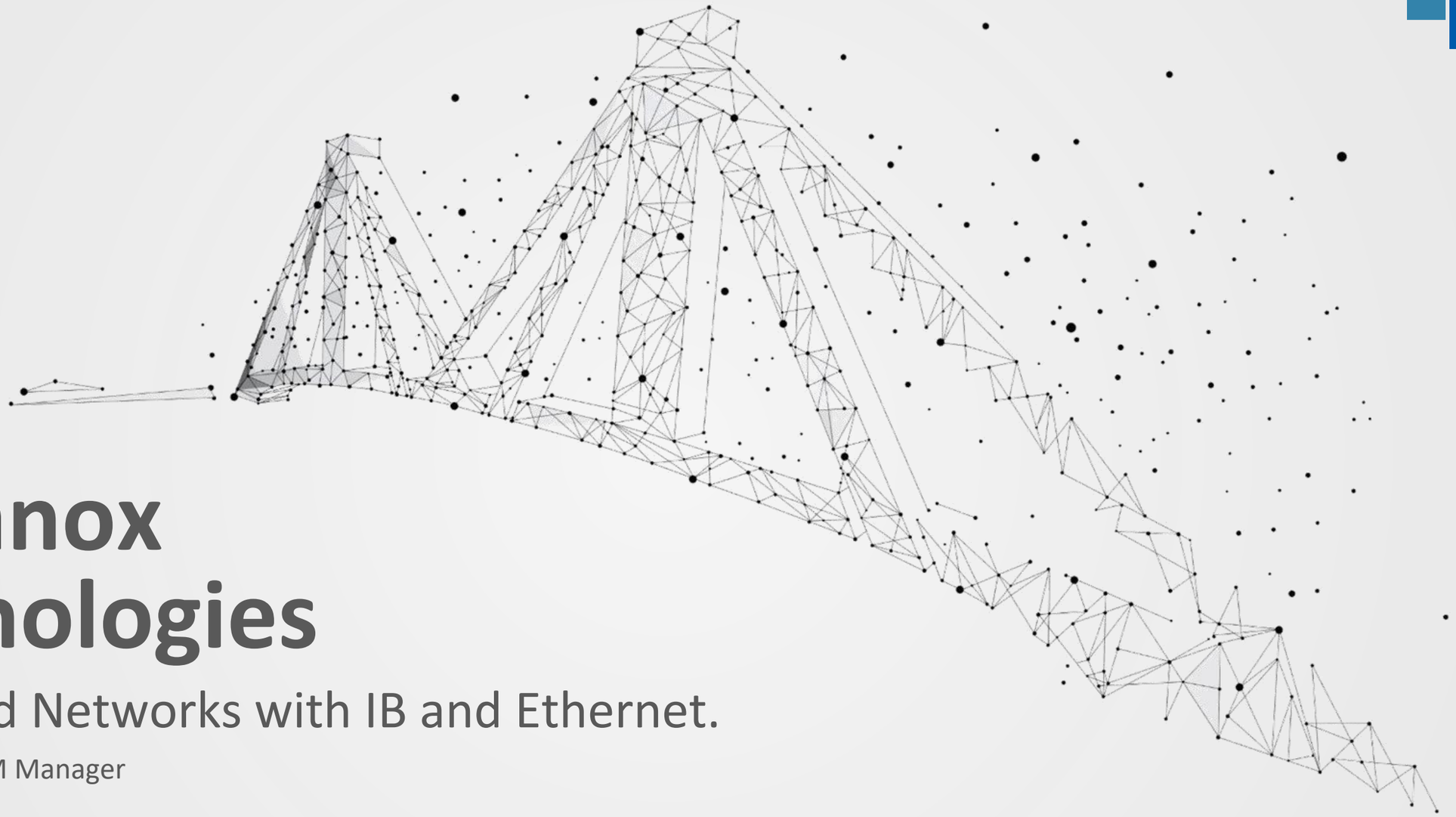
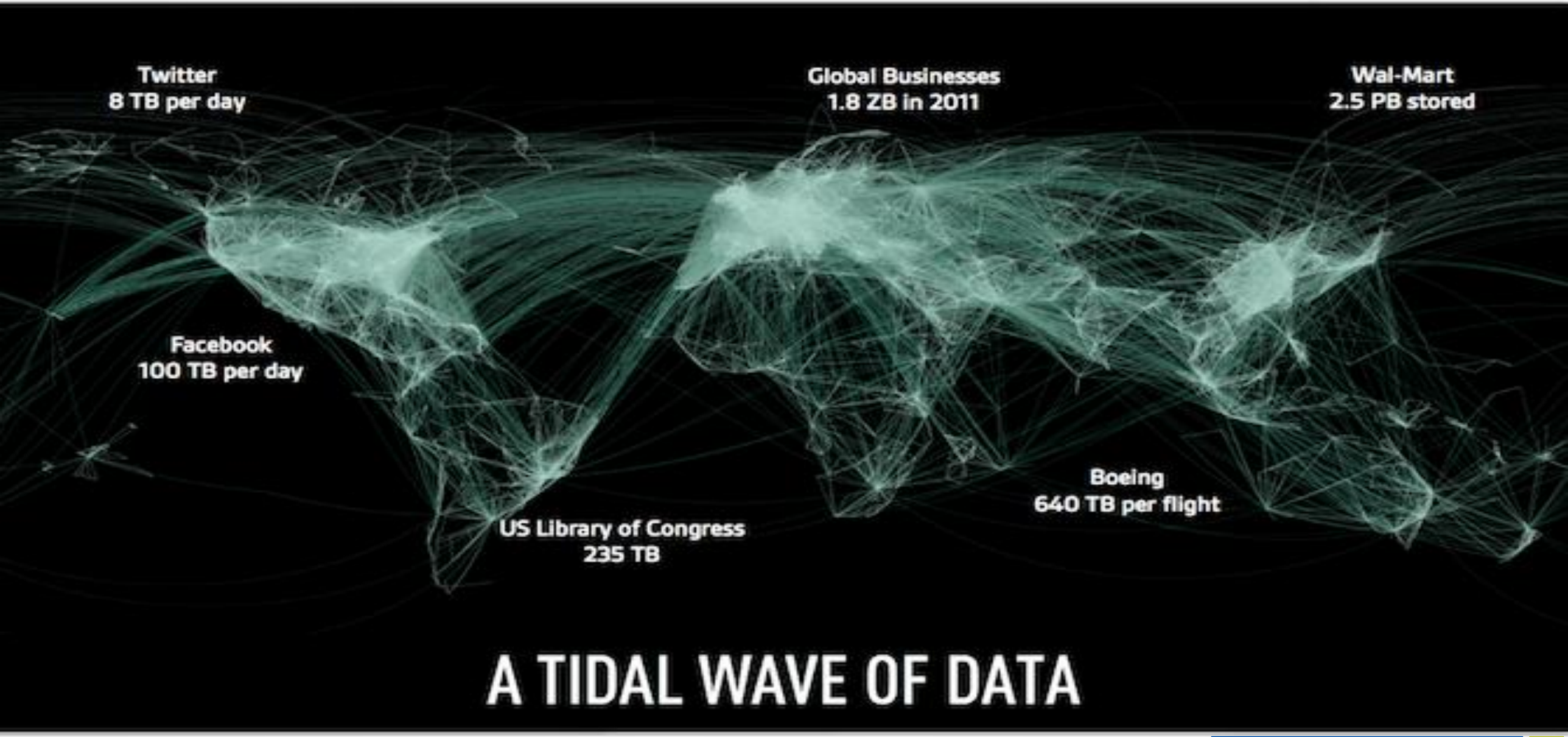


Mellanox Technologies

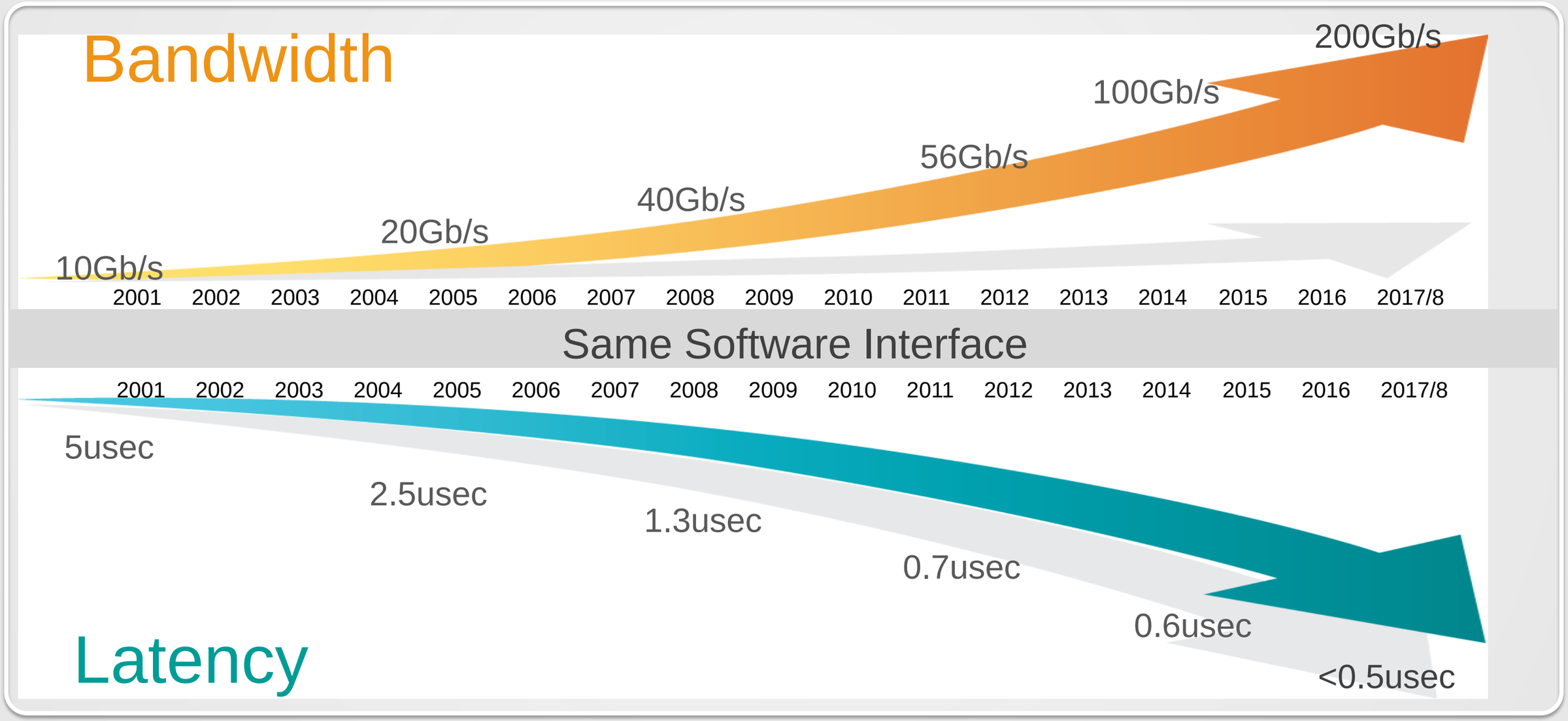
High Speed Networks with IB and Ethernet.

