Scaling Cloud-Native Virtualized Network Services with Flash Memory

Chloe Jian Ma (@chloe_ma)

Senior Director, Cloud Marketing

Mellanox Technologies

The Telco Industry is Going Through a Transformation

What if you could have this …

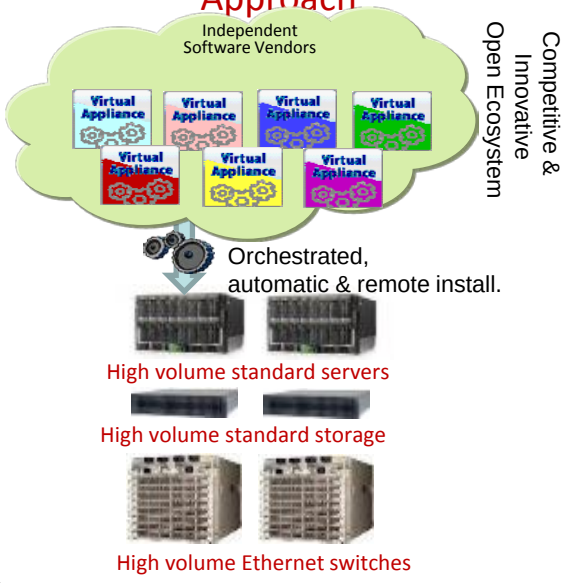
… at the metro service edge

NFV - The Move From Proprietary Appliances To COTS-Based Cloud

Classical Network Appliance Approach

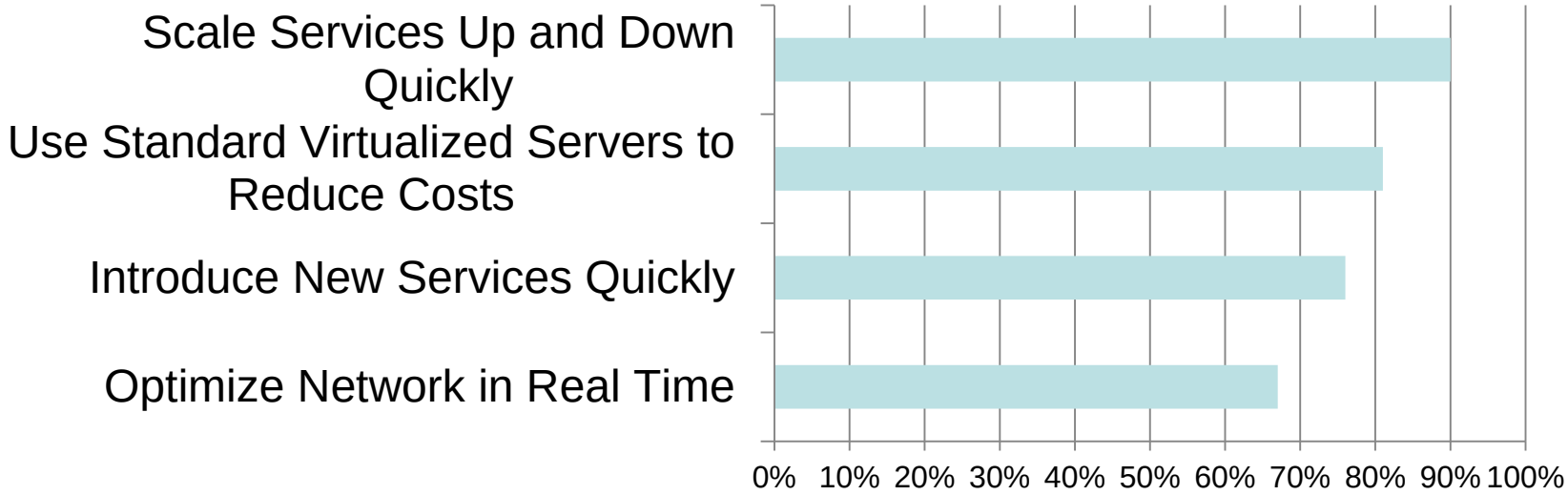
- Fragmented, purpose-built hardware.
- Physical install per appliance per site.
- Hardware development large barrier to entry for new vendors, constraining innovation & competition.

