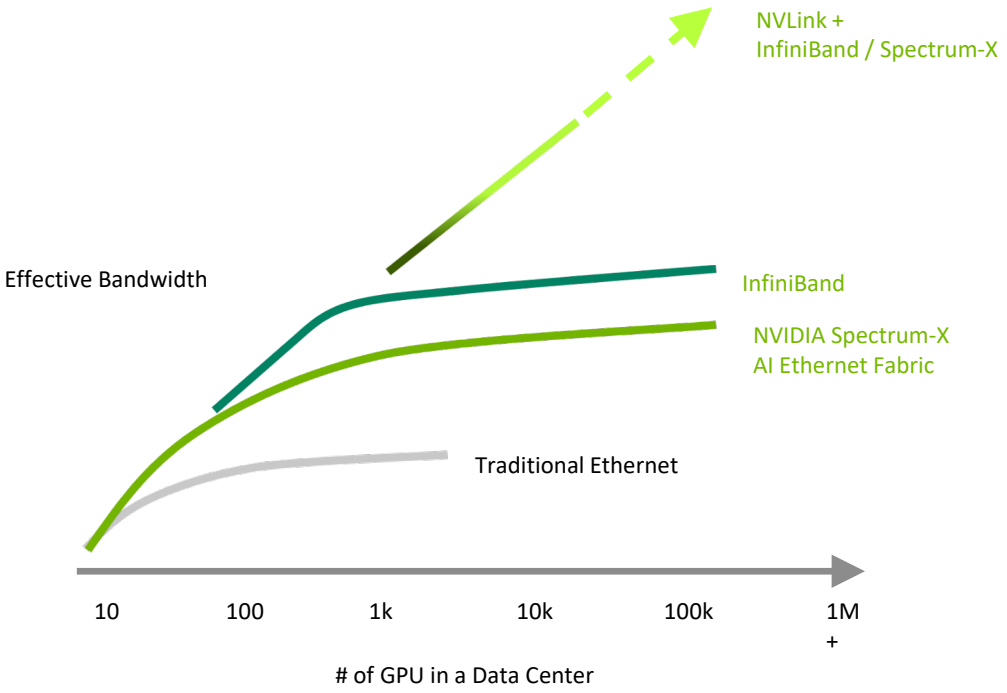
NVIDIA Networking for AI

October 2024

Generative AI is a Data Center Scale Computing Problem

AI factories emerge as a new class of generative AI data centers

Purpose-built high performance networking is necessary to effectively scale AI



NVIDIA NVLink

Fastest Compute Fabric Connectivity

NVIDIA InfiniBand

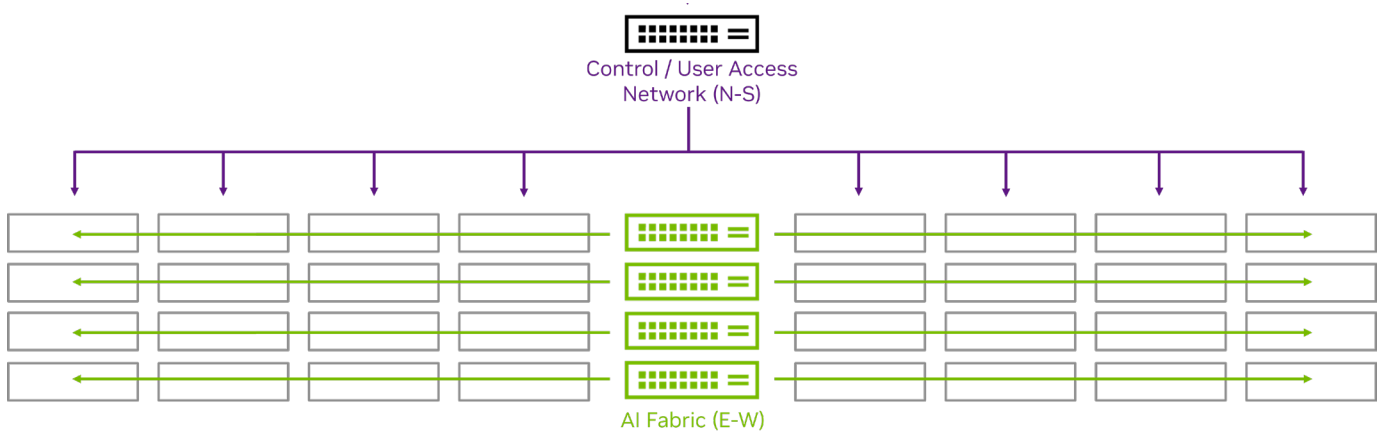
Gold Standard for Scale-Out AI Fabrics

NVIDIA Spectrum-X

Ethernet Optimized for Multi-Tenant AI Factories

Peak AI Performance Demands Optimized Networking

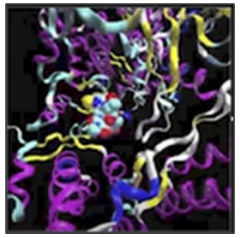
AI Workloads Require an AI Fabric



Control / User Access Network (N-S)		AI Fabric (E-W)
Loosely-Coupled Applications		Tightly-Coupled Processes
TCP (Low Bandwidth Flows and Utilization)		RDMA (High Bandwidth Flows and Utilization)
High Jitter Tolerance		Low Jitter Tolerance
Oversubscribed Topologies		Nonblocking Topologies
Heterogeneous Traffic, Statistical Multi-Pathing		Bursty Network Traffic, Predictive Performance