Hilmar Beck | OEM Manager

“90% of the World Data was Created in the Last 2 Years”



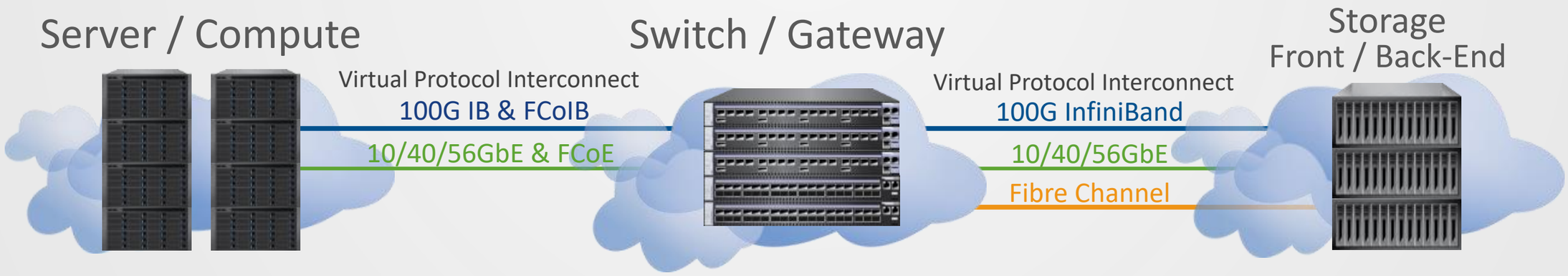
Leading Interconnect, Leading Performance



Infiniband or Ethernet?

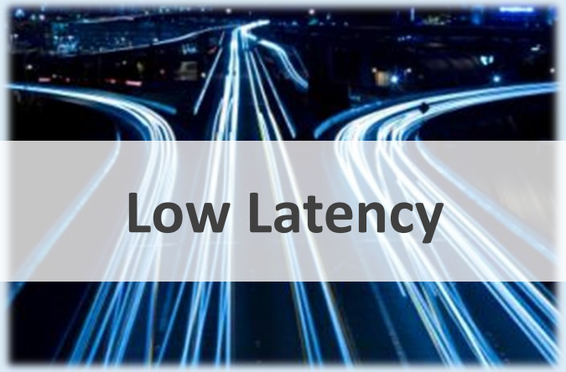
What is InfiniBand?

- Interconnect technology for interconnecting processor nodes and I/O nodes to form a system area network
- The architecture is independent of the host operating system (OS) and processor platform
- Open industry-standard specification
- Defines an input/output architecture
- Offers point-to-point bidirectional serial links
- IB enables a scaling of up to 48,000 Nodes in a single subnet



InfiniBand Key Features

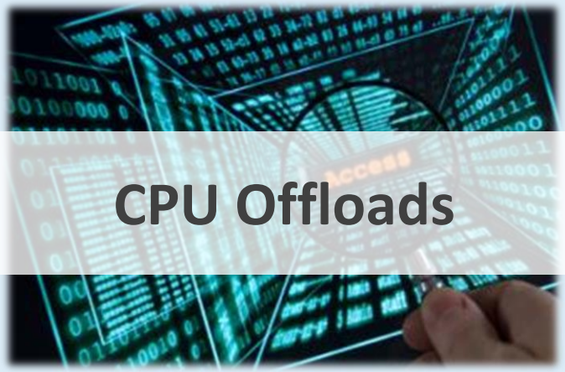
High Bandwidth



Low Latency

Quality of Service

Scalability/Flexibility



CPU Offloads

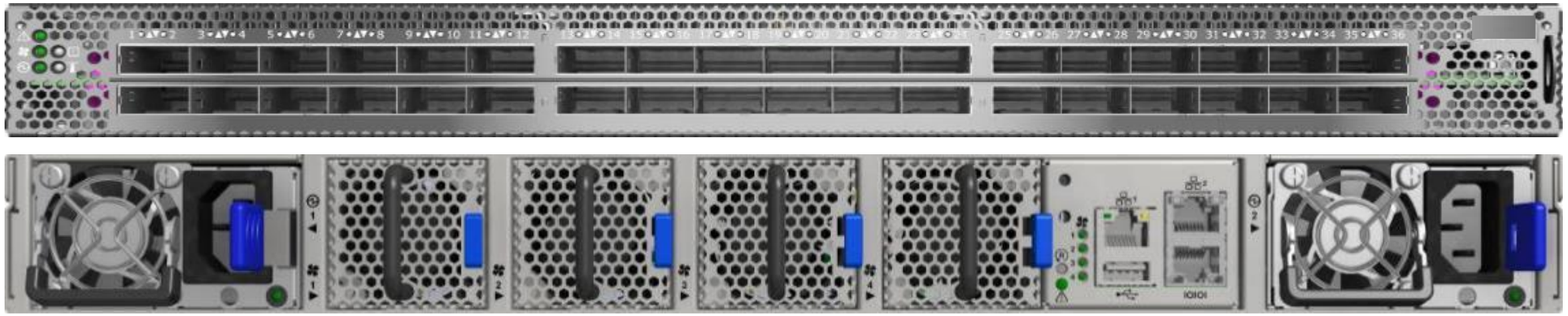
Simplified Management

SwitchIB™-2

- 36-ports QSFP28 EDR switch
  - LR4 (class 4) support
  - Same SwitchIB chassis - 1U height / 19’’ wide / 27’’ deep

- Software and image compatible with SX6036 (FDR) and SB7700 (EDR) switch systems
  - Same latency as SwitchIB

- **Collectives offload**
  - 5X reduction for MPI collective runtime ➔ 2u MPI E2E instead of ~10u today
  - Increase CPU availability and efficiency



Shattering The World of Interconnect Performance!

SwitchIB EDR 100G InfiniBand		
	Latency [ns]	BW[Gb/s]
@EDR Speed (Strong FEC enabled)	90	~100
@FDR Speed (LLR enabled, requires FDR systems FW upgrade)	130	52.5
@QDR Speed	114	31.3
InfiniBand Message Rate	149.5 Million/sec	

