Storage over Ethernet: What's In It for Me?

Stephen Foskett
stephen@fosketts.net
@SFoskett

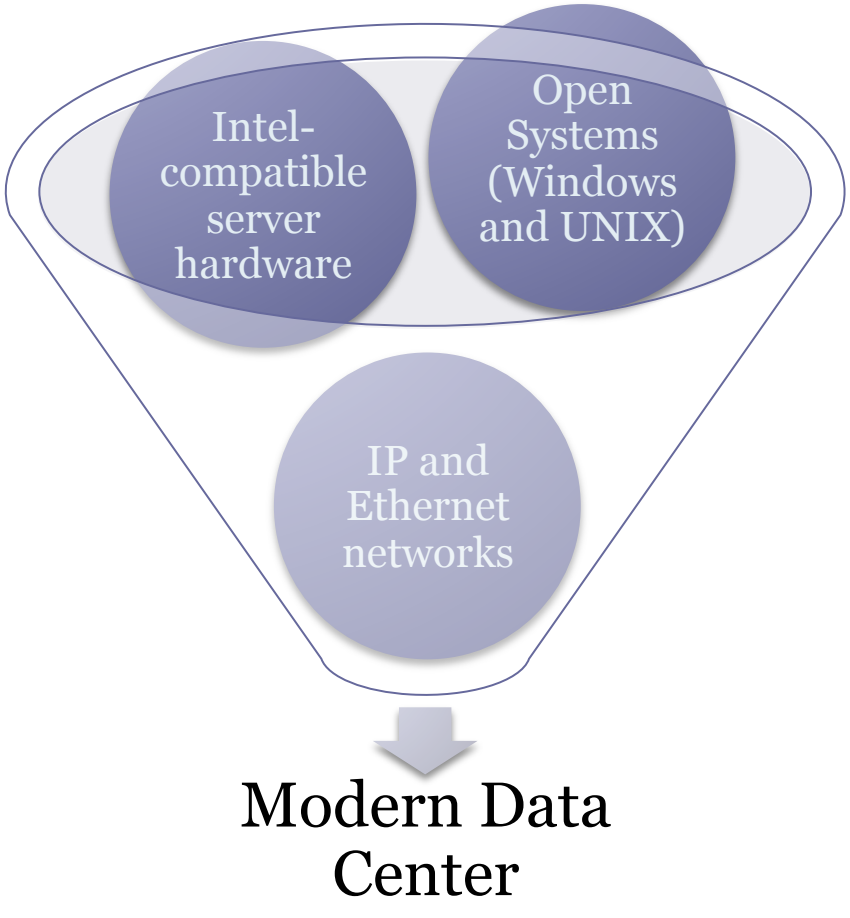
Gestalt IT

FoskettServices.com
Blog.Fosketts.net
GestaltIT.com

This is Not a Rah-Rah Session

Introduction: Converging on convergence

- Data centers rely more on standard ingredients
- What will connect these systems together?
- IP and Ethernet are logical choices



Drivers of Convergence

Virtualization

- Demanding greater network and storage I/O
- The “I/O blender”
- Mobility and abstraction

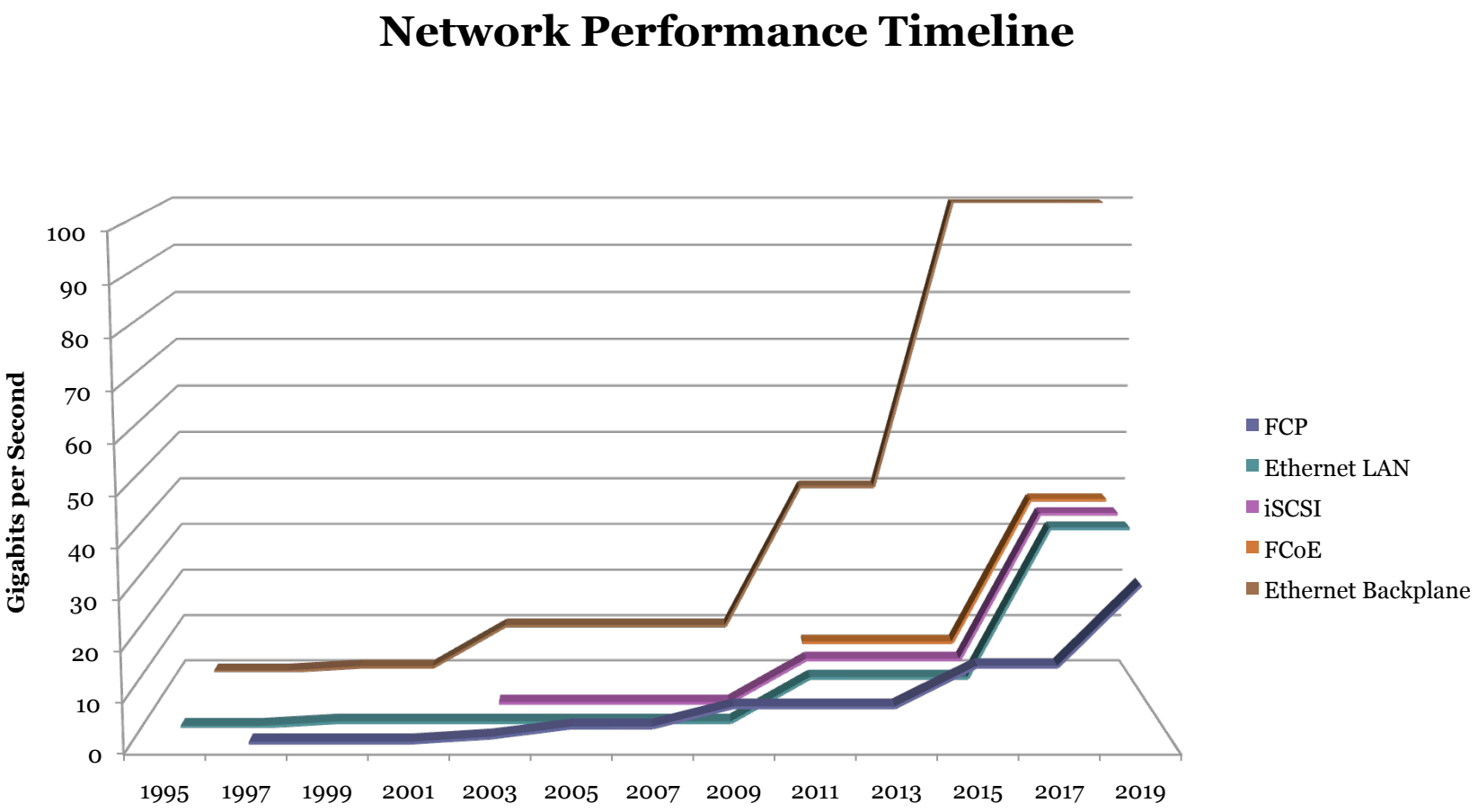
Consolidation

- Need to reduce port count, combining LAN and SAN
- Network abstraction features

