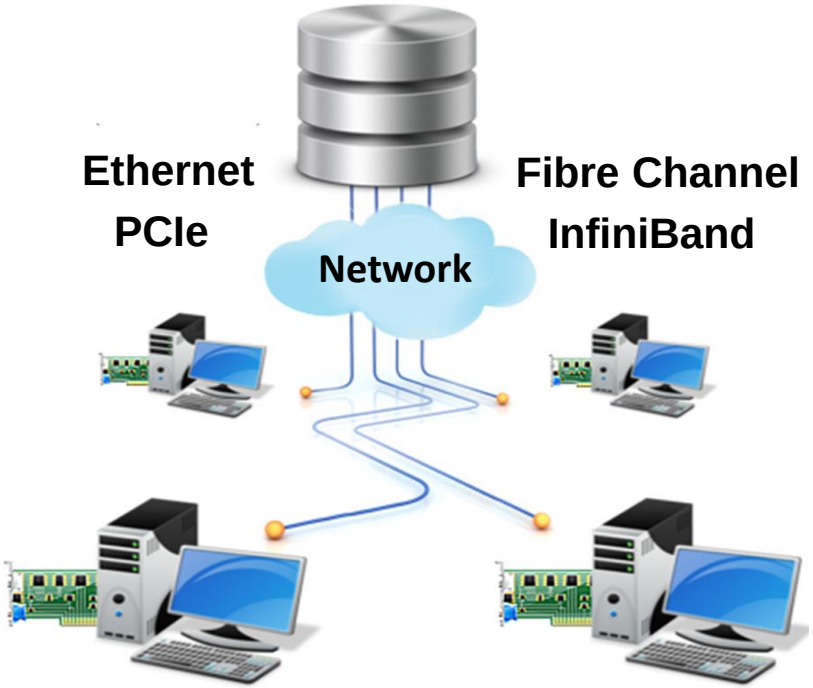
Networking Flash Technology "Showdown"

Rob Davis, J Metz, Motti Beck,
Brandon Hoff, Peter Onufryk

Why Network Flash Based Storage?

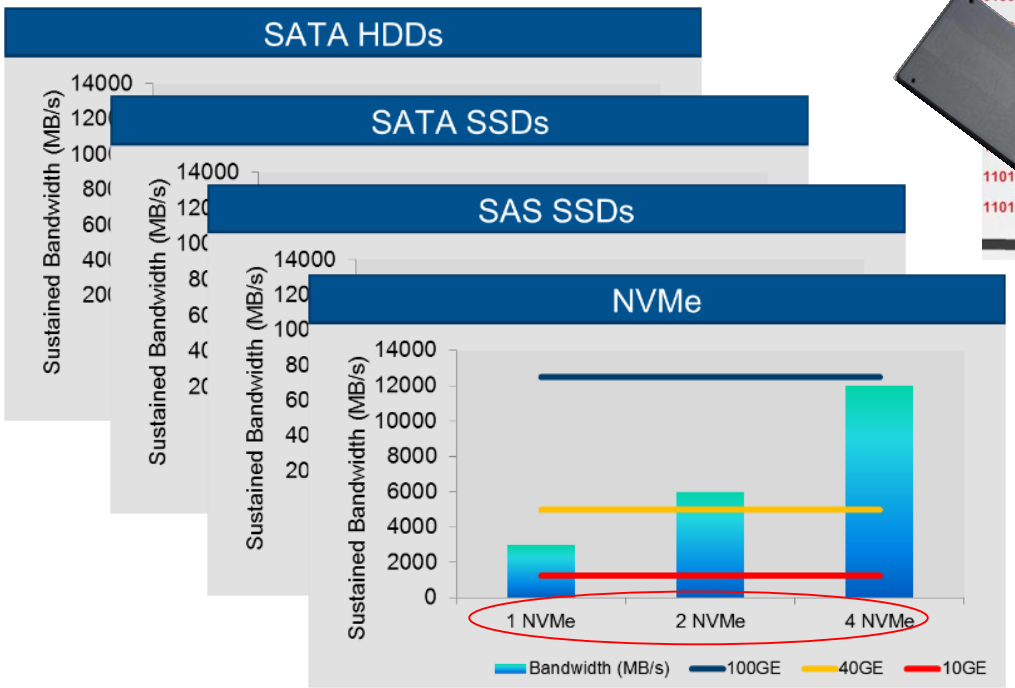
- There are advantages to shared storage
  - Better utilization:
    - capacity, rack space, power
  - Scalability
  - Manageability
  - Fault isolation
- Shared storage requires a Network

