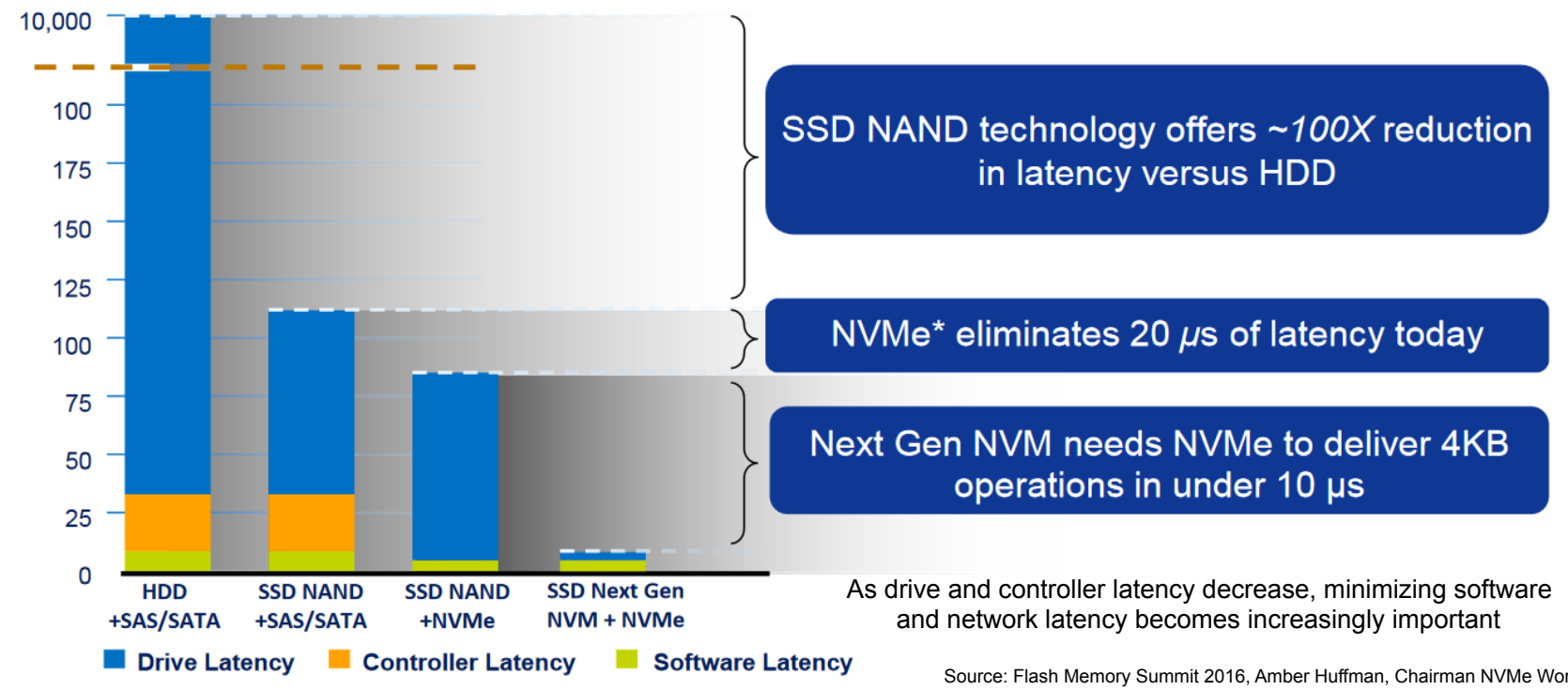
Storage Networking

Ethernet Offers the Speed and Affordability Required for Storage Networking

Presenter: David Iles

August 2018

Faster Storage Needs Faster Networks



As drive and controller latency decrease, minimizing software and network latency becomes increasingly important

