Flash Memory Summit

The Best Ethernet Storage Fabric

John F. Kim & Amit Katz

Flash Memory Summit

Storage Networking Background: From Fibre Channel to Ethernet

1997

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

2017

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

- Yesterday: Storage Network = FC
  - Fibre Channel offered best performance
  - All interesting storage is tier-1 block
  - No cloud or hyperconverged
- Today: Ethernet = Ideal Storage Fabric
  - Cloud! Hyperconverged! SDS!
  - Ethernet outperforms Fibre Channel
  - Block storage is less than 20% of capacity

Flash Memory Summit

Today’s Storage Landscape

The Storage Iceberg

Traditional FC SAN

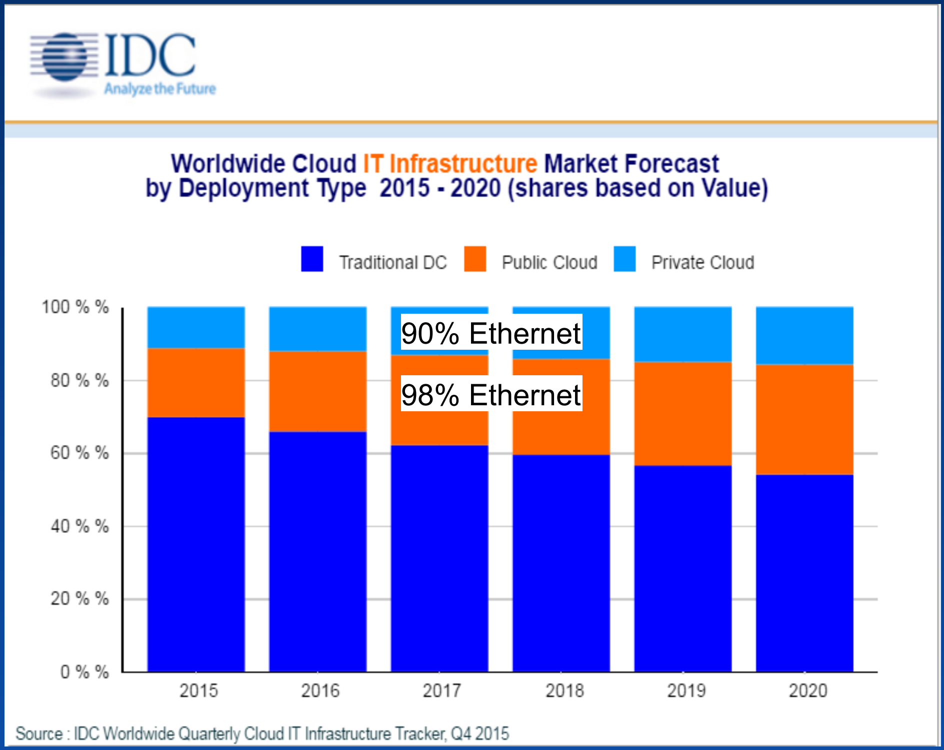
80% of your data is here

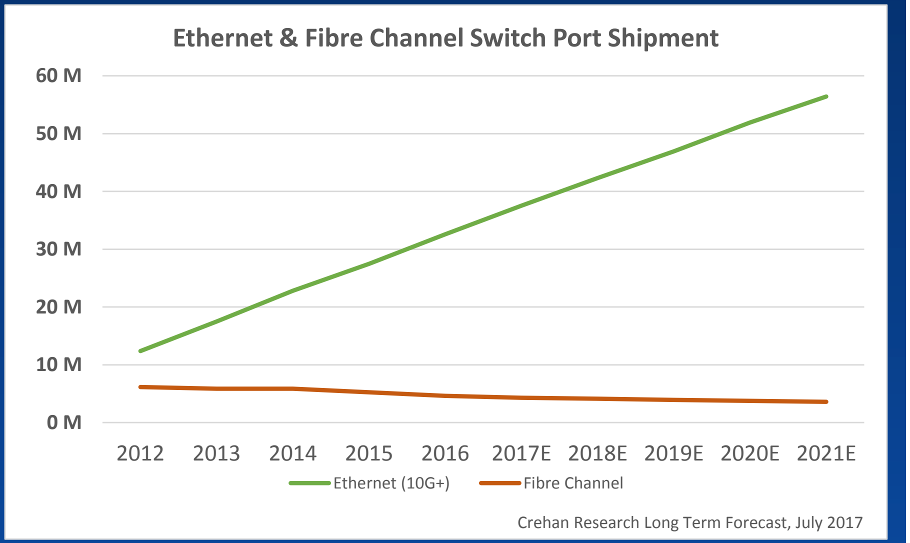
95% Ethernet

Storage Growth
All in Ethernet

Internal Use Only

Cloud Growth is Driving Scale-Out and Ehternet Storage

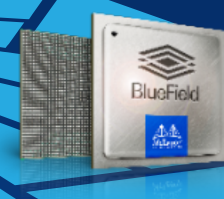




Flash Memory Summit

Ethernet Storage Fabric – We've Got You Covered

Ethernet Storage Fabric