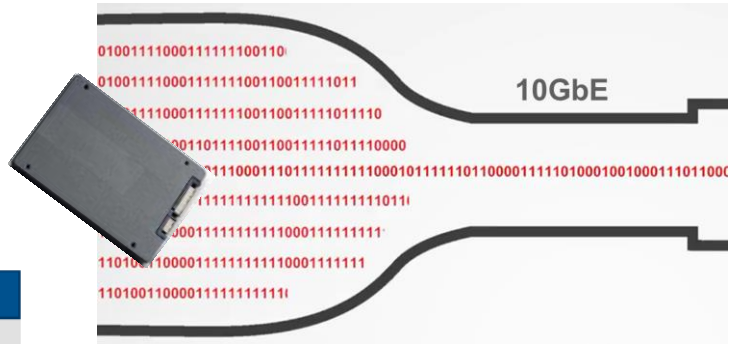


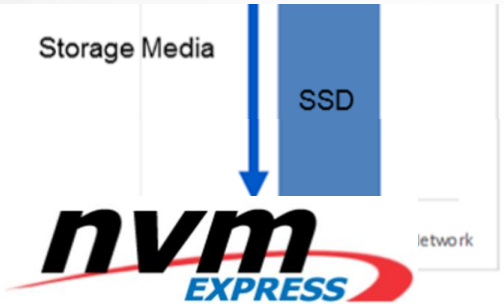
Agenda

- ***PCIe*** Networked Flash Storage
  - Peter Onufryk, **Microsemi(Microchip)**, NVM Solutions Fellow
- ***InfiniBand*** Networked Flash Storage
  - Motti Beck, **Mellanox**, Sr. Dir. Enterprise Market Development
- ***Fibre Channel*** Networked Flash Storage
  - Brandon Hoff, Principle Software Architect, **Emulex Connectivity Division, Broadcom**
- ***Ethernet*** Networked Flash Storage
  - J Metz, **Cisco**, R&D Engineer, Advanced Storage, Office of the CTO, UCS Systems Group
- Panel Discussion

