How to Network Flash Storage Efficiently at Hyperscale

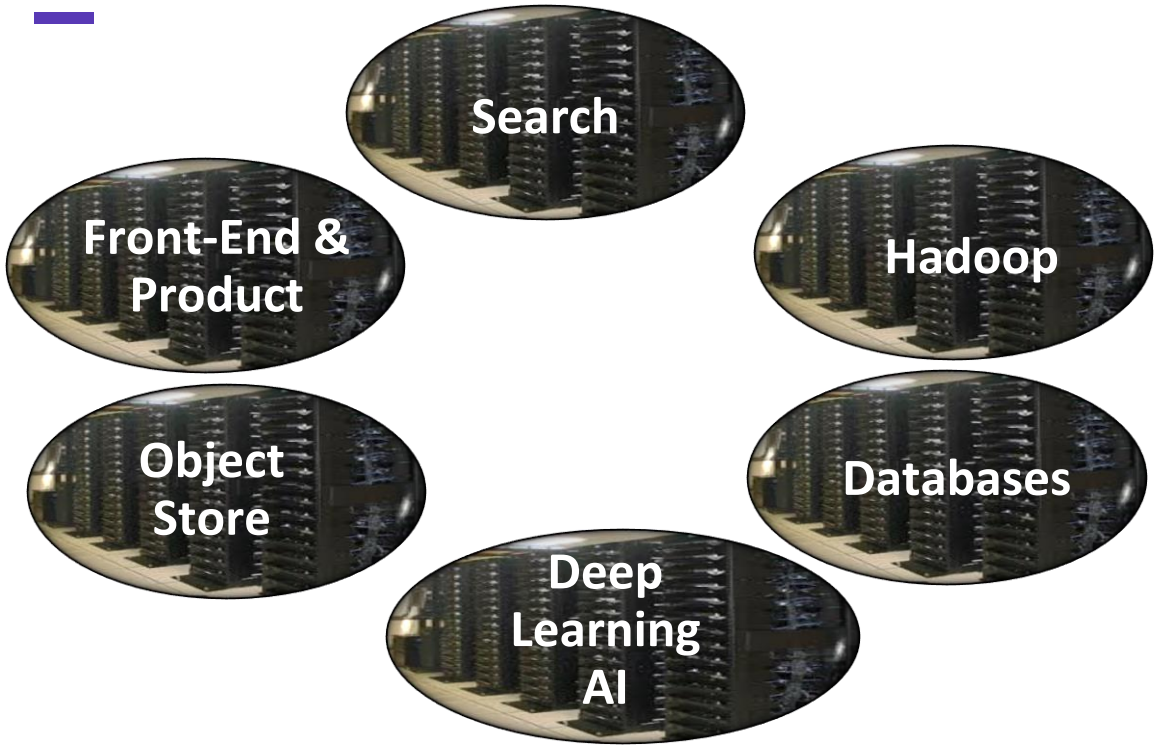
Manoj Wadekar

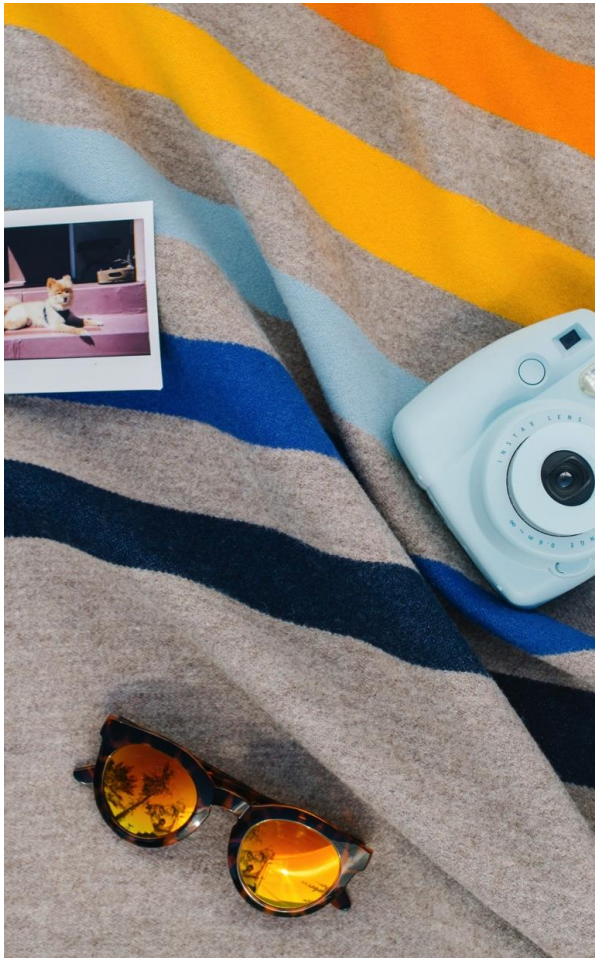
ebay

Michael Kagan

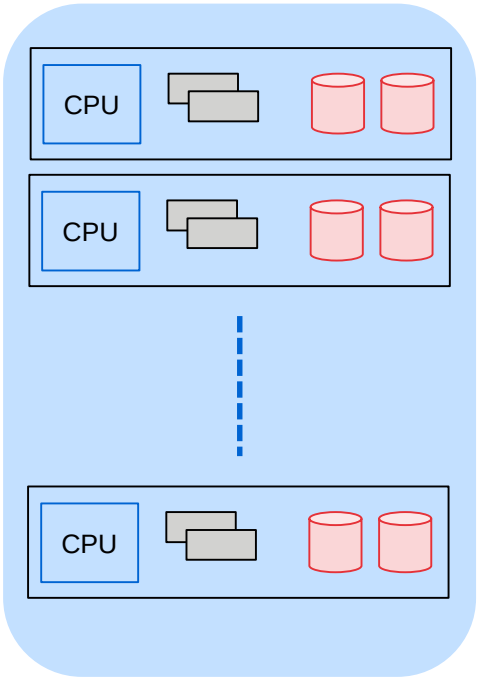
Mellanox TECHNOLOGIES

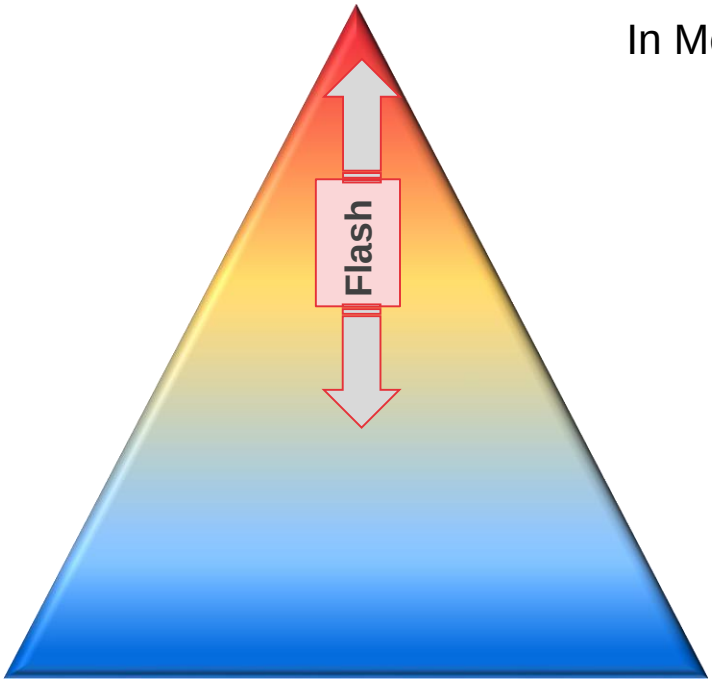
ebay Hyper scale Infrastructure





Typical hyper scale servers





Performance Needs

In Memory/Search

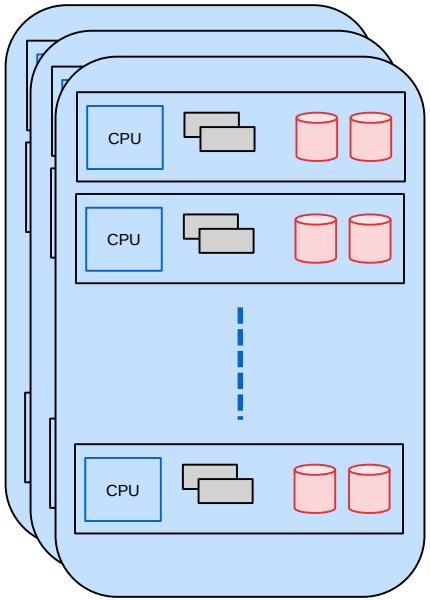
Databases

Hadoop

Object Store
FE/Dev

Archival/Cold

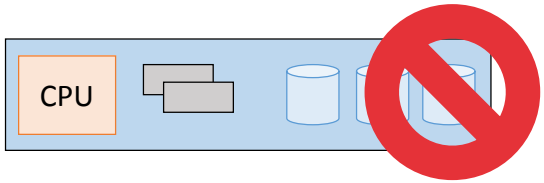
Typical hyper scale servers: Design goals



