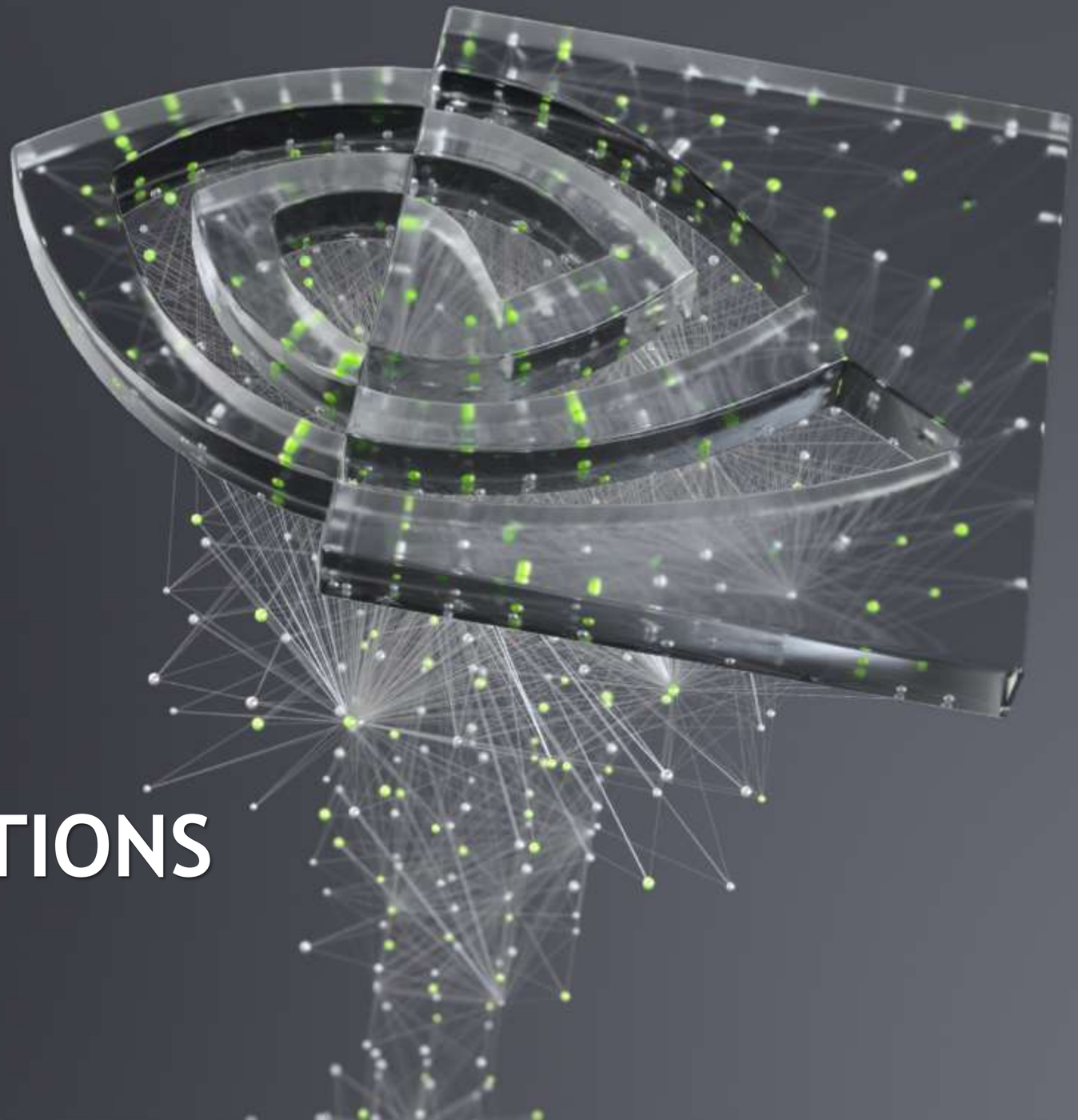