IBM Deep Learning Library with ResNet-50

IBM Research DDL software achieved an efficiency of 95% using Caffe

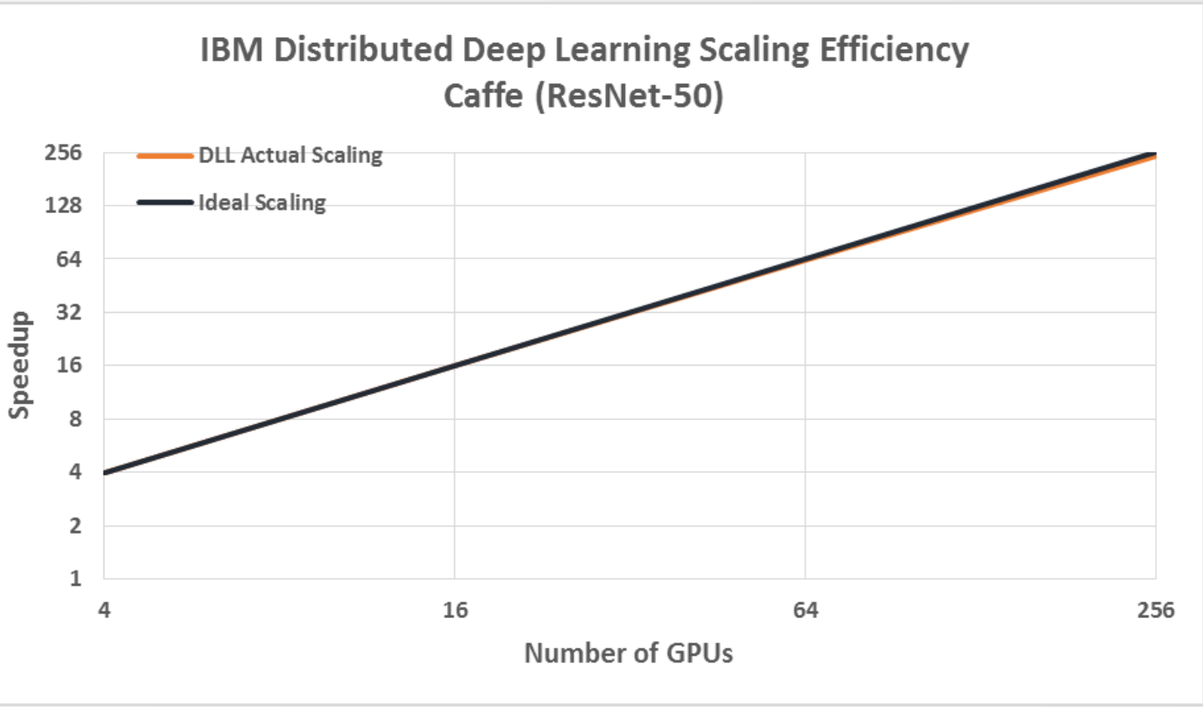
(64) “Minsky” Power S822LC systems

(4) NVIDIA Tesla P100 GPUs

IBM PowerAI

Mellanox InfiniBand

OpenPOWER Ecosystem



Mellanox InfiniBand Provides 95% Linear Scalability

The Ethernet Fabric That Fits Your Needs

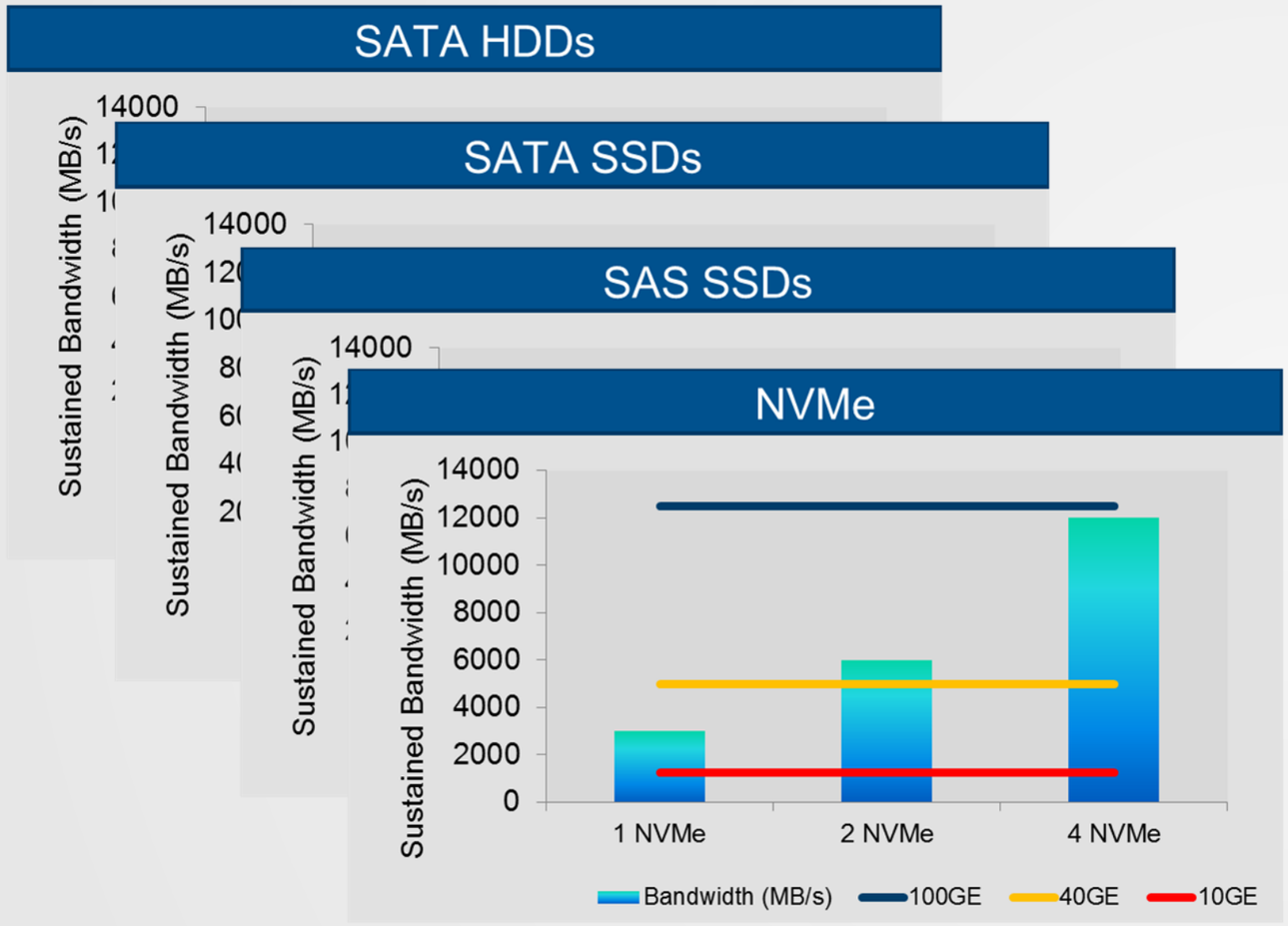
Storage Landscape

STORAGE ICEBERG

- **PRIMARY STORAGE**
- Traditional SAN
- Mission Critical

- **SECONDARY STORAGE**
- 80% of data
- Fragmented, inefficient
- Complex management
- Dark Data

New Storage Media Require Faster Networks



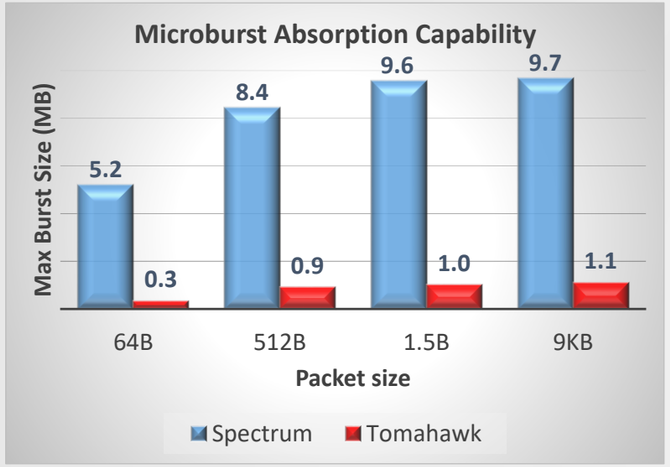
- Transition to faster storage media requires faster networks
- Flash SSDs move the bottleneck from the storage to the network
- What does it take to saturate one 10Gb/s link?
  - 24 x HDDs
  - 2 x SATA SSDs
  - 1 x SAS SSD
  - NVMe…

Ethernet Storage Switches vs. DC Switches

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

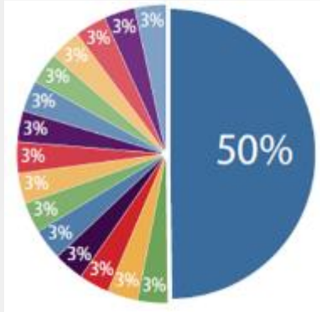
- ✓ None - it is a DC switch☹

Superior ASIC, Critical for Storage/HCI

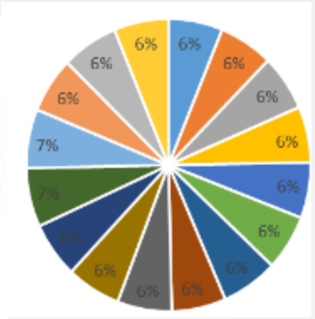


Microburst Absorption

Competition

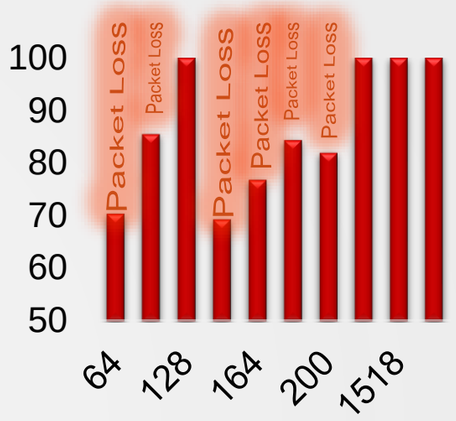


Spectrum

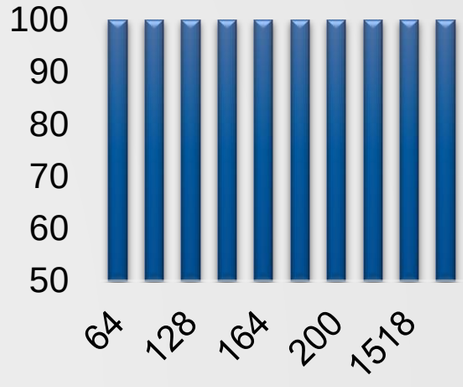


Fairness

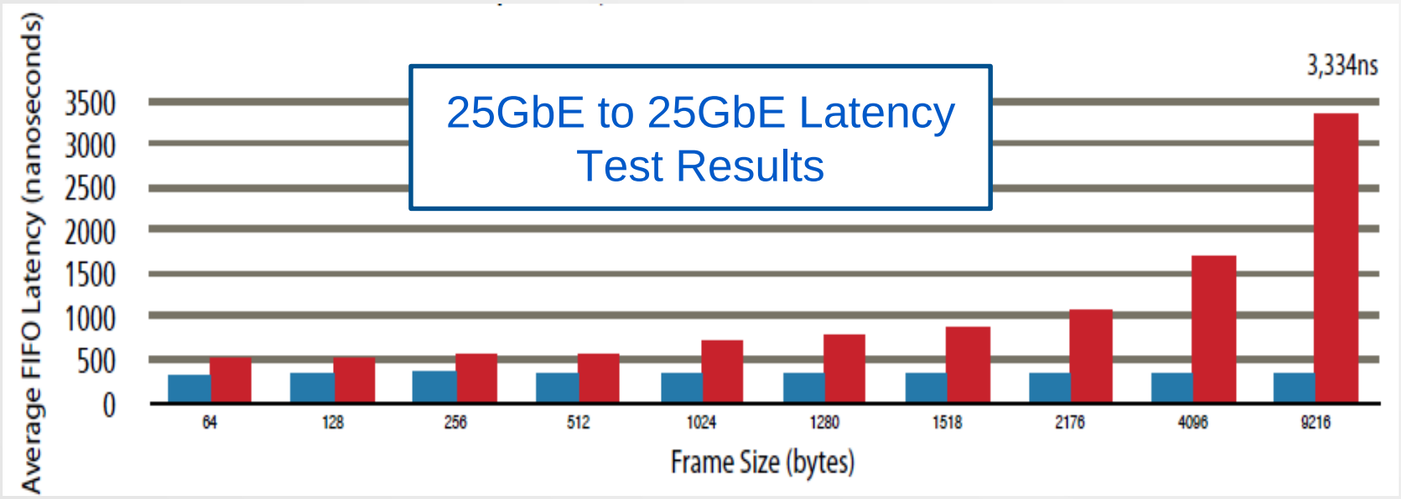
Competition



Spectrum



Avoidable Packet Loss



www.Mellanox.com/tolly
www.zeropacketloss.com

Spectrum™

Tolly.