Network Function Virtualisation Approach



Removing tightly coupled network function's software from underlying hardware

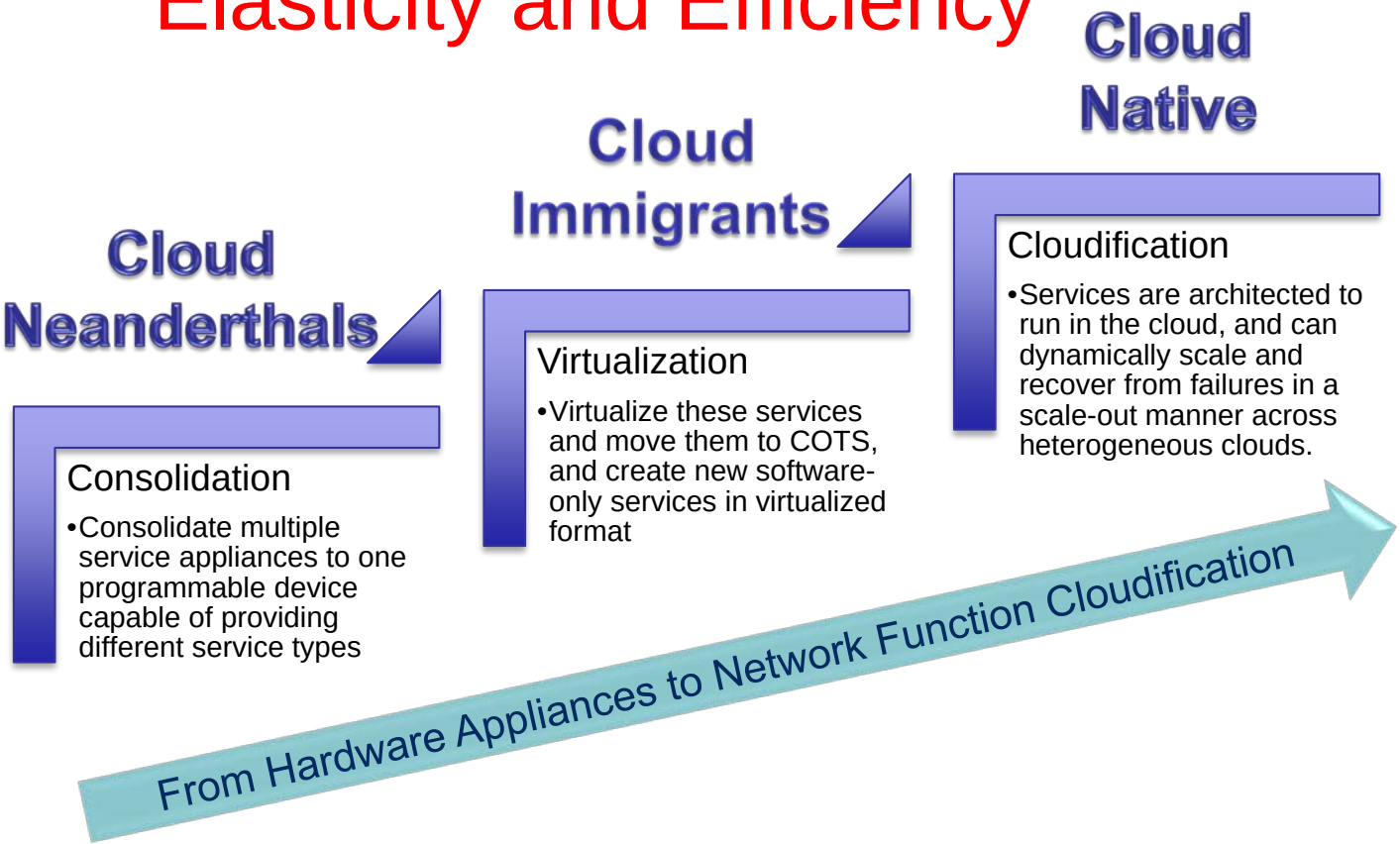
Why do Service Providers Want NFV?



Equally important as cost reduction, service agility and scalability are driving service providers to adopt NFV.

