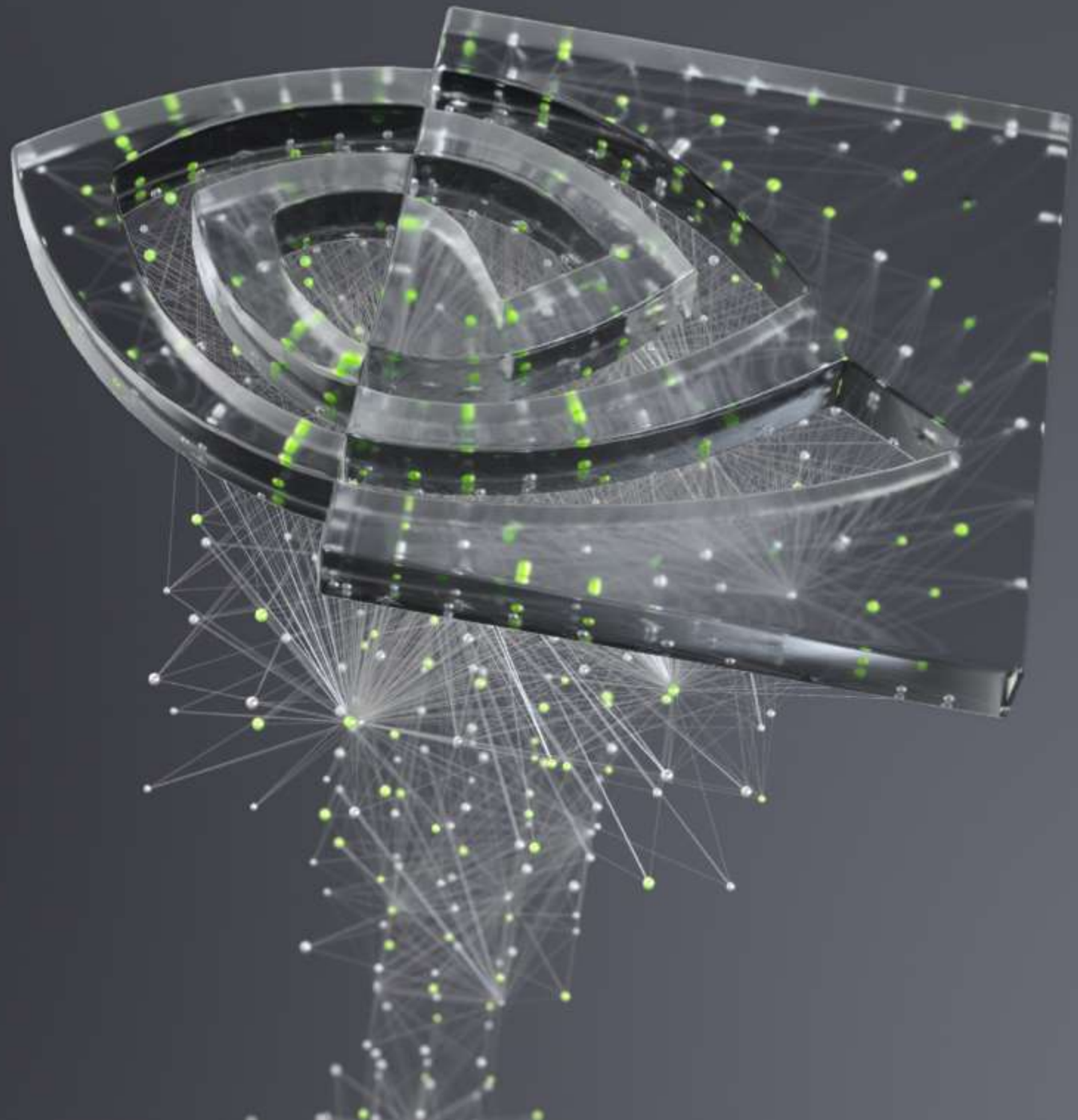
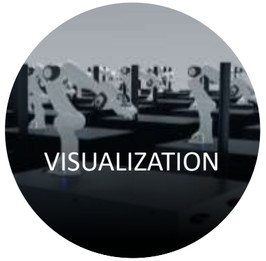
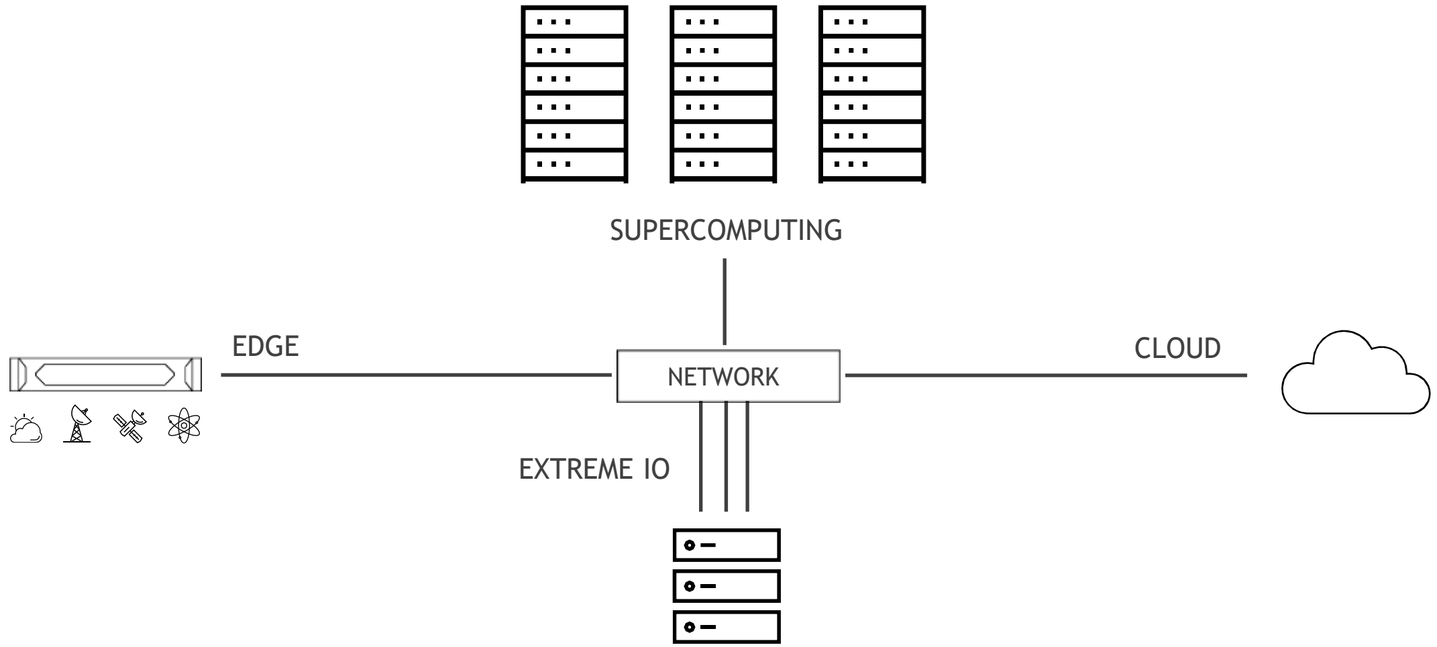