Faster Storage Needs a Faster Network

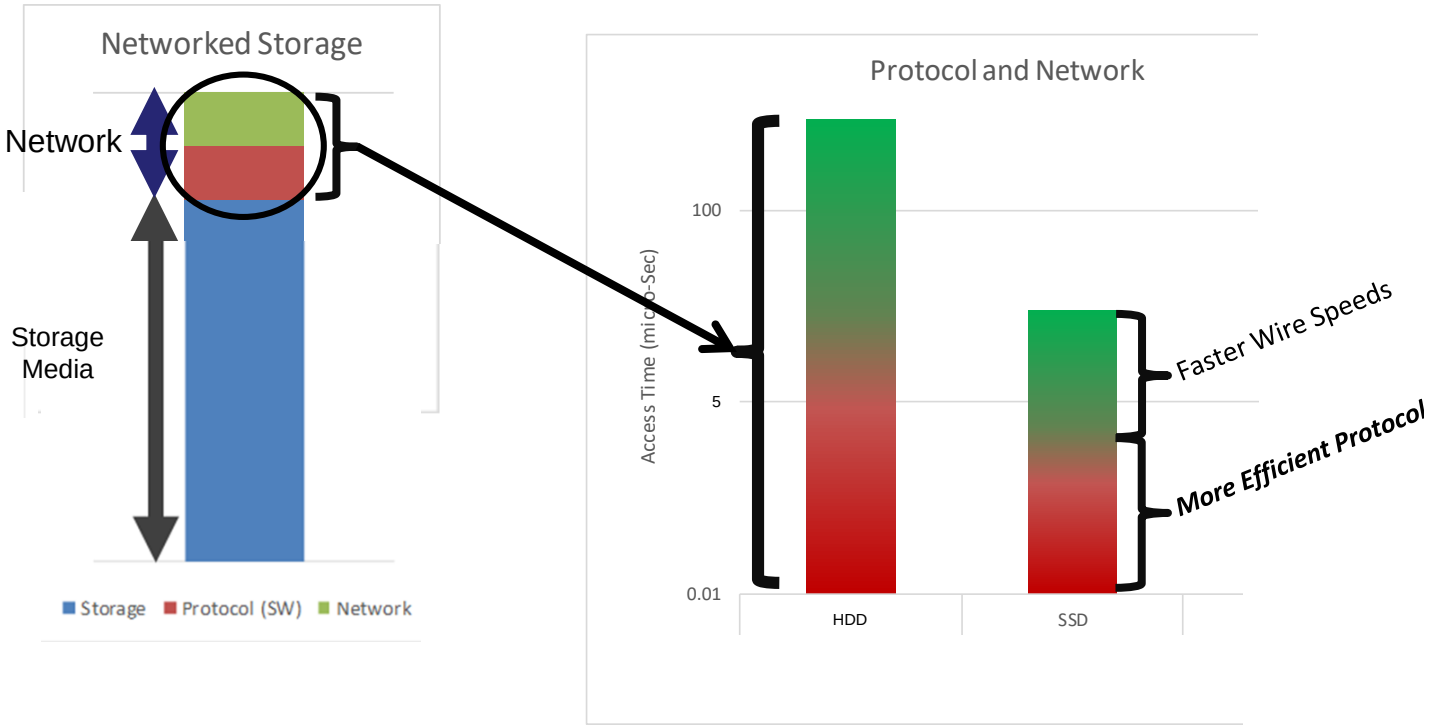






Flash SSDs move the Bottleneck from the Disk to the Network

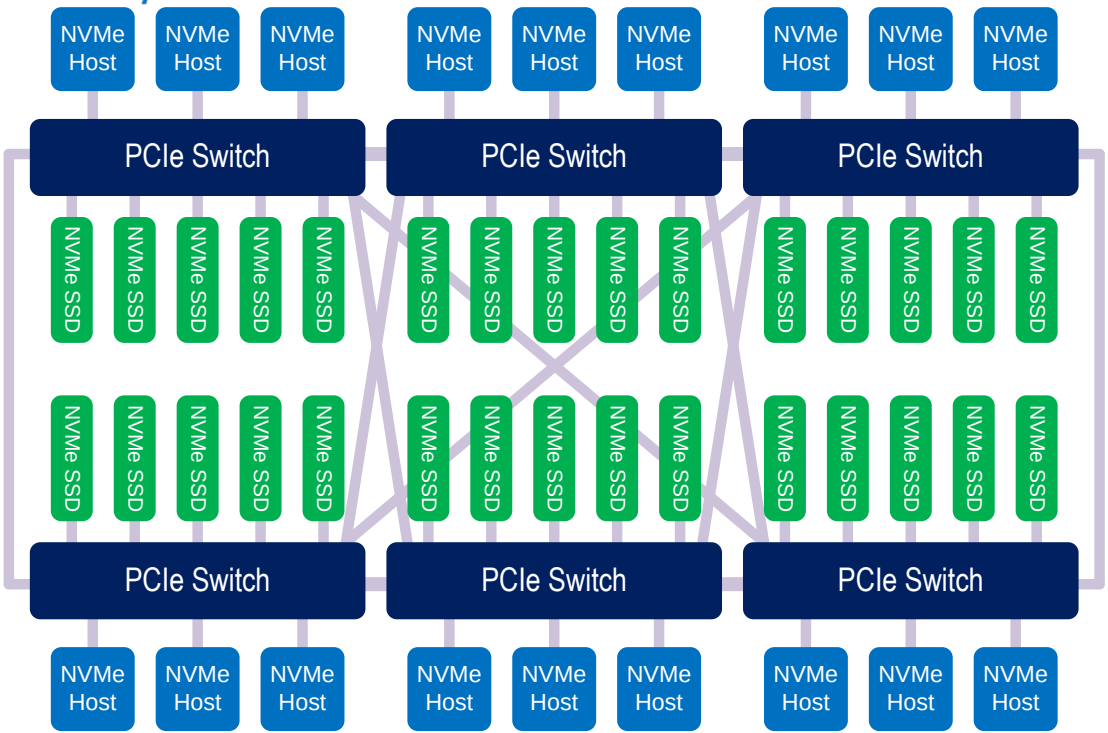
Eliminate the Bottleneck



PCIe® Fabric

Peter Onufryk
Microsemi Corporation

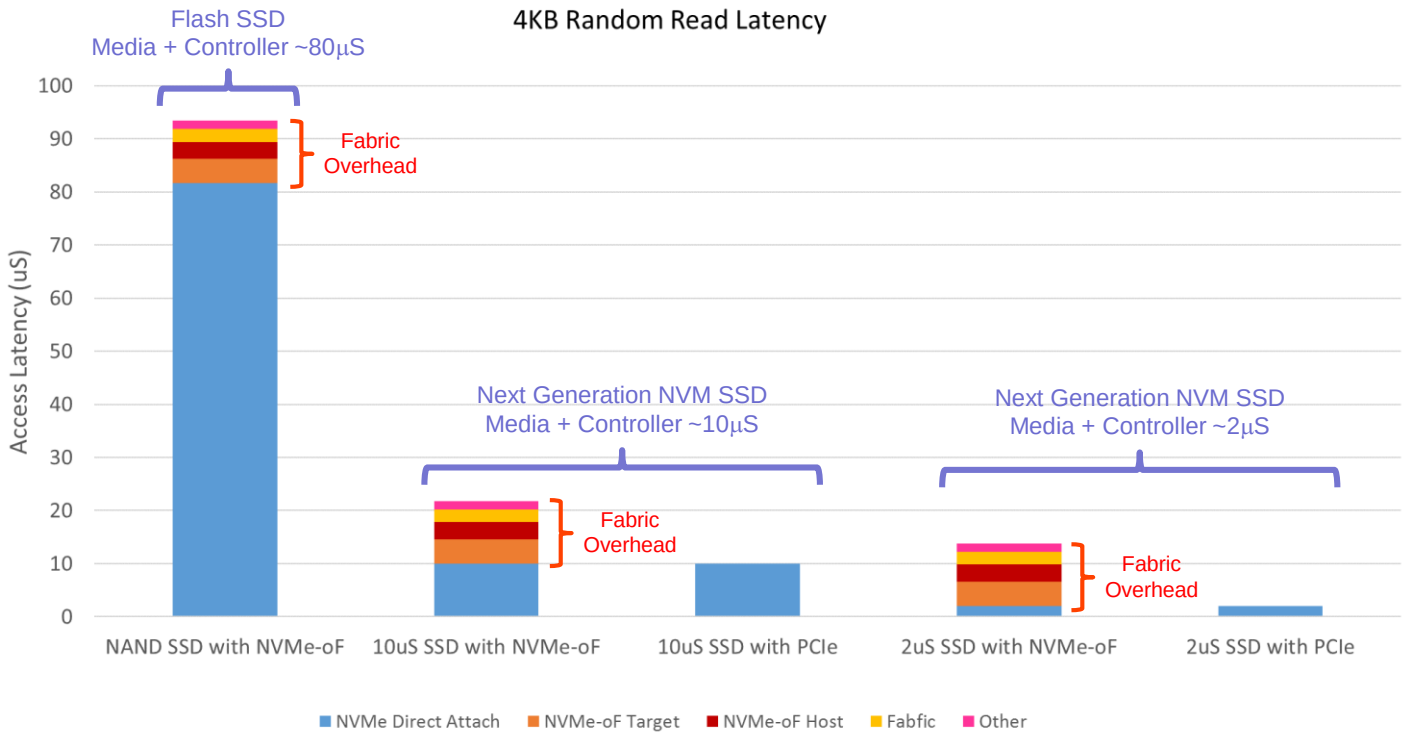
PCIe Fabric



- Storage Functions
  - Dynamic partitioning (drive-to-host mapping)
  - NVMe shared I/O (shared storage)
- Host-to-Host Communications