Efficiency:
Utilization, commonality

Growth:
Performance, Capacity

Converged Infrastructure: challenges



Mismatched App Needs:

- Compute/Storage needs can change for different clusters.
- Can result in under-performance or wastage

Inefficiency:

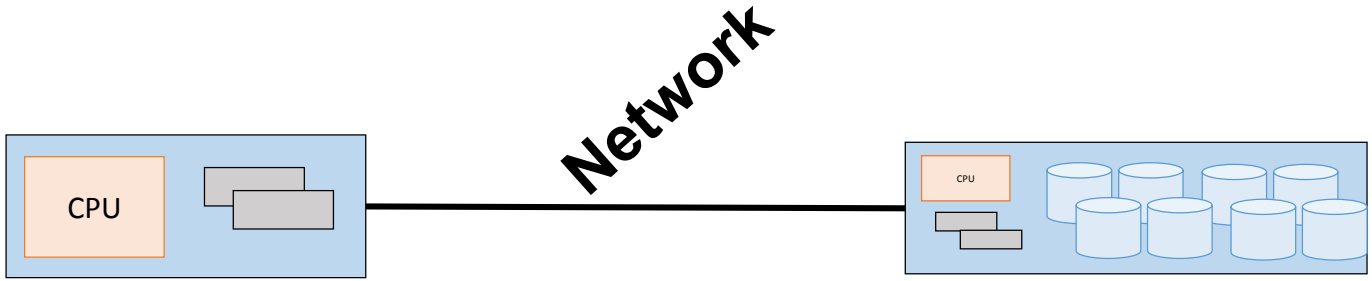
- Spend, infrastructure space, power utilization
- Further challenge to justify high density/high performance drives

Scale Challenge:

- “Cattle” use cases for data-heavy workloads may result in large data movement
- Complicated storage scheduler leads to constrained scaling