Performance

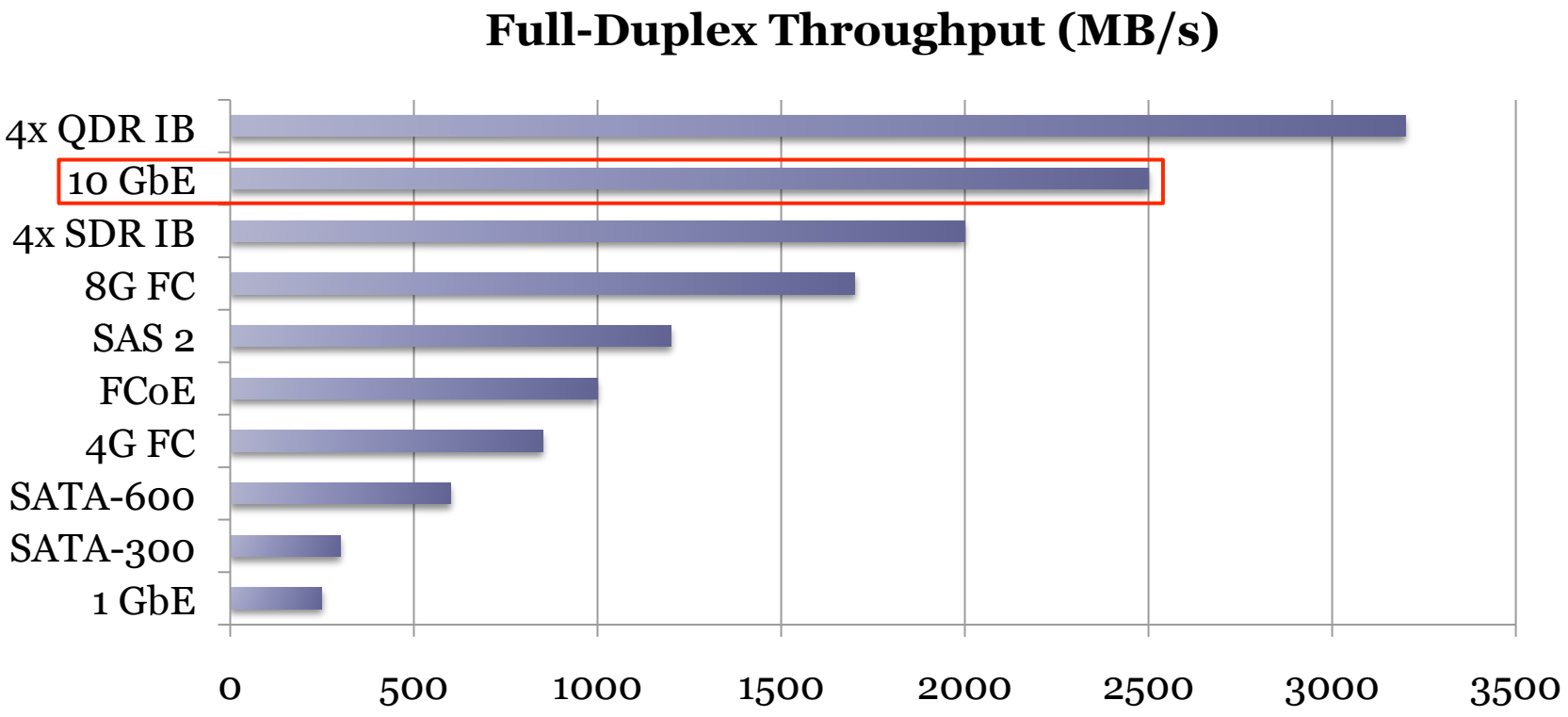
- Data-driven applications need massive I/O
- Virtualization and VDI

The Storage Network Roadmap



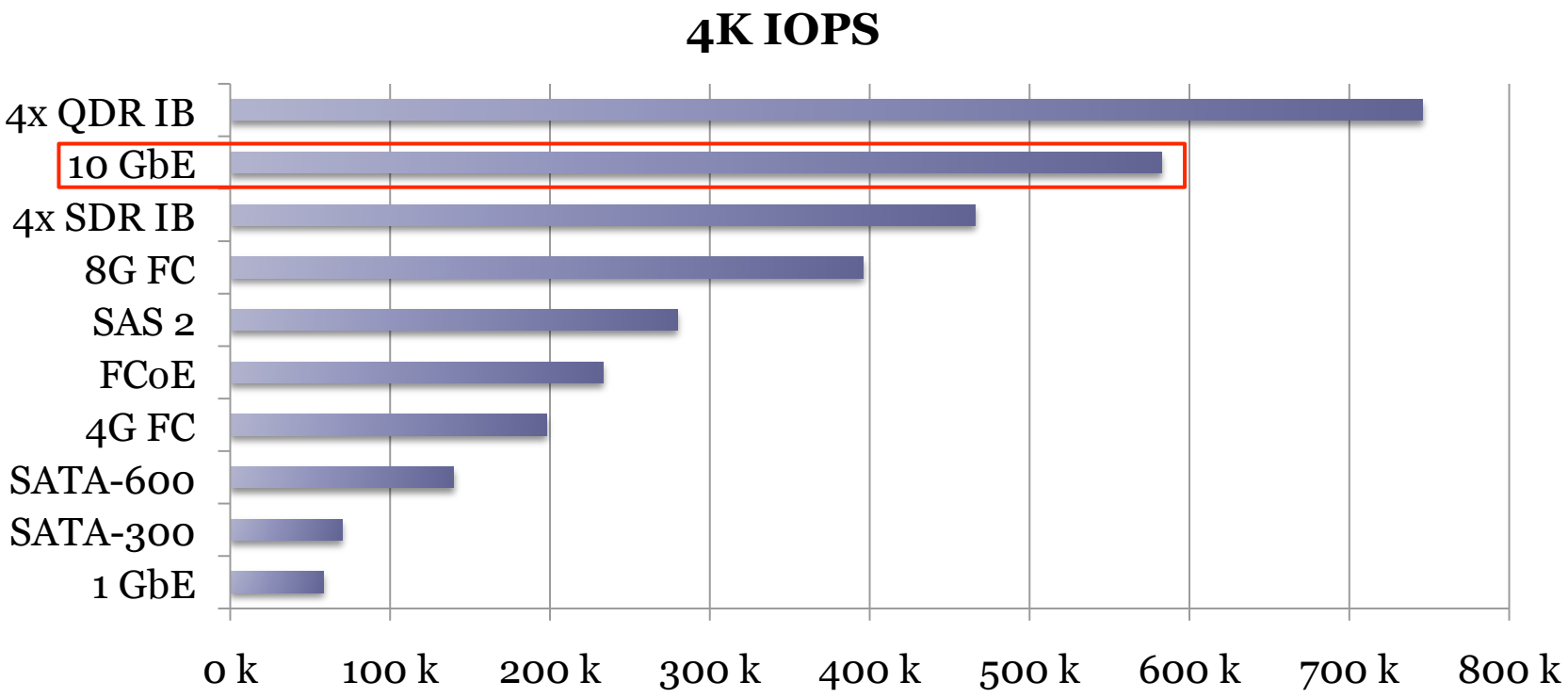
Serious Performance

- 10 GbE is faster than most storage interconnects
- iSCSI and FCoE both can perform at wire-rate