  - RDMA
  - Ethernet emulation
- Manageability
  - NVMe controller-to-host mapping
  - PCIe path selection
  - NVMe management
- Fabric Resilience
  - Supports link failover
  - Supports fabric manager failover

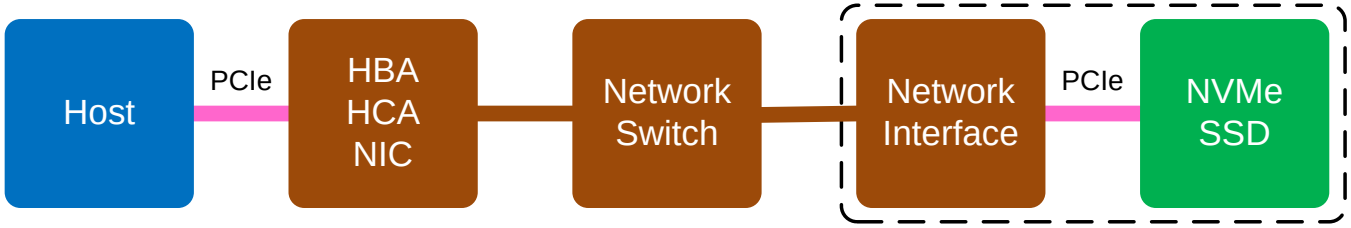
Flash Memory Summit

The PCIe Latency Advantage



Latency data from Z. Guz et al., "NVMe-over-Fabrics Performance Characterization and the Path to Low-Overhead Flash Disaggregation" in SYSTOR '17