Source: Flash Memory Summit 2016, Amber Huffman, Chairman NVMe Working Group

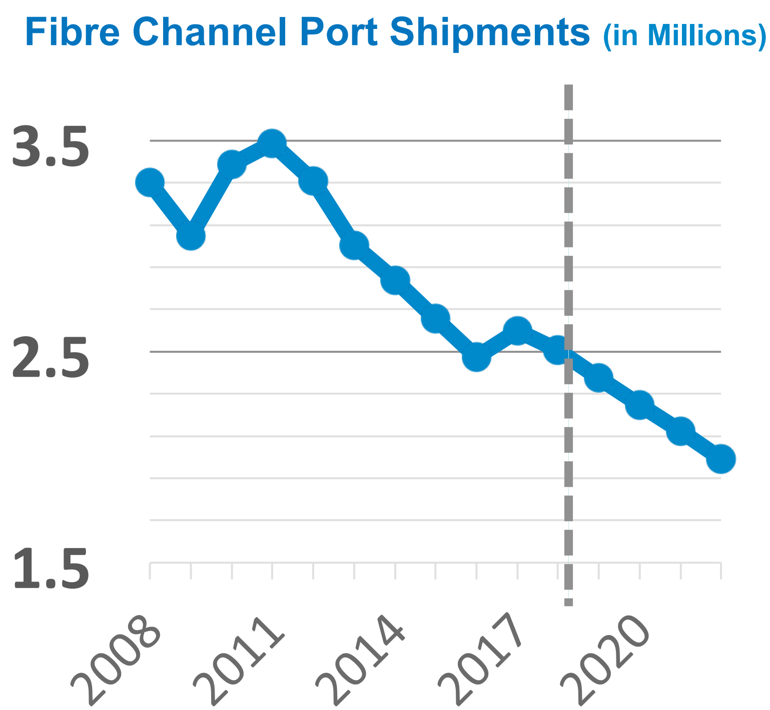
The Storage World is Changing

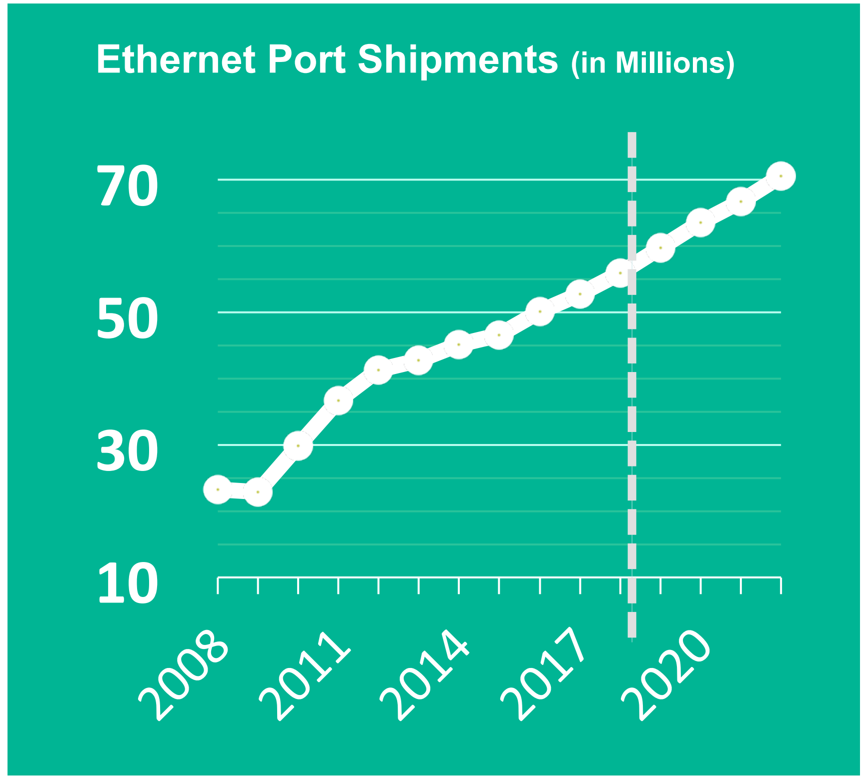
Changes	Effects
Flash, Faster servers	**Faster networking;** 10/25/40/50/100GbE
Social/Mobile/Video	**Huge data growth;** more file and object content
Hyperconverged	**Distributed "Server-SAN"** on Ethernet
Cloud	**Virtualization**, software-defined, price pressure
Big Data	**File and distributed** storage
Distributed applications	**More** east-west traffic

Bottom Line: More Ethernet Storage Traffic

Storage & Connectivity Evolution

Flash Memory Summit





Source: Crehan Research, Host Adapter Port Shipments, January 2018

Storage Networking Trend

1997

Feature	Fibre Channel	Ethernet
Bandwidth	1 G	100 M
Supports	Block	Block, file
Lossless	Yes	No
Cost	High $$$$	Medium $$
Cloud / HCI	No / No	No / No
Vendors	Several	Many
SDS / Scale-out	No / No	No / No

Yesterday: Storage Network = FC

- Fibre Channel offered best performance
- All interesting storage was tier-1 block
- No cloud or hyperconverged

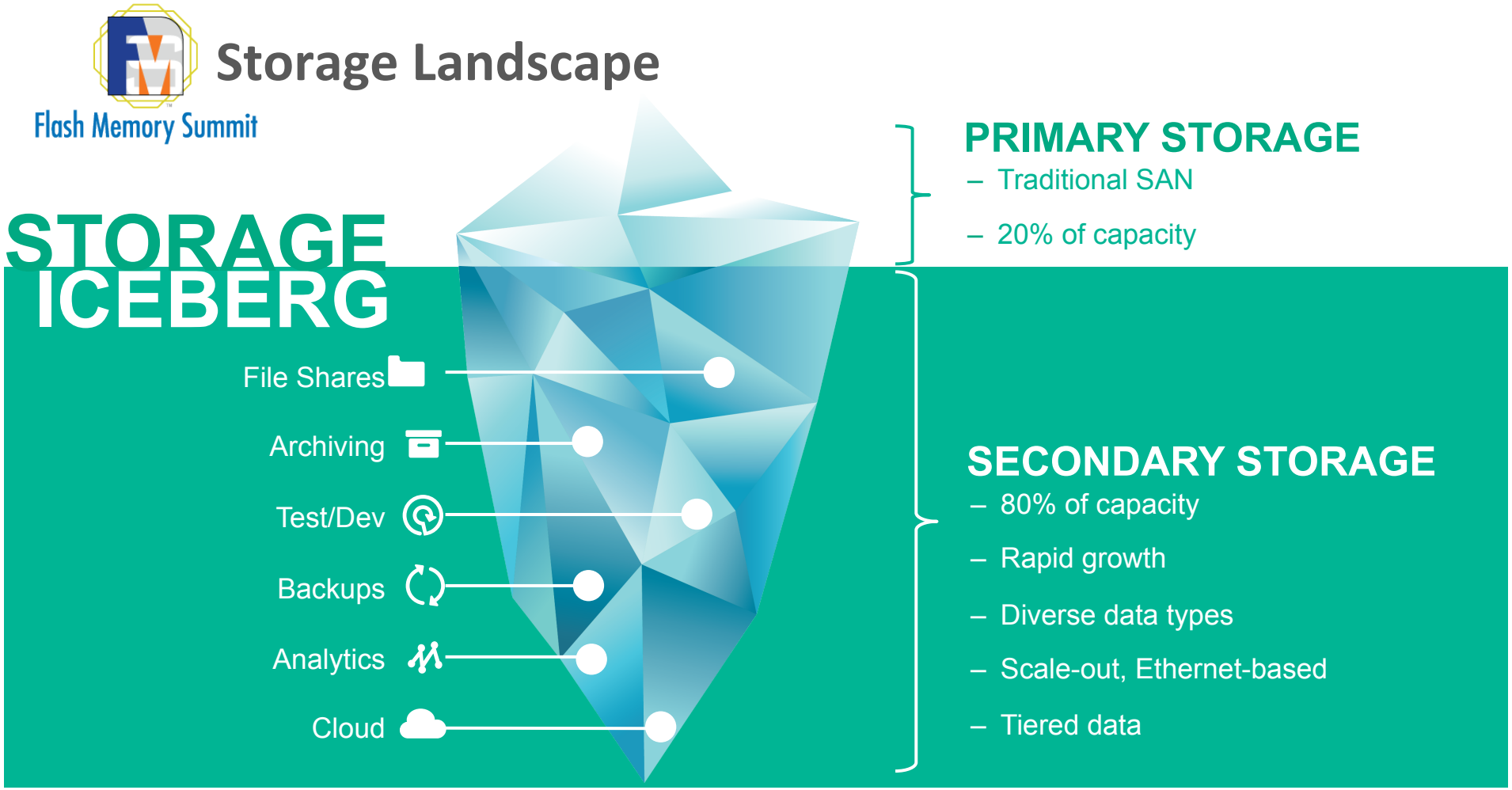
2017

Feature	Fibre Channel	Ethernet
Bandwidth	8/16/32 G	10/25/40/100 G
Supports	Block	Block, file, object
Lossless	Yes	Yes
Cost	Medium $$	Low $
Cloud / HCI	No / No	Yes / Yes
Vendors	2 / 2	Many / Many
SDS / Scale-out	Rare / No	Yes / Yes