Open Ethernet 10/25/40/50/100G Switch Portfolio

SN2010

Optimized 10/25GbE ToR for HCI and storage

- ½ width ToR
- 18x10/25GbE + 4x40/100GbE
- Supports 1GbE ports

SN2100

Ideal high-speed ToR for HCI and storage

- ½ width ToR
- 16x 40/100GbE
- 32x 50GbE or 64x 10/25GbE
- Supports 1GbE ports

SN2410

10/25GbE ToR for servers and storage

- 48x 10/25GbE + 8x 40/100GbE
- Supports 1GbE ports

SN2700/SN2740

40/100GbE aggregation for servers and storage

- 32x 40/100GbE
- 64x 10/25/50GbE
- Supports 1GbE ports

CUMULUS

MLNX-OS Mellanox OS

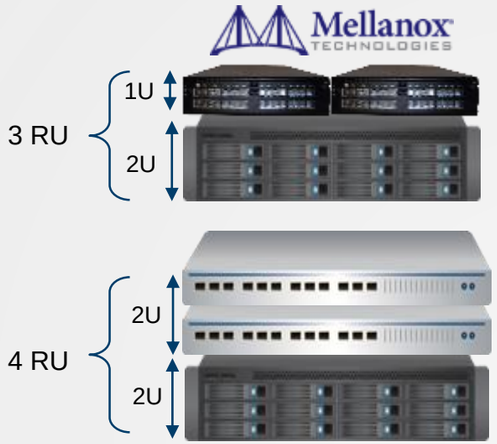
Spectrum™

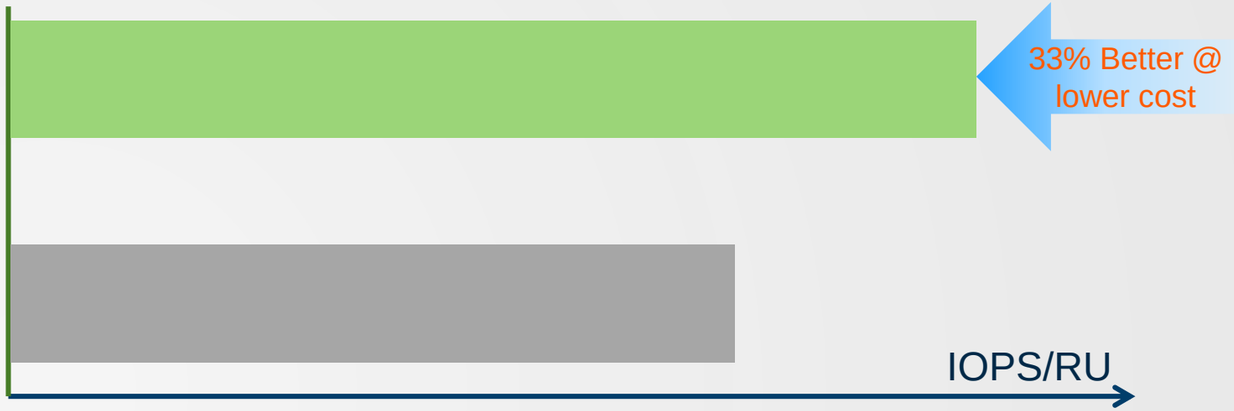
SN2100 Use Cases

Use Case

Mellanox Ethernet Alliances Portal

Storage Solution Performance Efficiency





High Performance Databases

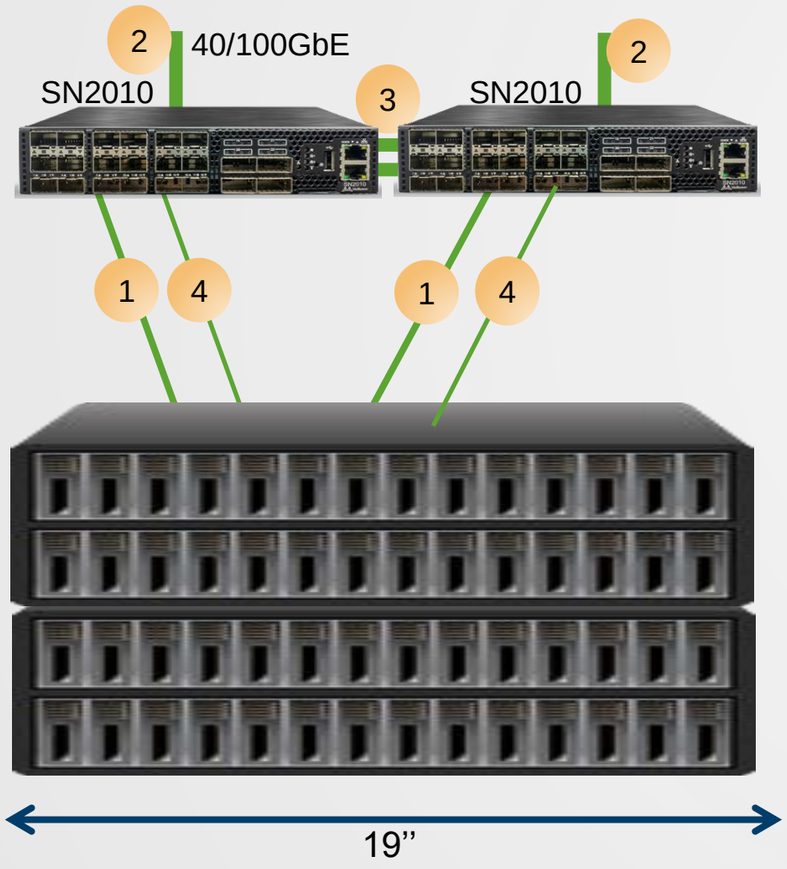
Cloud Infrastructure

4K+ Media Production

SN2010 – Optimized 10GbE Storage/HCI ToR

Mellanox Storage Rack Design

(1/10/25/40/100GbE supported)



1 rack = 18 nodes

1. 10/25GbE link: SFP+/28 to SFP+/28
2. 40/100GbE uplink: QSFP/28 to QSFP/28
3. 100GbE mLAG: QSFP28 to QSFP28
4. 1GbE link: 1GbE transceiver