The PCIe Advantage



Other Flash Storage Networks

PCIe Fabric

PCIe Fabric Characteristics

Property	Ideal Characteristic	PCIe Fabric	Notes
Cost	Free	Low	• PCIe built into virtually all hosts and NVMe drives
Complexity	Low	Medium	• Builds on existing NVMe ecosystem with no changes • PCIe fabrics are an emerging technology • Requires PCIe SR-IOV drives for low-latency shared storage
Performance	High	High	• High bandwidth • The absolute lowest latency
Power consumption	None	Low	• No protocol translation
Standards-based	Yes	Yes	• Works with standard hosts and standard NVMe SSDs
Scalability	Infinite	Limited	• PCIe hierarchy domain limited to 256 bus numbers • PCIe has limited reach (cables) • PCIe fabrics have limited scalability (less than 256 SSDs and 128 hosts)

InfiniBand Networked Flash Storage

Superior Performance, Efficiency and Scalability

Motti Beck – Sr. Director Enterprise Market Development, Mellanox Technologies

The Need for Intelligent and Faster Interconnect

Faster Data Speeds and In-Network Computing Enable Higher Performance and Scale

CPU-Centr (Offload)

CPU GPU CPU GPU

Onloa

CPU GPU

omputing

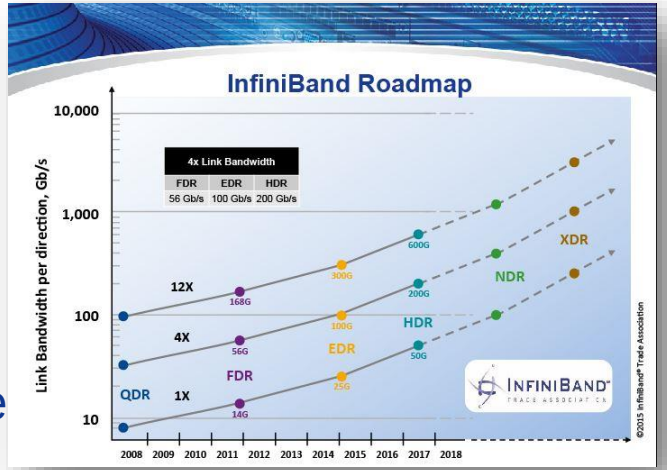
GPU CPU GPU CPU

Must Wait for the Data
Creates Performance Bottlenecks

Analyze Data as it Moves!
Higher Performance and Scale

InfiniBand Technical Overview

- What is InfiniBand?
  - InfiniBand is an open standard, interconnect protocol developed by the InfiniBand® Trade Association: http://www.infinibandta.org/home
  - First InfiniBand specification was released in 2000
- What does the specification includes?
  - The specification is very comprehensive
  - From physical to applications
- InfiniBand SW is open and has been develope
  - http://www.openfabrics.org/index.html