Everything a Traditional SAN Offers but …
Faster, Smarter, & Less Expensive

PERFORMANCE

- Highest Bandwidth
- Lowest latency
- RDMA and storage offloads
- Native NVMe-oF Acceleration

INTELLIGENCE

- Integrated & Automated Provisioning
- Hardware-enforced Security & Isolation
- Monitoring, Management, & Visualization
- Storage-aware QoS

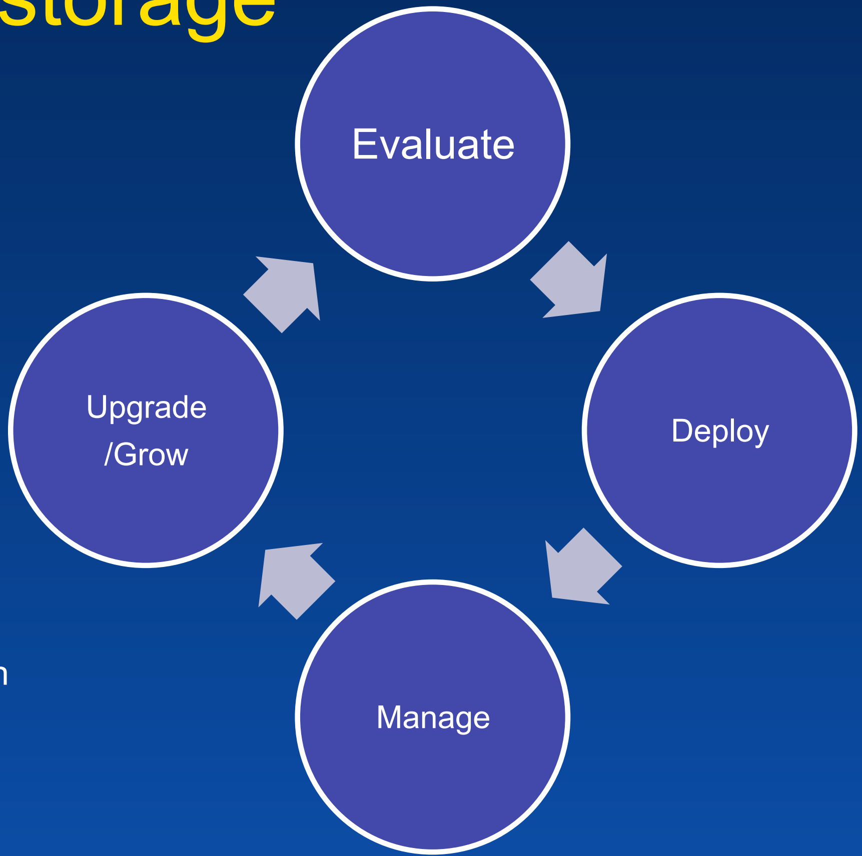
EFFICIENCY

- Just Works Out of the Box
- Flexibility: Block, File, Object, HCI
- Converged: Storage, VM, Containers
- Affordable: SAN without the $$

Network should not be an obstacle when deploying storage

Flash Memory Summit

- Evaluate / Decide
  - Lowest latency, highest bandwidth
  - Flexible configurations; Reasonably priced
- Deploy
  - Easy to deploy; Tested and verified End-2-End
  - Network cannot slow down storage traffic
  - Efficiency
- Manage
  - Easy to monitor / fix
  - Predictable, Fair performance
- Upgrade / Grow
  - Standard/interoperable, interchangeable, no lock-in
  - Easy network scale-out
  - Support the next speed