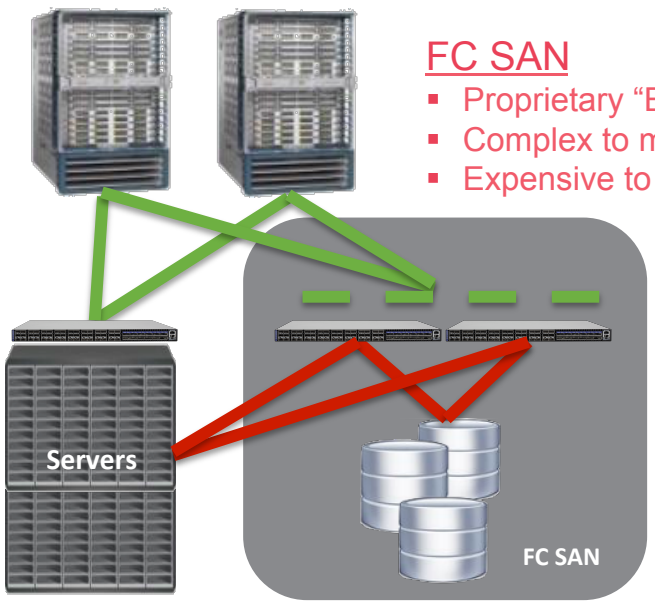
Today: Both FC & Ethernet for storage networks

- FC option for Primary Block Storage
- Ethernet only option for all Primary & Secondary Storage

(Block, Object, NAS, Cloud, Hyperconverged, Big Data)


Flash Memory Summit
Storage Landscape
STORAGE ICEBERG
File Shares
Archiving
Test/Dev
Backups
Analytics
Cloud
PRIMARY STORAGE
– Traditional SAN
– 20% of capacity
SECONDARY STORAGE
– 80% of capacity
– Rapid growth
– Diverse data types
– Scale-out, Ethernet-based
– Tiered data