Running AI Workloads on Traditional Networks

Sub-Optimal for Addressing Needs of AI



AI Workload

Significant Congestion

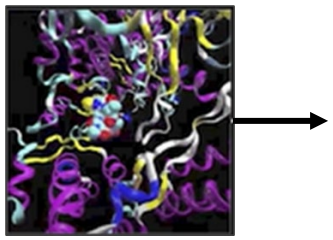
Increased Latency

Bandwidth Unfairness

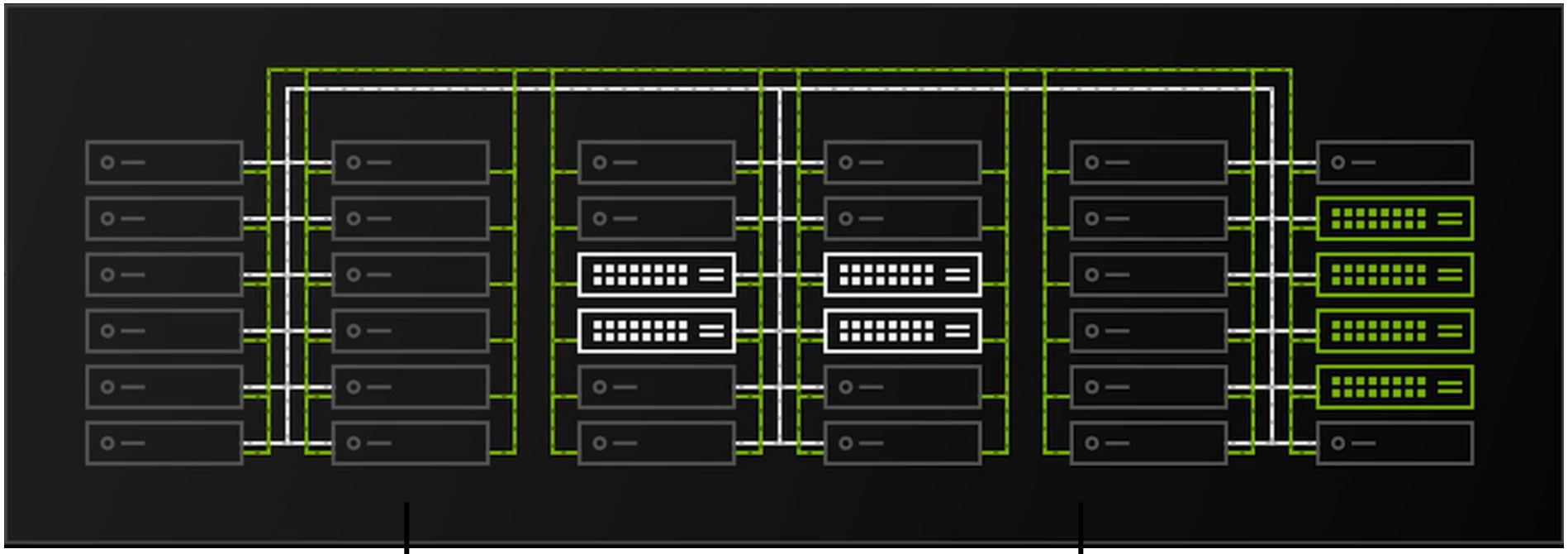
NVIDIA.

