NVMe-oF™ Enterprise Appliances

Sponsored by NVM Express® organization, the owner of NVMe™, NVMe-oF™ and NVMe-MI™ standards
Panelists

Kamal Hyder
Director of PLM
KumoScale
Toshiba Memory

Manoj Wadekar
Director HW Engineering
eBay

Yaniv Romem
CTO & Co-Founder
Excelero

Nishant Lodha
Product Marketing Manager
Marvell (Cavium)

Moderator:
Jeremy Werner, Sr. VP and GM SSD
Business Unit, Toshiba Memory America
Architected for Performance

NVMe™ Enabling the Future
Powering the next generations of storage

Kamal Hyder, Toshiba Memory America, Inc.
NVMe™ Excitement Continues!

- New Protocol, Exclusively for Flash
- Multiple Fabrics: RDMA, FC, TCP
- End-to-end support
- Native OS support
- Growing interest in Disaggregated Flash
- Suitable for Enterprise and Cloud Data Center Architectures
- Lowest Latency, Highest Performance ever! Storage no longer the Bottleneck
- Greenfield and Existing Environments
- Initiators to Switches to Targets
- Linux Kernel 4.9+, others in progress
- Multiple Vendors Supporting the Concept
- Bringing High Performance to Multiple Areas
We’ve Seen the Consolidation Movie Before

Homogenous “SAN Islands”

Consolidation over Multiprotocol Networks

Services at Multiple Layers – Fabrics and Storage

Phase I: Isolated Storage Islands

Tier1 Business Applications

Midrange Applications

VSANs

Multi-protocol

QoS

ACL

Mgmt

Scalability

Tier1 Business Applications

Midrange Applications

Host Services

Fabric Services

Disk Services

Array Services

Service Classes

Pooled Disk and Tape

Pooled Disk and Tape

Phase II: Consolidation/MultiProtocol Transport

Phase III: Differentiated Services
ebay Hyperscale Infrastructure

- Search
- Hadoop
- Databases
- Deep Learning AI
- Object Store
- Front-End & Product
Typical Hyperscale Servers: Design Goals

**Efficiency:** Utilization, commonality

**Growth:** Performance, Capacity
What’s needed: Rack-As-A-Compute

Node Local Storage

Rack Local Storage

CPU
CPU
CPU

Ethernet
Rack-As-A-Compute

Right Sizing:
- Clusters can use optimized ratio of compute and storage.
- Allows reducing wastage and improve performance

Independent Scaling:
Compute and storage capacities can be scaled per need
Distributed NVMe™ Architectures

Yaniv Romem, Excelero
How is flash deployed today?

All Flash Array

- Separate application servers & scale-out flash appliances
- Share capacity & performance across applications
- Fabric/Network hop involved

In Server SSDs

- Application, CPU & Flash in one appliance
- Capacity & Performance cannot be shared among isolated appliances
- Applications can take full advantage of NVMe™ performance
NVMe™ flash: So Many IOPs, So Much Bandwidth…

• NVMe solid state drives offer so much performance, one server struggle to make efficient use of a fully stuffed server

• This makes architectural choices even more important
  • Connectivity choice can impact performance

• Shared-nothing architectures have benefits
Distributed NVMe deployment options

**Local Shared Storage in Application Servers**
- Storage is unified into one pool
- Target Module & Client Block Driver run on all nodes
- Linearly scalable

**Storage is Centralized**
- Storage is unified into one pool
- Target Module runs on storage nodes
- Client Block Driver runs on server nodes
- Applications get performance of local storage
Hyperscale Challenges

Challenges for web-scale applications

• Maximize operational efficiency and architectural flexibility
• Achieve rigorous business objectives: 100% uptime, low TCO
• Meet complex application requirements: scalability, performance
• New application workloads such as real-time analytics and AI make hyper-scale challenges more onerous

Benefits of Converged Architectures

• SDS on standard servers enables hardware homogeneity
• Maximum utilization of NVMe™ SSD’s by creating a single pool of high-performance block storage
• No data localization for scale-out applications
• Can achieve predictable application performance – no noisy neighbors
The “well-connected” NVMe™!

Nishant Lodha, Marvell (Cavium)
What Do You Mean “Well-Connected” NVMe™?
Ethernet Speeds and Feeds!

Server Speed Transition in Enterprise

Server Speed Transition in Cloud

Server shipments into cloud continue while 25G/50G adoption grows

Ethernet Switch – Port Shipments (All)

Ethernet Switch - DC Port Shipments

100G started getting deployed at Scale in Data Centre

Source: Del’Oro Research, 2017
Trending all around the DC!

**Smart NICs** recognized as new adapter category

**Disaggregated Storage (AFA, JBOF)**

**SmartNICs**  

**RegularNICs**

**Industry embraces Open architectures**

**Rise of SDN/NFV** in Telco Cloud and birth of **Edge Compute**

**OCP 3.0 Mezz.**

**Software Defined Data Center (SDDC)**

**Emergence of Hybrid Cloud & secure Micro-services**
Scaling our NVMe™ Requires a (Real) Network

- Many options, plenty of confusion
- Fibre Channel is the transport for the vast majority of today’s all flash arrays
  - FC-NVMe Standardized in Mid-2017
- RoCEv2, iWARP and InfiniBand are RDMA-based but not compatible with each other
  - NVMe-oF™ RDMA Standardized in 2016
- FCoE fabric is an option
- NVMe/TCP - making it way through the standards process
NVMe-oF™: Making the “Well-Informed” Choice?

Not “just” about “fabrics” performance

Culture and Install Base

Use Cases