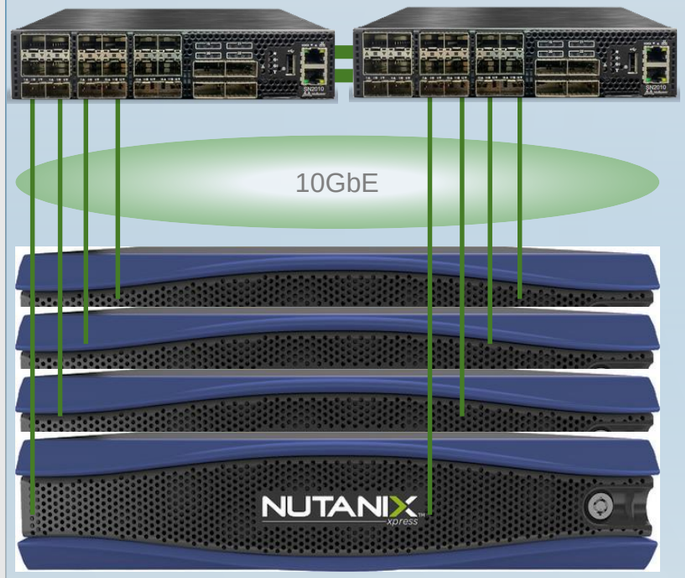
Branch Offices

Private Cloud

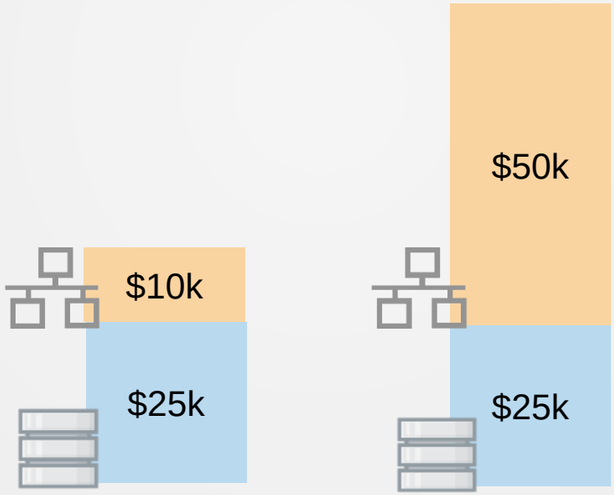
High Performance Database

SN2010 Advantage – Nutanix Xpress Example

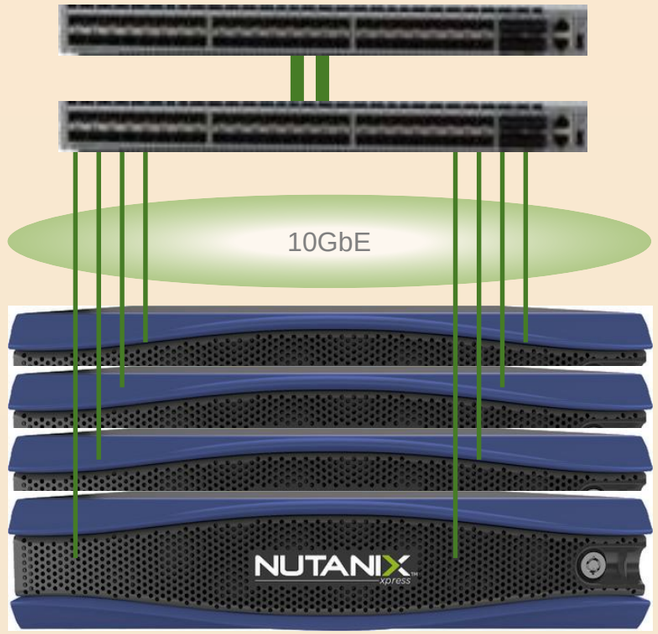
SN2010
New Better Type of ToR



$10k	Cost	$50k (5x)
1RU	Size	2RU
Spectrum	ASIC	Trident+

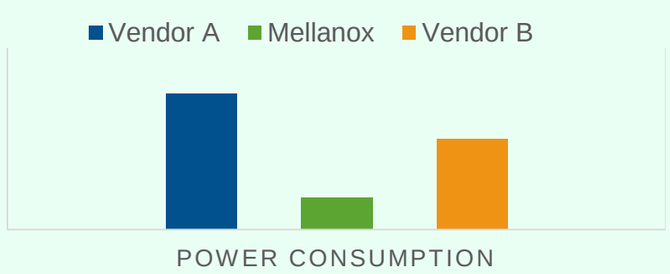


Competition
48*10G + 4x40G



Efficient and reduced size network lowers the cost by up to 80%

100GbE Virtual Modular Switch® for 25GbE Systems

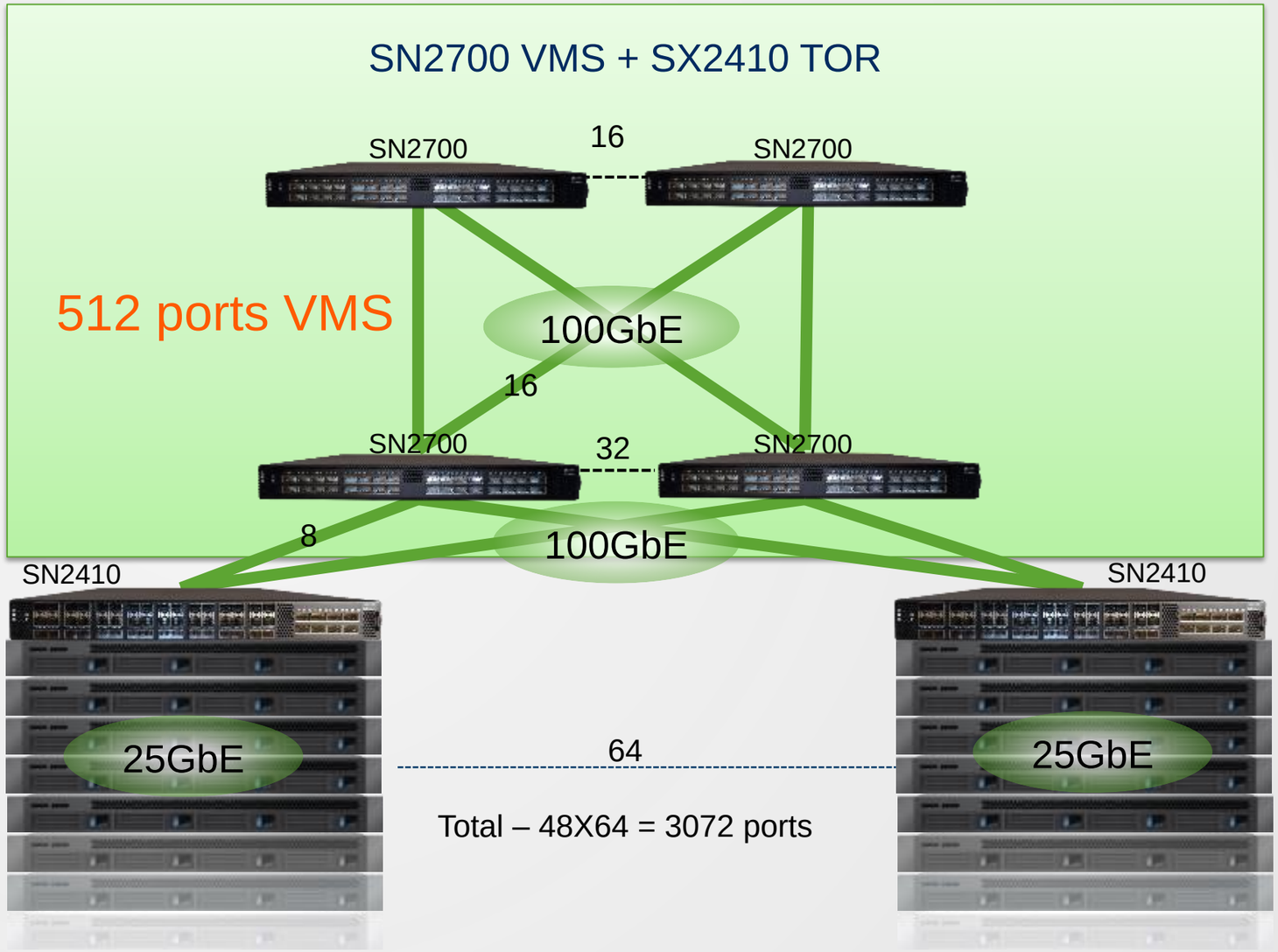


Cost Effective
- Under $1.3k per 100GbE port

Standard L3 Scale-out
- ECMP over OSPF/BGP

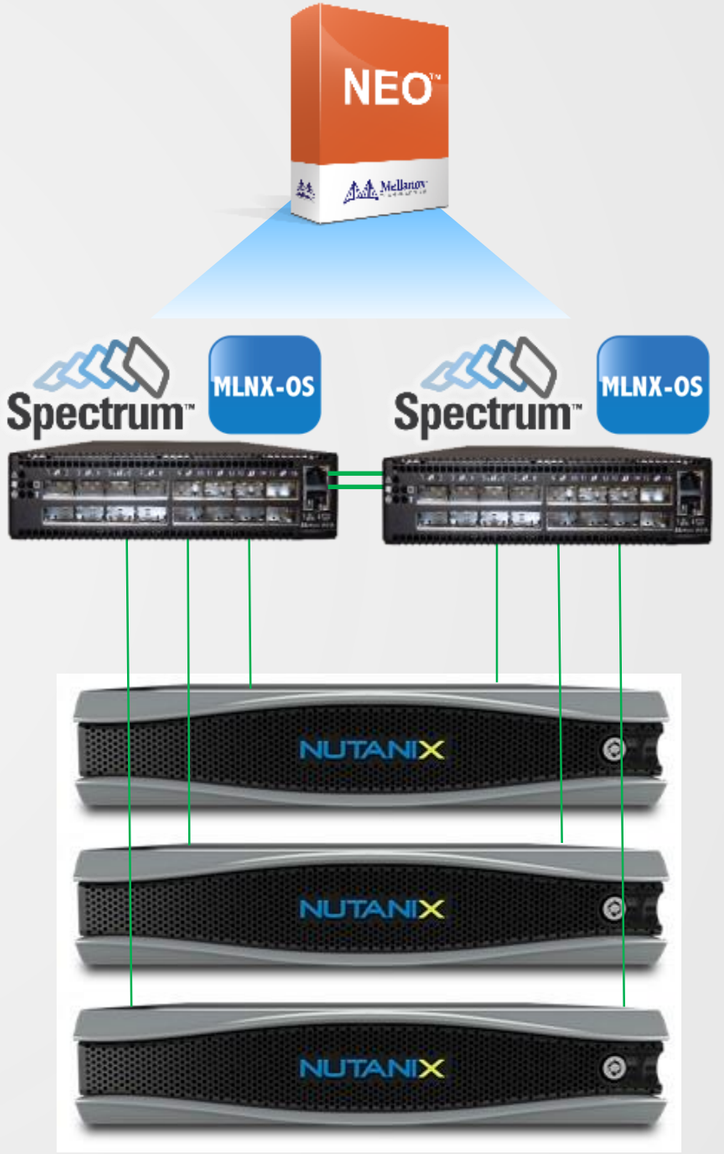
Automation
- *Configured in minutes (VMS Wizard)*

Flexible
- *10/25/40/50/100GbE ports*



Invisible Networking with Mellanox

- Integration adds network automation to Nutanix VM life-cycle:
  - VLAN auto-provisioning for VM creation.
  - VLAN auto-provisioning for VM migration.
  - VLAN auto-provisioning for VM deletion.
- Nutanix Ready integrated solution
- Fully transparent to the user – API-to-API integration.



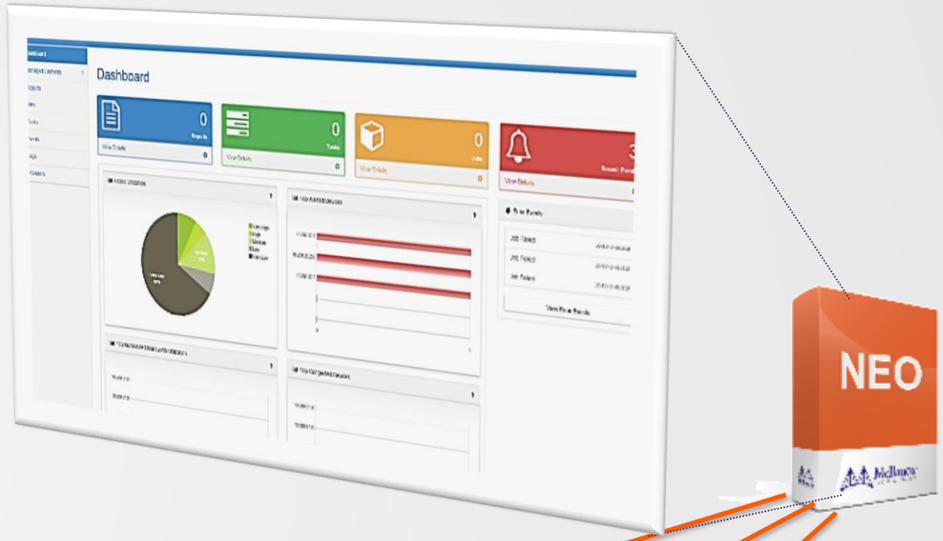
NEO™ Makes Networking Simple and Agile

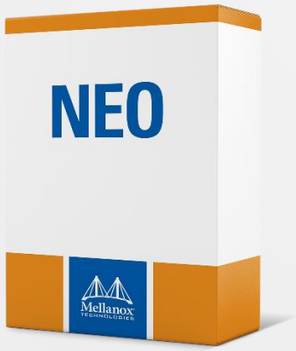
Fabric Bring-up Made Easy

Efficient Network Monitoring

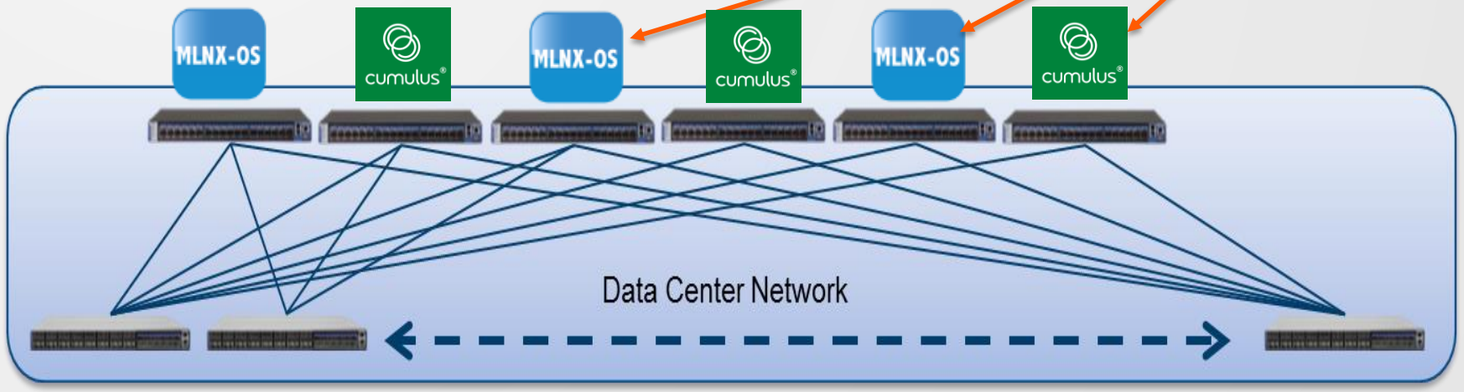
Save Time on Day-to-Day Operations

Network API, Built for Integration





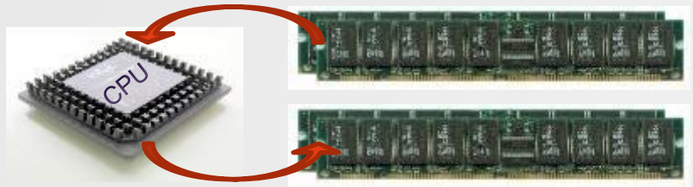
- ✓ MLNX-OS & Cumulus Linux
- ✓ End-to-End RoCE Automation
- ✓ Auto-Provisioning with Nutanix and OpenStack



Infiniband and Ethernet!

Remote Direct Memory Access

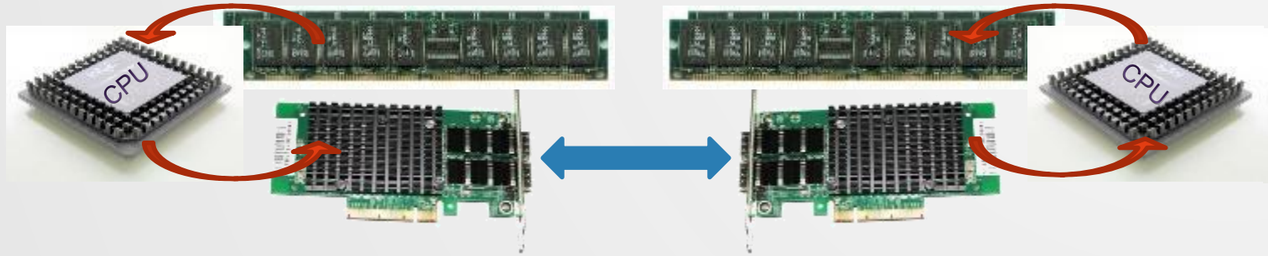
- *Direct Memory Access*
  - A CPU offloading engine that copy memory blocks from one address to another



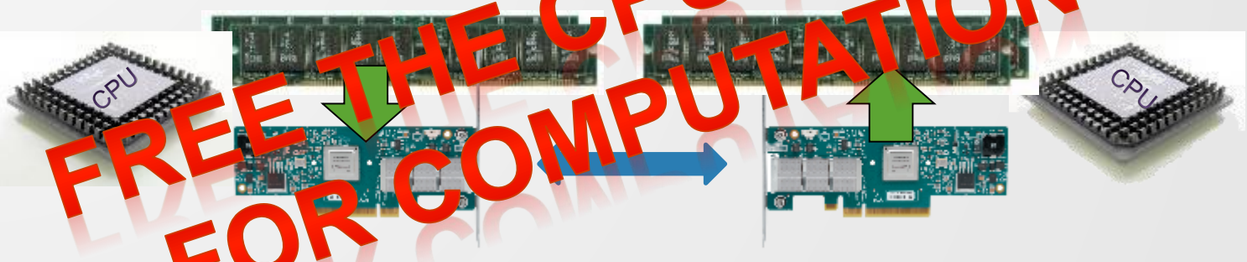
BEFORE

AFTER

- *Remote* Direct Memory Access
  - Offload the copy of memory blocks from REMOTE machine