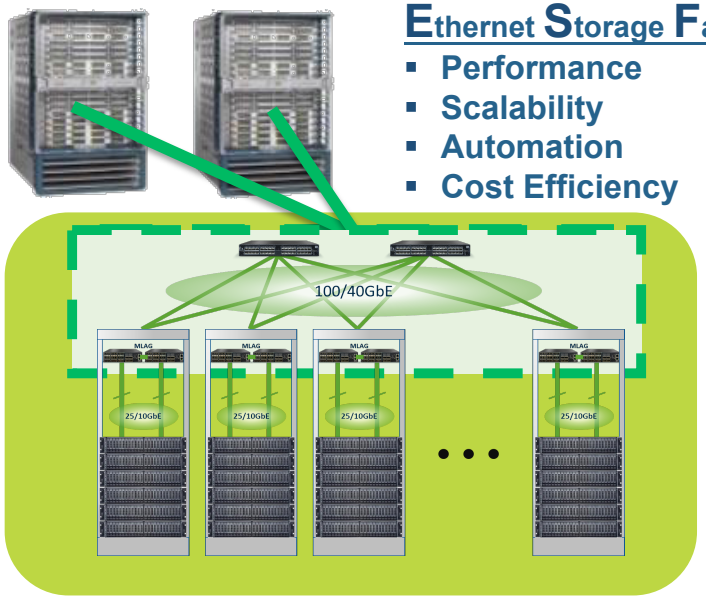
Modern Datacenter Storage Networking



FC SAN

- Proprietary "Big Box"
- Complex to manage
- Expensive to scale

Legacy DC – FC SAN



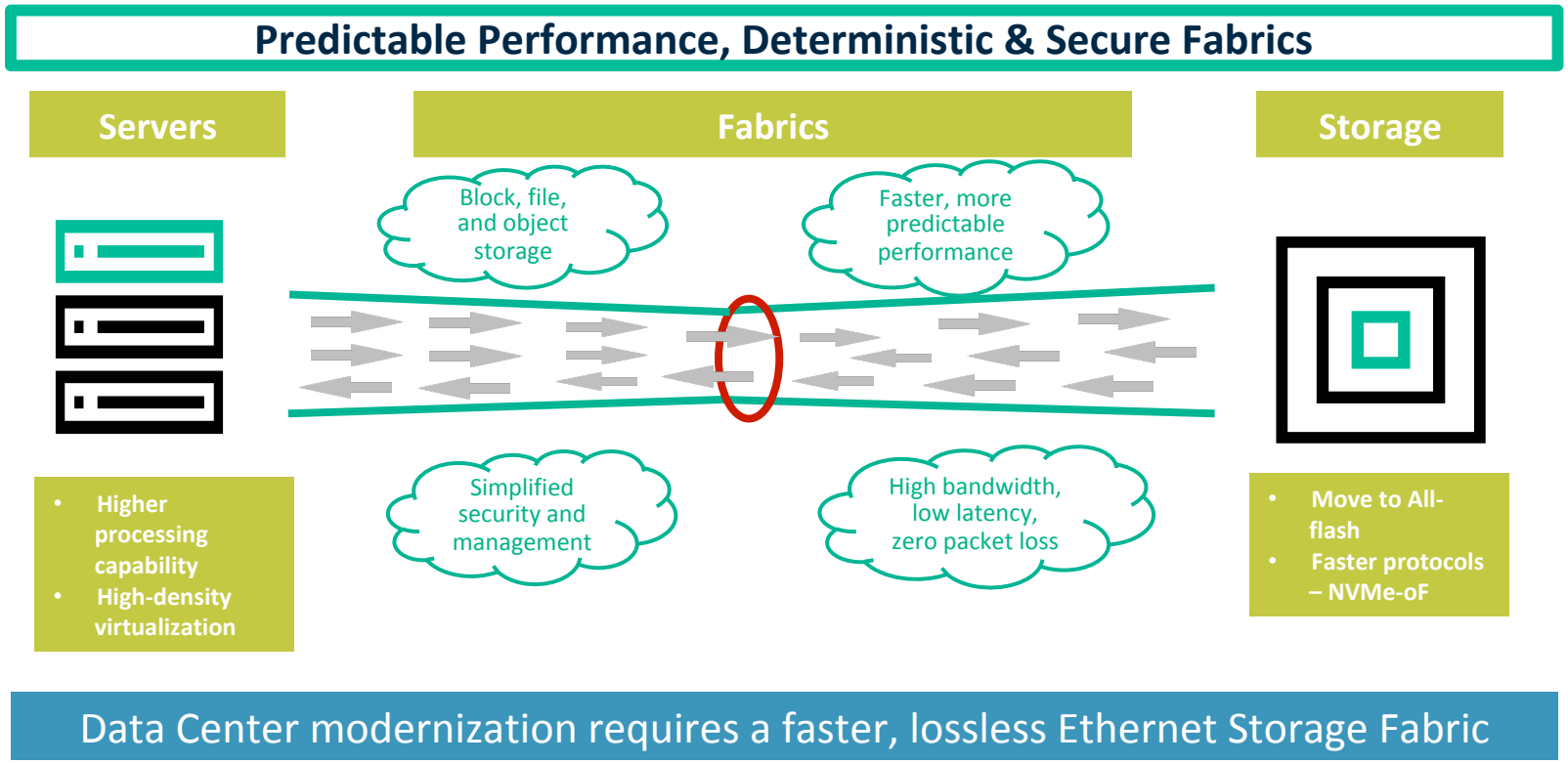
Ethernet Storage Fabric

- Performance
- Scalability
- Automation
- Cost Efficiency

Modern Datacenters/Clouds

Flash Memory Summit

Re-defining Modern Data Centers

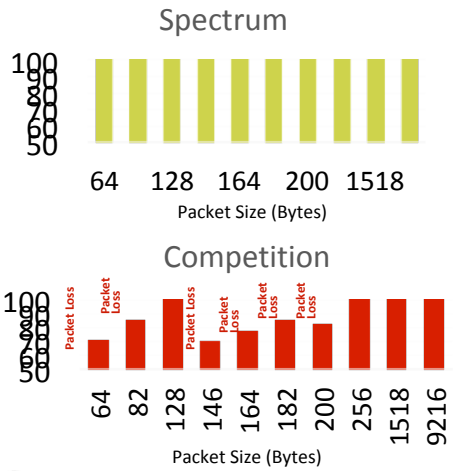


Data Center modernization requires a faster, lossless Ethernet Storage Fabric

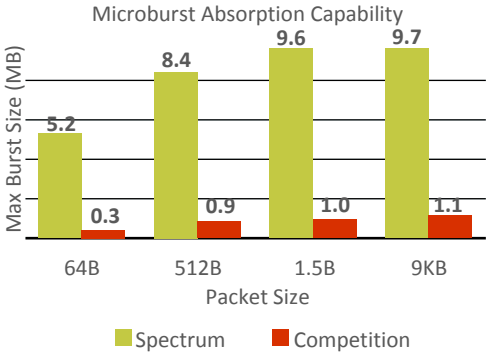
Not All Switches Were Created Equal

Flash Memory Summit

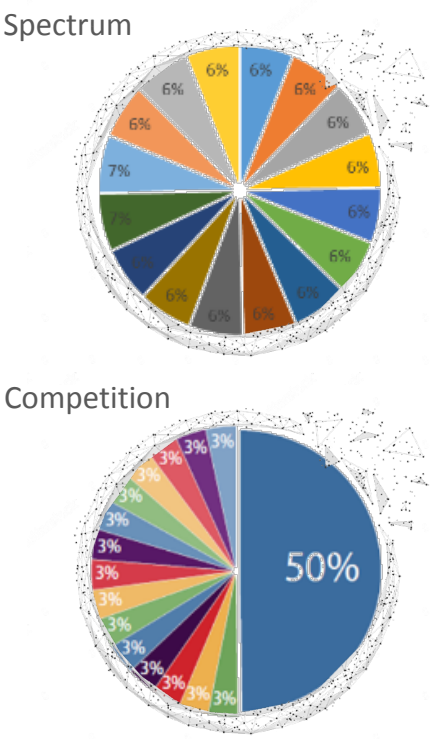
Avoidable Packet Loss



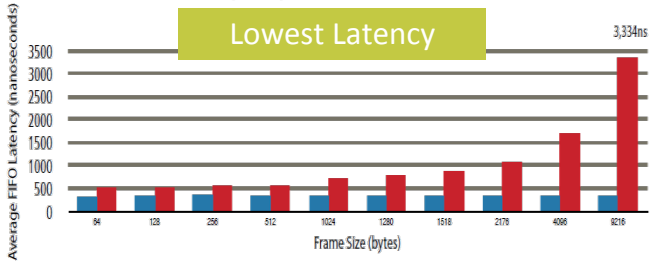
Congestion Management



Fairness & QoS



Lowest Latency



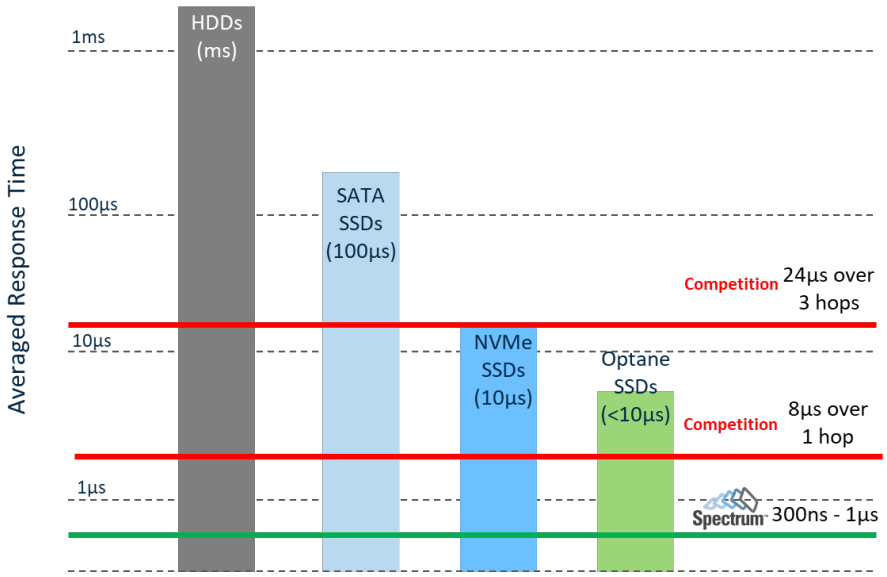
TEST REPORT

Tolly.