CLOUD NATIVE SUPERCOMPUTING

NVIDIA InfiniBand Networking, June 2021

EXPANDING UNIVERSE OF SCIENTIFIC COMPUTING





NVIDIA HDR 200G INFINIBAND INFRASTRUCTURE

In-Network Computing Accelerated Network for Supercomputing

METROX-2 LONG-HAUL

SKYWAY GATEWAY

UFM CYBER-AI

CONNECTX ADAPTER

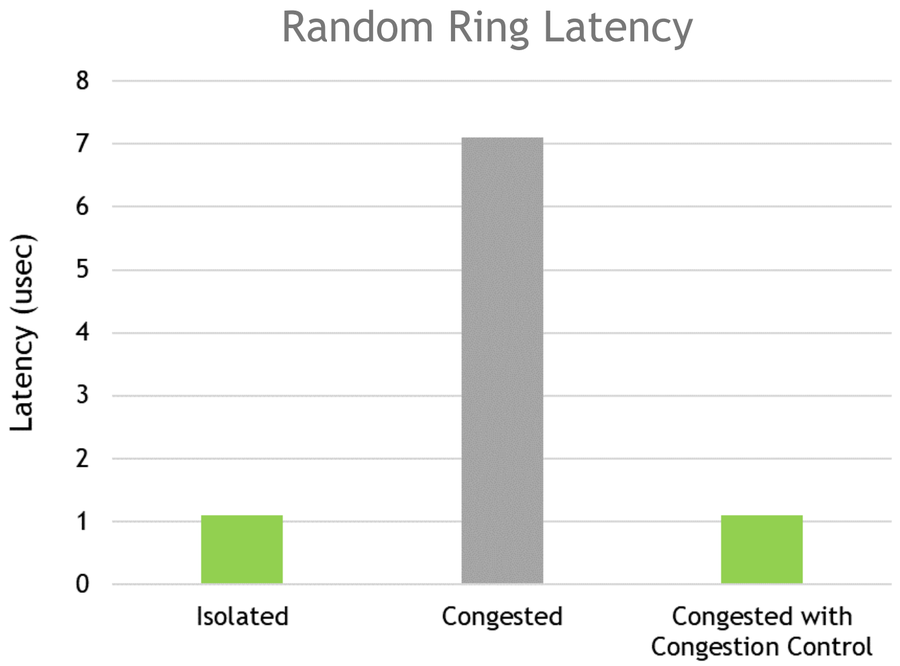
BLUEFIELD DPU

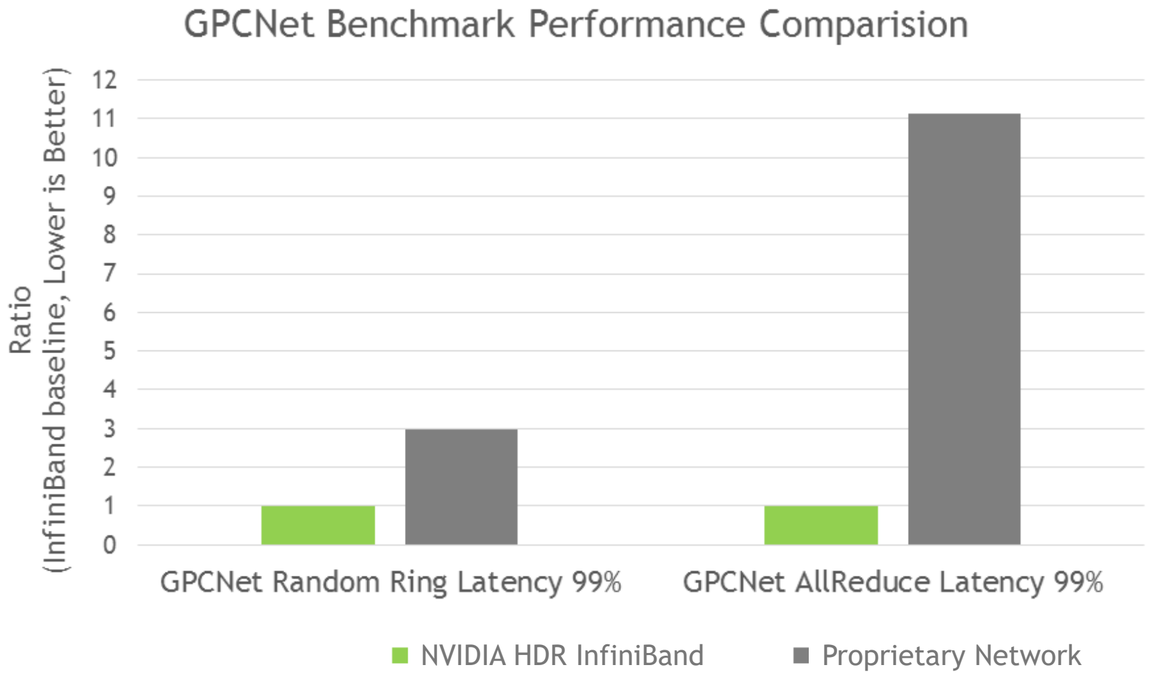
QUANTUM SWITCH

LINKX

INFINIBAND CONGESTION CONTROL

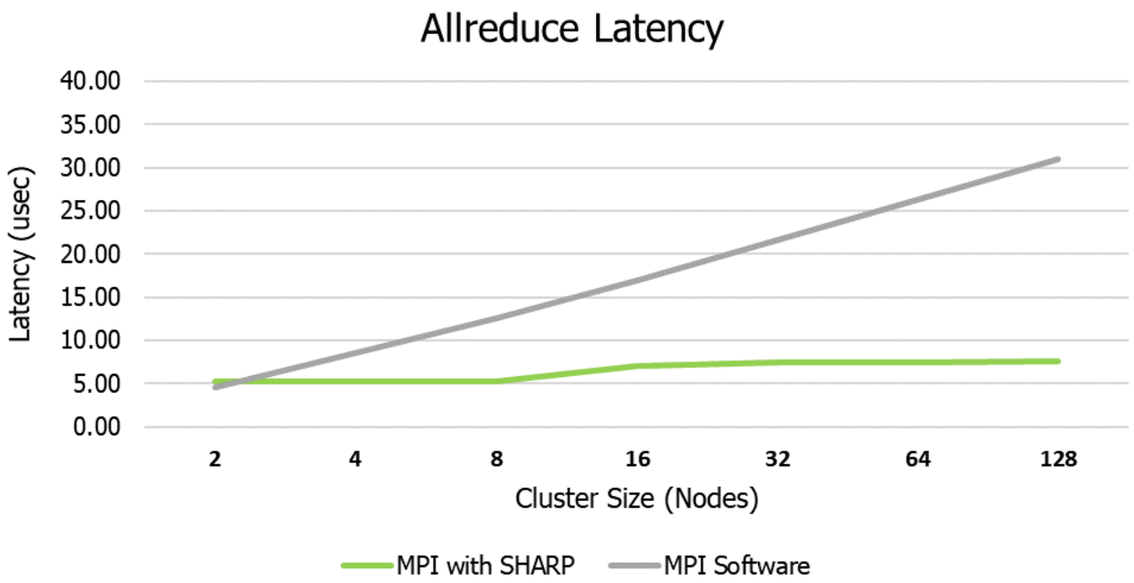
Eliminates Congestion Hot Spots



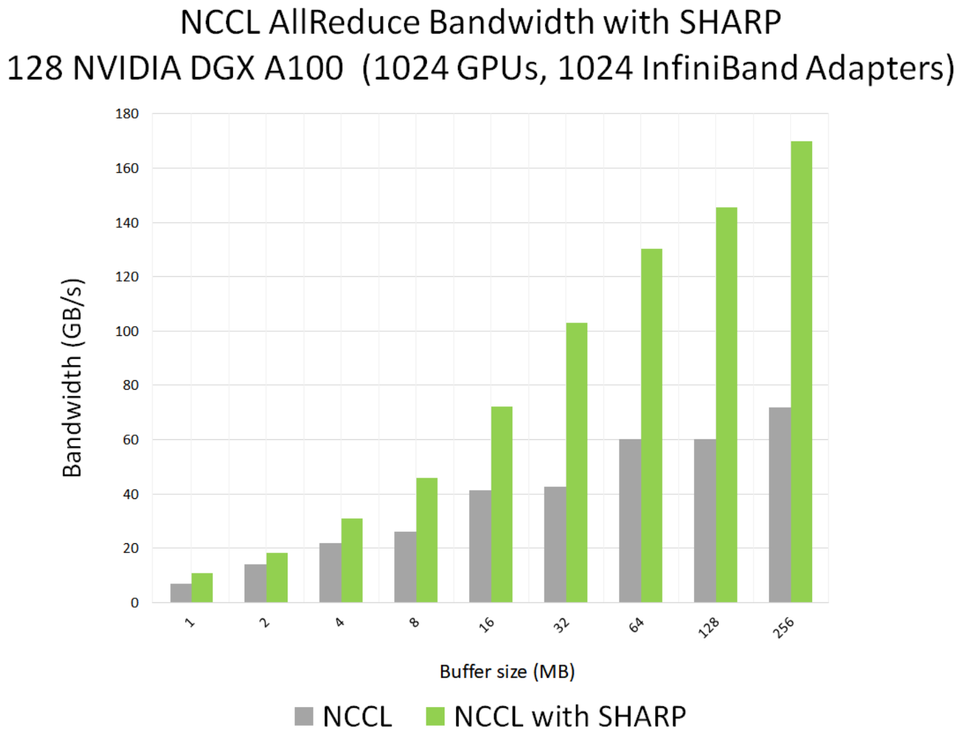


SHARP ALLREDUCE PERFORMANCE ADVANTAGES