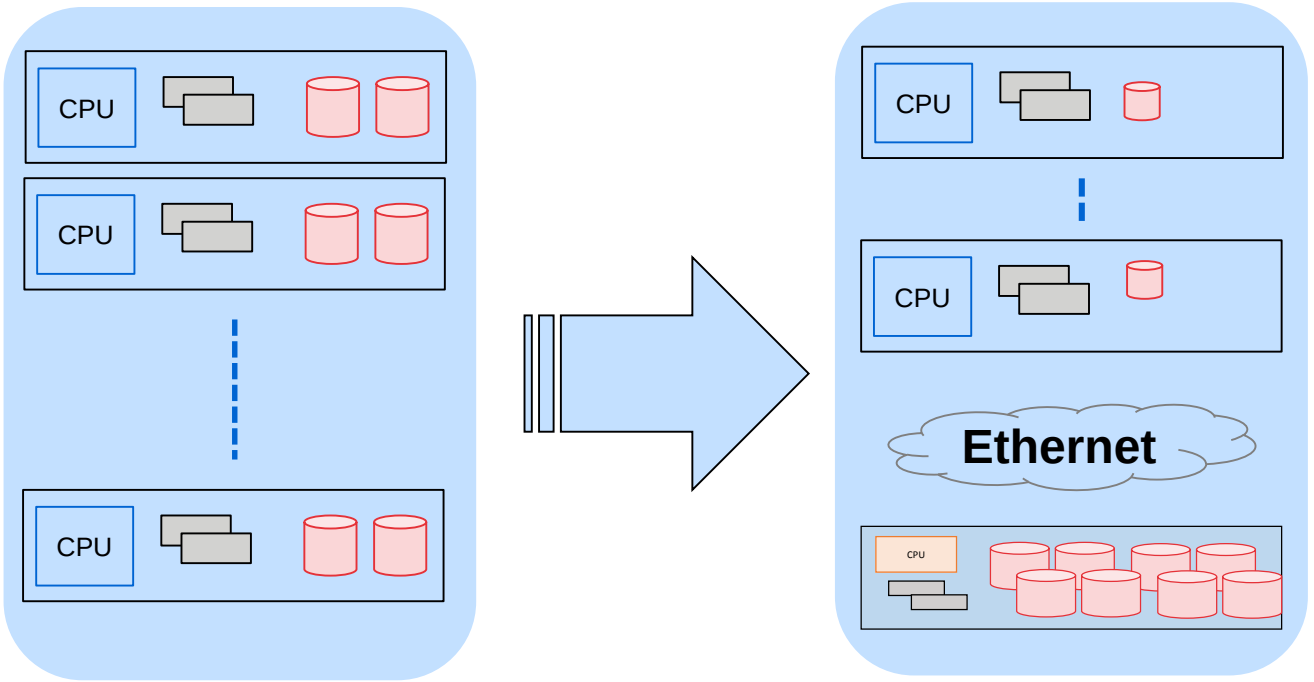
Server Platform: Shackled to local storage

What's needed: Disaggregated Storage



Separate out storage and compute resources

What's needed: Rack-As-A-Compute



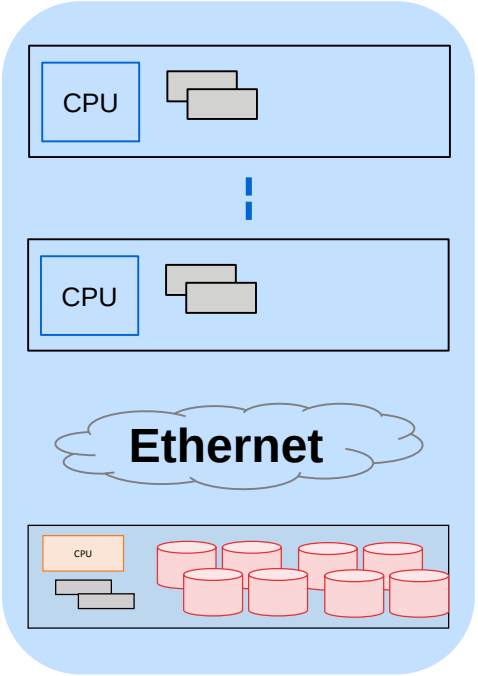
Why: Rack-As-A-Compute

Right Sizing:

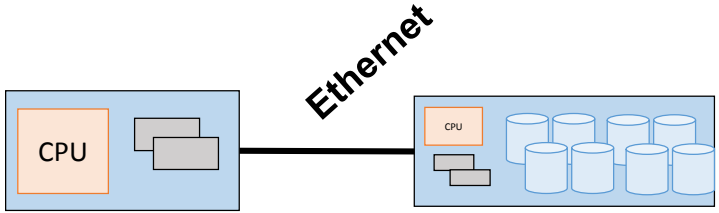
- Clusters can use optimized ratio of compute and storage.
- Allows reducing wastage and improve performance

Independent Scaling:

Compute and storage capacities can be scaled per need



Disaggregated Storage: Interconnect Needs



Throughput:

- Sequential workloads driven by throughput
- Aggregated storage drives higher needs

Latency:

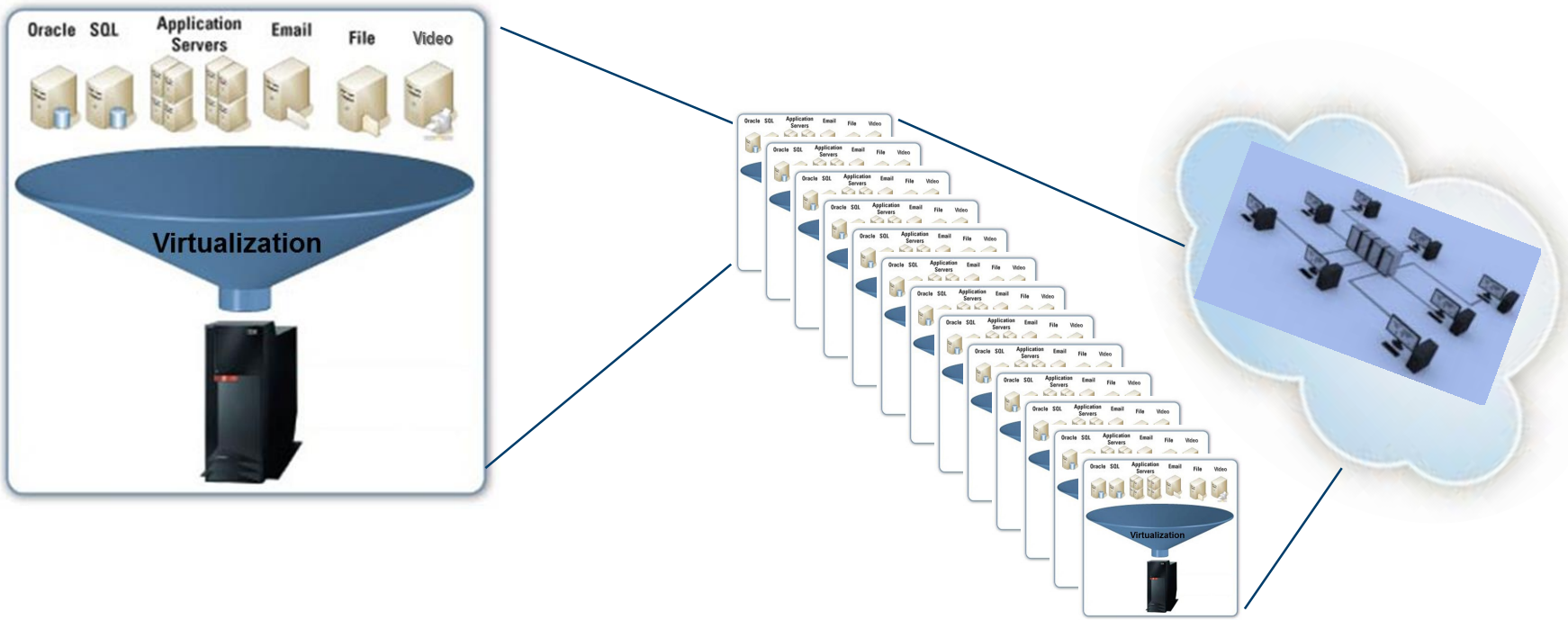
- IOPs sensitive workloads
- Appropriate deployment topologies