Source: Infonetics Survey

NFV Evolution to Leverage Cloud Elasticity and Efficiency



New Way to Scale Services in the Cloud

Scale Up

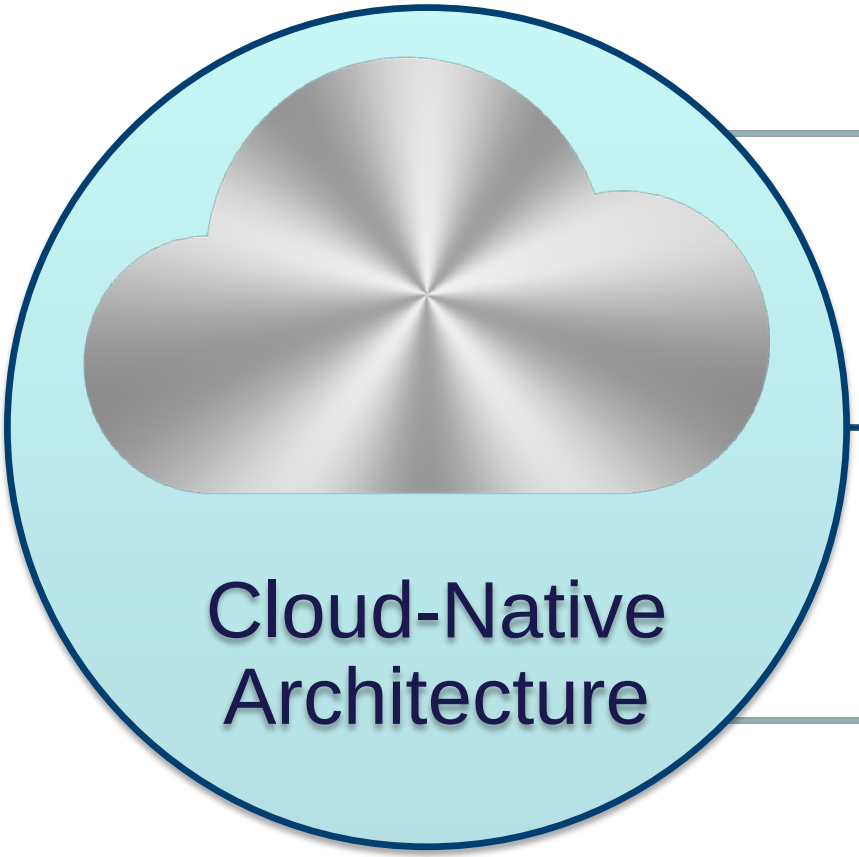
- Pets are given names like Sweetie Pie
- They are unique, lovingly hand raised and cared for
- When they get ill, you nurse them back to health

Scale Out

- Cattles are given numbers like cow101
- They are almost identical to other cattle
- When they get ill, you get another one

"Future application architecture should use Cattle but Pets with strong configuration management are viable and still needed." - Tim Bell, CERN