Latency is Critical Too

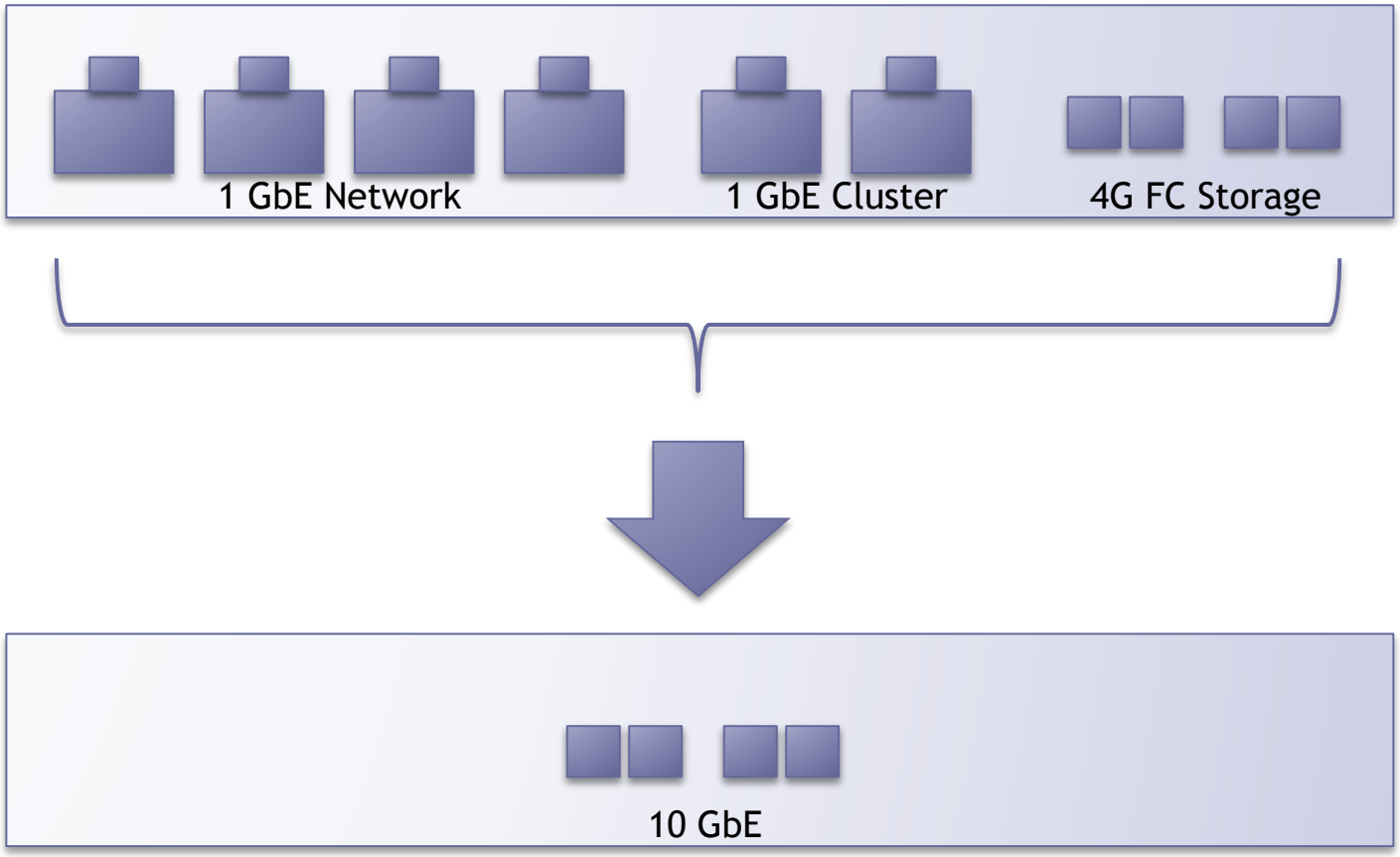
- Latency is even more critical in shared storage



Benefits Beyond Speed

- 10 GbE takes performance off the table (for now…)
- But performance is only half the story:
  - Simplified connectivity
  - New network architecture
  - Virtual machine mobility

Server Connectivity



(Plus 6 Gbps
extra capacity)

Flexibility

- No more rats-nest of cables
- Servers become interchangeable units
  - Swappable
  - Brought on line quickly
  - Few cable connections
- Less concern about availability of I/O slots, cards and ports
- CPU, memory, chipset are deciding factor, not HBA or network adapter

Changing Data Center

- Placement and cabling of SAN switches and adapters dictates where to install servers
- Considerations for placing SAN-attached servers:
  - Cable types and lengths
  - Switch location
  - Logical SAN layout
- Applies to both FC and GbE iSCSI SANs
- Unified 10 GbE network allows the same data and storage networking in any rack position

Virtualization: Performance and Flexibility

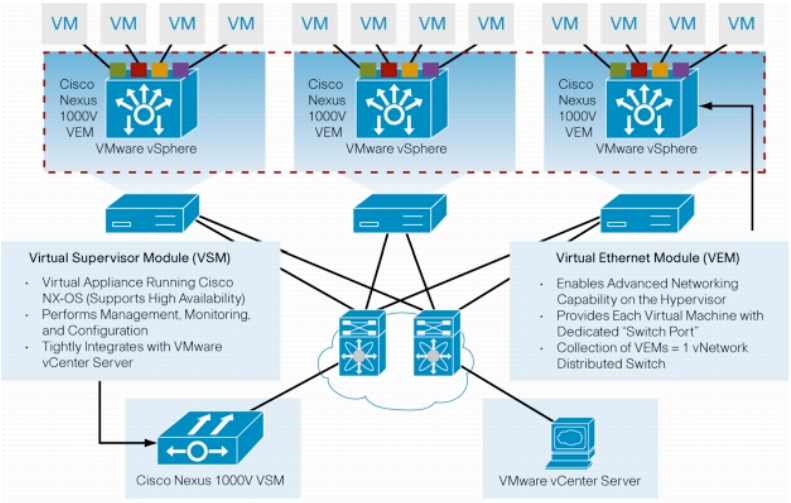
- Performance and flexibility benefits are amplified with virtual servers
- 10 GbE acceleration of storage performance, especially latency – “the I/O blender”
- Can allow performance-sensitive applications to use virtual servers