NVIDIA Networking Platforms Accelerate AI Workloads

Highest Performance Networking for AI



AI Workload



95%

Highest Effective Bandwidth

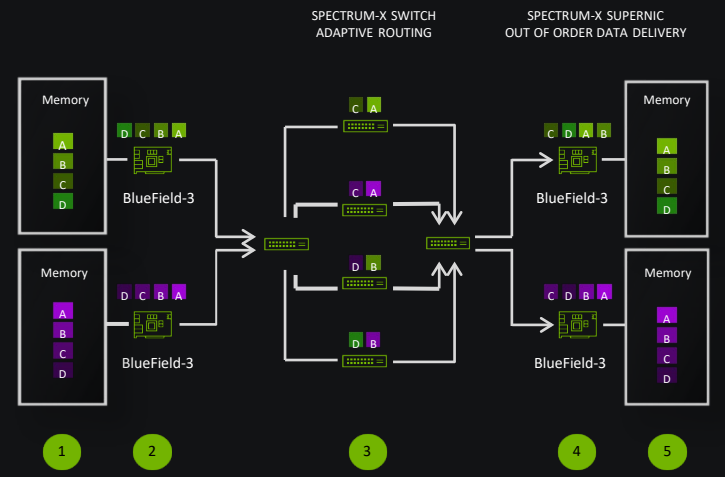
1.6X

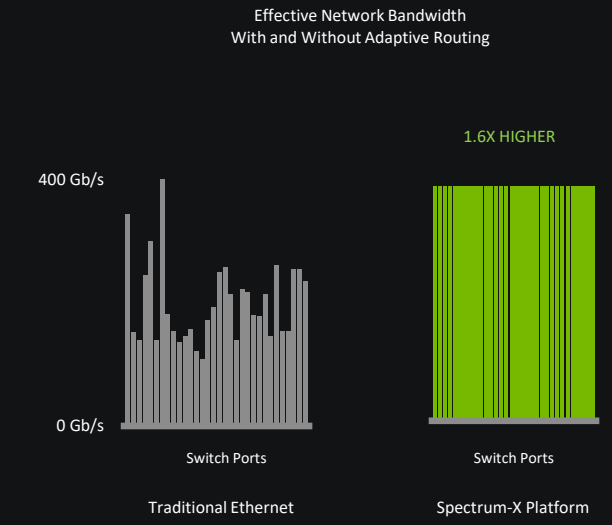
Increased AI Network Performance

End-to-End Adaptive RDMA Routing With Lossless Network

Increases effective data throughout by 1.6X

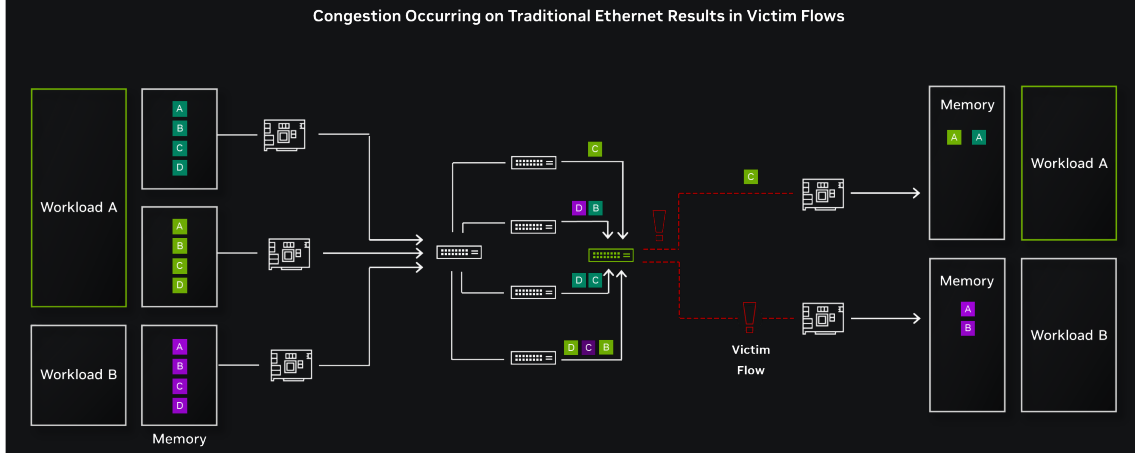
- The SuperNIC sends data into the switch network
- The switch spreads the data packets across all available routes
