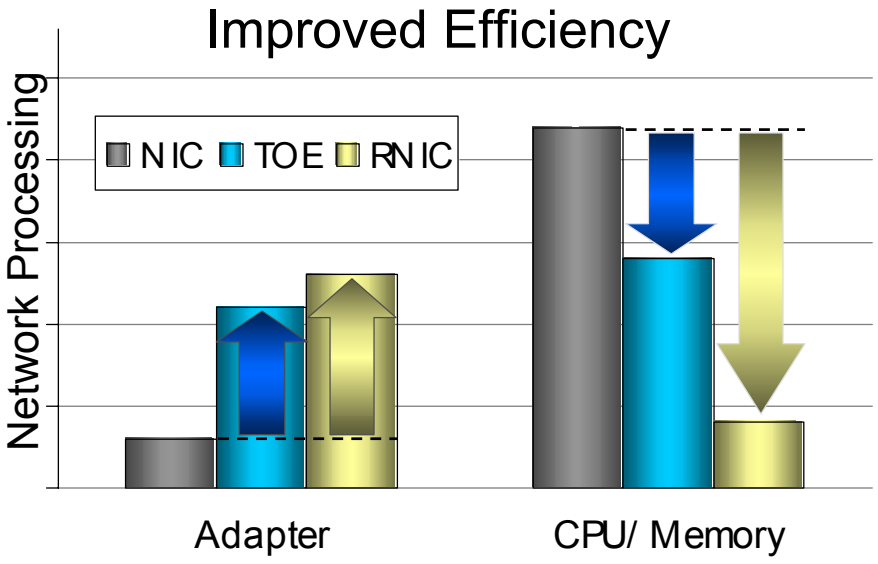
Converge functions

- Multiple functions can be consolidated over a common type of network
- Blade server storage connectivity (low cost)
- Packaged “end-to-end” Ethernet solutions

Broad connectivity - Ethernet Everywhere

- Bridge storage & network “islands”
- Extend geographic reach globally
- Centralize management

RNICs - Just Better Networking



Networking Benchmarks	BW Mbps	CPU Util %	Perf. Index
1Gb/s Enet	1000	60%	17
TOE	1000	40%	25
1Gb/s RDMA	1250	15%	74
10Gb/s RDMA	8500	10%	850

4X

50X

Note: Based on internal HP projections

RDMA enabled NICs (RNICs)

- More efficient network communications
- TOE moves TCP/IP work from the CPU
- RDMA reduces the communication work

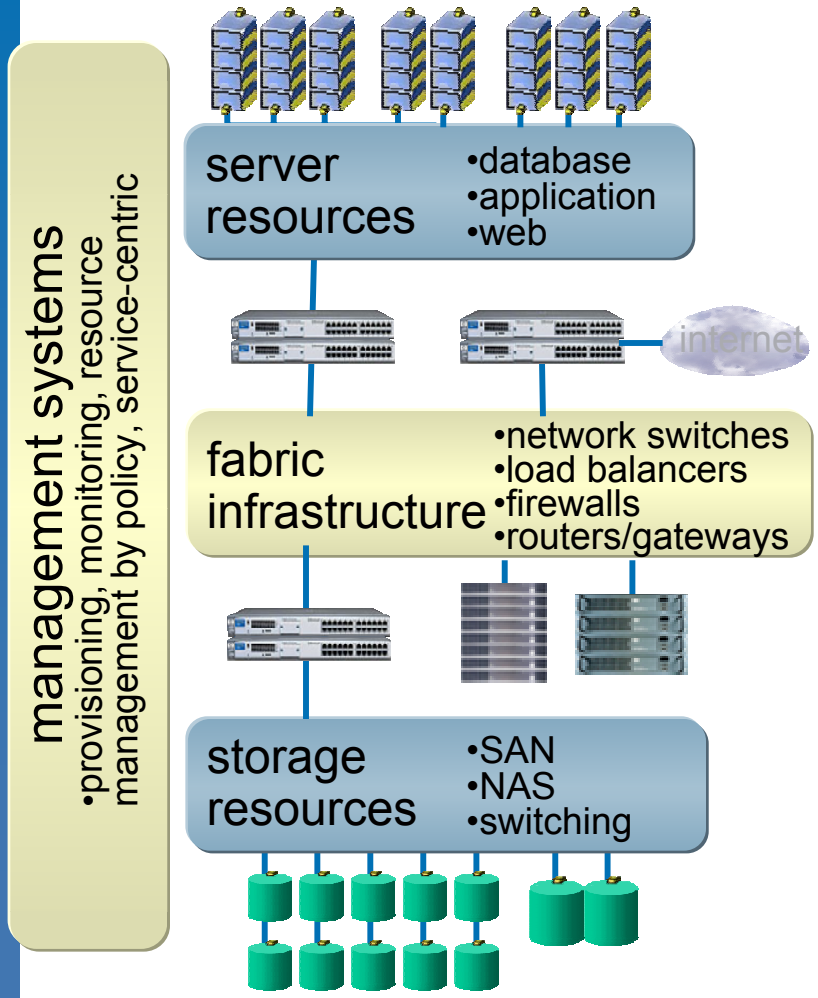
CPU/memory freed up for applications

- Zero-copy RDMA protocol conserves valuable memory bandwidth
- Much lower CPU utilization
- Per message communication overhead

Improved application performance

- Opportunity for increased application throughput or server consolidation
- Improved scalability for streaming applications or large data files

Future Data Centers



Data Center Fabrics

- Fabrics connecting heterogeneous "islands" of compute & storage resources
- "Ethernet everywhere" scaling across the datacenter with RDMA/TCP efficiency
- Routers & bridges translate between heterogeneous infrastructure islands

Enabling Utility Computing

- Static n-tier architecture (DISA) enhanced by dynamic resource pools
- Resource access managed by UDC, ProLiant Essentials and HP OpenView
- Functions virtualized over the fabric

Summary

HP is demonstrating pragmatic leadership in networked storage with the delivery of greater support for the integration of IP connectivity in enterprise storage solutions

- Move beyond traditional boundaries with multi-protocol capabilities and greater flexibility as part of the hp StorageWorks SAN
- Merge NAS and SAN environments with the fusion of a unified network storage architecture for improved return on your overall storage investment
- HP delivers tested, proven products and services to maximize the capability and efficiency of the storage network


invent

HP WORLD 2004

Solutions and Technology Conference & Expo

Co-produced by:

interex

shared knowledge • shared power

encompass

AN HP USER GROUP

RECOMMENDED TRAINING VENUE FOR THE

HP Certified Professional