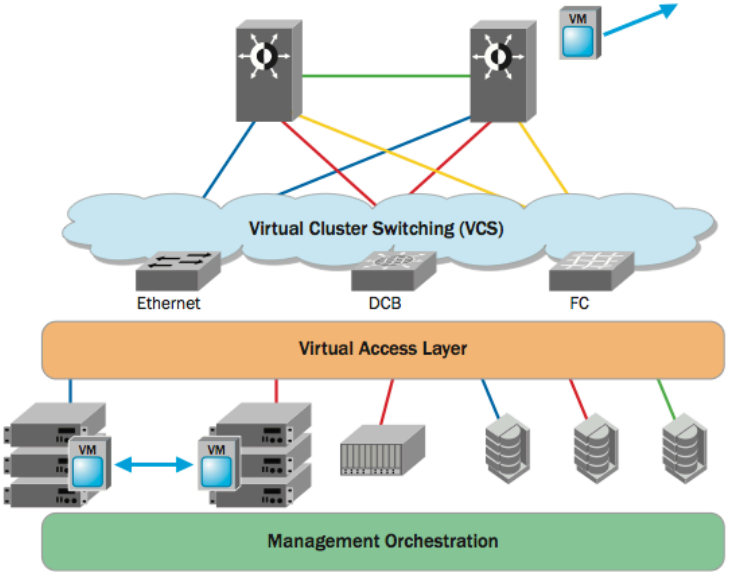
Virtual Machine Mobility

- Moving virtual machines is the next big challenge
- Physical servers are difficult to move around and between data centers
- Pent-up desire to move virtual machines from host to host and even to different physical locations

Virtualization-Aware Networks

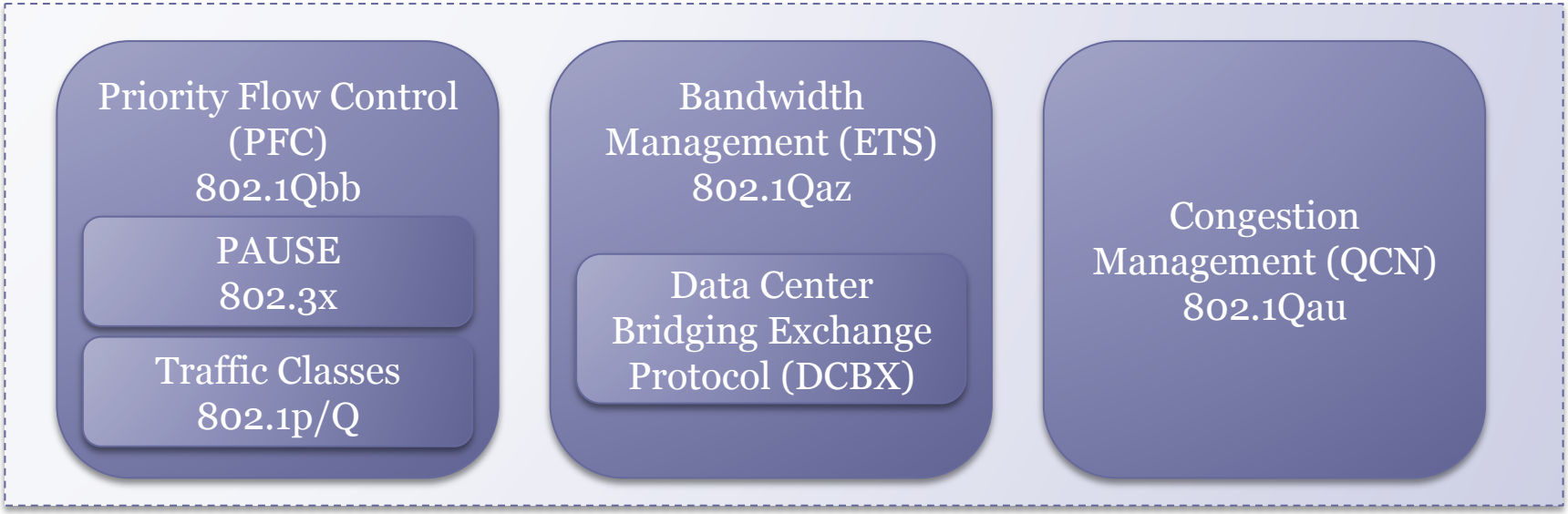
- Two schools of thought have emerged:
  - Extend the network inside the virtual environment (e.g. Cisco)
  - Rely on smart and virtualization-aware physical network switches (e.g. Brocade)
- Both enable seamless movement of virtual machines around the LAN





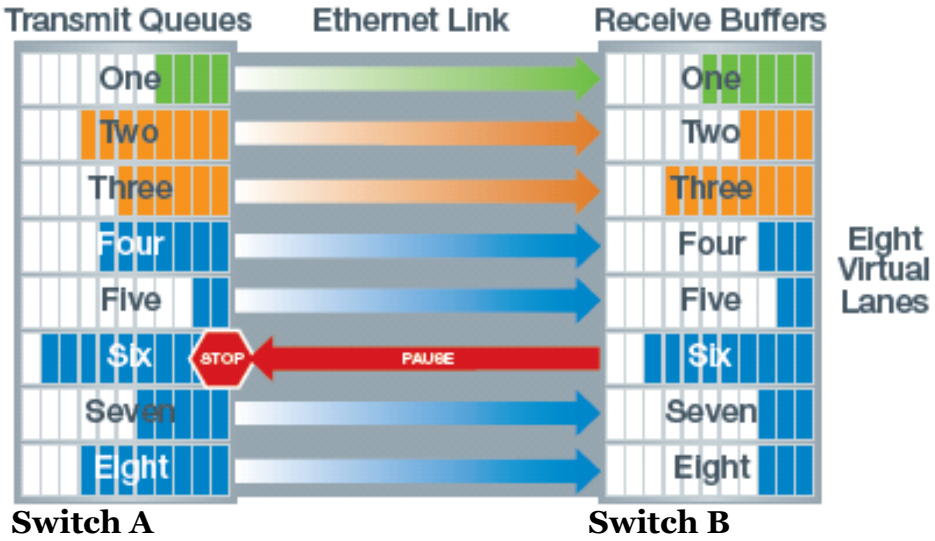
Data Center Ethernet

- Ethernet and SCSI were not made for each other
  - SCSI expects a lossless and transport with guaranteed delivery
  - Ethernet expects higher-level protocols to take care of issues
- Data Center Bridging is a project to create lossless Ethernet
  - IEEE name is Data Center Bridging (DCB)
  - Cisco trademarked Data Center Ethernet (DCE)
  - Many vendors used to call it Converged Enhanced Ethernet (CEE)