- The SuperNIC ensures in-order data delivery
- Increase from typical 60% to 95% effective bandwidth

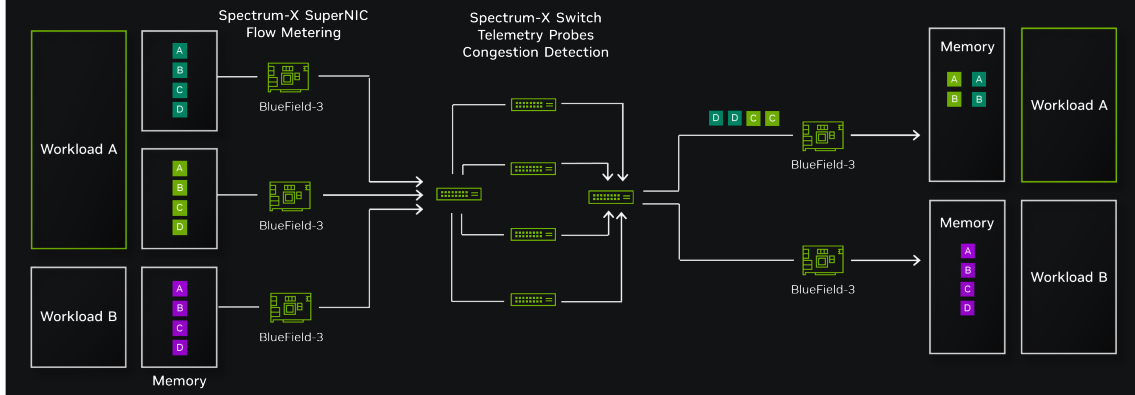




Noise Isolation With Programable Congestion Control

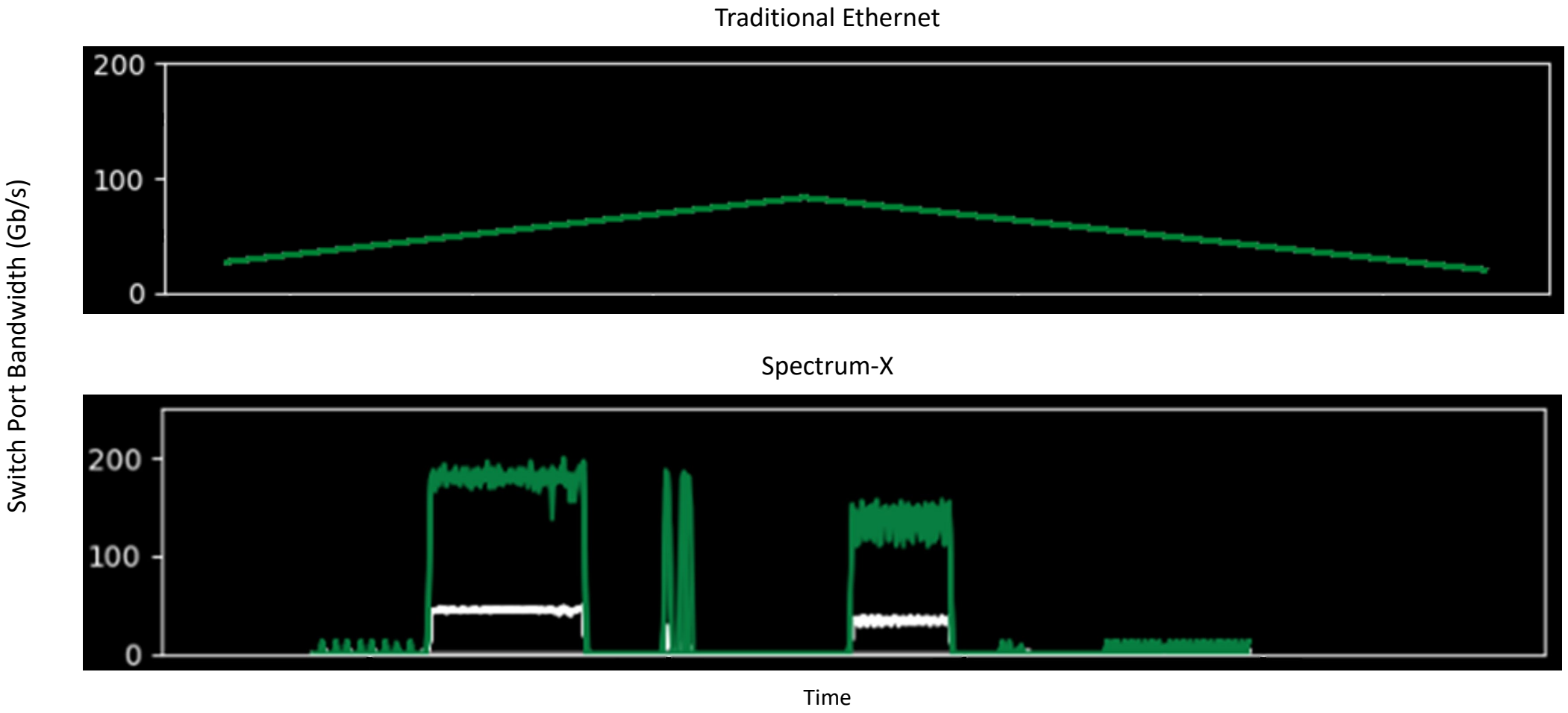
- Diverse workloads can impact each other’s performance
- Spectrum-X detects congestion spots in real time
- Programmable congestion control meters the data flow
- Results in performance isolation across workloads