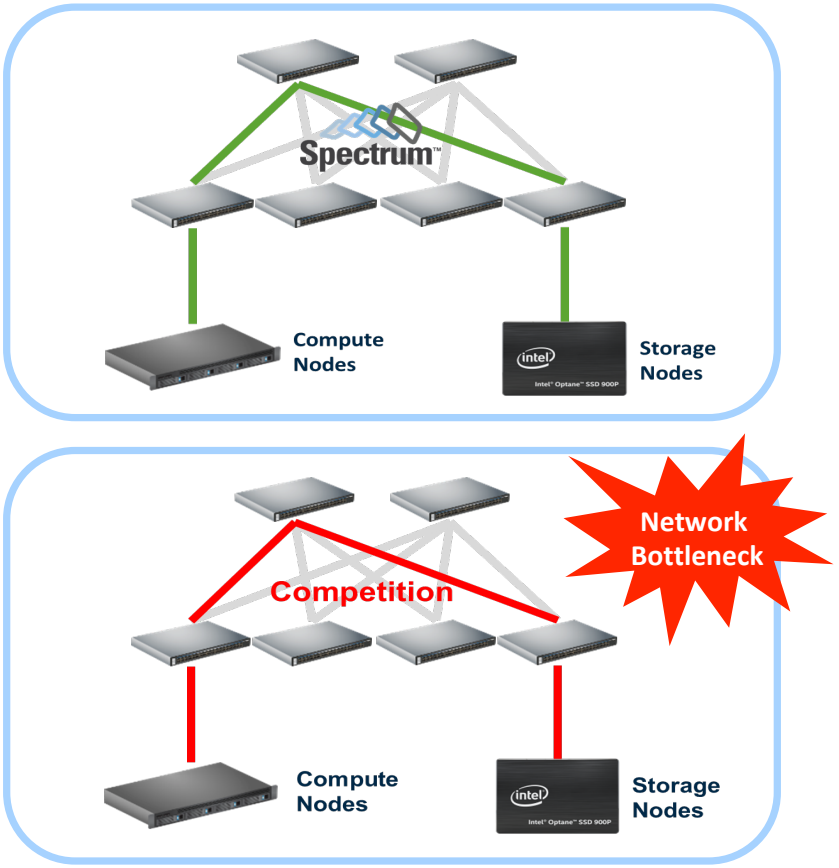
www.zeropacketloss.com

Unlock the Maximum Flash Performance

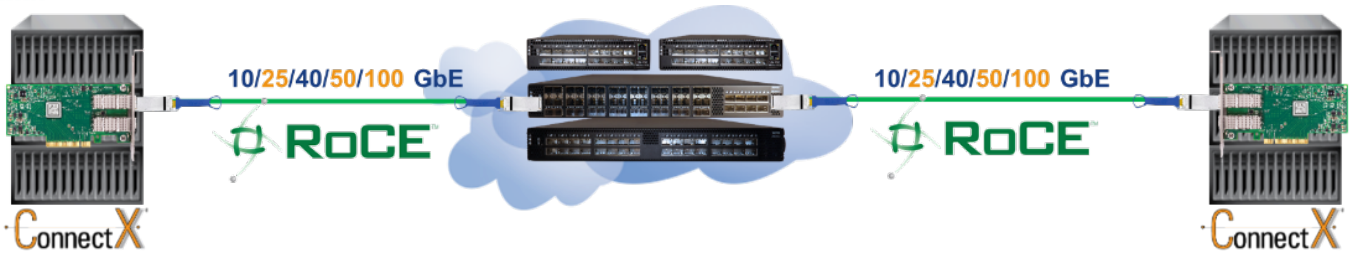
Flash Memory Summit



Storage is Getting a Lot Faster!



End-to-End RoCE Acceleration



Spectrum™

- Zero packet loss, line-rate performance at all packet sizes and port combination
  - 30% loss in competitive solution
- Predictable buffer allocation to any port & packet sizes
  - Competitive variance spreading by ~600%
- Low latency, up to 90% latency in a typical TOR deployment scenario

ConnectX®

- Highest performance and lowest latency
- Hardware RDMA offload
- Hardware offload of RoCE congestion control
- Hardware offload of data path and NVMe command offload

RoCE Enabled Storage

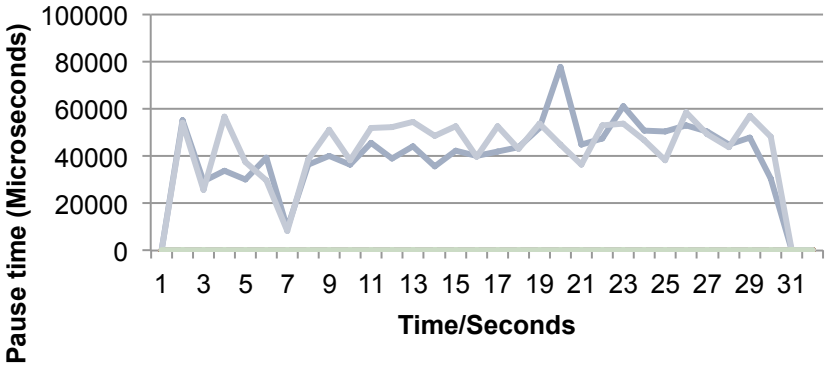
- **iSER**
- **NVMe-oF**
- **Microsoft SMB 3.0**
- **vSphere 6.5**
- **Ceph**
- **Spark**