Flow Control

- PAUSE (802.3x)
  - Reactive not proactive (like FC credit approach)
  - Allows a receiver to block incoming traffic in a point-to-point Ethernet link
- Priority Flow Control 802.1Qbb)
  - Uses an 8-bit mask in PAUSE to specify 802.1p priorities
  - Blocks a class of traffic, not an entire link
  - Ratified and shipping

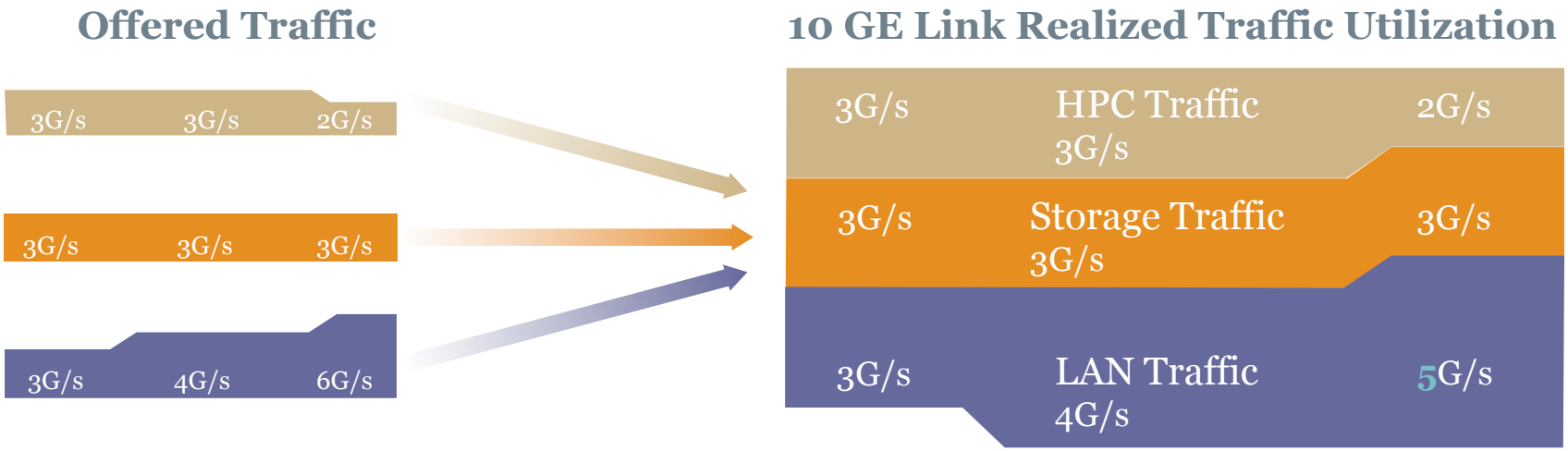


Graphic courtesy of EMC

- Result of PFC:
  - Handles transient spikes
  - Makes Ethernet lossless
  - Required for FCoE

Bandwidth Management

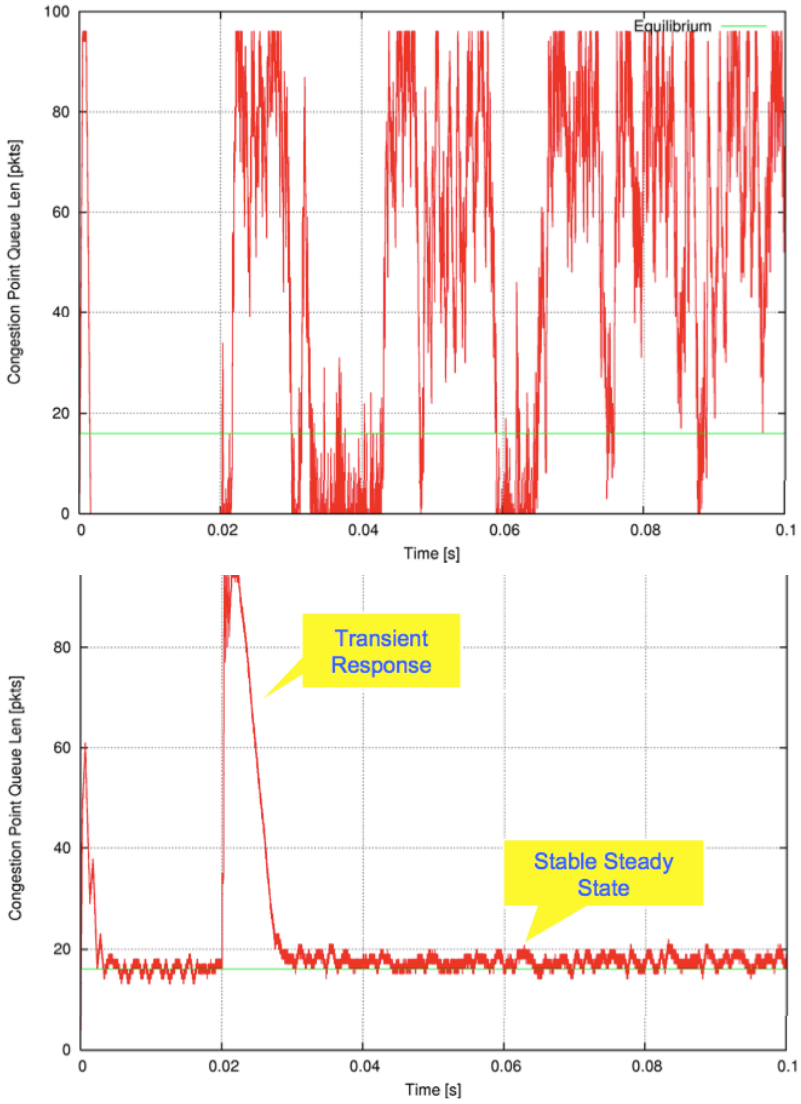
- Enhanced Transmission Selection (ETS) 802.1Qaz
  - Latest in a series of attempts at Quality of Service (QoS)
  - Allows "overflow" to better-utilize bandwidth
- Data Center Bridging Exchange (DCBX) protocol
  - Allows devices to determine mutual capabilities
  - Required for ETS, useful for others
- Ratified and shipping



Graphic courtesy of EMC

Congestion Notification

- Need a more proactive approach to persistent congestion
- QCN 802.1Qau
  - Notifies edge ports of congestion, allowing traffic to flow more smoothly
  - Improves end-to-end network latency (important for storage)
  - Should also improve overall throughput
- Not quite ready



Graphic courtesy of Broadcom

SAN History: SCSI

- Early storage protocols were system-dependent and short distance
  - Microcomputers used internal ST-506 disks
  - Mainframes used external bus-and-tag storage
- SCSI allowed systems to use external disks
  - Block protocol, one-to-many communication
  - External enclosures, RAID
  - Replaced ST-506 and ESDI in UNIX systems
  - SAS dominates in servers; PCs use IDE (SATA)