NVIDIA Networking Enables Highest Network Utilization

Traditional Ethernet is designed for average traffic and provides insufficient resolution



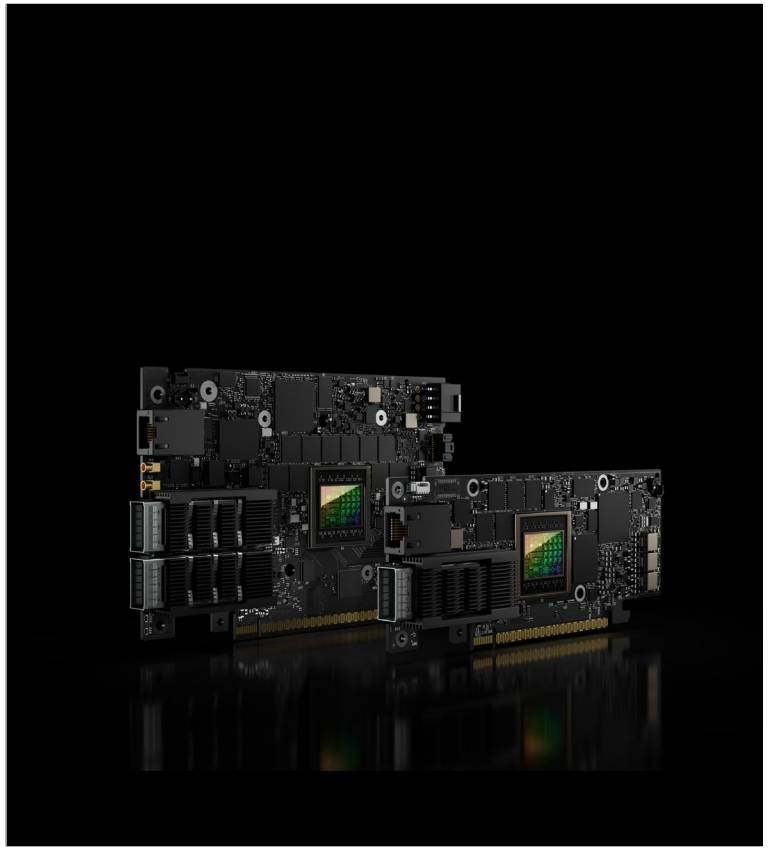
GPT3 Network Utilization

NVIDIA Networking Platforms

Accelerated Networking Solutions for Powering a New Era of Computing

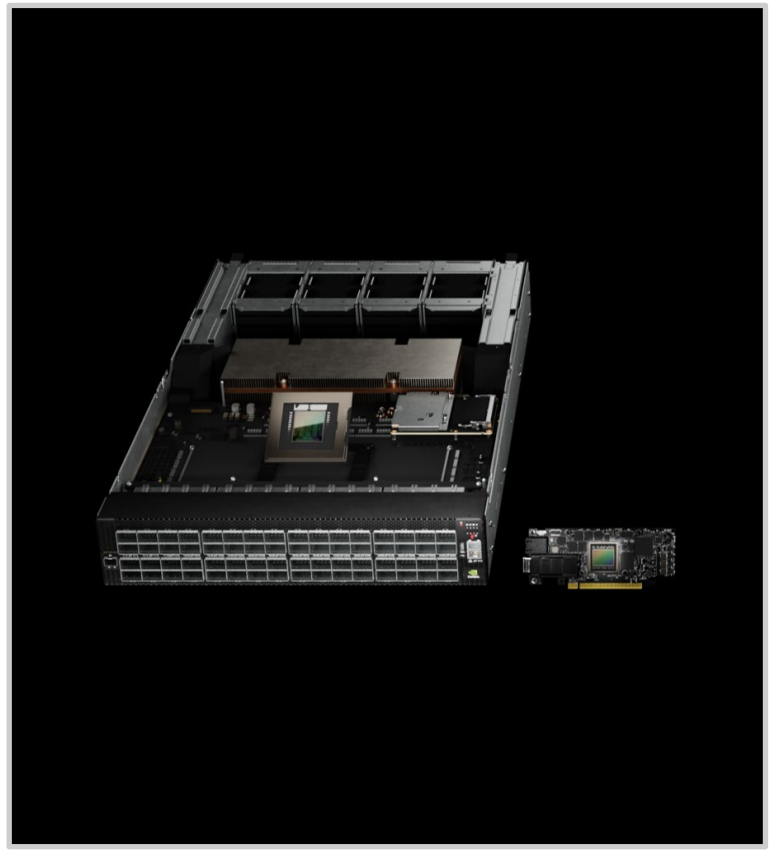
Quantum-X800 InfiniBand

Highest Performance for AI-Dedicated Infrastructure



BlueField-3 Networking

Accelerated Infrastructure Computing



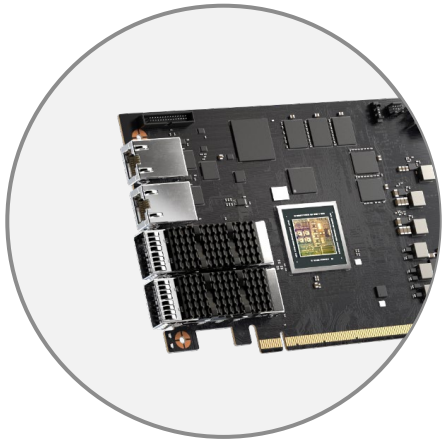
Spectrum-X Ethernet

Bringing High-Performance AI to Ethernet

NVIDIA Quantum & Spectrum-X Platform Components

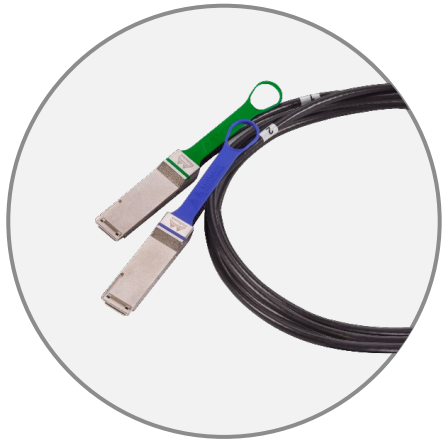
NVIDIA Quantum-2 InfiniBand Platform

ConnectX-7 Adapter
400G InfiniBand
PCIe Gen5
Programmable Datapath
In-Network Computing



BlueField-3 DPU
400G InfiniBand with Arm Cores
PCIe Gen5, DDR5
AI Application Accelerators
Programmable Datapath
In-Network Computing

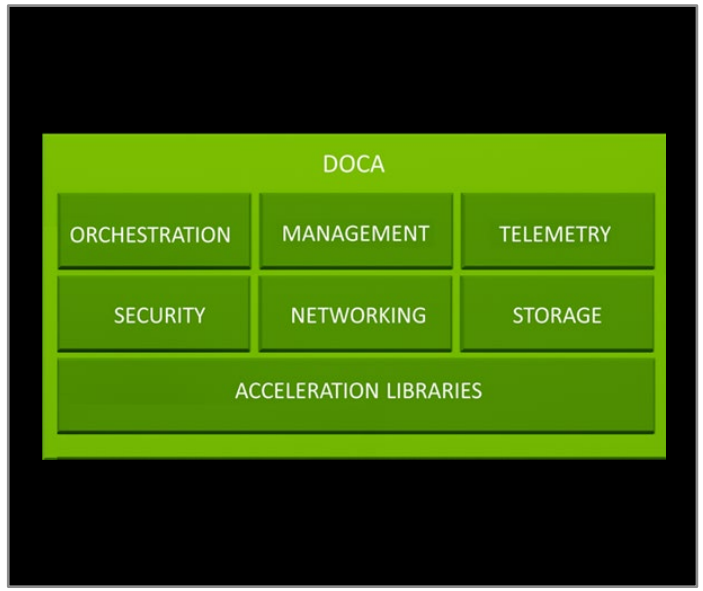
Quantum-2 Switch
64-ports 400G InfiniBand
128-ports 200G
In-Network Computing



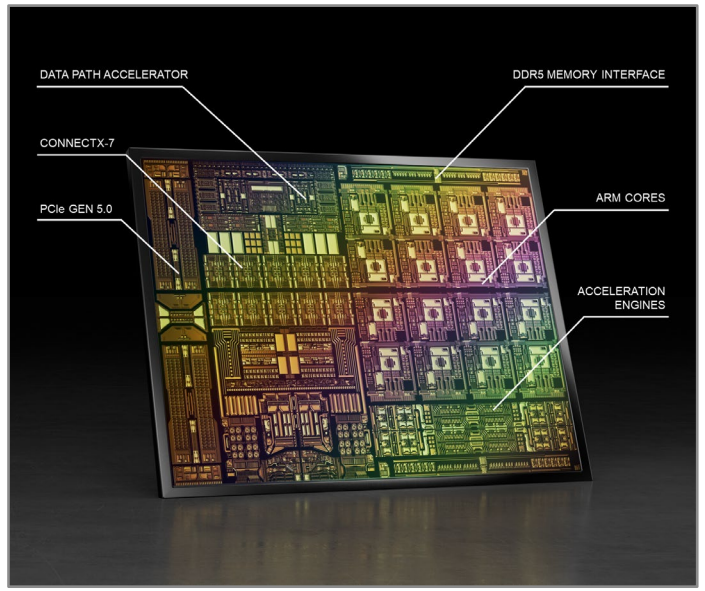
Cable
Copper Cables
Active Copper Cables
Optical Transceivers

From Supercomputers to SuperClouds:

Cloud-Native Supercomputers



DOCA Enabling Growing Partner Ecosystem



Bluefield-3 Next Generation 400G Data Center Infra Processor

NVIDIA Quantum-2 400G InfiniBand In-Network Computing Interconnect

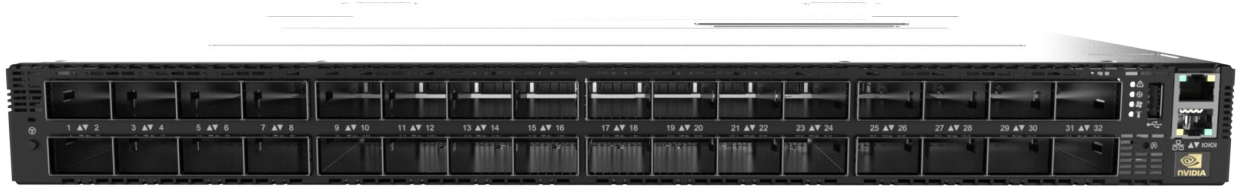
Quantum-2 Switch

QM9700 and QM9790 Family of 1U Switches

- 64 ports of 400Gb/s (NDR) over 32 OSFP cages
- 128 ports of 200Gb/s (NDR200)
- Secured Boot

- 51.2Tb/s aggregate bandwidth
- 66.5 billion packets per second
- SHARPv3 - low latency data reduction and streaming aggregation

- Internally managed (QM9700), and externally managed (QM9790) SKUs
- 26’’ depth, Air cooled
- 2 power supplies (1+1), hot swappable



InfiniBand Router

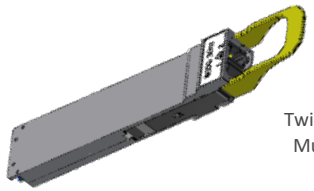
NVIDIA Quantum-2 QM9700 Series Switch + Router

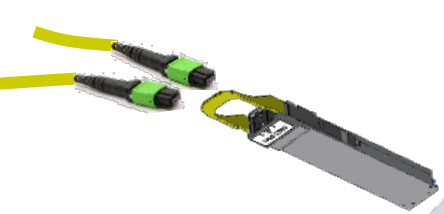
- Increasing resiliency and ease of scale
- Enables segmenting a very large network (42K HCA's) into smaller subnets
- Subnet isolation
- Each subnet has a dedicated Subnet Manager (SM)
- Subnets sharing a common storage network
- NDR generation, part of Quantum 2 managed switch
- Single-hop router
- Ease of use - no need for manual configuration files

400G InfiniBand Cabling Overview

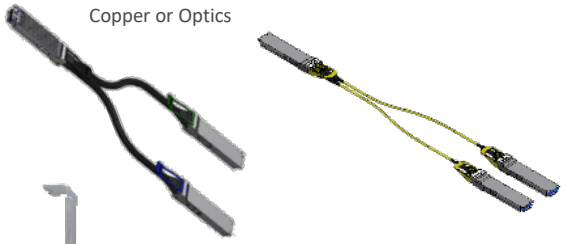
Switch
64 ports of 400Gb/s (4x100Gb/s PAM4)
32 OSFP connectors – 2 ports per connector

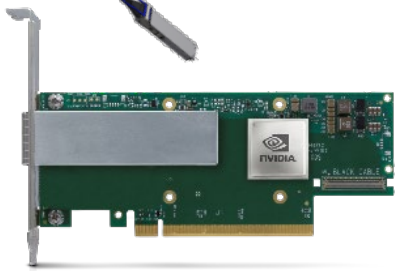
Switch-to-switch
Twin-port Transceivers & MPOs
Multi-Mode and Single-Mode





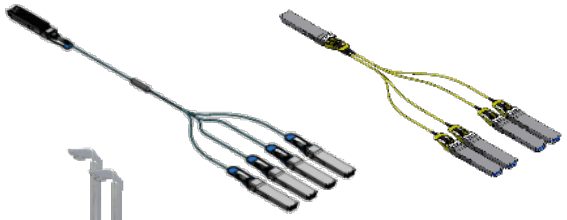
Switch-to-HCA
OSFP -> 2x OSFP or 2x QSFP112
Copper or Optics

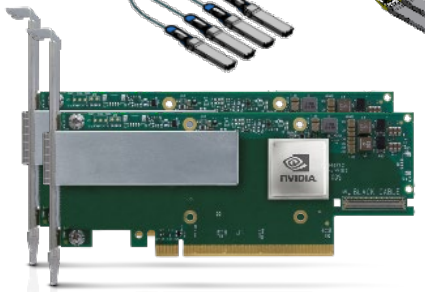




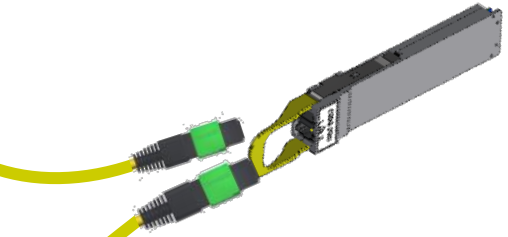
400G Adapter

Switch-to-HCA
OSFP -> 4x OSFP or 4x QSFP112
Copper or Optics





200G Adapter



Images for illustration only

ConnectX-7

400G to Data-Centric Solutions

400Gb/s ports using 100Gb/s SerDes

32 lanes of PCIe Gen5 (compatible with Gen4/Gen3)