Flash Memory Summit

NVMe-OF Benchmarks

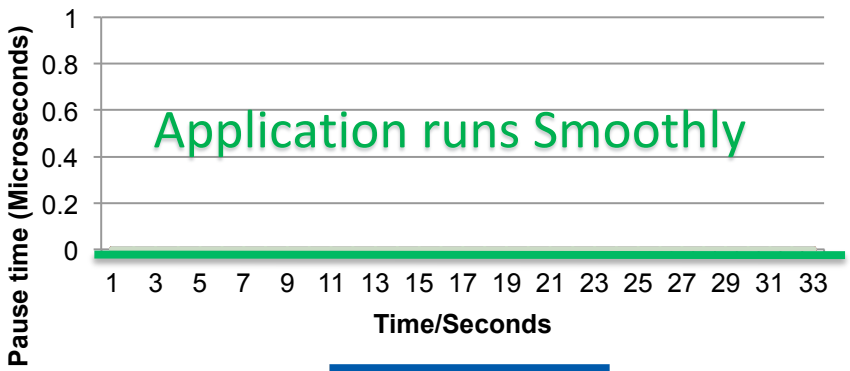
RoCE™

Application Blocked by the Switch



Other Switches

Application Blocked by the Switch



Mellanox TECHNOLOGIES

Storage-Optimized Ethernet Switches

Ethernet Storage Fabric needs dedicated storage switches

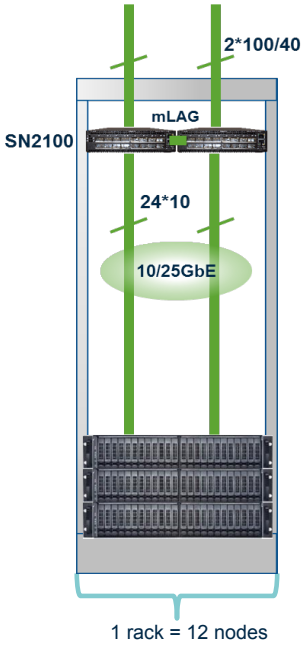
- Performance
- High Availability
- Simple
- Automated
- Scalable
- Cost Efficient

Spectrum™

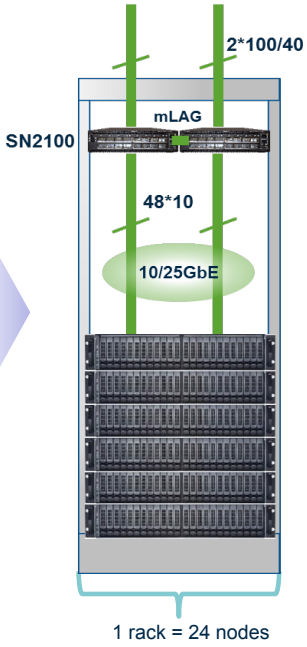
- ✓ **2 Switches in 1RU**
- ✓ **Storage/HCI port count**
- ✓ **Zero Packet Loss**
- ✓ **Low Latency**
- ✓ **RoCE optimized switches (NVMe-oF)**
- ✓ **NEO for Network automation/visibility**
- ✓ **Native SDK on a container**
- ✓ **Cost optimized**
- ✓ **NOS alternatives**