Cloud-Native Architecture



Application

Open, Stateless
Microservices

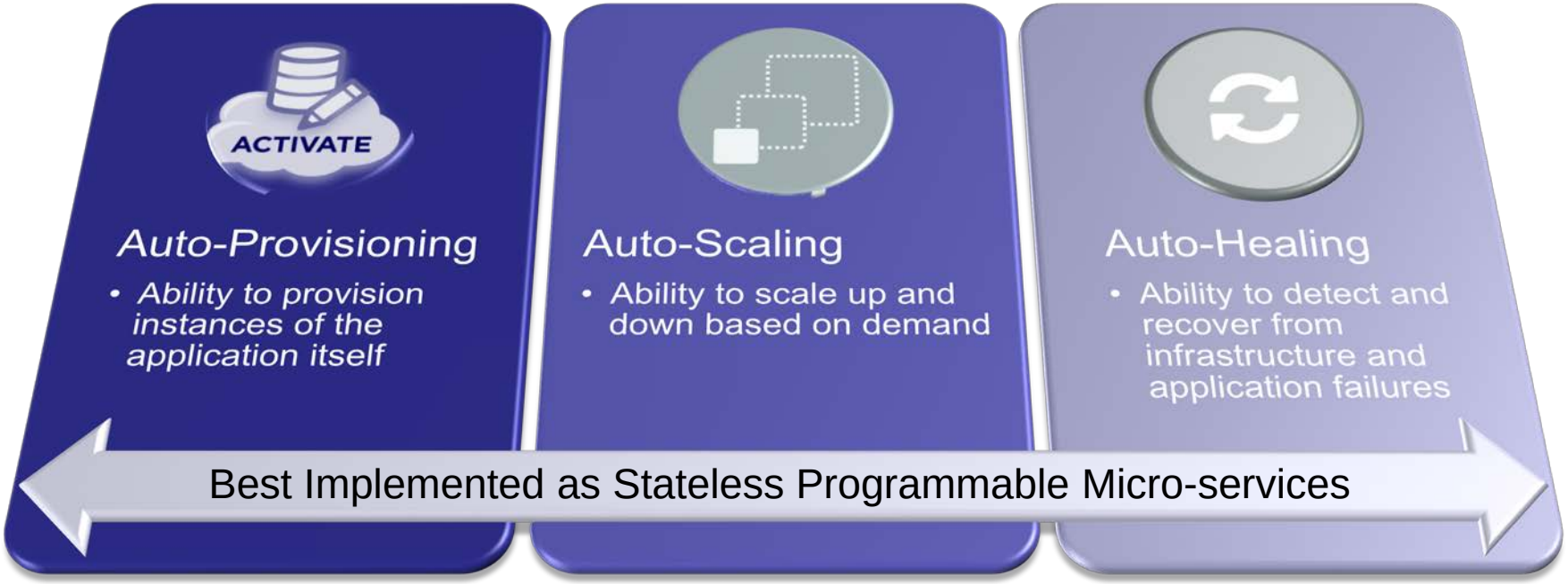
Orchestration

Intelligent
Management

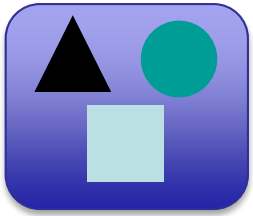
Infrastructure

Efficient
Infrastructure

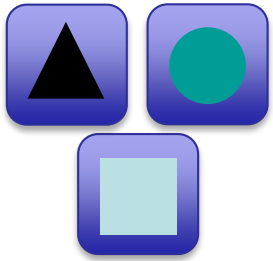
Cloud-Native Applications



Transform to Cloud-Native Applications



Monolithic



Micro-services

Stateful

Stateless

Closed

Open

Benefits

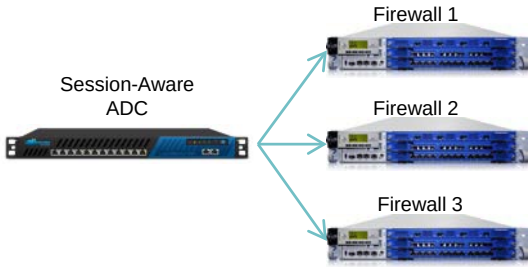
- Smaller Failure Domain
- Better Scalability and Resiliency
- Business Agility with CI/CD

Impact on Infrastructure

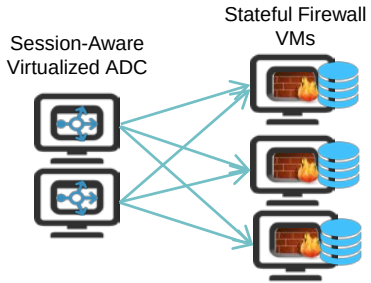
- Much denser virtual machine or container instances on a server
- Much higher requirements on storage performance
- Much higher volume of east-west traffic between application VMs

An Example: VNF Transformation to be Cloud Native

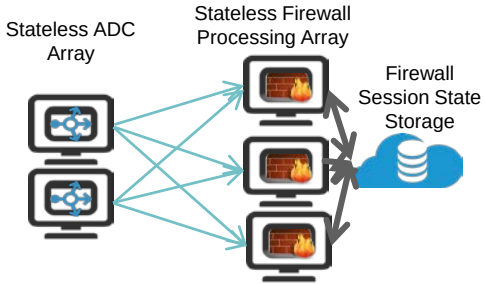
- Pre-Virtualization
  - Complex Appliances
  - Hard to scale
  - Hard to recover from box failure
  - Over provisioning and waste of resources



- Post-Virtualization
  - Complex stateful software
  - Hard to scale
  - Hard to recover from VM failure
  - Automated provisioning possible
  - If you virtualize complex system, you get virtualized complex system



- Into the Cloud
  - Simple virtual appliances with stateless transaction processing coupled with state storage access
  - Scale almost infinitely
  - Fast recovery from VM failure
  - On-demand provisioning and consolidation