Simplicity:

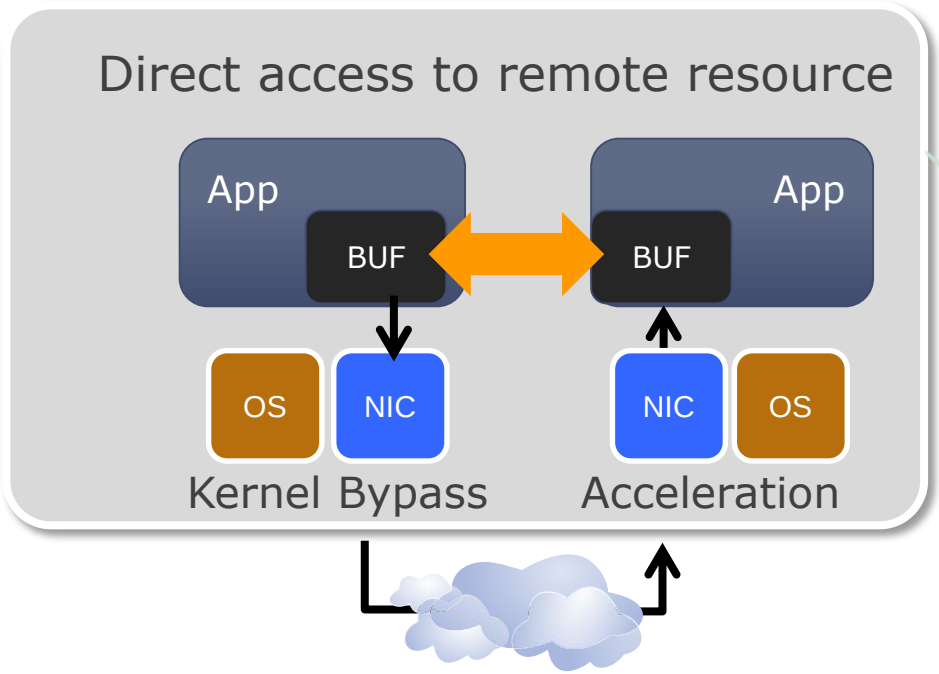
- Known, ubiquitous network

Is Ethernet Ready?