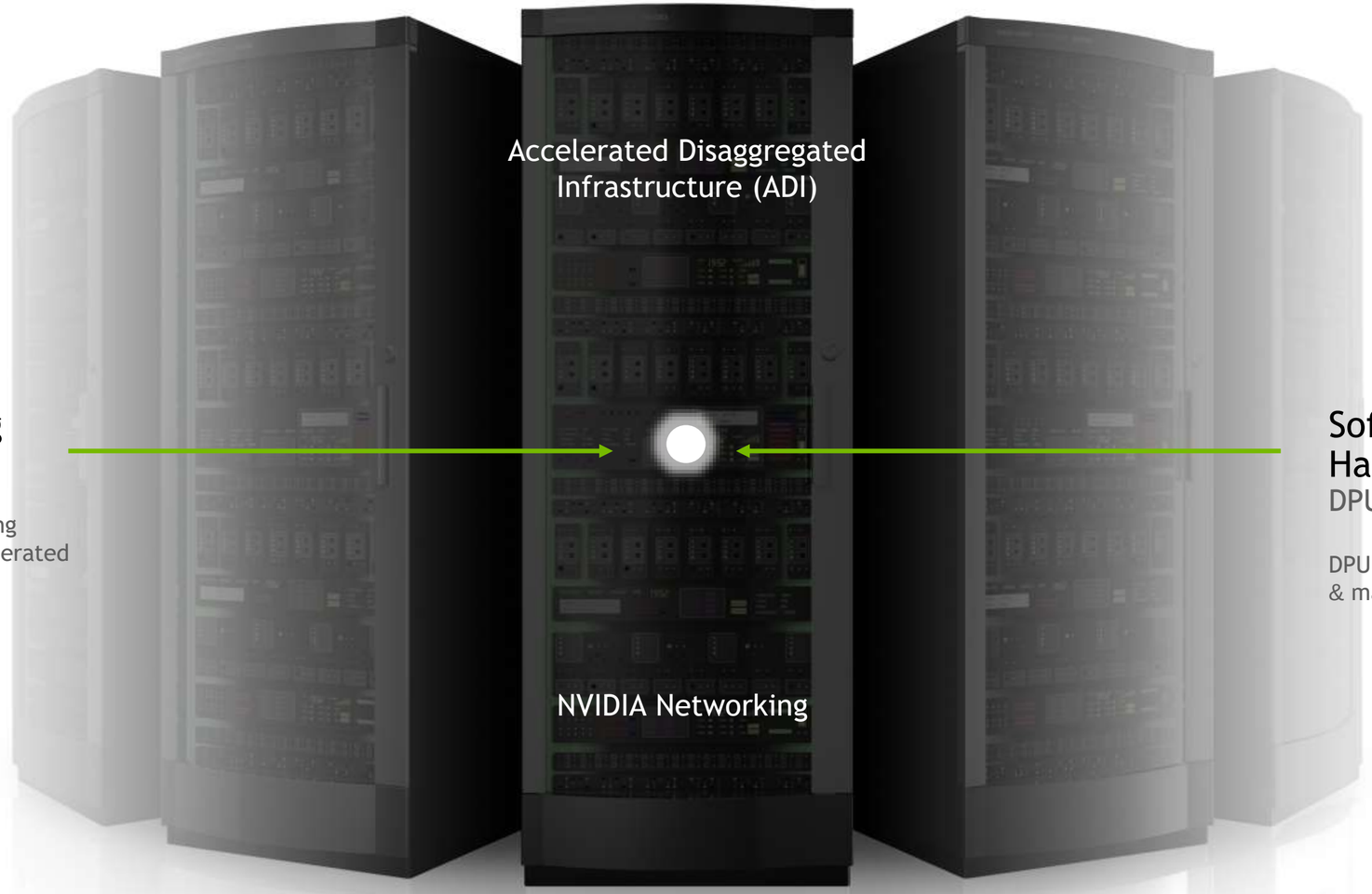