BEFORE



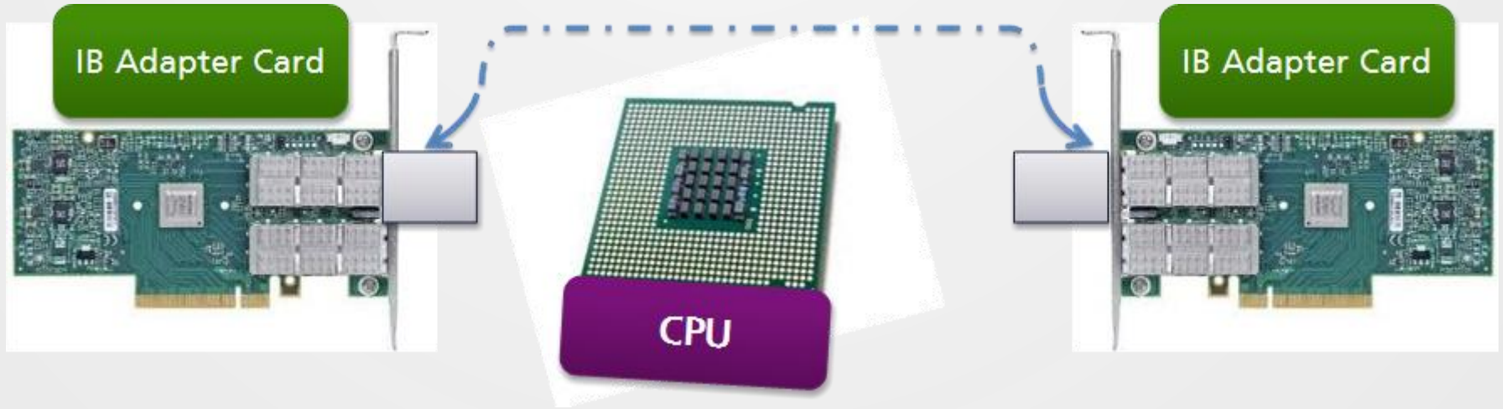
FREE THE CPU CORES FOR COMPUTATION

AFTER

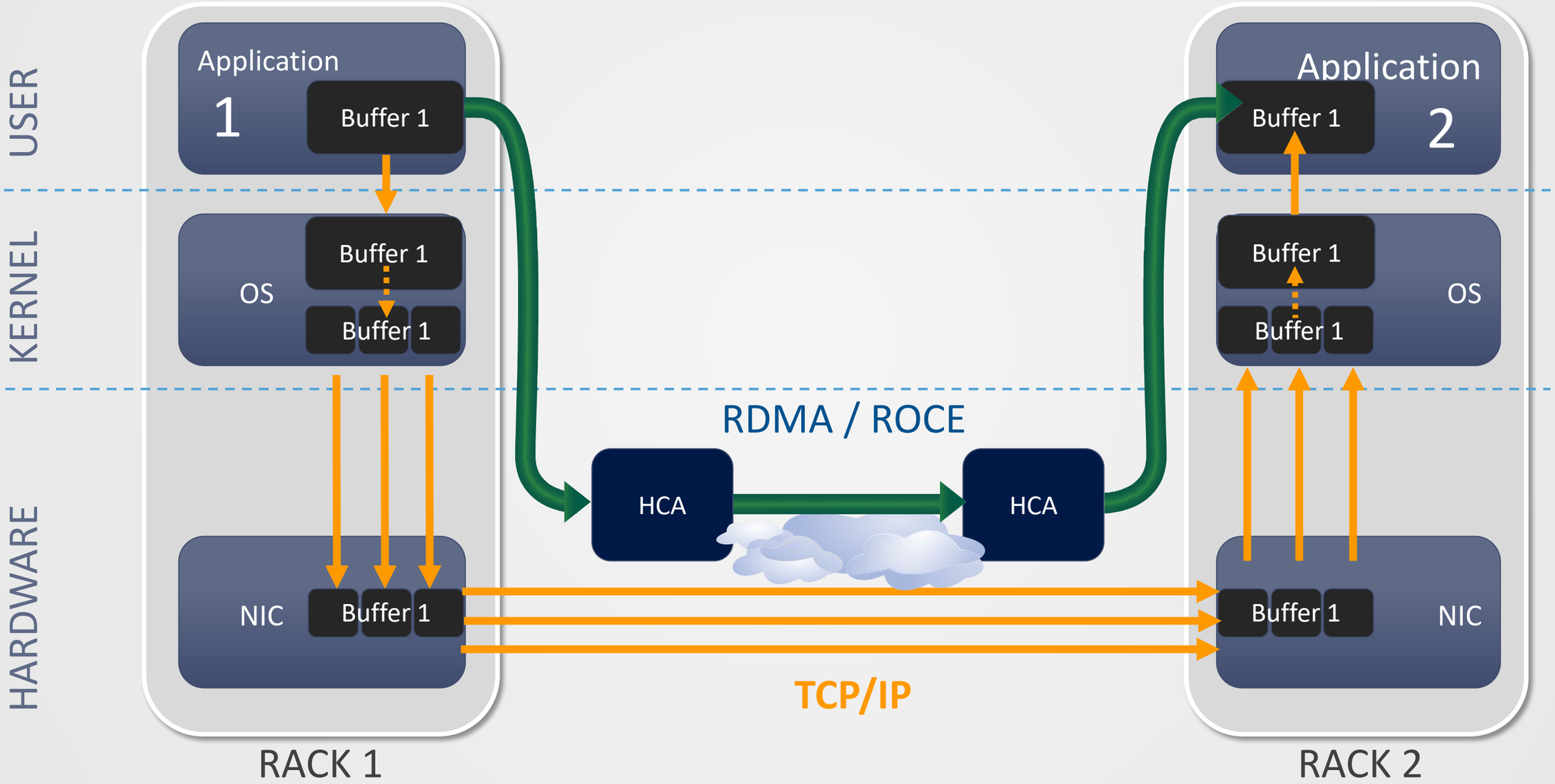
CPU Offloads

- The IB architecture supports packet transportation with minimal CPU intervention
- This is achieved thanks to:
  - Hardware-based transport protocol
  - Kernel bypass
  - Reliable transport
  - RDMA support

Mellanox adapter cards

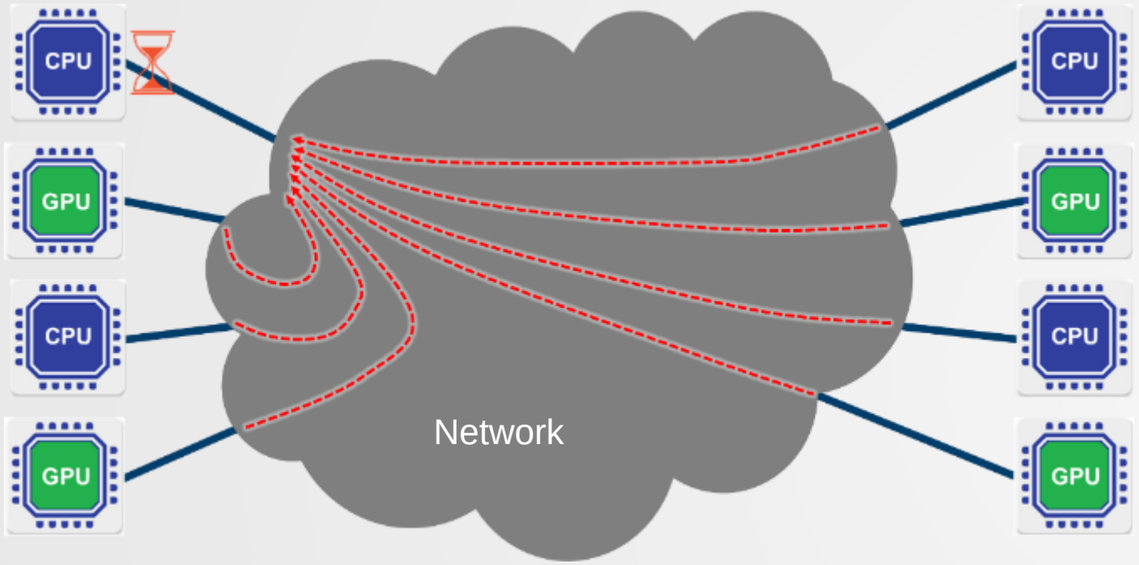


RDMA – How does it Work



Data Centric Architecture to Overcome Latency Bottlenecks

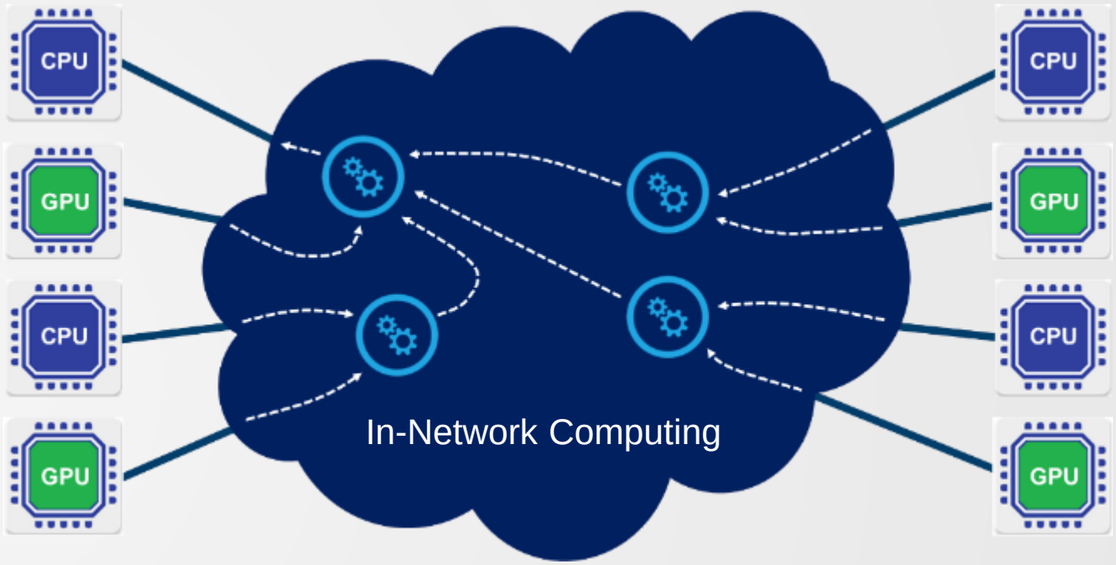
CPU-Centric (Onload)



Machine Learning

Communications Latencies of 30-40us

Data-Centric (Offload)



Machine Learning

Communications Latencies of 3-4us

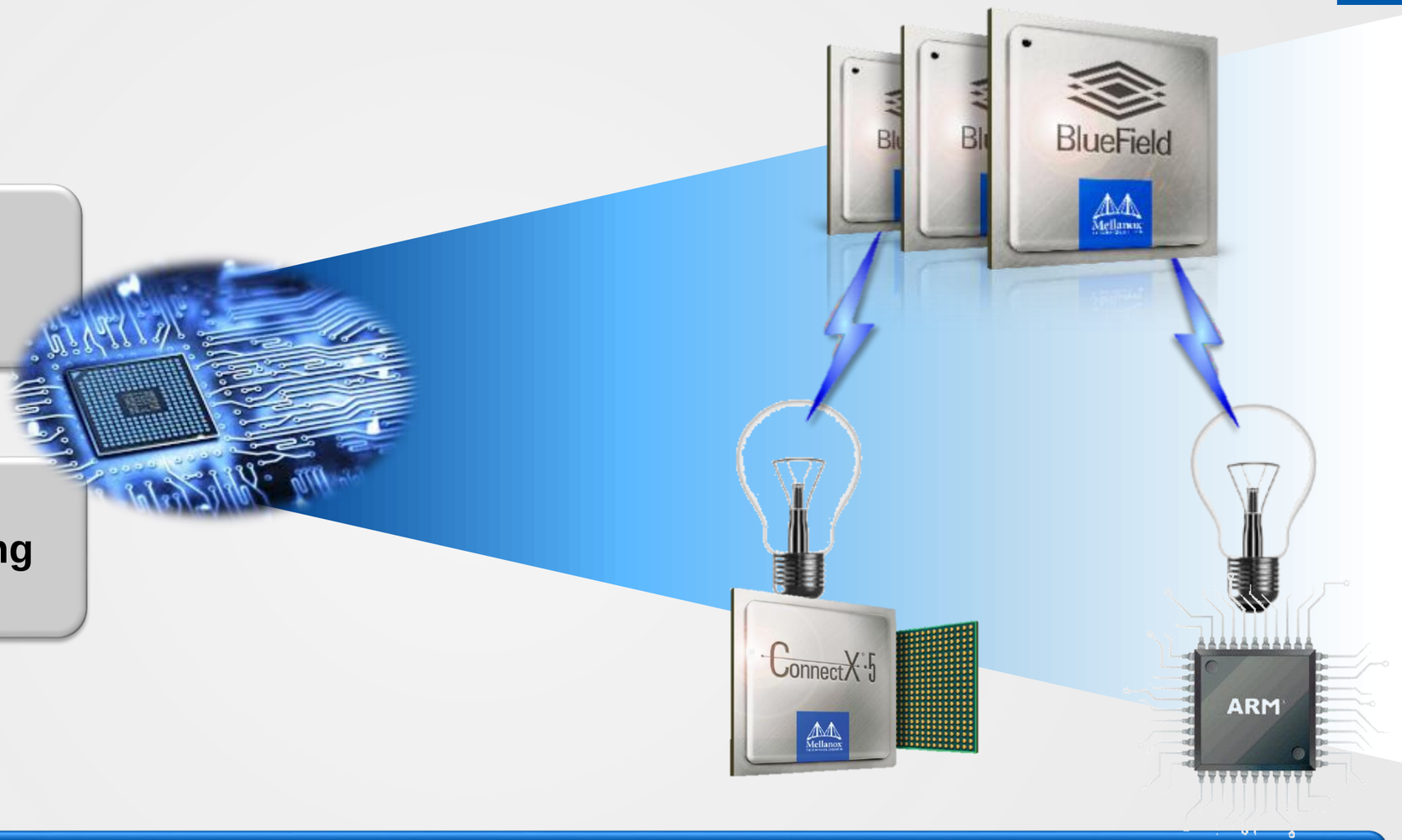
Intelligent Interconnect Paves the Road to Exascale Performance