The Many Faces of SCSI

Comparing Protocols

	iSCSI	FCoE	FCP
DCB Ethernet	Optional	Required	N/A
Routable	Yes	No	Optional
Hosts	Servers and Clients	Server-Only	Server Only
Initiators	Software and Hardware	Software* and Hardware	Hardware
Guaranteed Delivery	Yes (TCP)	Optional	Optional
Flow Control	Optional (Rate-Based)	Rate-Based	Credit-Based
Inception	2003	2009	1997
Fabric Management	Ethernet Tools	FC Tools	FC Tools

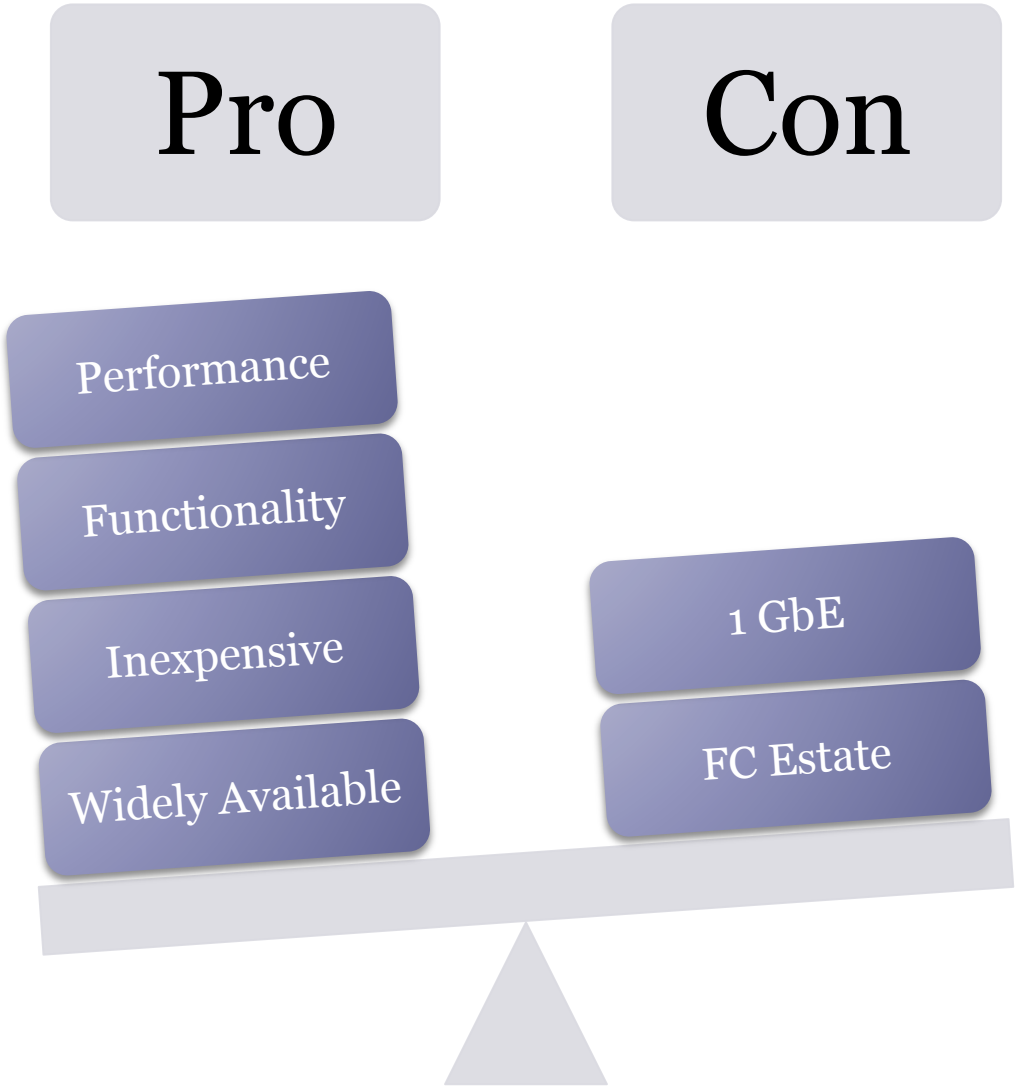
iSCSI: Where It's At

- iSCSI targets are robust and mature
  - Just about every storage vendor offers iSCSI arrays
  - Software targets abound, too (Nexenta, Microsoft, StarWind)
- Client-side iSCSI is strong as well
  - Wide variety of iSCSI adapters/HBAs
  - Software initiators for UNIX, Windows, VMware, Mac
- Smooth transition from 1- to 10-gigabit Ethernet
  - Plug it in and it works, no extensions required
  - iSCSI over DCB is rapidly appearing

iSCSI Support Matrix

	Certified?	Initiator	Multi-Path	Clustering
Windows	Yes	HBA/SW	MPIO, MCS	Yes
Sun	Yes	HBA/SW	Trunking, MPIO	Yes
HP	Yes	SW	PV Links	?
IBM	Yes	SW	Trunking	?
RedHat	Yes	HBA/SW	Trunking, MPIO	Yes
Suse	Yes	HBA/SW	Trunking, MPIO	Yes
ESX	Yes	SW	Trunking	Yes

Why Go iSCSI?


Pro
Con
Performance
Functionality
Inexpensive
Widely Available
1 GbE
FC Estate

The Three-Fold Path of Fibre Channel

End-to-End Fibre Channel

- The traditional approach – no Ethernet
- Currently at 8 Gb
- Widespread, proven