Flash Memory Summit

RDMA enabled Networking Powers Modern Storage Platforms

DataDirect NETWORKS

dataON DataONSTORAGE

DELL EMC

FUJITSU

Hewlett Packard Enterprise

ORACLE

Microsoft

Micron

NetApp

NIMBUS DATA

IBM xiv tms

SEAGATE

TOSHIBA

TERADATA

Western Digital

VIRIDENT

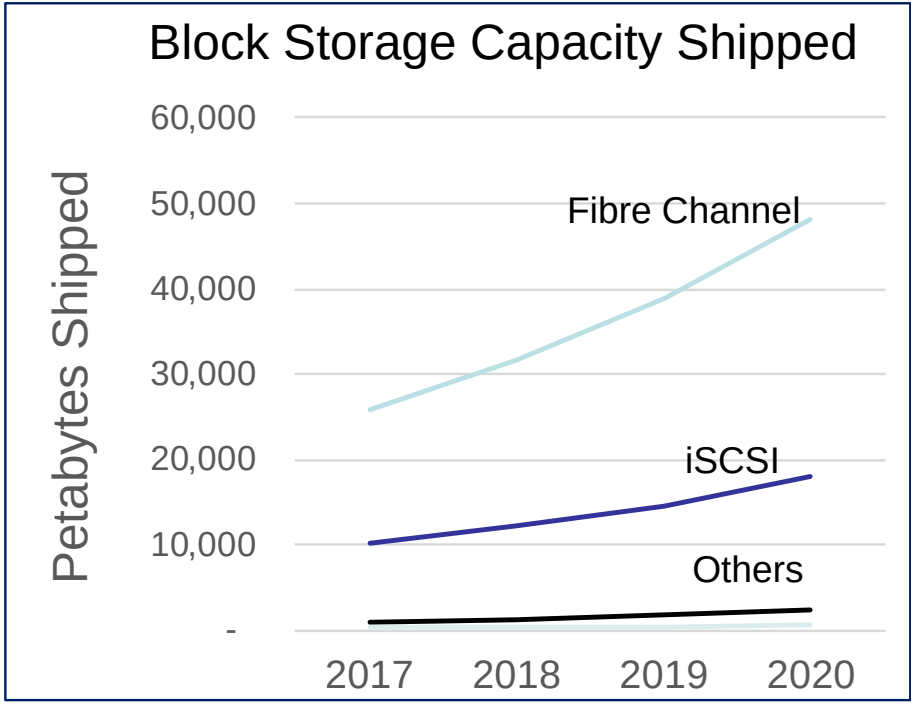
Higher Performance, Higher Efficiency and Higher Scalability

NVMe over Fibre Channel

Brandon Hoff
Principal Software Architect
Emulex Connectivity Division, Broadcom Inc.

Enterprise Storage Market

- Fibre Channel storage shows strong growth in capacity
  - Fibre Channel Storage capacity shipped is larger than all other types of external storage combined
- The adoption of All Flash Arrays and NVMe storage will drive the need for faster networks
- iSCSI is the dominate technology block over Ethernet
- Logical transition
  - SCSI FCP transitions to NVMe/FC
  - iSCSI transitions to NVMe/TCP