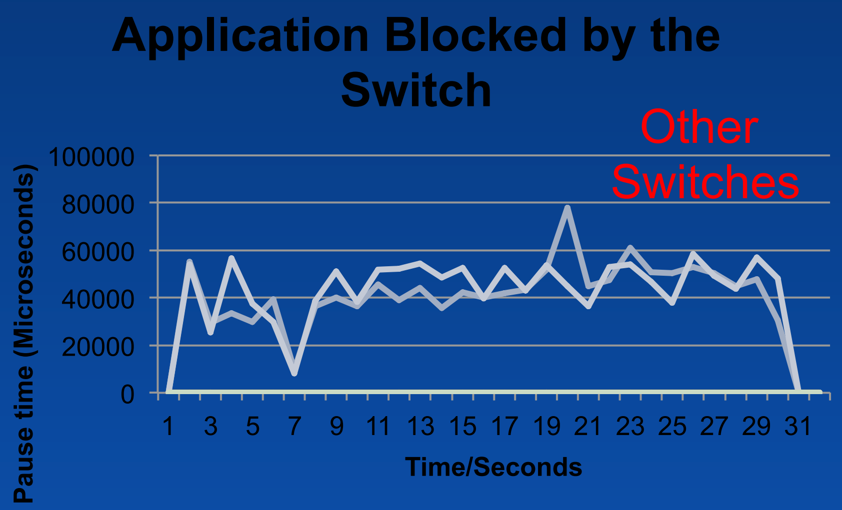


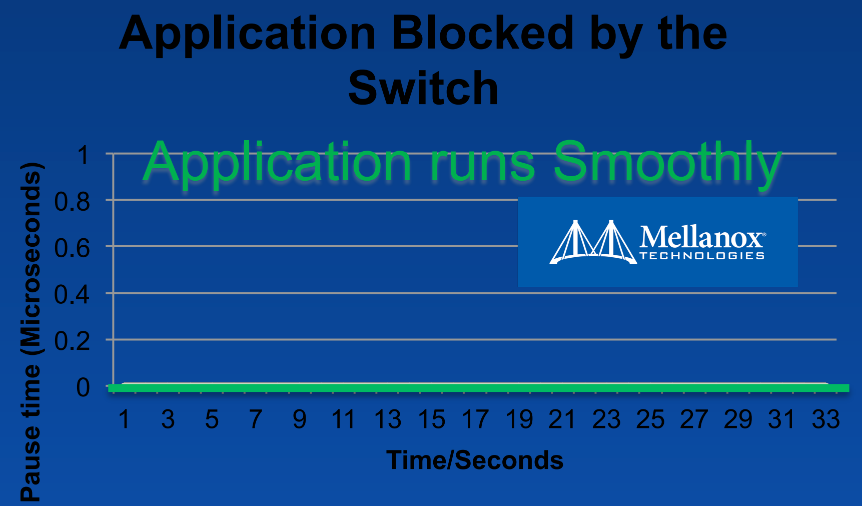
Deployment phase

- Easy to deploy (easy for storage guys/channels…)
  - Match settings
  - Set up high availability
- Network cannot slow down the storage traffic
  - Faster speeds, line rate, non-blocking, RDMA
  - Option to use QoS or bandwidth or both
  - Eliminate long-tail latency / jitter on storage network
- Tested and verified End-2-End
- Reasonably priced – right-sized configuration
  - Don't buy 48 ports when you need 4 of them, there are low port count switches in the market
- Don't underestimate efficiency – space and power
  - There are 100GbE switches today running <100W, all ports, full wire speed
  - Why using 2RU if you can use 1RU (2 x ½ 19'')?