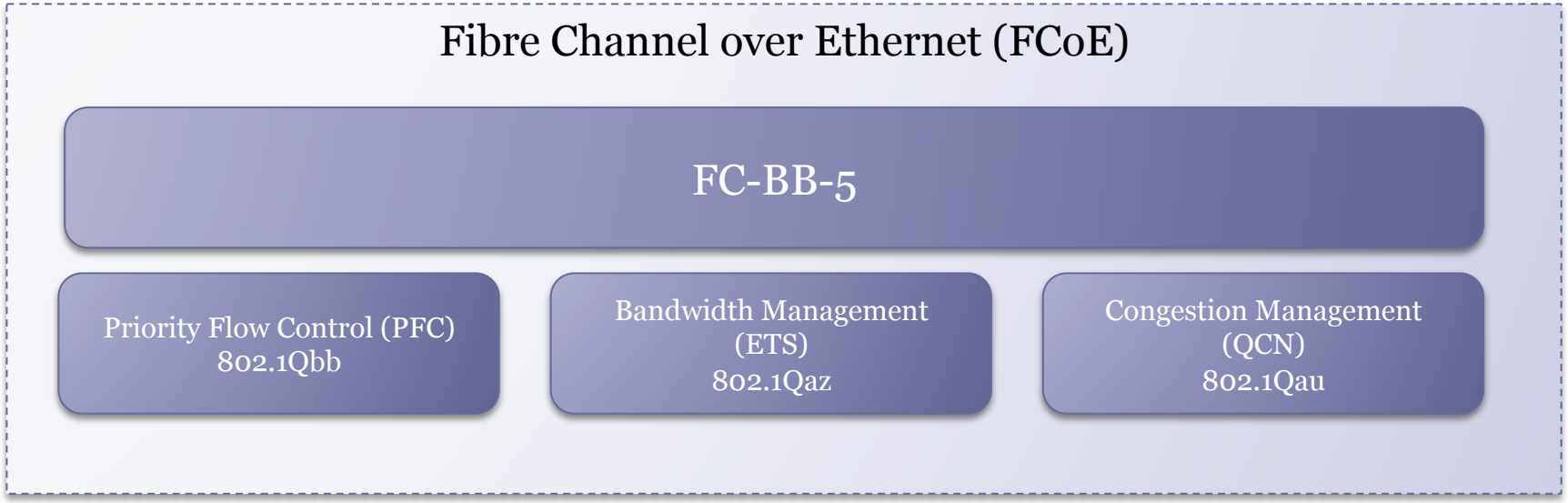
FC Core and FCoE Edge

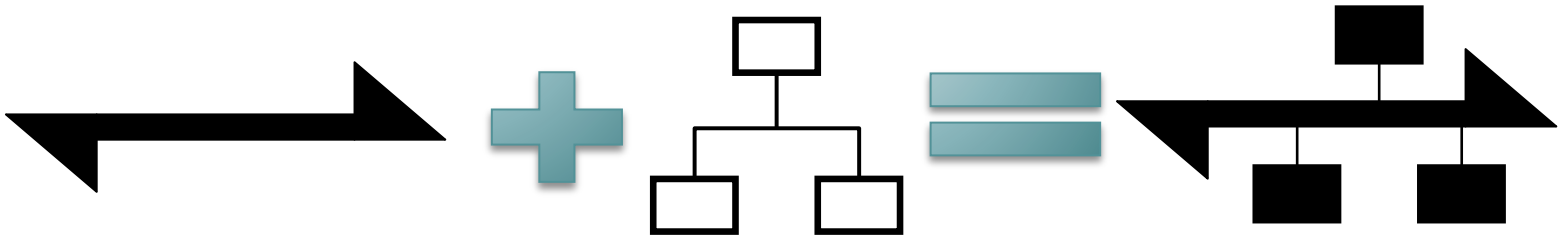
- Common "FCoE" approach
- Combines 10 GbE with 4- or 8-Gb FC
- Functional
- Leverages FC install base

End-to-End FCoE

- Extremely rare currently
- All-10 GbE
- Should gain traction
- Requires lots of new hardware

FCoE Spotters' Guide





Fibre Channel

Ethernet

Fibre Channel over Ethernet (FCoE)

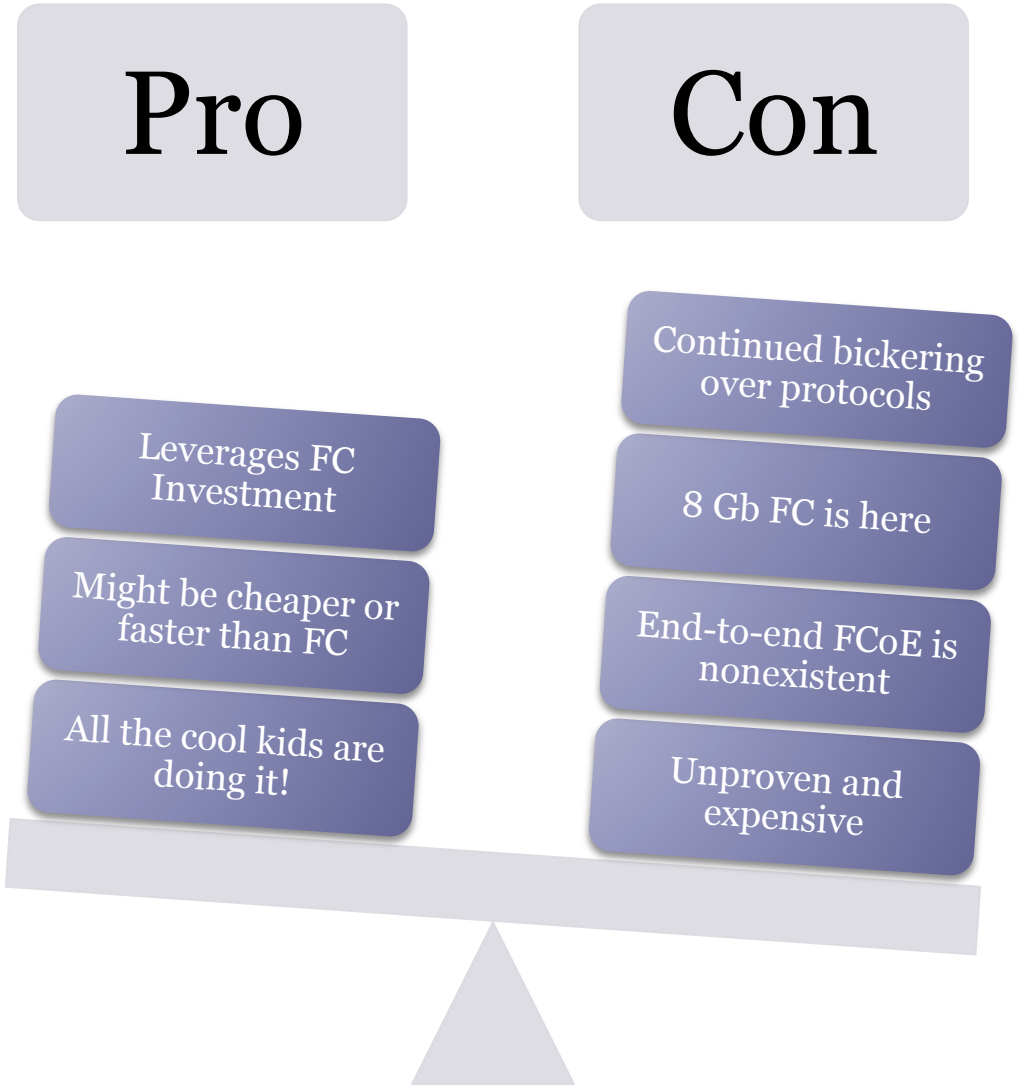
Why FCoE?

- Large FC install base/investment
  - Storage arrays and switches
  - Management tools and skills
- Allows for incremental adoption
  - FCoE as an edge protocol promises to reduce connectivity costs
  - End-to-end FCoE would be implemented later
- I/O consolidation and virtualization capabilities
  - Many DCB technologies map to the needs of server virtualization architectures
- Also leverages Ethernet infrastructure and skills

Who's Pushing FCoE and Why?