The Paradigm Shift – Resource as a Service



RDMA: Data Center Infrastructure Foundation

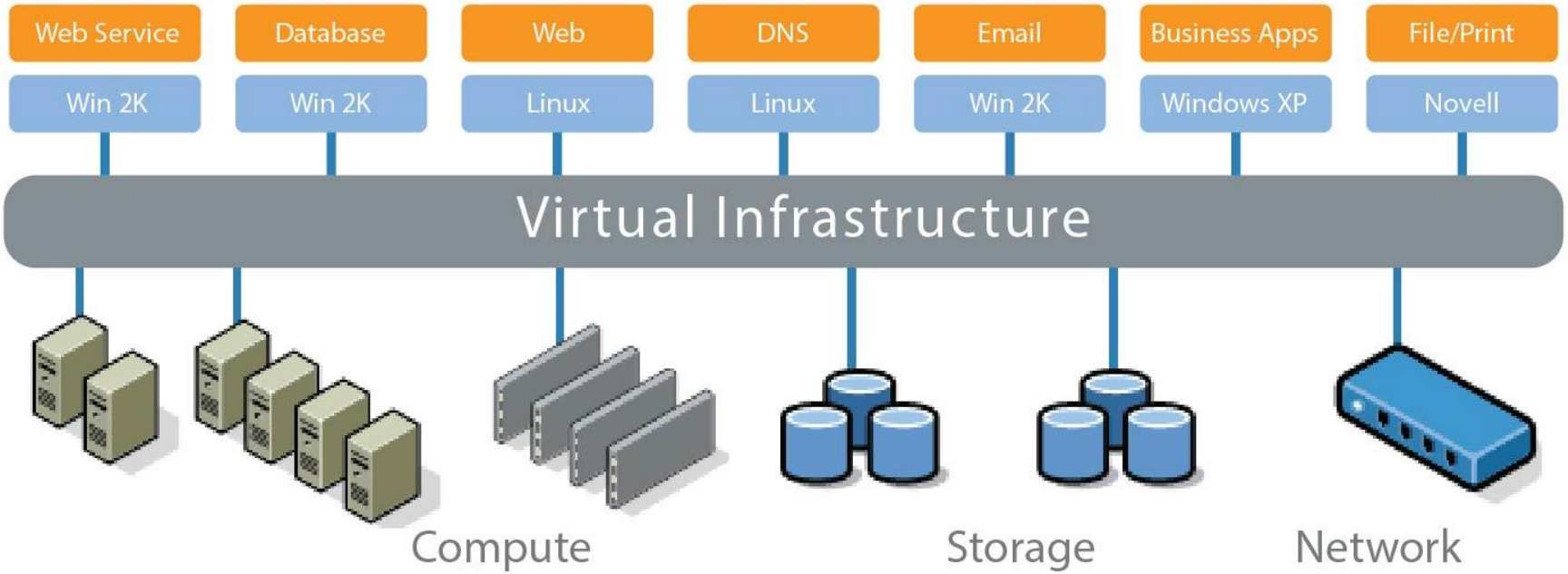


RoCE™

RDMA over Converged Ethernet – RoCE

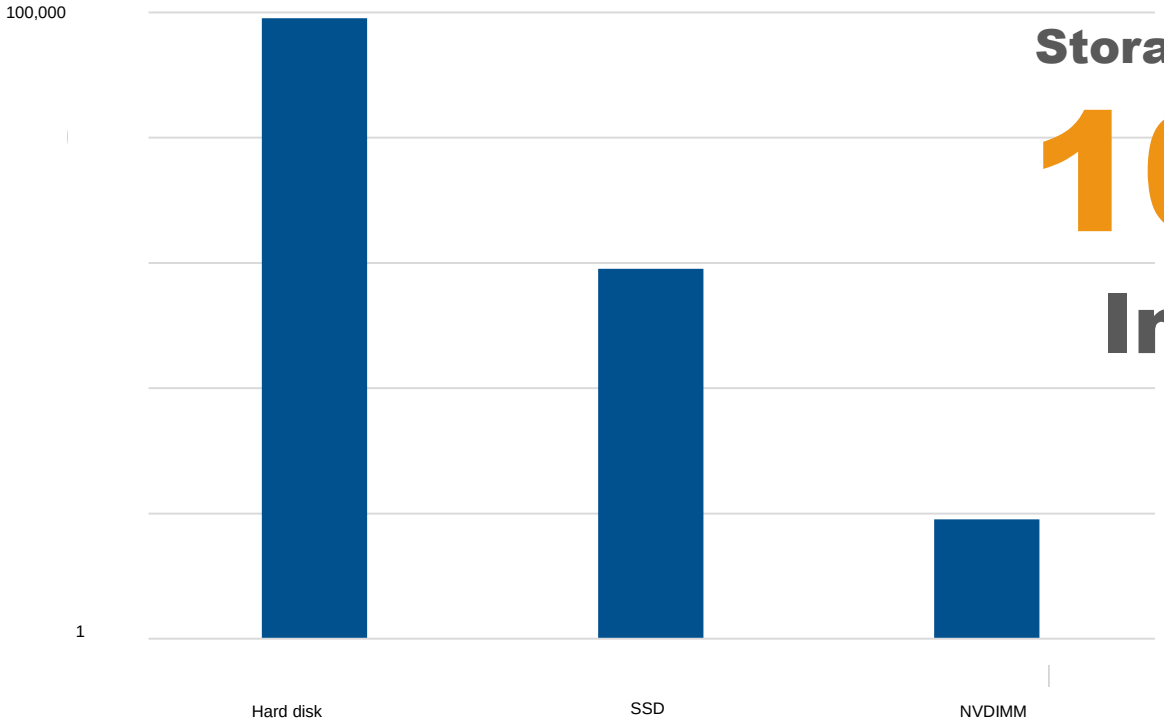
© 2017 Mellanox Technologies

RDMA-Enabled Cloud Infrastructure



© 2017 Mellanox Technologies

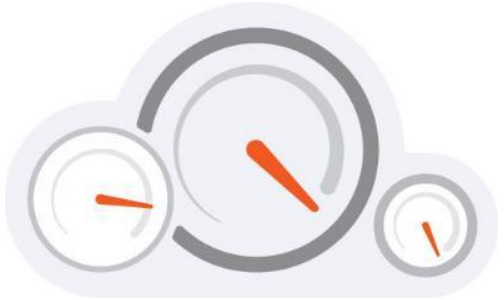
Storage Media Technology



Storage Media Access Time

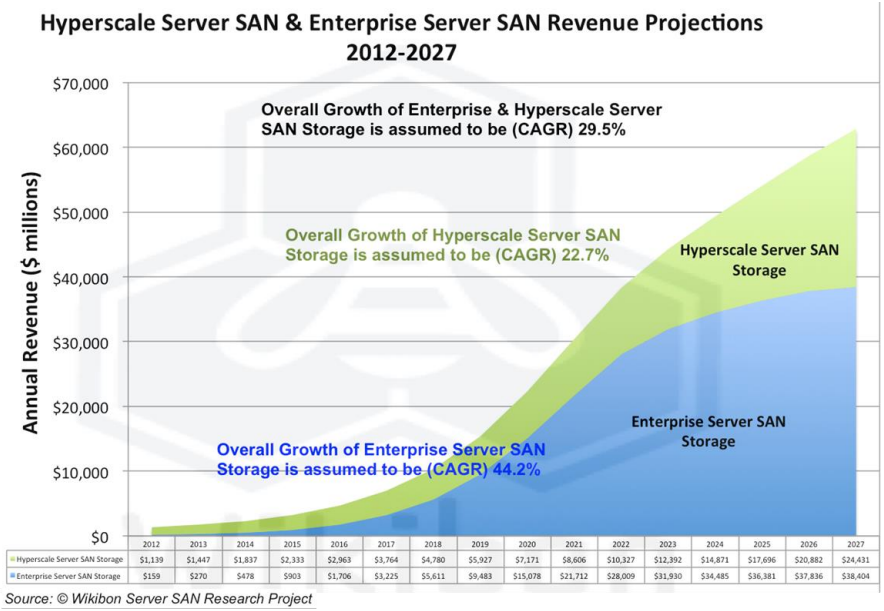
10,000X

Improvement



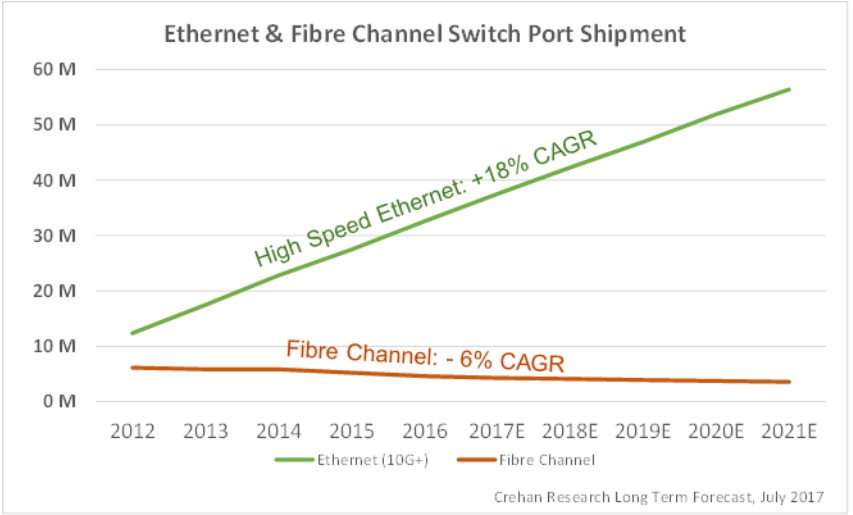
© 2017 Mellanox Technologies

Networked Storage Continues Growth and Movies to Ethernet



	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
Hyperscale Server SAN Storage	$1,139	$1,447	$1,837	$2,333	$2,963	$3,764	$4,780	$5,927	$7,171	$8,606	$10,327	$12,392	$14,871	$17,696	$20,882	$24,431