NVIDIA MELLANOX NETWORKING SOLUTIONS

January 2021

THE DATA CENTER IS THE NEW UNIT OF COMPUTING



Accelerated Computing
GPU: AI & machine learning

GPU critical for AI & machine learning
Every workload will become AI Accelerated

**Software defined,
Hardware-accelerated**
DPU (data processing unit)

DPU essential to disaggregate resources
& make composable ADI

THE NEW LEADER IN DATACENTER

New name, same dedication to open networking

Leaders in Linux Networking Software

CUMULUS

+

Leaders in Open Networking Hardware

Mellanox® TECHNOLOGIES

NVIDIA®

NVIDIA DATACENTER PLATFORM

BUSINESS APPLICATIONS

Customer Engagement | Patient Diagnostics | Fraud Detection | Quality Assurance | Industrial Automation | Precision Marketing | Molecular Simulations | ++

NGC

SOFTWARE HUB

HELM

Certified Containers

Pre-trained Models

SDKs

APPLICATION FRAMEWORKS

SMART CITY	CONVERSATIONAL AI	AUTONOMOUS VEHICLES	RECOMMENDATION SYSTEMS	HEALTHCARE	++
Metropolis	*Jarvis*	*Drive*	*Merlin*	*Clara*	…

DEVELOPER TOOLKITS

ML & DATA ANALYTICS	AI TRAINING & INFERENCE	HIGH PERFORMANCE COMPUTING	RENDERING & VISUALIZATION
RAPIDS, dmlc XGBoost	TensorFlow, TensorRT, PYTORCH, mxnet	*NVIDIA HPC SDK*	*IndeX, OptiX, CloudXR, MDL*