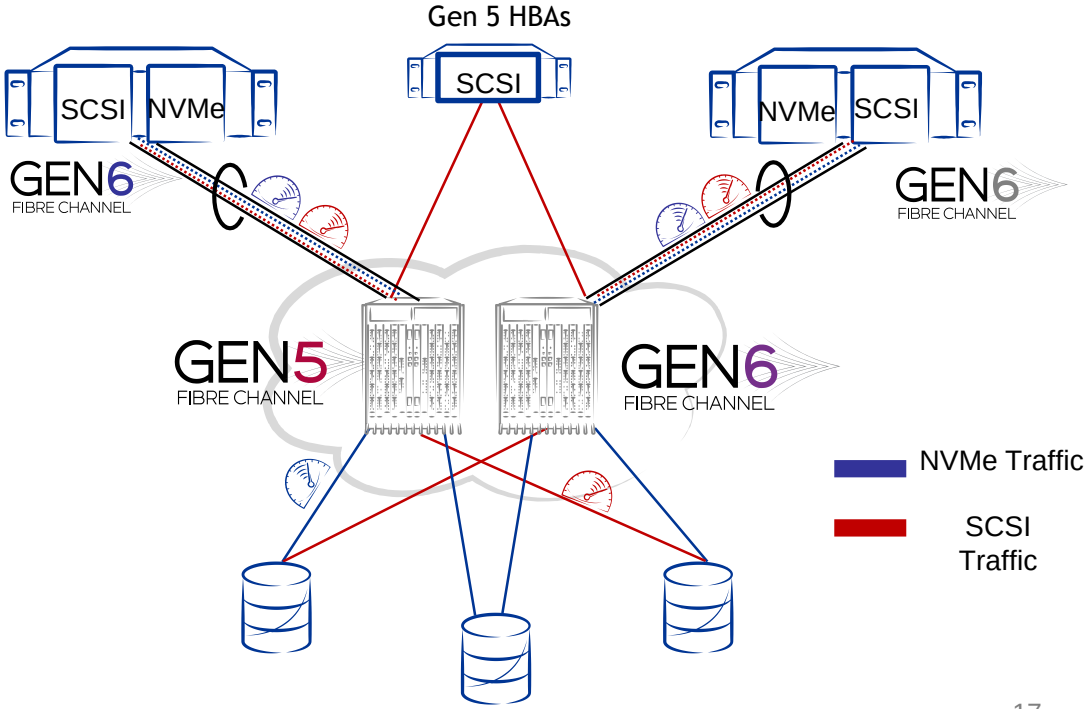


Other Includes: FICON, FCoE, Infiniband, External SAS

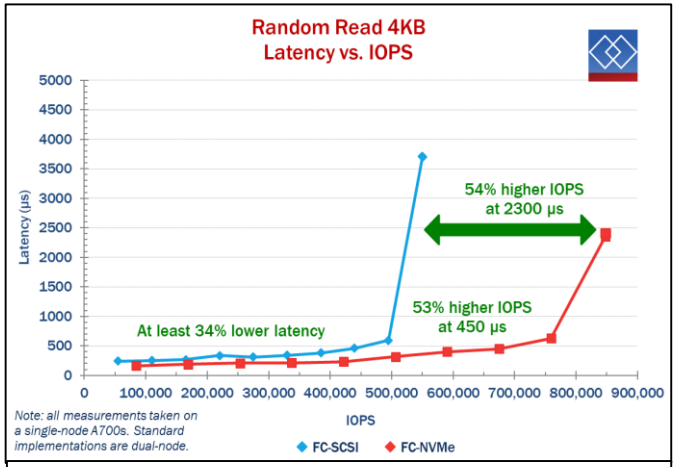
IDC WW Capacity Shipped, 2016

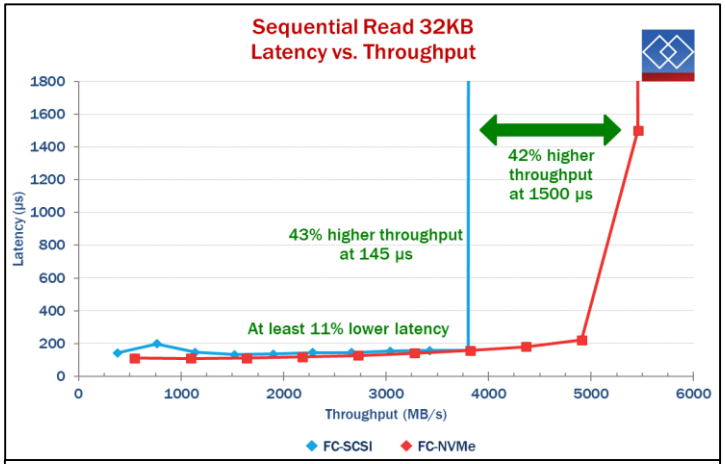
Dual protocol SANs enable low risk NVMe adoption

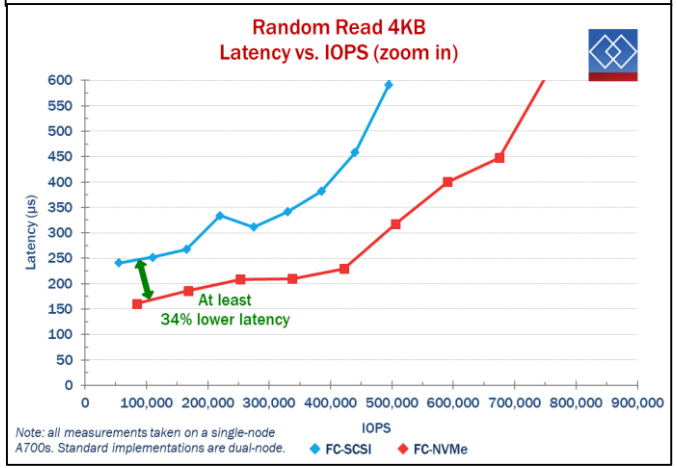
- Get NVMe performance benefits while migrating incrementally “as-needed”
- Migrate application volumes 1 by 1 with easy rollback options
- Interesting dual-protocol use cases
- Full fabric awareness, visibility and manageability with existing Brocade Fabric Vision technology