Enterprise Server SAN Storage	$159	$270	$478	$903	$1,706	$3,225	$5,611	$9,483	$15,078	$21,712	$28,009	$31,930	$34,485	$36,381	$37,836	$38,404

Source: © Wikibon Server SAN Research Project

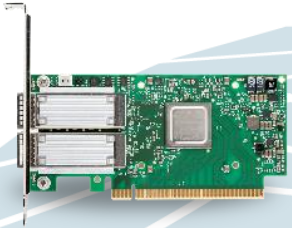


- Networked Storage (SANs)
  - Better utilization: capacity, rack space, power
  - Scalability, management, fault isolation

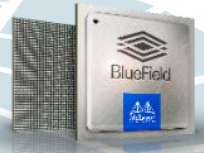
- Ethernet growing very rapidly driven by:
  - Cloud & Hyper Converged Infrastructure (HCI)
    - No Fibre Channel in the Cloud
  - NVMe Over Fabrics
  - Software Defined Storage

© 2017 Mellanox Technologies

Ethernet Storage Fabric – We've Got You Covered



Ethernet Storage Fabric

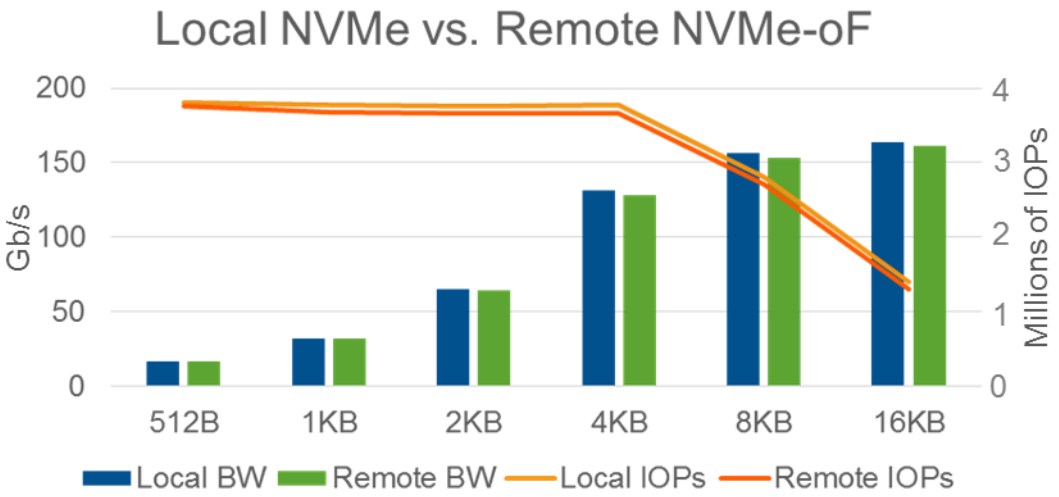


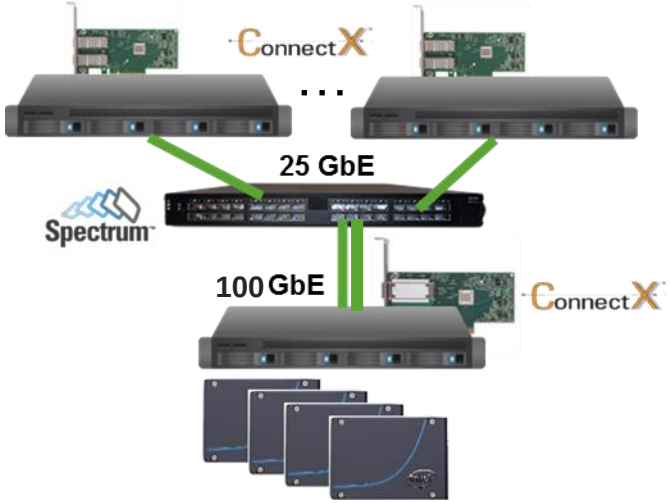
Everything a Traditional SAN Offers but …
Faster, Smarter, & Less Expensive