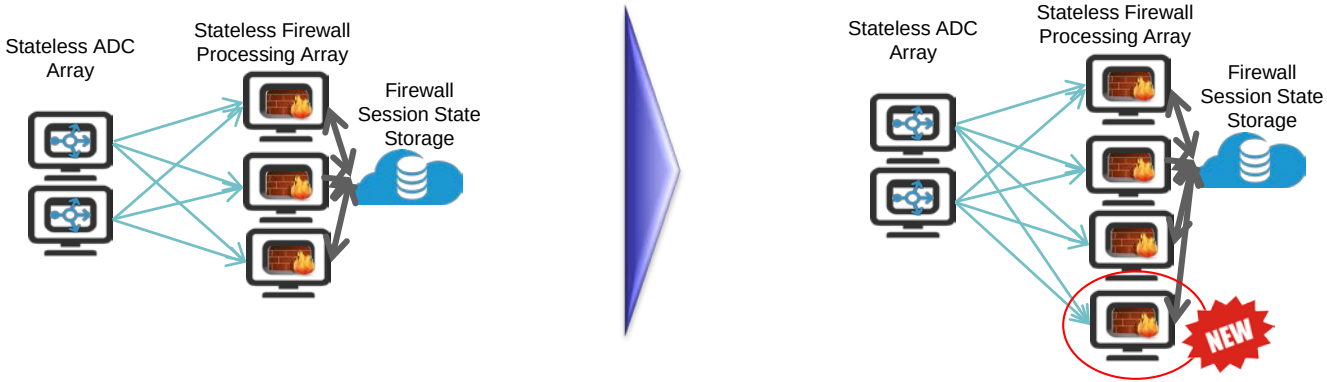
Storage Performance Requirements

Table 4: vSRX Services Gateway Key Performance Metrics

Performance*	VMware	KVM
Firewall (UDP 1514 byte puts)	4.35 Gbps	2.6 Gbps
Firewall (IMIX)	1.05 Gbps	620 Mbps
Firewall ramp rate (TCP)	22,000 cycles/second	22,000 cycles/second
Firewall latency (512 byte UDP)	107 ms	87 ms
Firewall IPv6 (UDP 512 byte packets)	1.46 Gbps	829 Mbps

- 4.35 Gbps is roughly 2.8 Million packets per second
- Each packet needs at least one read from cache to identify the session info.

SCALE OUT CASE

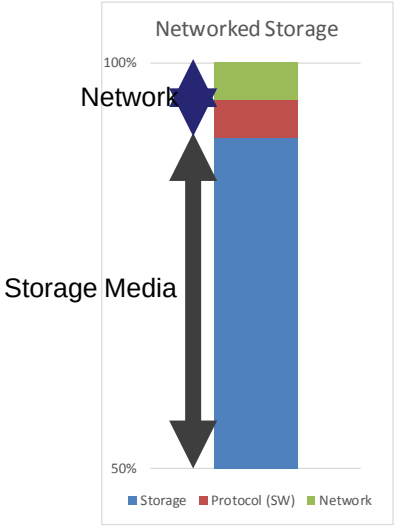


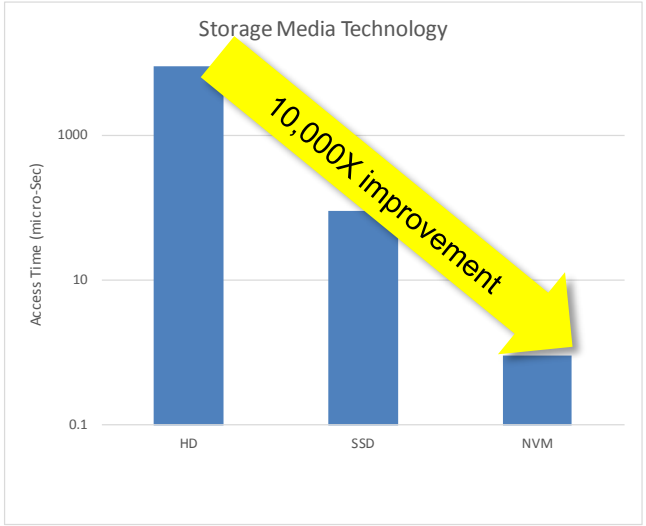
- New Firewall VM can potentially get 2.8X(3-1)/4=1.4 Million packets per second.
- Assuming 1 session read for every 5 packets, we are looking at IOPS of 280K

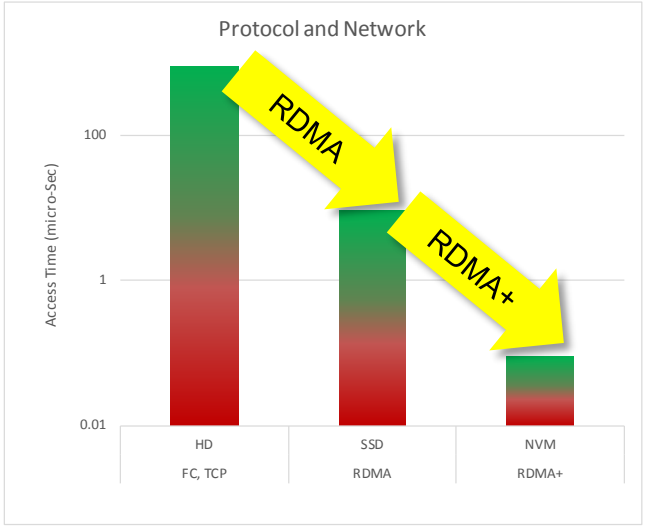
AUTO-HEALING CASE

- Every firewall VM carries 2.8X(3-1)/3=1.87 Million packets per second.
- Death of one firewall VM will add 0.94 Million pps to the other two remaining. Assuming 1 session read for every 5 packets, we are looking at IOPS of 188K