PCIe switch and Multi-Host (up to 4 hosts) technology

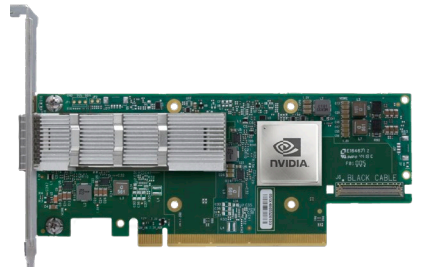
Secure boot

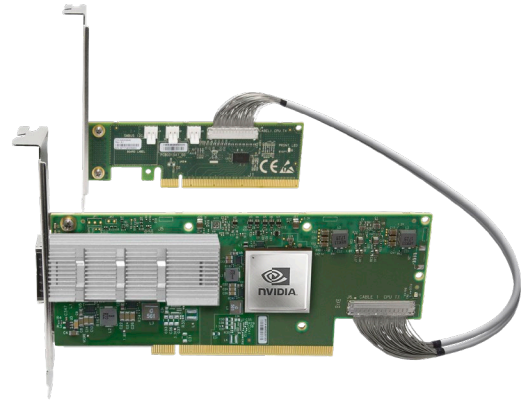
400Gb/s (NDR) throughput

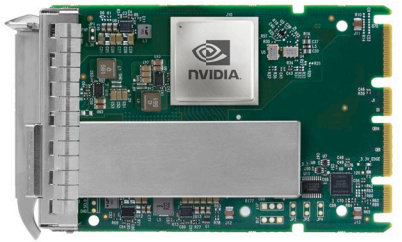
330-370M msg/sec rate

In-Network Computing

- MPI All-to-All hardware engine
- MPI Tag Matching hardware engine
- Programmable acceleration units



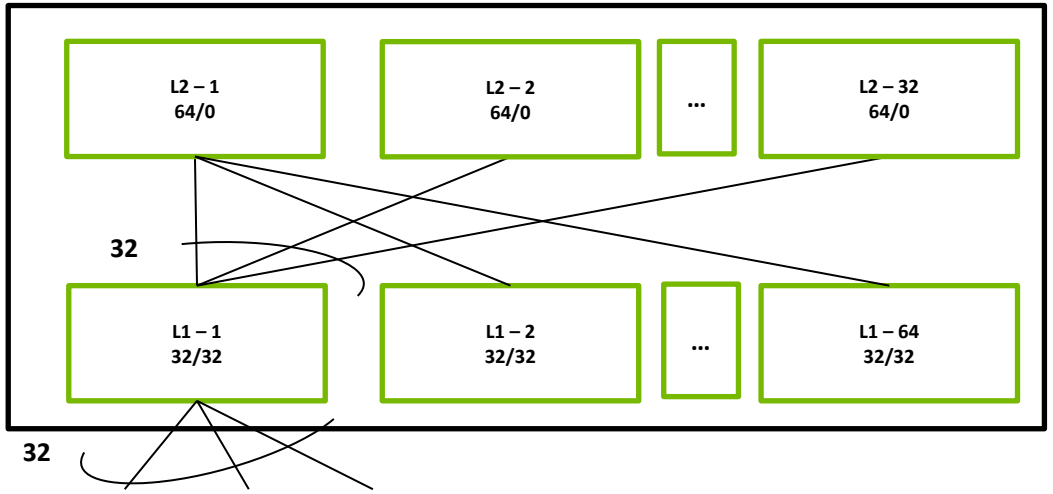


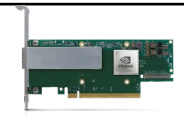


Images for illustration only

2,048 Nodes Cluster, NDR 400G per Node

Fat Tree



NDR InfiniBand, 2,048 node, 400Gb/s per node			
Product	**OPN (example)**	**Qty**	
Quantum™ 2 managed	MQM9700-NS2F	2	
Quantum™ 2 unmanaged	MQM9790-NS2F	94	
Twin-port transceiver	MMS4X00-NS	2,048	
MPO Cable	MFP7E30-N015	2,048	
Passive Copper OSFP to 2xOSFP (NDRx2 to 2 NDR)	MCP7Y00-N001	1,024	
ConnectX-7 Single port NDR, OSFP, PCIe Gen5 x16	MCX75510AAN-NEAT	2,048	

Images for illustration only

NVIDIA Spectrum-X Ethernet Platform

High-Performance Ethernet for AI

Spectrum SN5600 Switch

- 51.2 Terabits per second bandwidth (4X higher)
- 128x 400Gb/s; 64x 800Gb/s
- Adaptive routing, congestion control, high frequency telemetry