Predictable Performance when Running NVMe-oF/RoCE

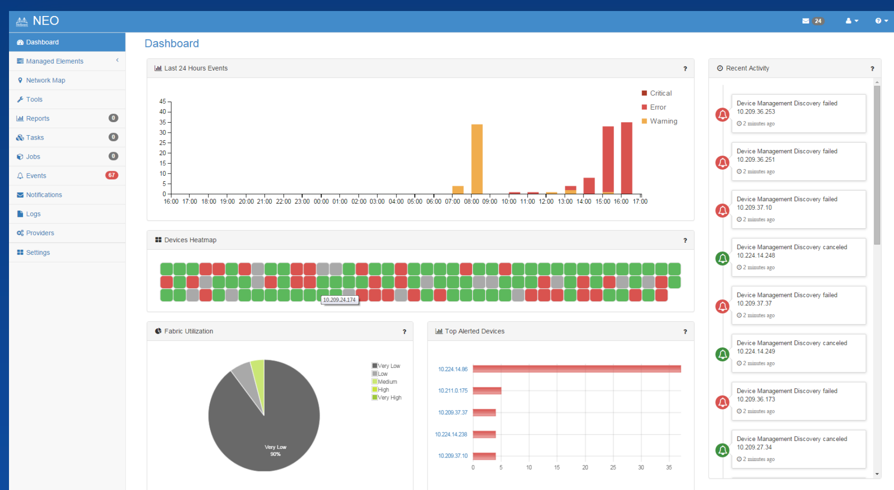
Switch is configures with PFC/ECN running a basic RoCE benchmark





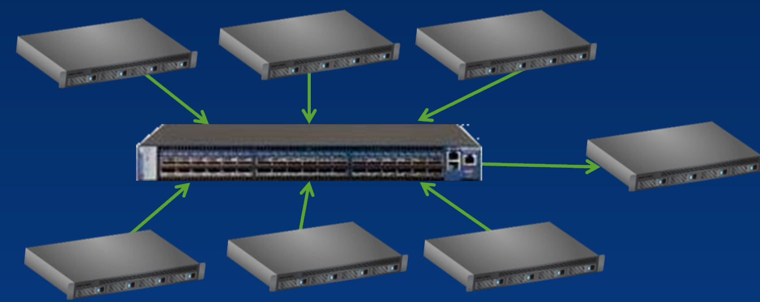
Easy to Monitor / Manage

- Identify hardware failures quickly
- Identify configuration differences – know what changed lately
- Avoid human mistakes – upgrade all switches/NICs as a fabric
- Buffer histograms, telemetry, alerts, storage integration

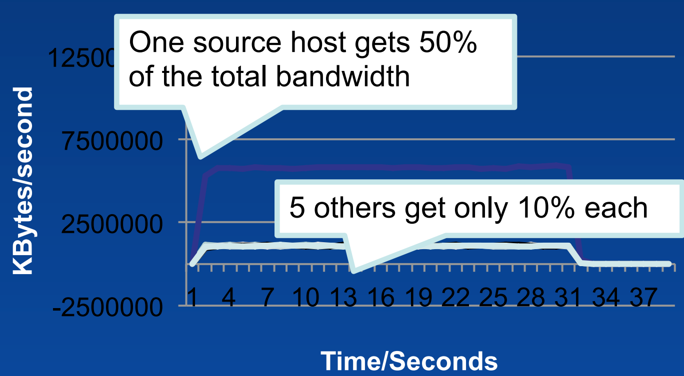


Manage

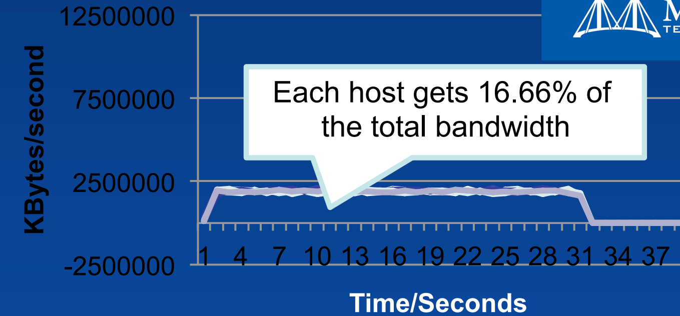
- Be fair to all tenants – it's not trivial…