- Cisco wants to move to an all-Ethernet future
- Brocade sees it as a way to knock off Cisco in the Ethernet market
- Qlogic, Emulex, and Broadcom see it as a differentiator to push silicon
- Intel wants to drive CPU upgrades
- NetApp thinks their unified storage will win as native FCoE targets
- EMC and HDS want to extend their dominance of high-end FC storage
- HP, IBM, and Oracle don't care about FC anyway

FCoE Reality Check

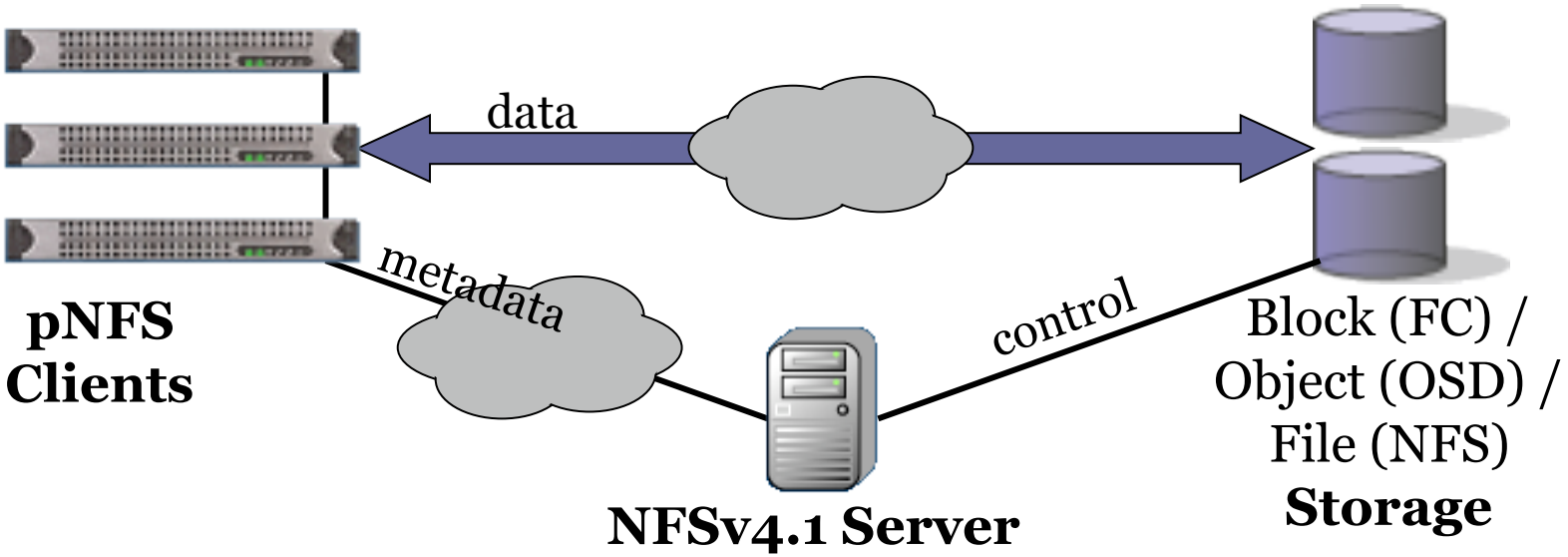


NFS: The Other Ethernet Storage Protocol

- NFS has grown up and out
  - NFS v4 is a much-improved NAS protocol
  - pNFS (v4.1) does it all - file, block, and object
- Do you hate NFS? NFS v4 should fix that!
  - One protocol on a single port
  - Stateful with intelligent leases
  - Strong, integrated authentication
  - Better access control
  - Strong, integrated encryption (Kerberos V5)
  - No more UDP!

Then There's pNFS…

- "What if we added everything to NFS?"
  - pNFS is the child of SAN FS and NFS
  - Focused on scale-out
  - Developed by Panasas, EMC, Sun, NetApp, IBM



Graphic courtesy of SNIA

What You Should Know About pNFS

- General pNFS protocol is standardized in NFS v4.1
- File access is standardized in NFS v4.1
- Block access is not standardized but will use SCSI (iSCSI, FC, FCoE, etc)
- Object access is not standardized but will use OSD over iSCSI
- Server-to-server control protocol isn't agreed on
- OpenSolaris client is file-only
- Linux client supports files, and work on blocks and objects is ongoing
- Single namespace with single metadata server

What's in it for you?

Server Managers

- More flexibility and mobility
- Better support for virtual servers and blades
- Increased overall performance

Network Managers

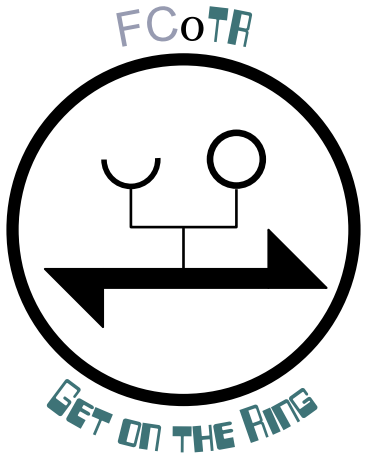
- Wider sphere of influence (Ethernet everywhere)
- More tools to control traffic
- More to learn
- New headaches from storage protocols

Storage Managers

- Fewer esoteric storage protocols
- New esoteric network protocols
- Less responsibility for I/O
- Increased focus on data management

Counterpoint: Why Ethernet?

- Why converge on Ethernet at all?
  - Lots of work just to make Ethernet perform unnatural acts!
- Why not InfiniBand?
  - Converged I/O already works
  - Excellent performance and scalability
  - Wide hardware availability and support
  - Kinda pricey; another new network
- Why not something else entirely?
  - Token Ring would have been great!



Conclusion

- Ethernet will come to dominate
  - Economies of scale = lower cost
  - Focal point of development
  - Excellent roadmap
  - DCB is here (PFC, ETS, DCBX)
  - Further extensions questionable (QCN, TRILL)

Conclusion

- iSCSI will continue to grow
  - Easy, cheap, widely supported
  - Grow to 10 Gb seamlessly
- FCoE is likely but not guaranteed
  - Relentlessly promoted by major vendors
  - Many areas still not ready for prime-time
- The future of NFS is unclear
  - NFS v4 is an excellent upgrade and should be adopted
  - pNFS is strongly supported by storage vendors
  - Scale-out files are great, but do we need block and object in NFS, too?

Thank You!

Stephen Foskett
stephen@fosketts.net
twitter.com/sfoskett
+1(508)451-9532

FoskettServices.com
blog.fosketts.net
GestaltIT.com
Tomorrow and Friday: TechFieldDay.com