Performance | **Efficiency** | **Ubiquity**

© 2017 Mellanox Technologies

NVMe-oF Performance with Open Source Linux





© 2017 Mellanox Technologies

BlueField: High Performance Yet Cost Effective Storage

200 Gb/s of throughput

BlueField

More than 5 million IOPS

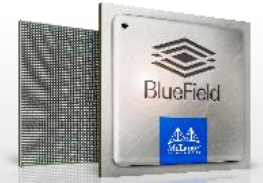
Industry-Leading ConnectX Intelligent Offload

Tile Multicore ARM Architecture

© 2017 Mellanox Technologies

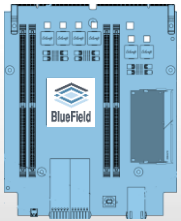
BlueField Building Blocks for Storage Platforms

ASIC



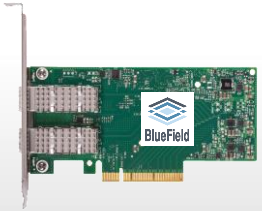
- **Family of products**
  - **4, 8, 16 cores**

Storage Controller



- **NMVe storage solutions**
- **Customized per-OEM**

Adapter

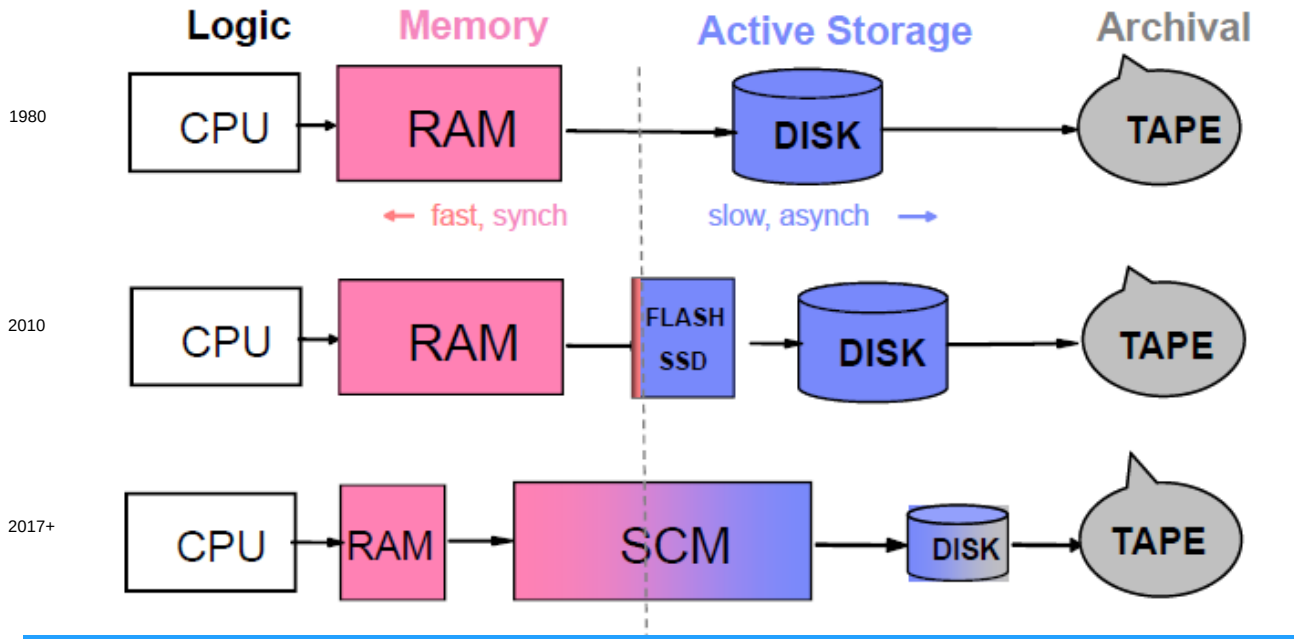


- **Smart NIC**
- **Standard PCIe slot**

Open Platform Software – Standards Based
BlueOS Linux – Built from kernel.org
Standard Linux development tools to build your application

© 2017 Mellanox Technologies

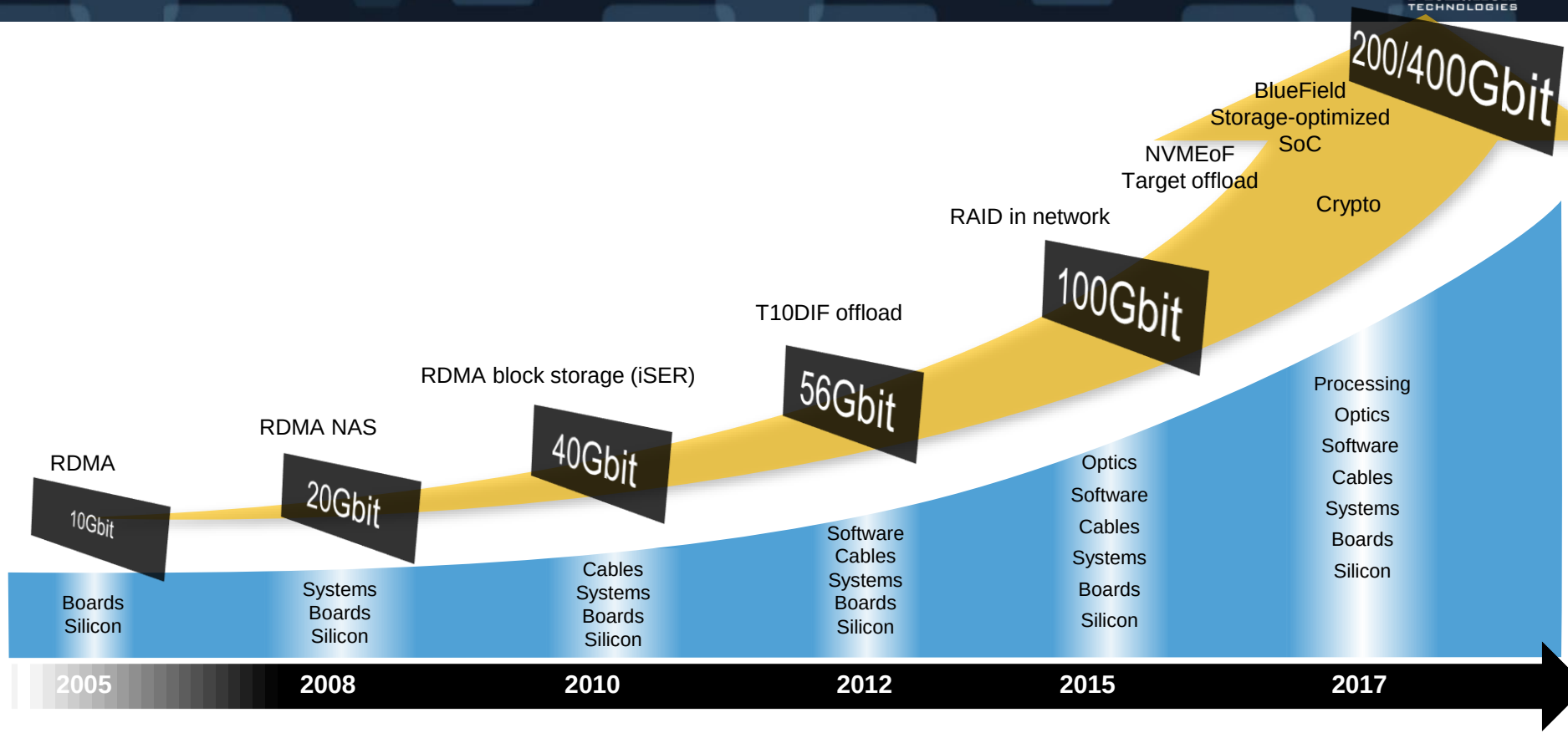
Storage Class Memory (SCM)