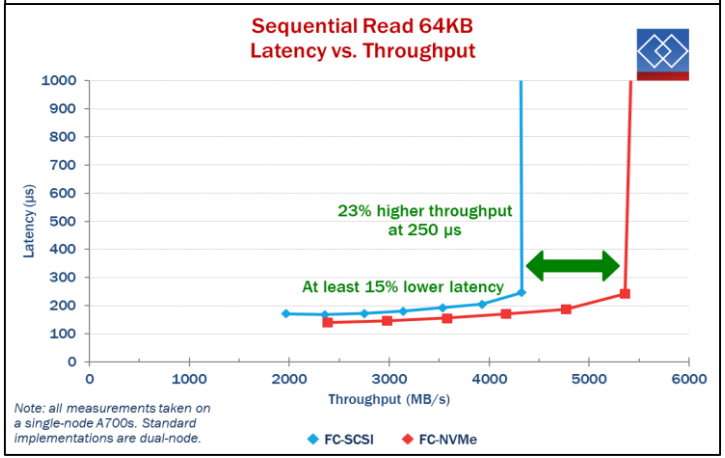


NVMe over Fibre Channel Performance on a A700s single node









Ethernet-Networked Flash Storage

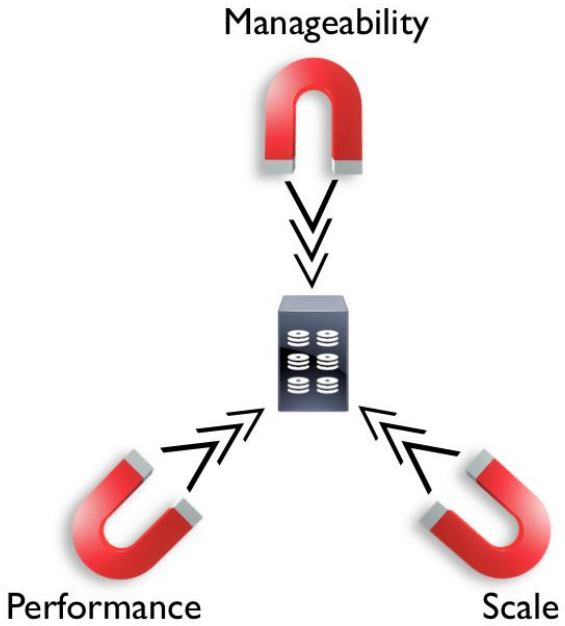
J Metz, Ph.D

R&D Engineer, Advanced Storage

Cisco Systems

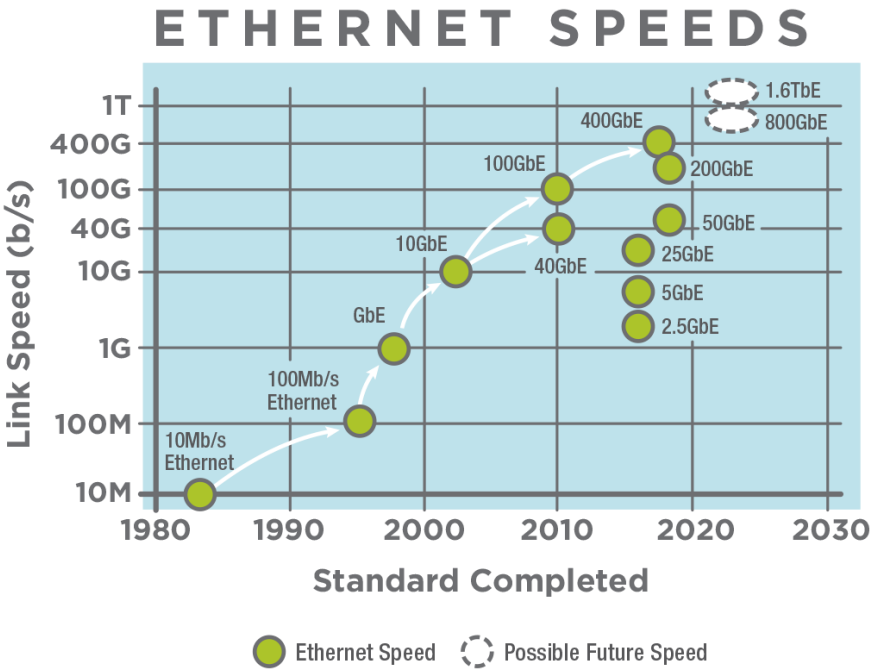
@drjmetz

Storage Perspective



- There is a “sweet spot” for storage
  - Depends on the workload and application type
  - No “one-size fits all”
- What is the problem to be solved?
  - Deterministic or non-deterministic?
  - Highly scalable or highly performant?
  - Level of manageability?
- Understanding “where” the solution fits is critical to understanding “how” to put it together

Ethernet Roadmap



- How to go faster
  - Different modulation techniques
  - Different data rate/lanes chosen
- New Signaling methods
  - Pulse Amplitude Modulation 4 vs. Non Return to Zero (NRZ)
- New Form Factors
  - Multi-lane interfaces

Source: ethernet alliance

Summary

- Ethernet
  - General Purpose network designed to solve many, many problems and do it well
  - Flexible for all but the most extreme conditions
  - Largest ecosystem of developers, vendors, and users
  - From the smallest system to the largest, there is no other networking technology more suited, or best understood