Faster Storage Needs Faster Network

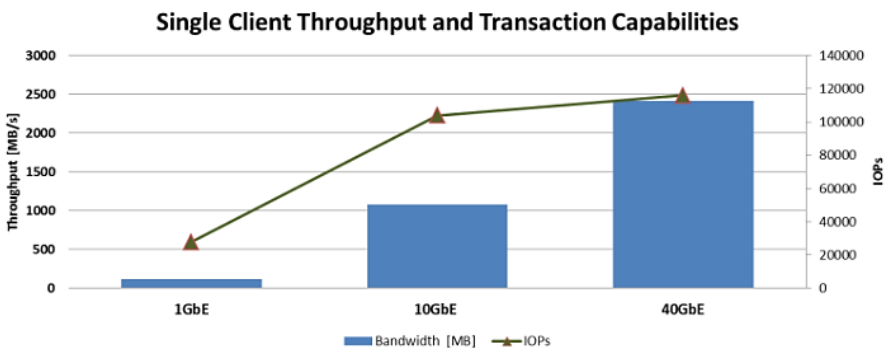


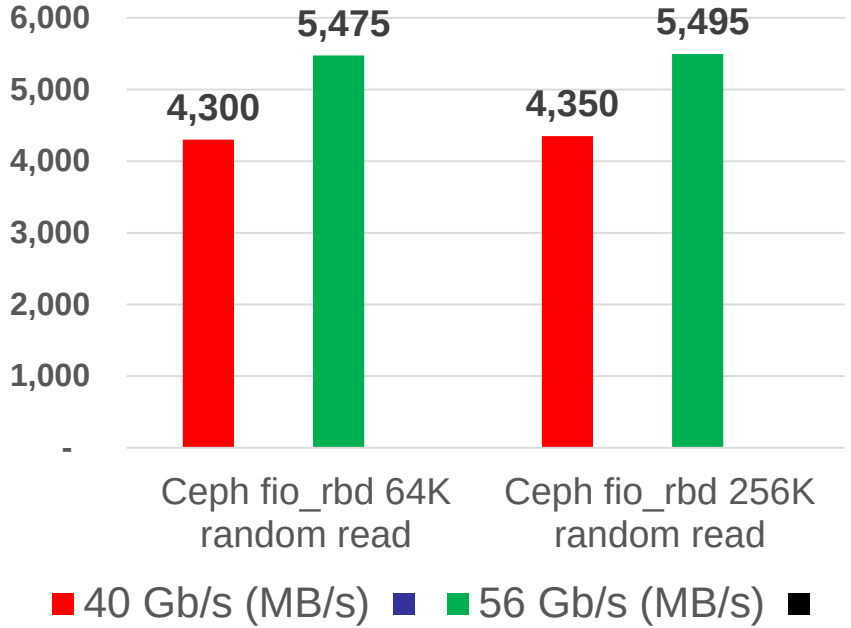




Advanced Networking and Protocol Offloads Required to Match Storage Media Performance

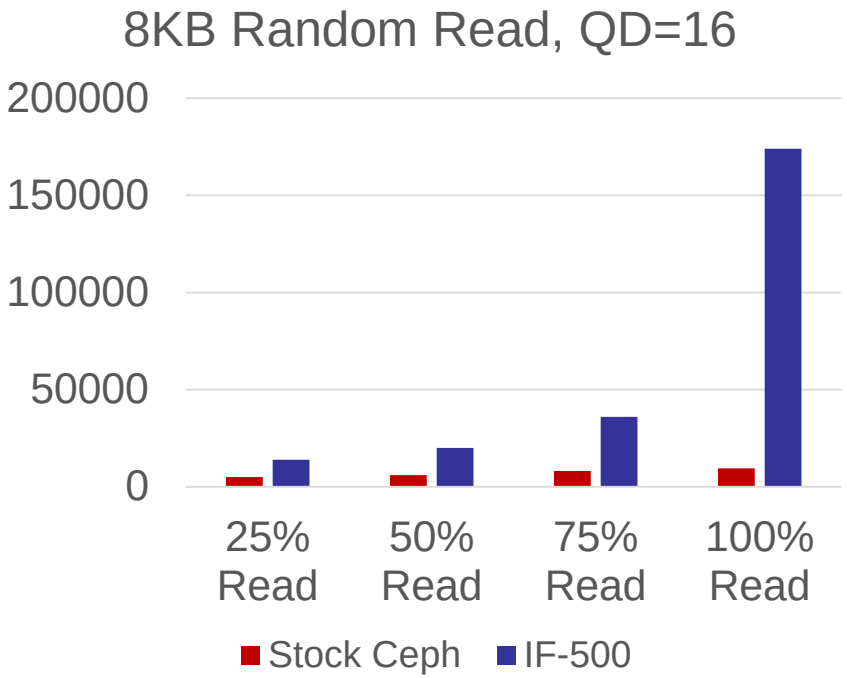
Scale-Out Flash Storage Needs High-Speed Networking