Smart NICs

BlueField: Delivering Advanced Networking

Powerful Adapter Capabilities

Powerful Core Processing



Providing Tomorrow's Software Defined Controllers

BlueField Product Family Enables Growing Market Segments

Storage

Enabling the Path to NVMe-over-Fabrics Storage Platforms

- NVMe Flash Storage
- Scale-Out Storage Array
- Flexible Storage HBA

Security

Data Center Infrastructure Protection

- IDS/IPS
- Anti-DDoS
- Firewall
- Crypto Protocols

Machine Learning

The Most Efficient Machine Learning Systems

- Network-attached GPUs and Neural Network Processors

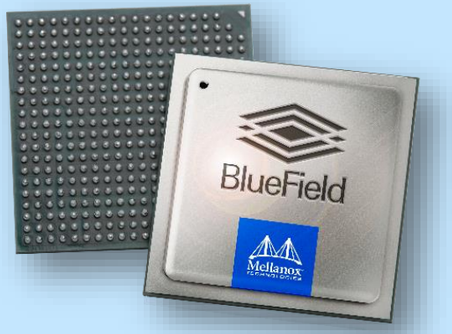
NFV

Performance and Flexibility at Lower Cost

- VNF Acceleration
- Open vSwitch (OVS), SDN
- Overlay Networking

BlueField Product Line

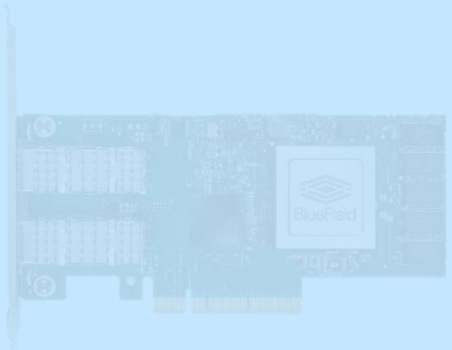
A solid product line in different form factors



BlueField IC

Dual Port 100Gb/s Controller Card

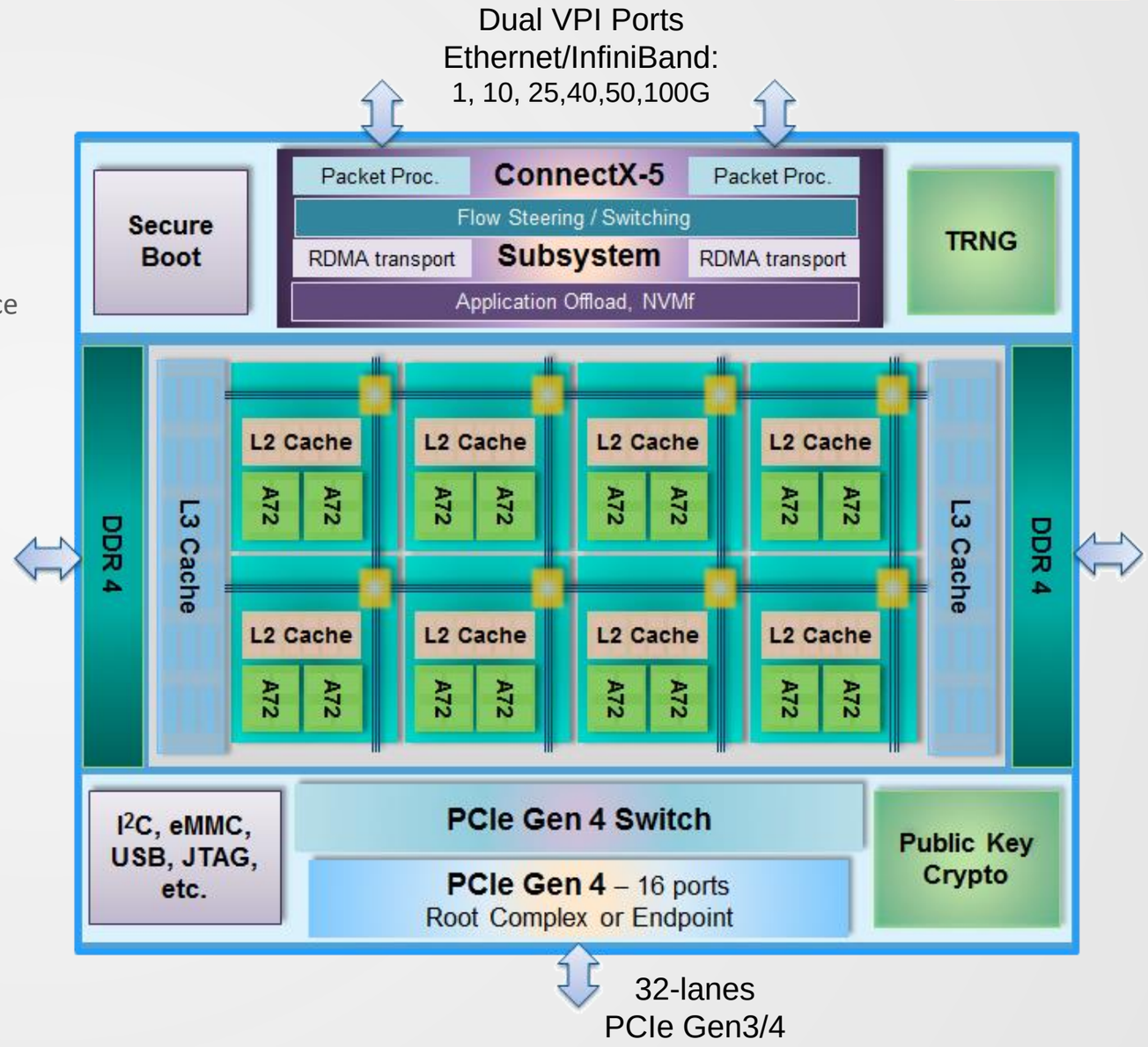
Reference Platform



SmartNIC

BlueField Block Diagram

- Tile Architecture - 16 ARM® A72 CPUs
  - SkyMesh™ fully coherent low-latency interconnect
  - 8MB L2 Cache, 8 Tiles
  - 48KB I-Cache, 32KB D-Cache per core
  - 12MB L3 Last Level Cache
  - ARM Frequency: 1.1GHz for Effective flavor, 1.3GHz for High Performance
- Integrated ConnectX-5 subsystem
  - Dual 100Gb/s Ethernet/InfiniBand, compatible with ConnectX-5
  - NVMe-oF hardware accelerator
  - High-end Networking Offloads: RDMA, Erasure Coding, T10-DIF
- Fully Integrated PCIe switch
  - 32 Bifurcated PCI Gen3/4 lanes (up to 200Gb/s)
  - Root Complex or Endpoint modes
  - 2x16, 4x8, 8x4 or 16x2 configurations
- Crypto Engines
  - Bulk crypto by A72 Neon ISA (AES, SHA)
  - Public Key acceleration, True RNG
- Memory Controllers
  - 2x Channels DDR4 Memory Controllers w/ ECC
  - NVDIMM-N Support



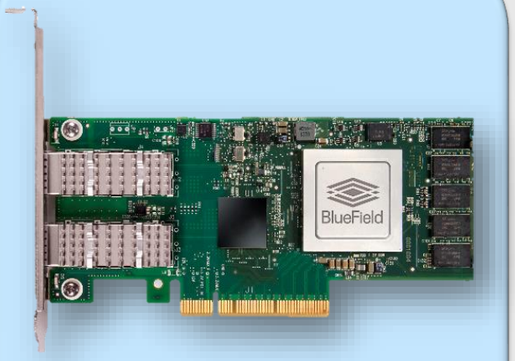
BlueField Product Line

Full Product Range – Shorten Time to Market

BlueField IC

Dual Port 100Gb/s Controller Card

Reference Platform

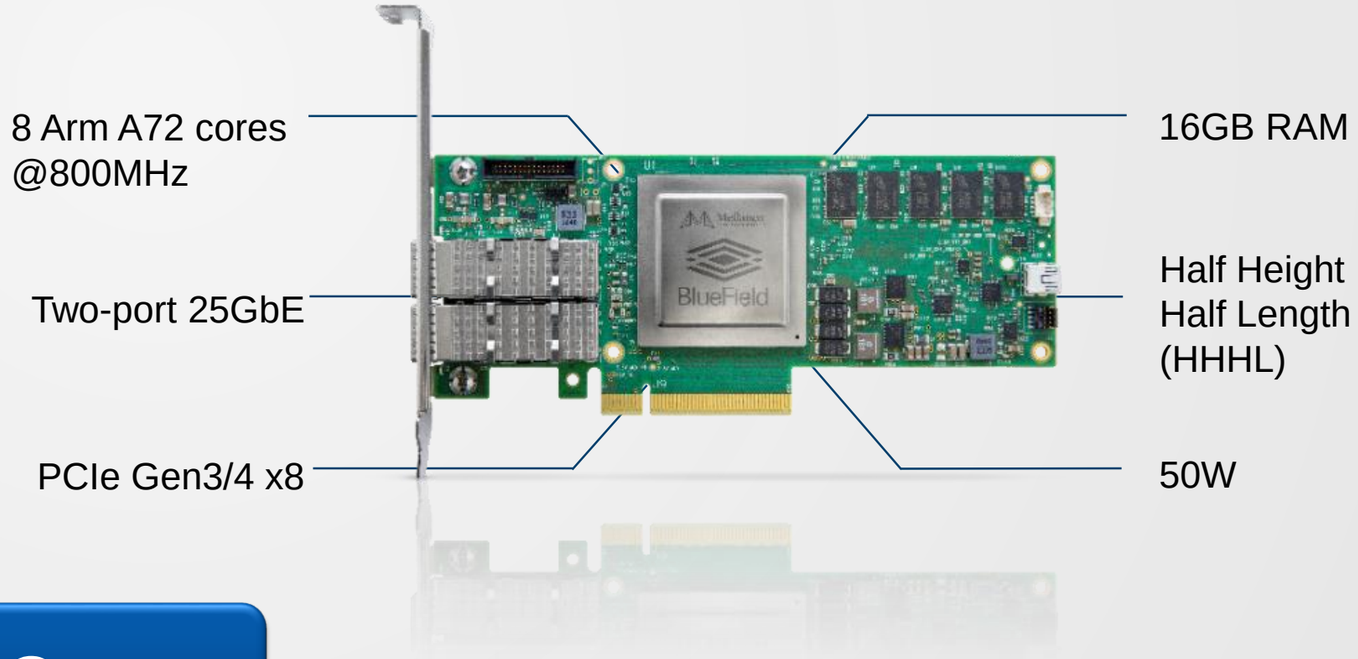


SmartNIC

BlueField SmartNIC

- Combines best-in-class hardware network offloads with ARM processing power
- Accelerates wide range of security, networking and other workloads
- Reduces TCO by offloading main CPU
  - Main CPU is left for compute and applications rather than security or networking functions
- Standard embedded Linux software stack

- Card characteristics
  - PCIe gen3 x8
  - 2-ports 25GbE
  - Up to 16GBytes DDR
  - HHHL



Next Generation Cloud NIC

Why BlueField SmartNIC?