Mellanox ESF is Easy to Scale

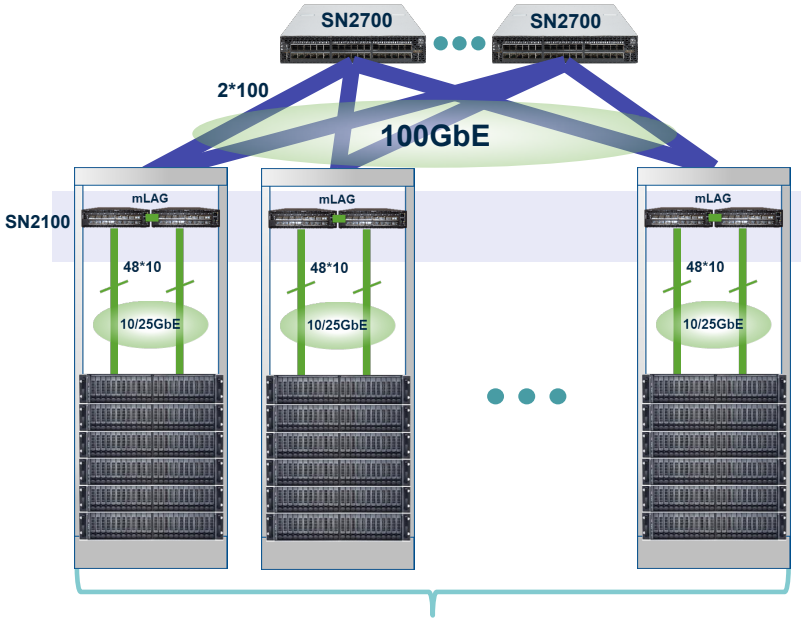
Half Rack



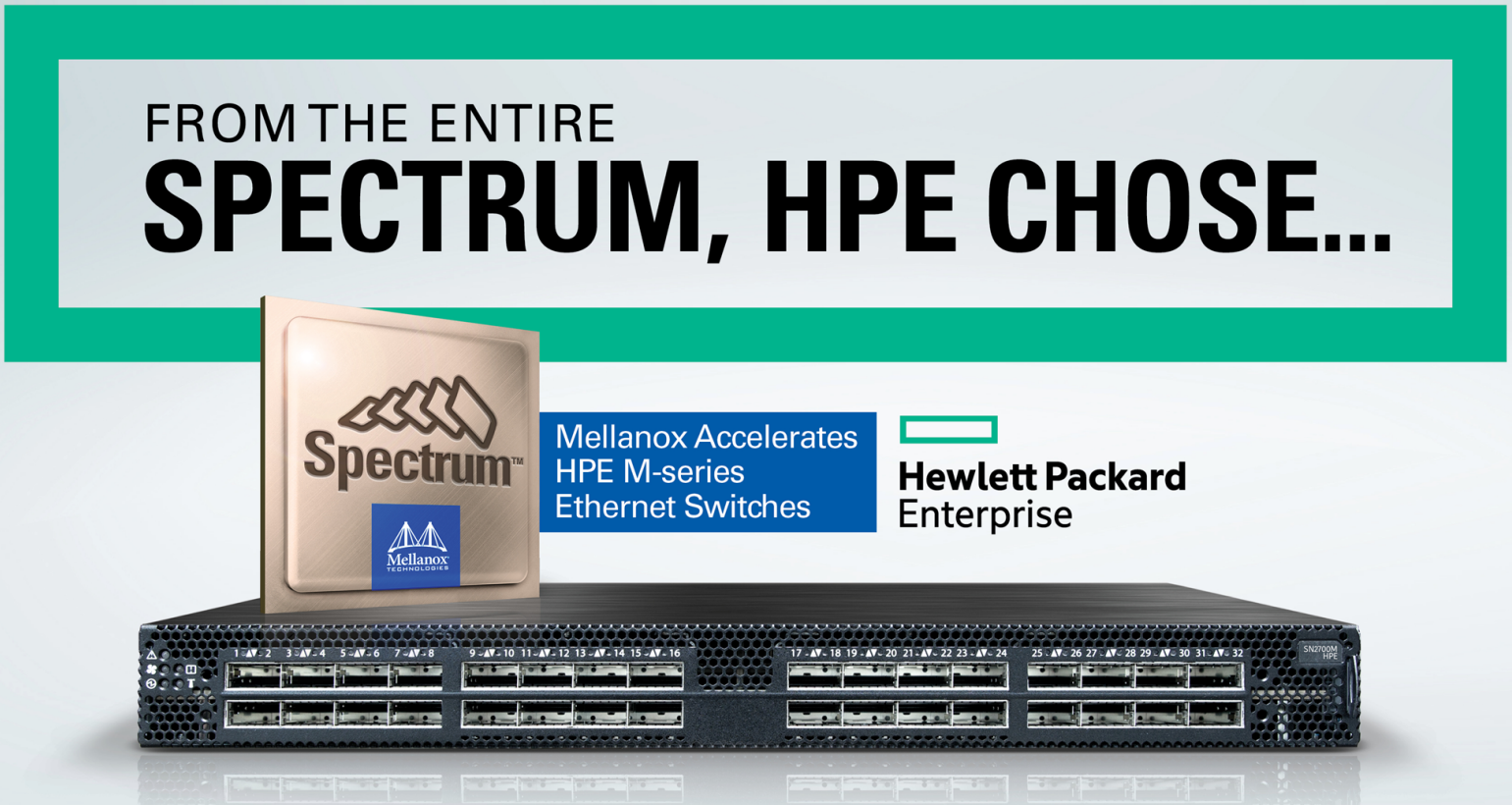
Full Rack



Multiple Racks



Mellanox Powers HPE Storage Fabric



Source: HPE Chooses Mellanox Spectrum™ To Power StoreFabric M-series Switches

Additional Information

- Nutanix Solution Note
- NEO automated network provisioning for Nutanix AHV VM operations
- Accelerate RedHat Ceph Storage
- Microsoft S2D Performance with 100GbE Mellanox solution
- Microsoft SQL Performance with 100GbE Mellanox solution
- VMware vMotion with 40GbE Mellanox RoCE
- Micron SolidScale™ NVMe platform
- E8 high-performance NVMe storage
- Excelero “NVMesh” software-defined NVMe flash platform
- Scale Computing config guide

- Selected Solution Briefs

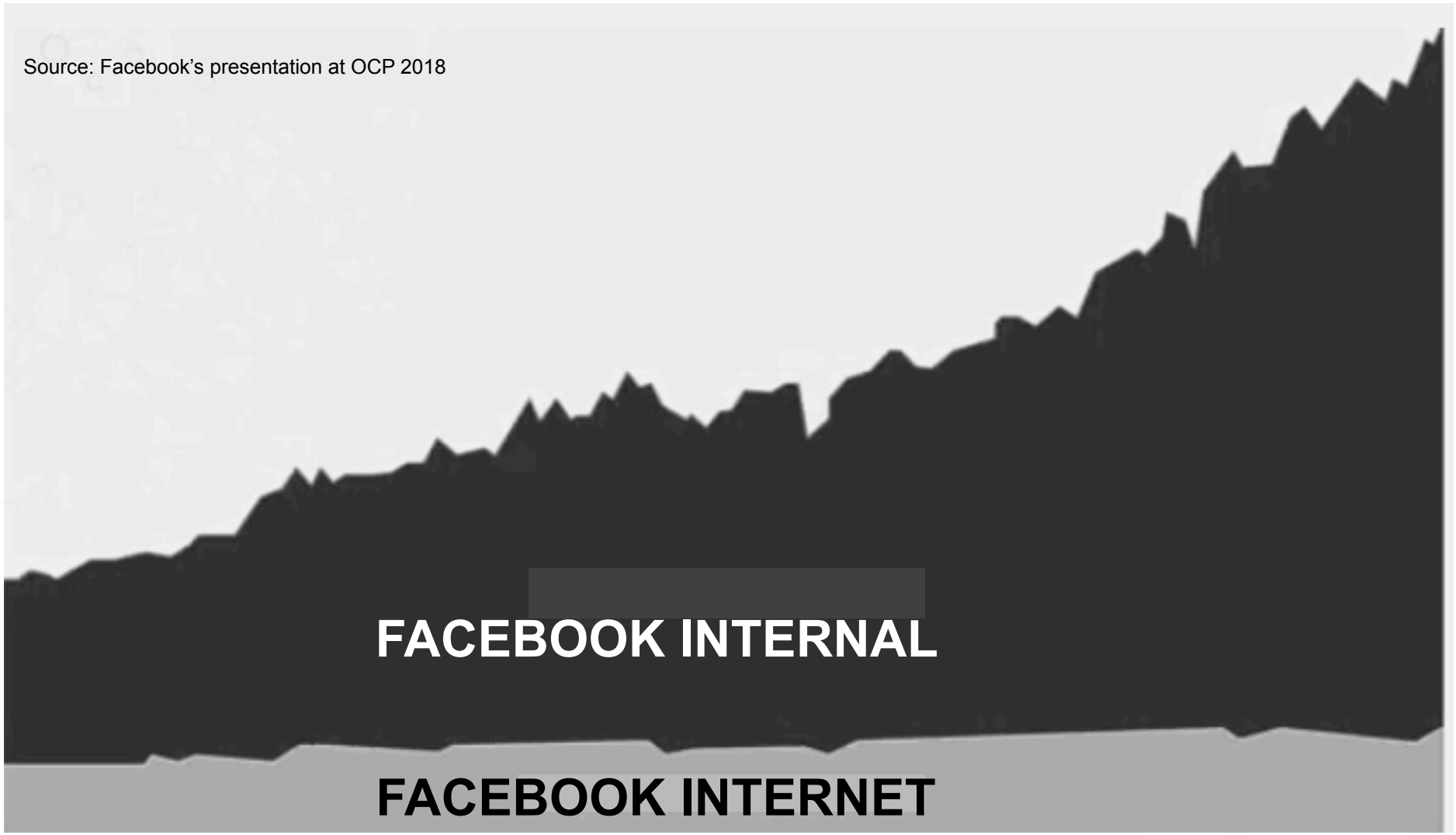
CONCURRENT dataON NUTANIX SCALE COMPUTING ZADARA STORAGE