7x Higher MPI Performance, 2.5x higher AI Performance



Lower is better



Higher is better

CLOUD-NATIVE SUPERCOMPUTING

Bare-Metal secured infrastructure

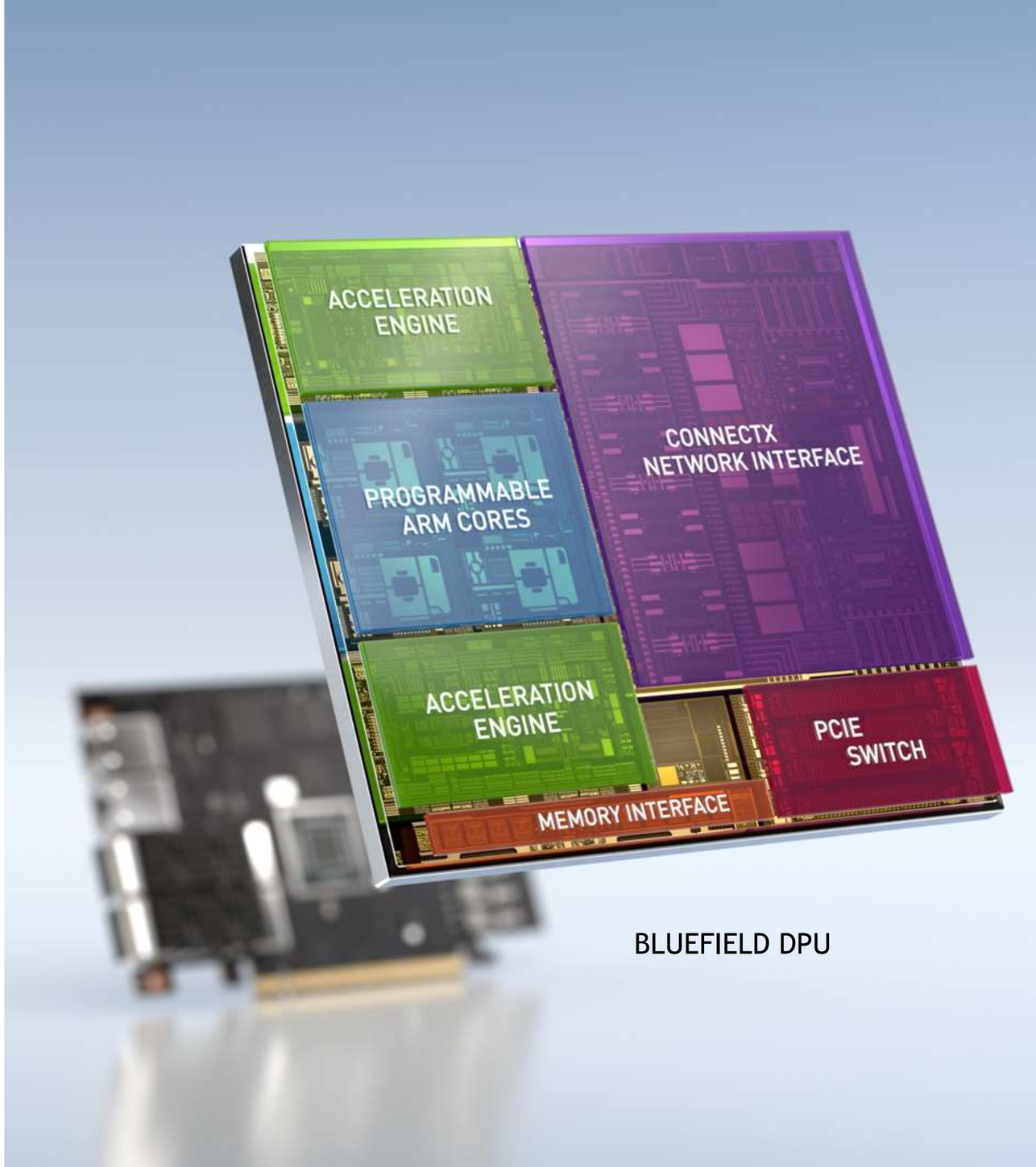
Higher application performance

From the edge to the main data center

BARE-METAL PERFORMANCE

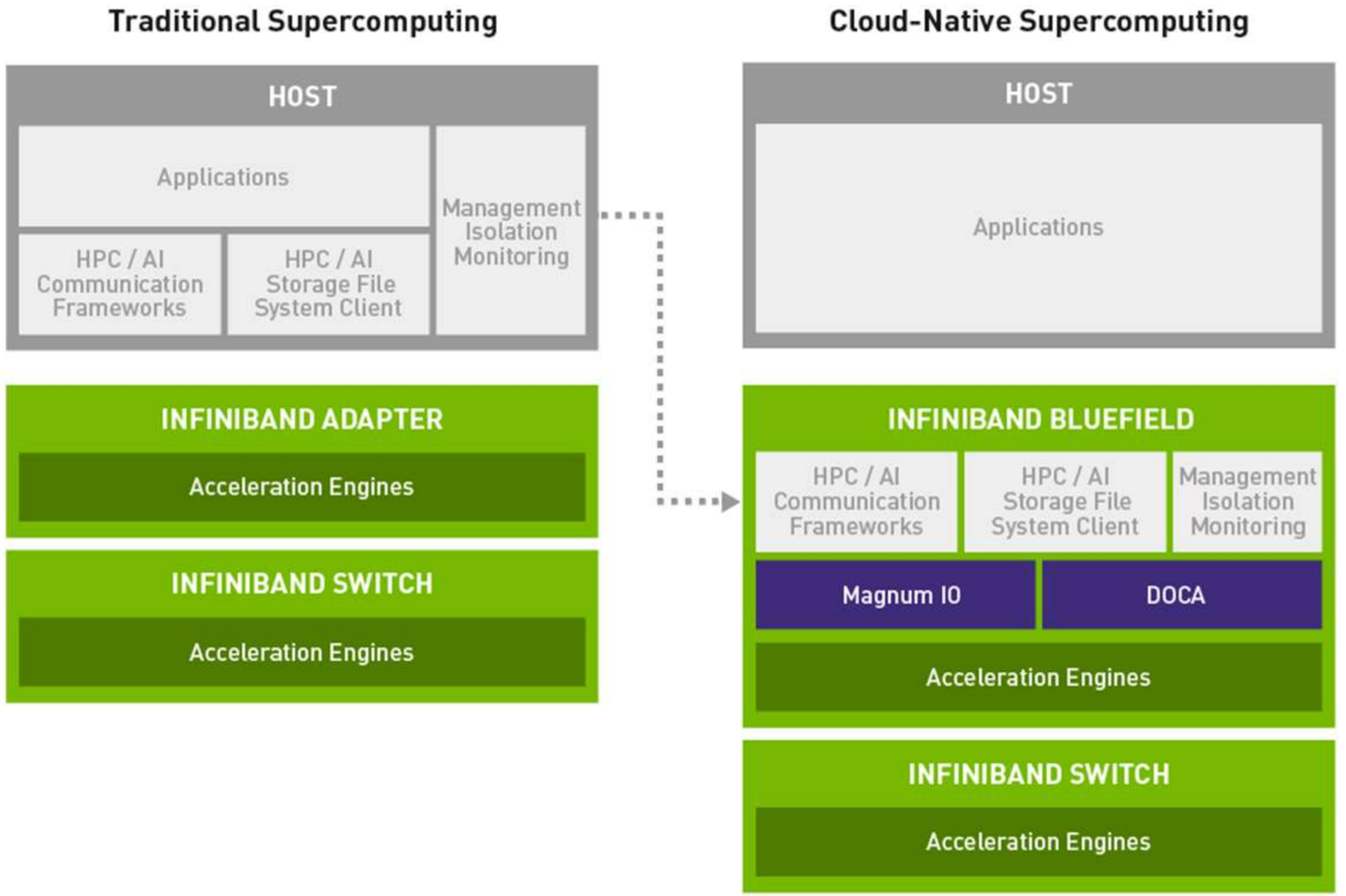