Spectrum SN5600 Switch

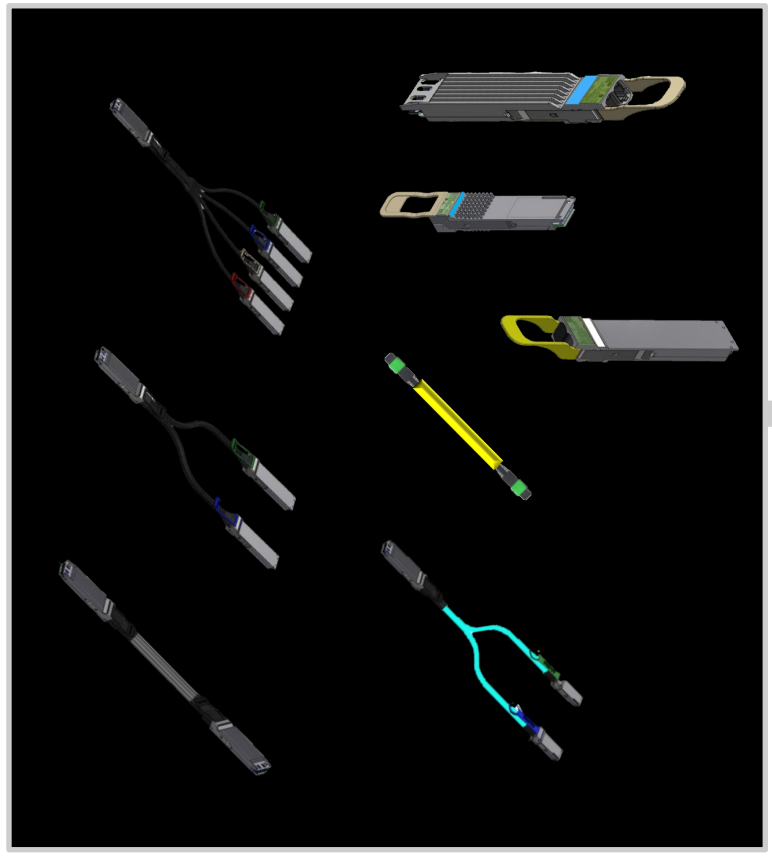
BlueField-3 SuperNIC

- Best-in-class RoCE for AI workloads
- Multi-tenancy at massive scale
- Power efficient, low-profile design

BlueField-3 SuperNIC

NVIDIA AI Networking: LinkX Cables and Transceivers

Optimized End-to-End Connectivity for NVIDIA AI Solutions



LinkX Cables & Transceivers

Performance Optimized for NVIDIA AI Solutions

Minimal bit errors, low error correction delays, high data rate performance, and optimal thermal management

- 25G/40G/100G/200G/400G/800G/1.6T
- Ethernet & InfiniBand

Best Signal Integrity & Reliability

Work seamlessly out of the box with rigorous design, simulation, and extensive live testing processes

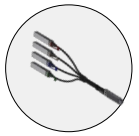
- Rigorous quality control
- Multi-source supply chain

Higher ROI with End-to-End NVIDIA Network

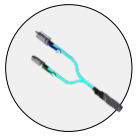
Works in unison with NVIDIA's specialized hardware, software and firmware to monitor links ensuring maximum uptime

- UFM Integration
- Remote In-Service Software Installation

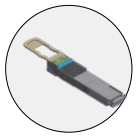
Passive Copper



Active Copper



Active Optical



Multimode transceivers

Single Mode Transceivers

Additional Resources

Networking for AI

Networking for AI Video

Spectrum-X Video

Networking for AI Whitepaper

Spectrum-X Whitepaper

Spectrum-X Webpage

Quantum-X800 Webpage

NVIDIA Quantum-X800 InfiniBand Platform

Highest-Performance for Scientific Computing

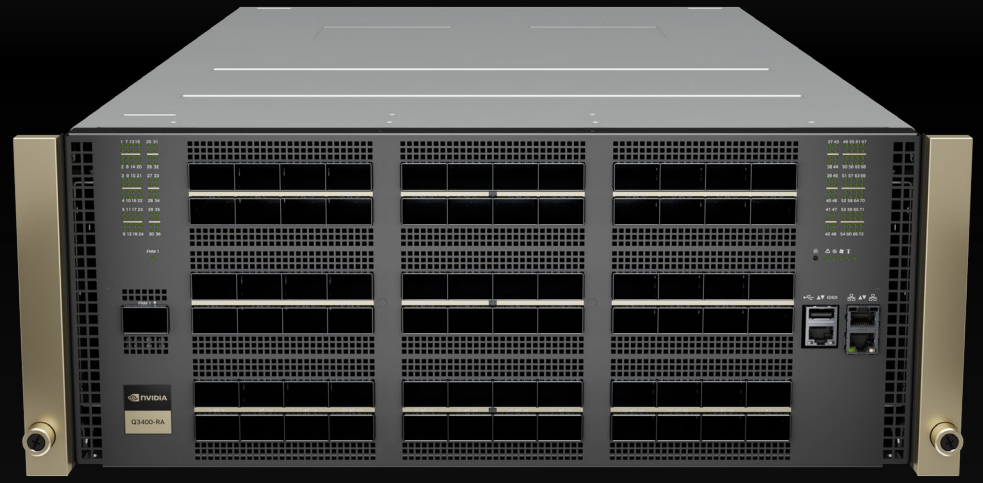
800G End-to-End

Quantum-X800 Q3400-RA Switch

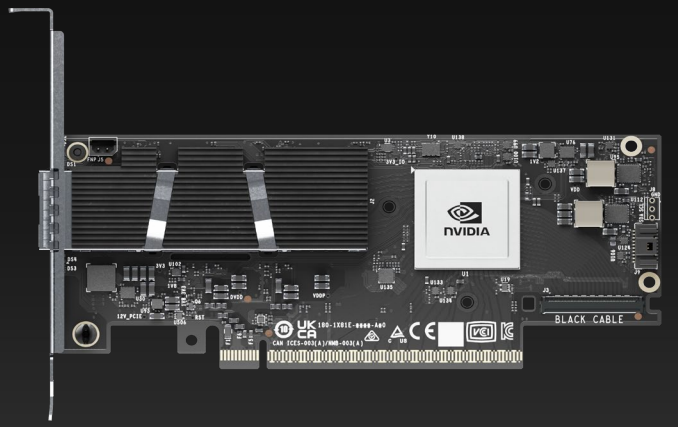
- 115 Terabits per second bandwidth (5X higher)
- SHARP v4 with 14.4 TFlops of In-Network Computing (9X higher)
- Adaptive routing, congestion control, advanced power management

ConnectX-8 SuperNIC

- In-Network Computing
- 32 lanes of PCIe Gen6.1, PCIe switch, Socket Direct



Quantum-X800 Q3400-RA Switch



ConnectX-8 SuperNIC