- Technical References
  - Get started with RoCE configuration
  - Understanding RoCE v2 Congestion Management
  - Bring up Ceph RDMA – Developer’s Guide
  - How to configure NVMe-OF

Thank You

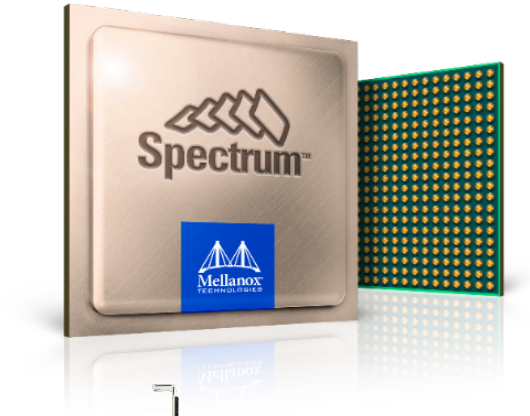
Backup slides

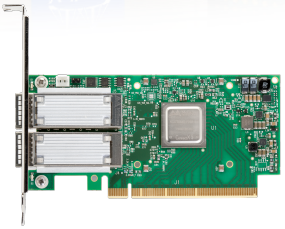
Source: Facebook's presentation at OCP 2018



Summary: Mellanox ESF Switches

Flash Memory Summit

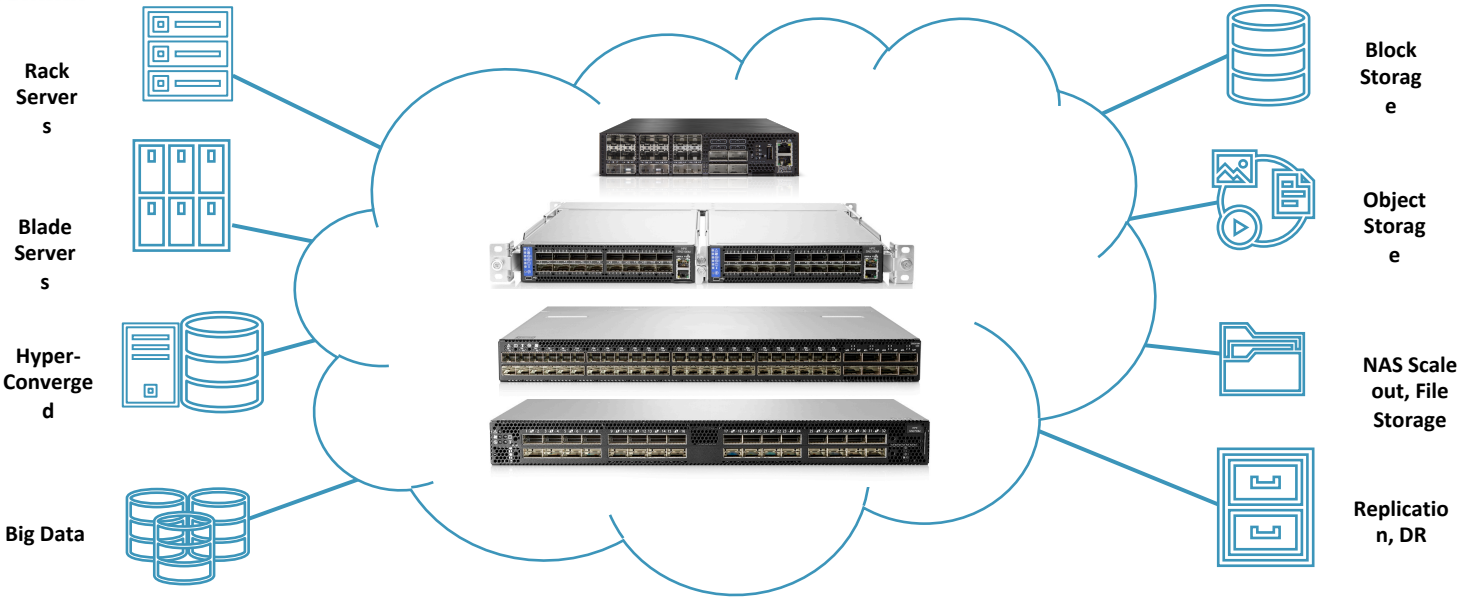




- Better Performance
- Enough ports in 1RU
- Easy Setup
- Better Visibility
- Tested End-2-End

ESF: Storage Networking Done Right

Flash Memory Summit



Deliver Performance and Efficiency for Scale-out Storage and Hyperconverged Infrastructures

http://www.mellanox.com/ethernet-storage-fabric/