Programmability

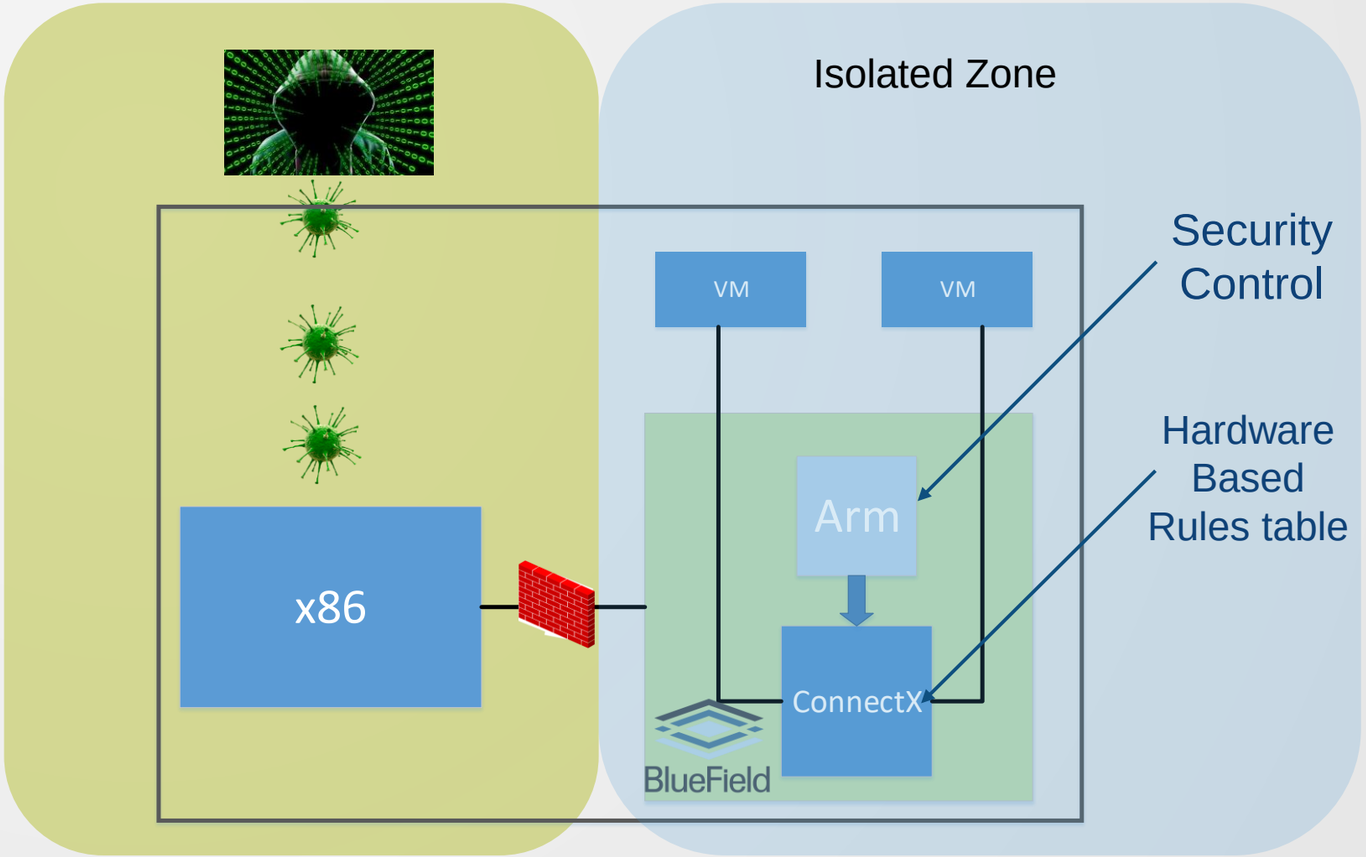
One Size fits all

Isolation

Performance

BlueField SmartNIC: Function Isolation and Security

- Host isolated from security control
- A reduced impact of a breach
- Security moved from ToR to Host
  - Increase scalability
  - Increase Switch performance (less ACLs)
- Host based SDN model improves scalability
- Value add security services for tenants
- No impact of security services on host CPU
- Make use of ASAP^2 Direct (SR-IOV)



Why BlueField SmartNIC: Performance

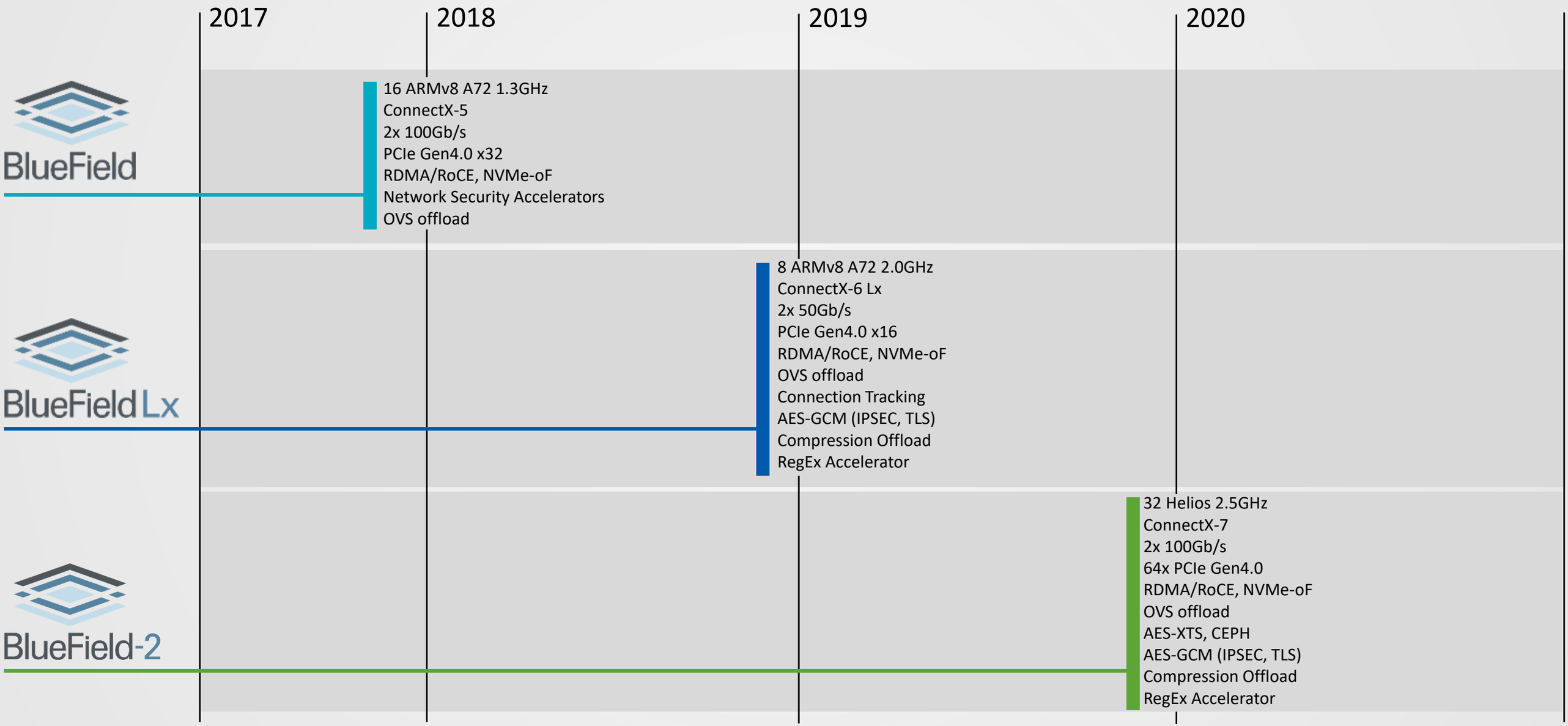
- Zero host CPU utilization for Control or Data
- CPU serves customers\apps instead of itself
- More I/O capacity per server
- Increased infrastructure scalability

BlueField SmartNIC: Programmability

- Customer and 3rd party can add value by adding application on top of Arm
- The Arm allows implementing new data or control plane features on top of silicon capabilities
  - Action or Packet classification not supported by silicon today can be added

[root@BlueField ~]# No Software Limits

BlueField Roadmap - Samples



Innova-2 – The 2nd Generation of Innova SmartNICs

Innova 2 FLEX

Programmability

Innova 2 SEC

Smarter Adapter. Greater Security.

Crypto Offloads

A Configurable Adapter to Fit a Range of Applications

Innova and Innova-2 Flex Adapters Family

Innova flex

Innova2 FLEX

Product	Innova Flex – Direct 40G (POC Only)	Innova-2 Flex – P² 25G	Innova-2 Flex – Open 25G	More to Come
Shell Logic	Direct: In Line Acceleration	Programmable Pipeline	None	
Form Factor	Stand-up HHHL, Active heatsink			
Speed	1-port QSFP 40Gb/s Ethernet	2-port SFP+ 25Gb/s Ethernet		
PCIe	PCIe Gen4.0 x8			
openCAPI, CAPI 2.0	✖	✓	✓	
FPGA	Kintex UltraScale – KU060	Kintex UltraScale+ KU15P		
DDR	2GB	4GB /8GB		
OPN	MNV101512A-BCAT	MNV303212A-ADAT	MNV303212A-ADLT	

Innova-2 – Programmable Adapters

Encryption Offload

Network Offload

RDMA

Storage Offload

Peer Direct

OVS Offload

Innova2

Smarter Adapter. Greater options.

Ethernet and InfiniBand

ConnectX-5

4GB-8GB DDR4 Memory

Xilinx Kintex UltraScale+

PCIe Gen4 x8

25Gb/s

100Gb/s Interfaces

Series of Programmable Adapters with Flexible FPGA Architecture

Thank You