MULTI TENANCY

EDGE COMPUING



BLUEFIELD DPU

CLOUD-NATIVE SUPERCOMPUTING INFRASTRUCTURE



MULTI-TENANT ISOLATION

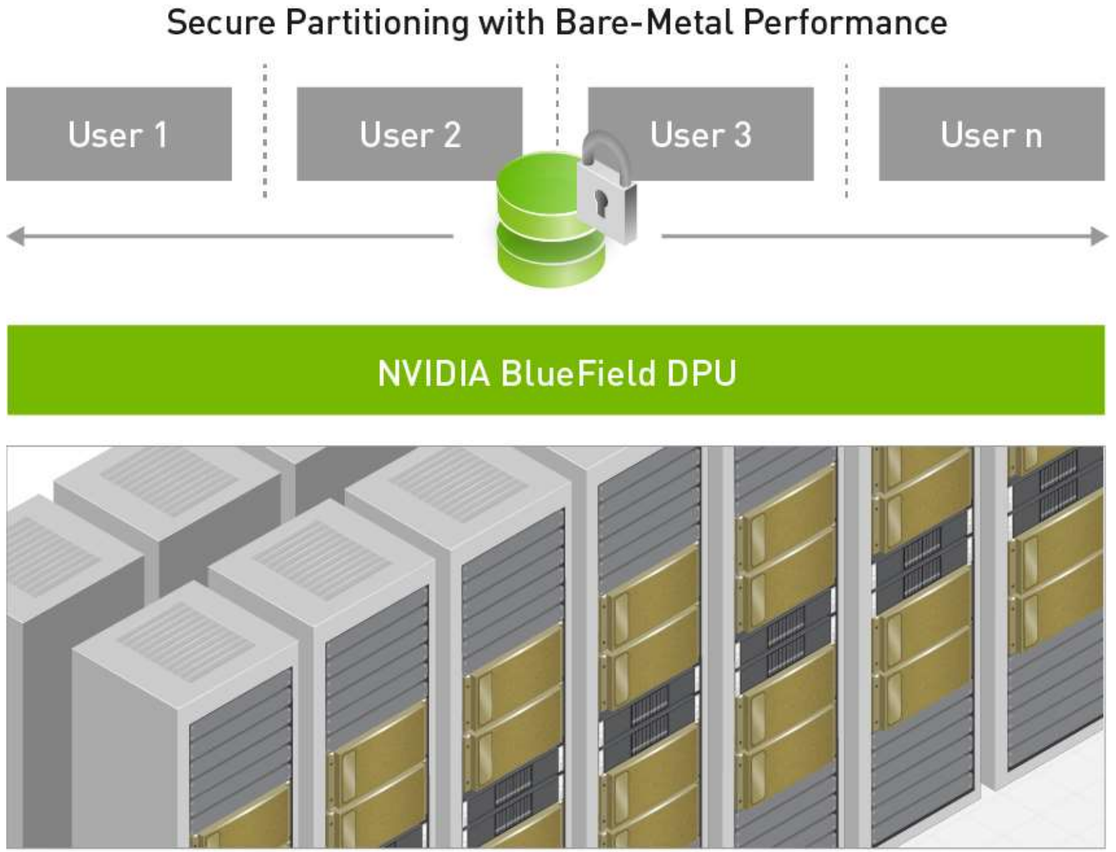
Zero-Trust Architecture

Secured network infrastructure and configuration

Storage virtualization

Tenant Service Level Agreement (SLA)