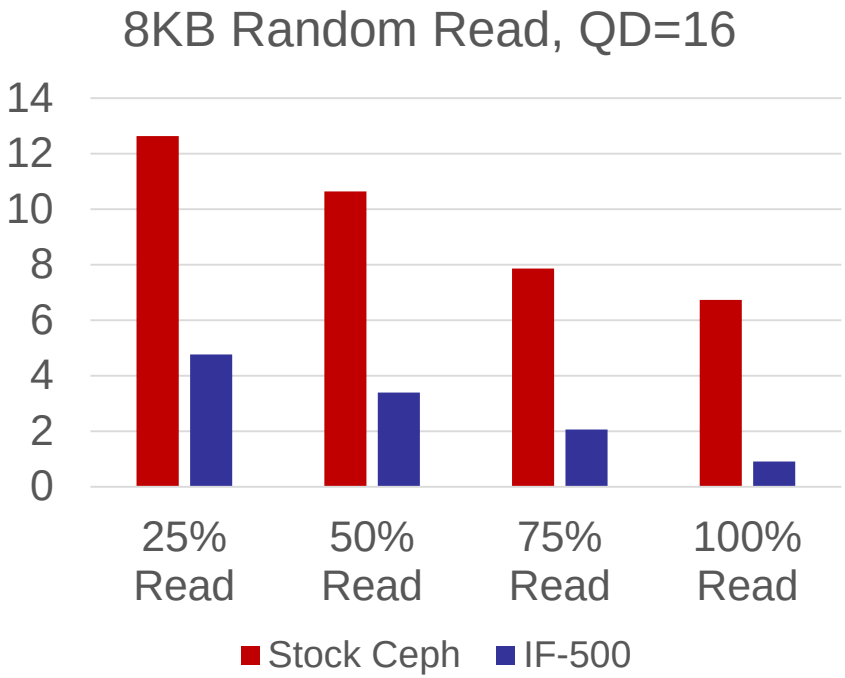


SanDisk InfiniFlash, Maximizing Ceph Random Read IOPS

Random Read IOPs

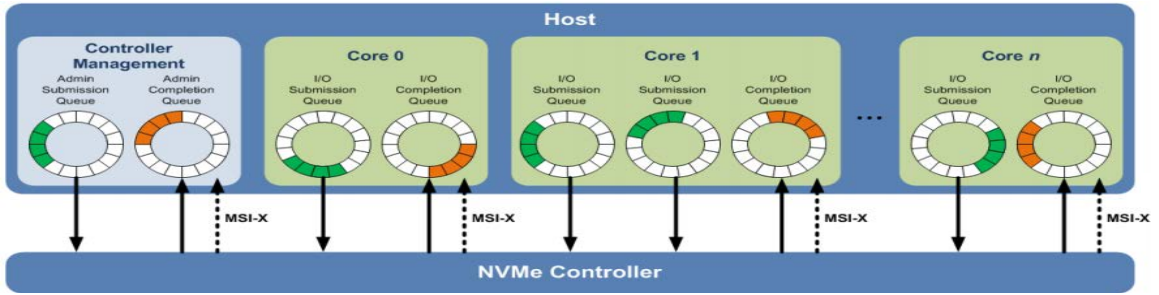


Random Read Latency (ms)



Local NVMe is available TODAY inside Servers

- Now completed standard PCIe host controller interface for solid-state storage
  - Developed by industry consortium, 80+ members
  - Replaces SAS/sATA for flash
- Focus on efficiency, scalability and performance
  - Simple command set (13 required commands)
  - Multiple and deep ques, separate cmd and data
- Already in box driver in many OSes