Tomahawk Based Switch



Spectrum Switch



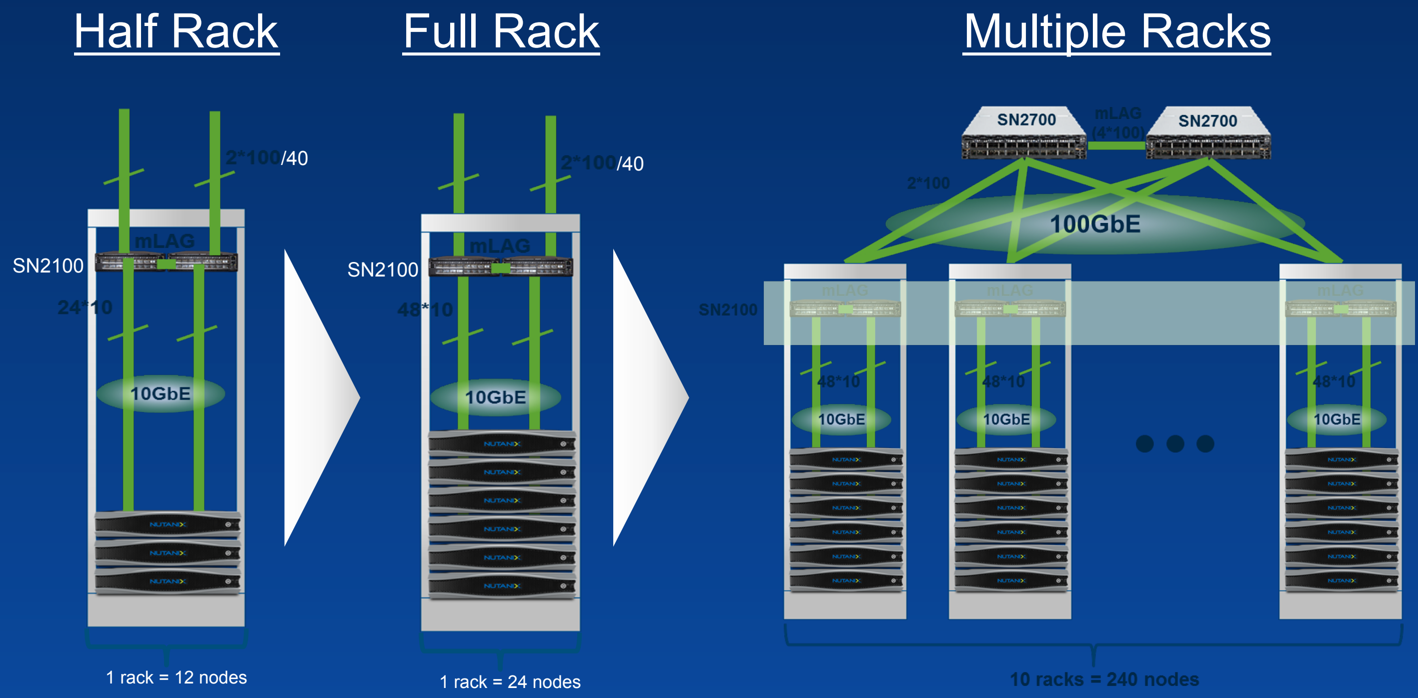
Can you afford unfair traffic distribution? Who will the customer blame???

Upgrade / Expand

- Standard/interoperable, interchangeable, no lock-in
  - Avoid proprietary switches, looks simple, at some point people typically regret for taking this path
- Easy scale out
  - Make sure there are enough available uplinks with the right speeds as storage tend to grow over time
  - Architecture must be simple to scale (L3 vs. MLAG from switch to switch)
- Support the next speed
  - Don't be limited by choosing 10GbE as 25/40/50/100GbE is gaining momentum
  - Mellanox 40GbE/100GbE is often lower costs than incumbent's 10GbE switch

Flash Memory Summit

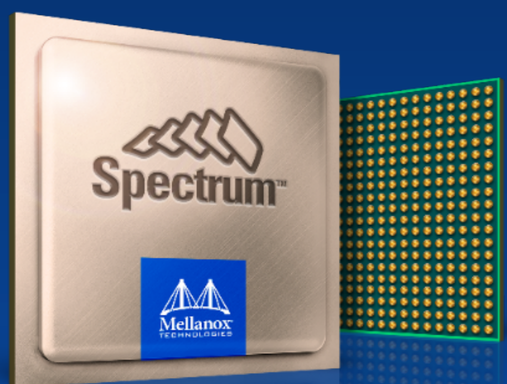
Networking is Easy to Scale



Nutanix solution note: https://www.nutanix.com/go/mellanox-networking-with-nutanix.html

Flash Memory Summit

Leverage Spectrum Features to Make the Best Storage Switch



- 1RU Better Performance
- 2 x 16 ports in 1RU
- Easiy Setup
- Better Visibility
- Tested End-2-End

Thank You

John F. Kim & Amit Katz