32K concurrent isolated users on single subnet



HIGHER APPLICATION PERFORMANCE

DPU-Accelerated HPC Communications, Higher Application Performance

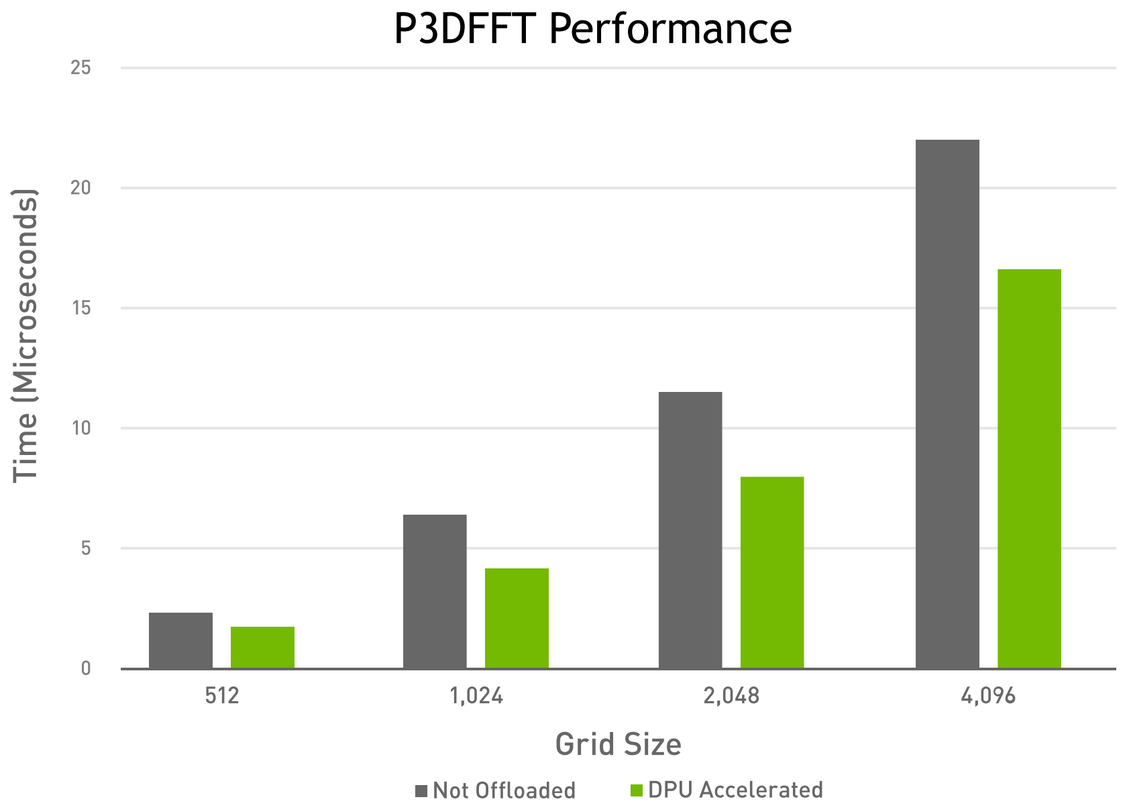
Collective offload

Active messages

Smart MPI progression

Data compression

User-defined algorithms



NVIDIA DOCA

Enabling Broad DPU Partner Ecosystem

Software application framework for BlueField DPUs

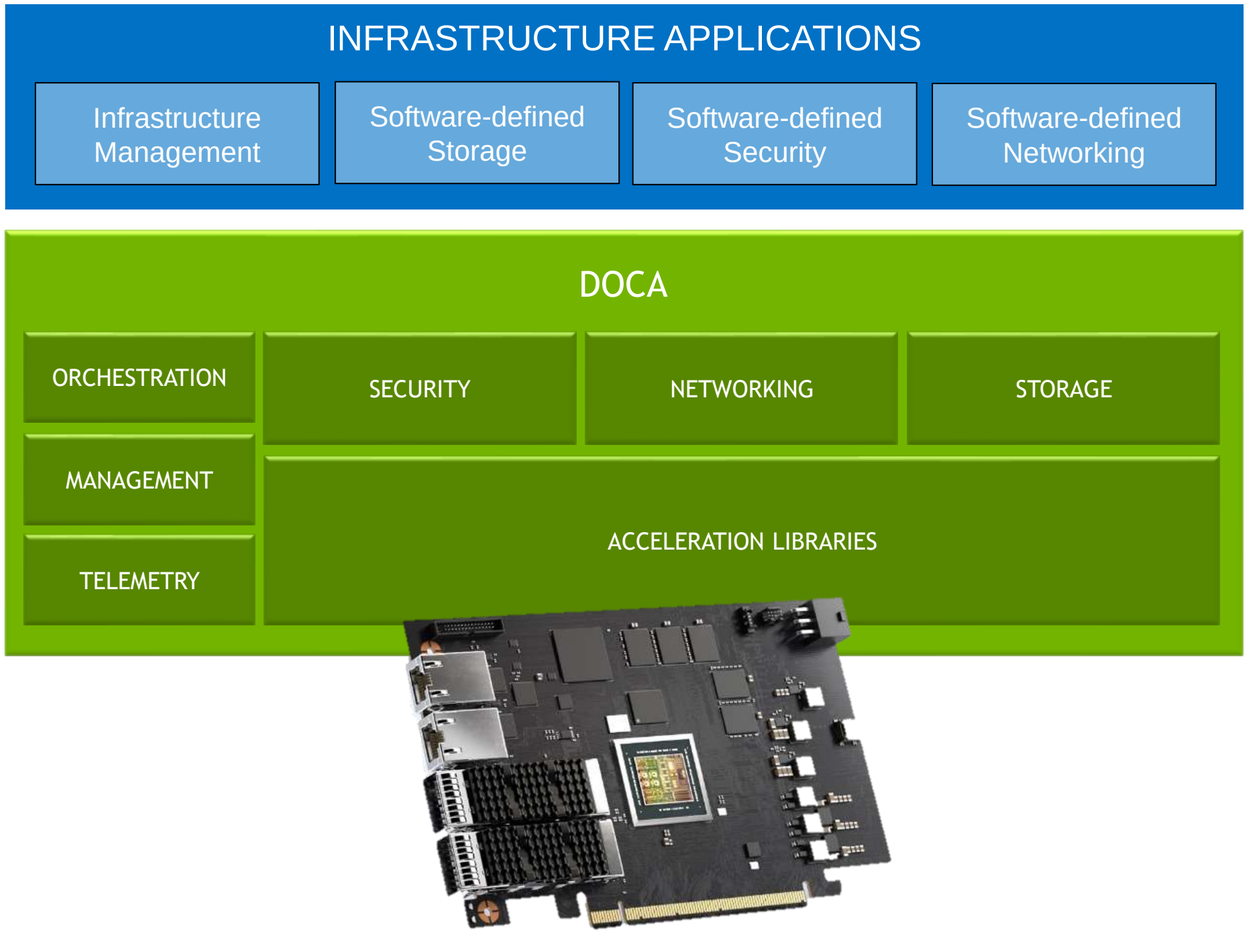
DOCA is for DPUs what CUDA is for GPUs

Protects developer investment for future DPUs

Certified reference applications, APIs & partner solutions

Rich partner ecosystem across industries and workloads

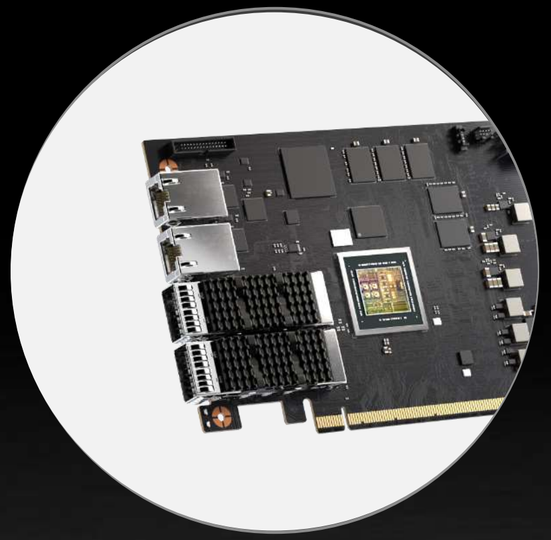
https://developer.nvidia.com/networking/doca