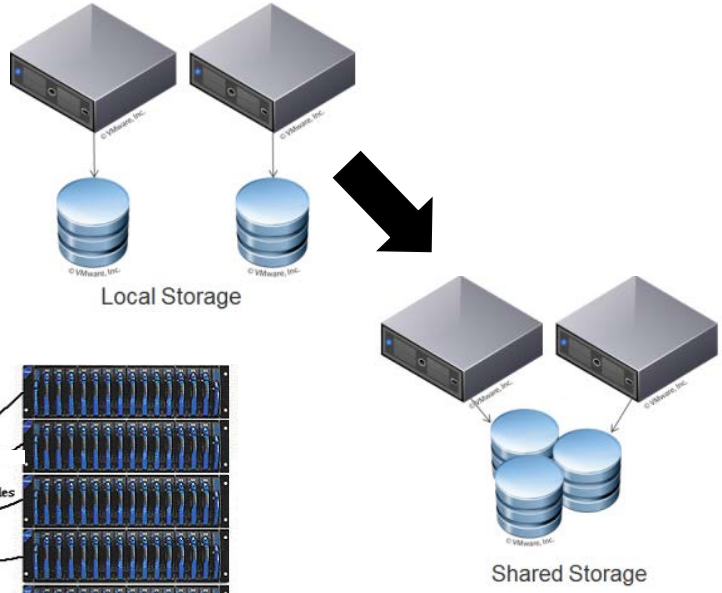
Driver Development on Major OSes

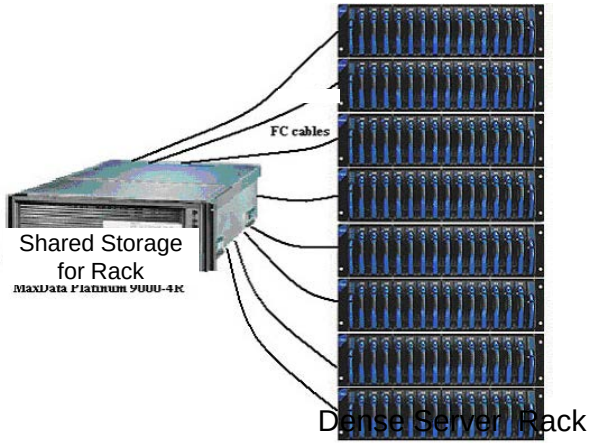
Windows*	• Windows* 8.1 and Windows* Server 2012 R2 include native driver • Open source driver in collaboration with OFA
Linux*	• Stable OS driver since Linux* kernel 3.10
Unix	• FreeBSD driver upstream
Solaris*	• Solaris driver will ship in S12
VMware*	• vmklinux driver certified release in 1H, 2014
UEFI	• Open source driver available on SourceForge

Native OS drivers already available, with more coming!

“NVMe over Fabrics” is the Logical and Historical next step

- Sharing NVMe based storage across multiple CPUs is the next step
- Driven by the need for the best possible compute efficiency and shared advantages
- Shared storage requires a Network/Fabric
- NVMe over Fabrics standard in development
  - Mellanox Architects Contributing
  - Version 1.0 4Q15
- RDMA protocol is required

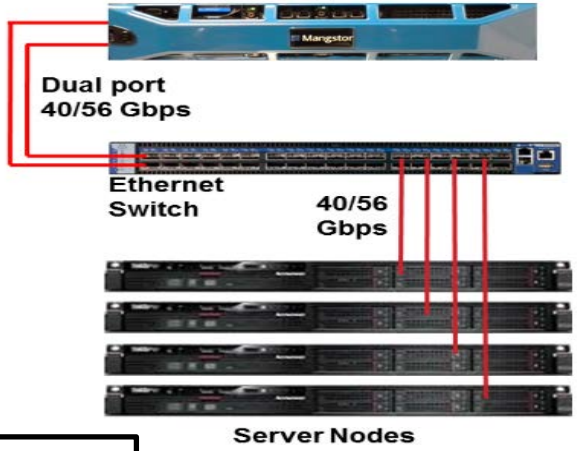




NAB Demo of Early NVMe over Fabric

NAB Show April 11 - 16, 2015 Las Vegas

10GBytes/s reads, 8GB writes, 2.5M random read IOPS (4KB block size), Max latency 6-8us over local



From PR

The Mangstor NMX Series appliances address the need of video editing and shared game development applications requiring high R/W bandwidth while maintaining low latency to shared flash storage in cluster server configurations. The solution uses Mellanox VPI adapters, supporting NVMe over Fabrics using either RoCE or IB to provide high throughput at low latency.

Mangstor's storage appliance provides a non-proprietary solution which outperforms traditional SAN-attached all-flash and hybrid arrays, and enables customers to seamlessly evolve their traditional server attached storage to growing server-SAN storage environments.

Flash Memory SUMMIT

Thank You!

Contact: Chloe Jian Ma (@chloe_ma)

chloe@Mellanox.com