SCM Over Fabrics Definition WIP in IBTA

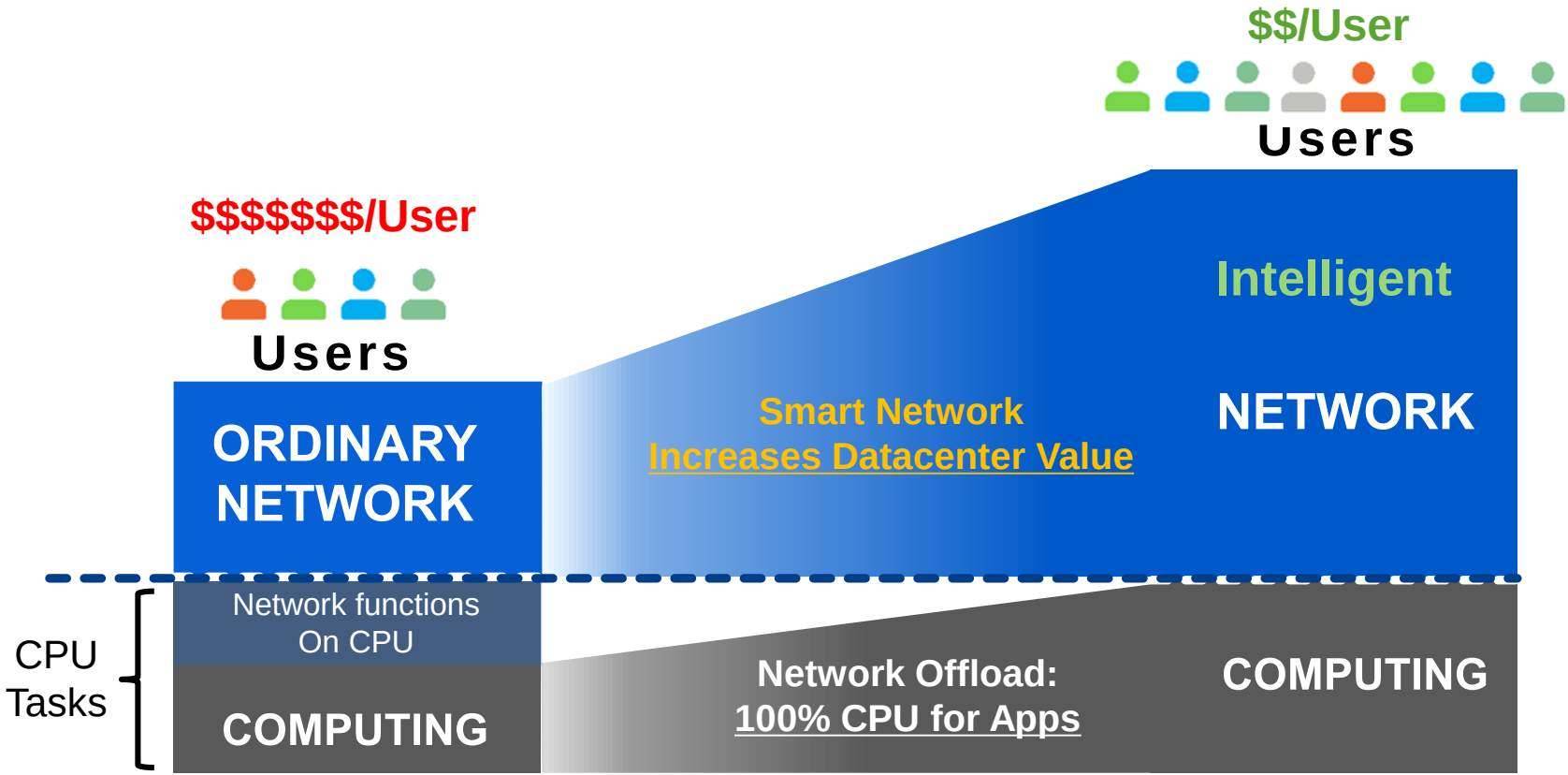
© 2017 Mellanox Technologies

Technology Leadership



© 2017 Mellanox Technologies

Network is a Computer – Deliver Value



Questions?

Manoj Wadekar

ebay

Michael Kagan

Mellanox TECHNOLOGIES