NVIDIA NDR 400G INFINIBAND: NEXT-GENERATION INFINIBAND ARCHITECTURE

ConnectX-7 Adapter

NDR 400G InfiniBand
In-Network Computing



BlueField-3 DPU

NDR 400G InfiniBand with Arm Cores
AI Application Accelerators
Programmable Datapath
In-Network Computing

Quantum-2 Switch

64-ports NDR 400G InfiniBand
128-ports 200G NDR200
In-Network Computing

Cable

Copper Cables
Active Copper Cables
Optical Transceivers

ANNOUNCING NVIDIA NDR 400G INFINIBAND SYSTEMS

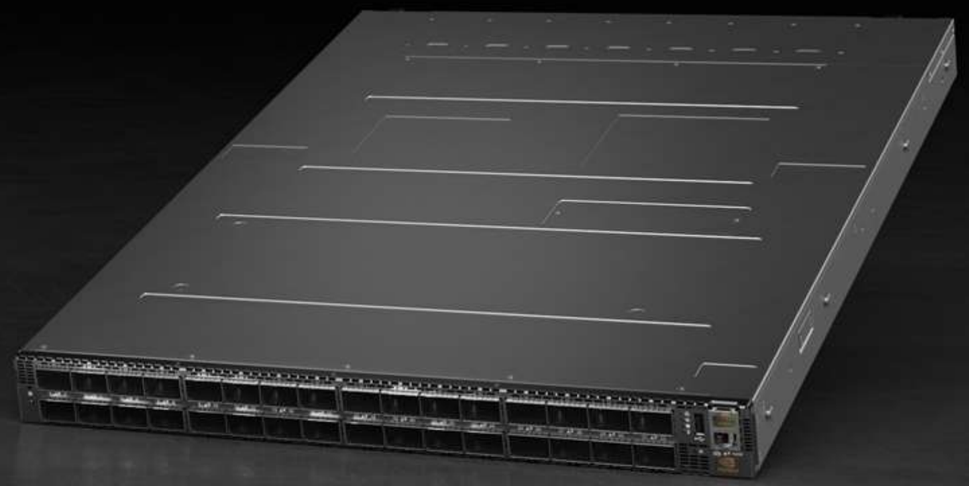
In-Network Computing Accelerated Network for Cloud-Native Supercomputing at Any Scale

2x
Data Throughput
400 Gigabits per Second

32x
More AI Acceleration
SHARP In-Network Computing

6.5x
Higher Scalability
>1M nodes with DF+ 3 hops

5x
Switch System Capacity
>1.6 Petabit per Second



ANNOUNCING NVIDIA NDR 400G INFINIBAND SYSTEMS

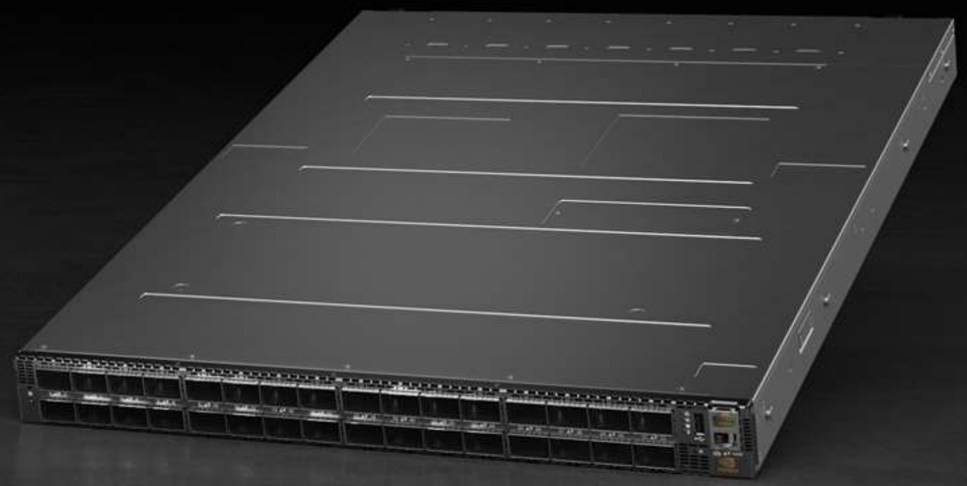
In-Network Computing Accelerated Network for Cloud-Native Supercomputing at Any Scale

15%
Faster Deep Learning Recommendations

17%
Faster Natural Language Processing

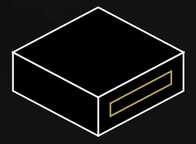
15%
Faster Computational Fluid Dynamics Simulations

60%
Lower Power Consumption

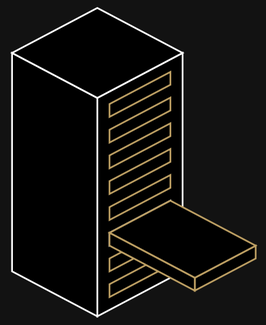


ANNOUNCING NVIDIA NDR 400G INFINIBAND SYSTEMS

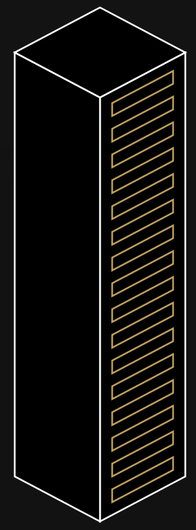
In-Network Computing Accelerates Cloud-Native Supercomputing at Any Scale



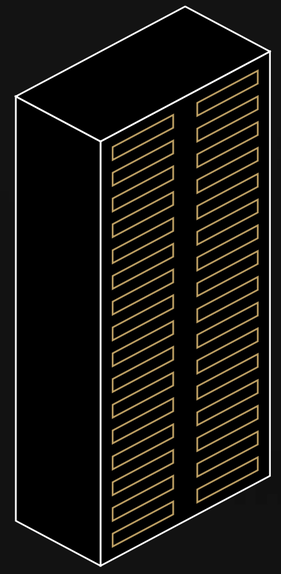
64 NDR Ports
128 NDR200 Ports



512 NDR Ports
1024 NDR200 Ports



1024 NDR Ports
2048 NDR200 Ports

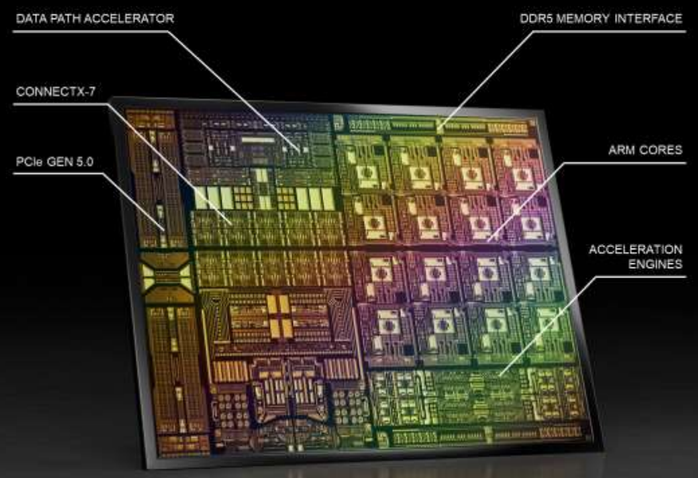


2048 NDR Ports
4096 NDR200 Ports

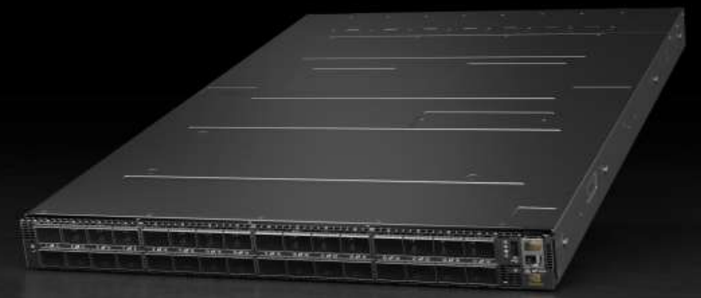
FROM SUPERCOMPUTERS TO SUPERCLOUDS: CLOUD-NATIVE SUPERCOMPUTERS

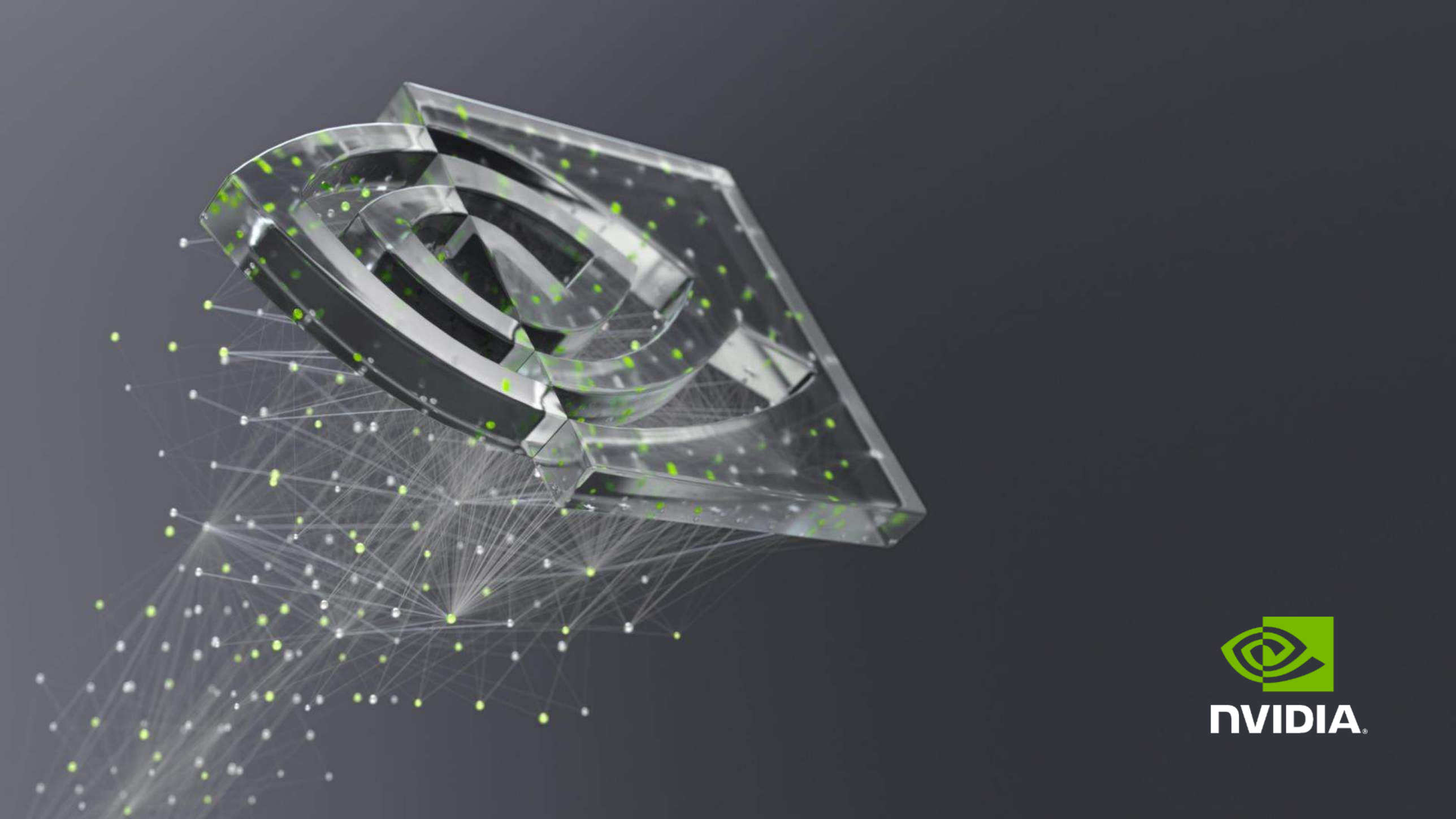
DOCA ENABLING GROWING PARTNER ECOSYSTEM

BLUEFIELD-3 NEXT GENERATION 400G DATA CENTER INFRA PROCESSOR



NVIDIA NDR 400G INFINIBAND IN-NETWORK COMPUTING INTERCONNECT




NVIDIA