ACCELERATION LIBRARIES

COMPUTE	NETWORKING, STORAGE & SECURITY
CUDA-X	*DOCA, MAGNUM IO*

NVIDIA CERTIFIED

VALIDATED SOLUTIONS

SERVERS & CLOUD

DGX HGX — Purpose Built

EGX — Mainstream & Edge

CSP Instances

HARDWARE TECHNOLOGIES

GPU

NVSwitch

BF DPU

SMART NIC

Mellanox Switch

OPERATIONS

TRITON INFERENCE SERVER

FLEET COMMAND

NVIDIA GPU Operator

Red Hat

vmware

MONITORING

DCGM

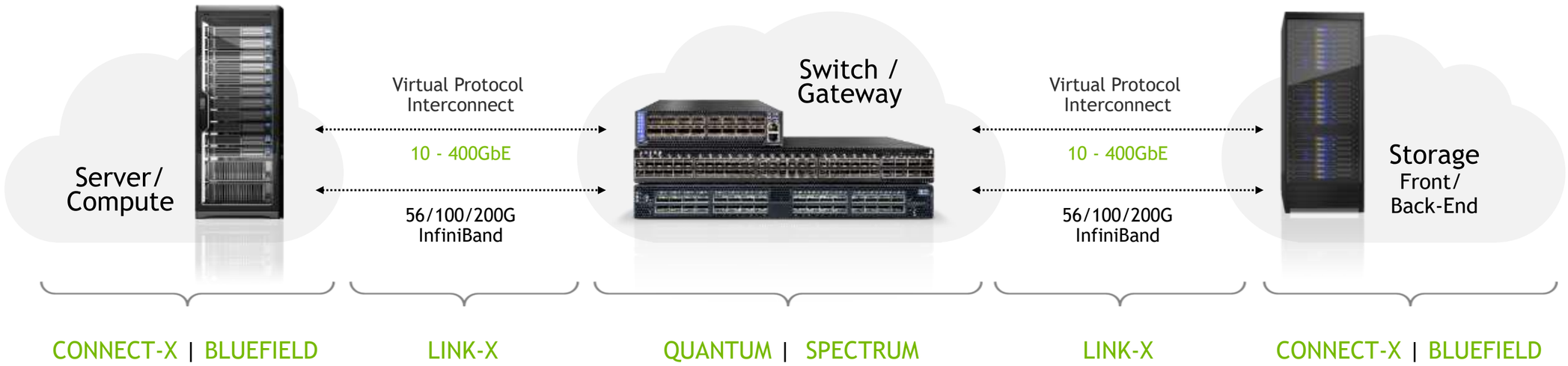
Grafana

Prometheus

UFM

LEADING SUPPLIER OF INFINIBAND & ETHERNET END-TO-END INTERCONNECT SOLUTIONS

The Smart Choice for Intelligent Compute & Storage Platforms

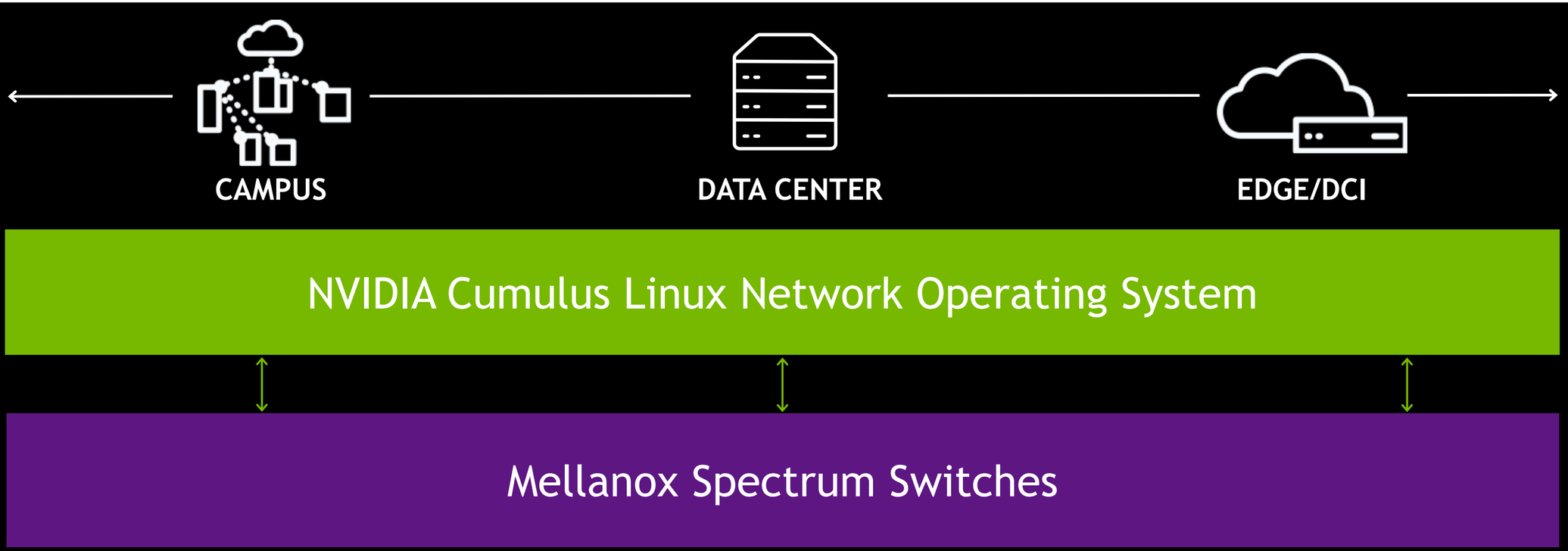


ONE NOS, MULTIPLE USE CASES

Unified configuration and operations

Enterprise Support

End-to-end simulation and automation



NVIDIA NETWORKING

CONNECTX NICs

World's Leading Ethernet NICs & Adapters

BLUEFIELD DPUs

World's Most Secure Data Processing Unit for Visual Computing

SPECTRUM

World's Leading Open Ethernet Switches

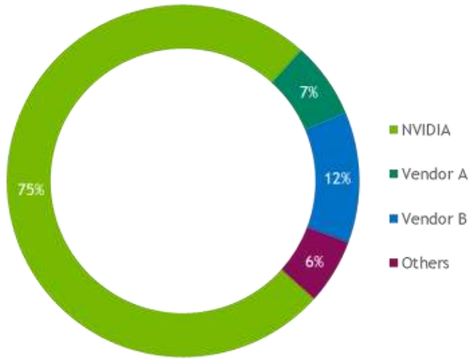
INFINIBAND

World's Highest Performance HPC & AI Networking

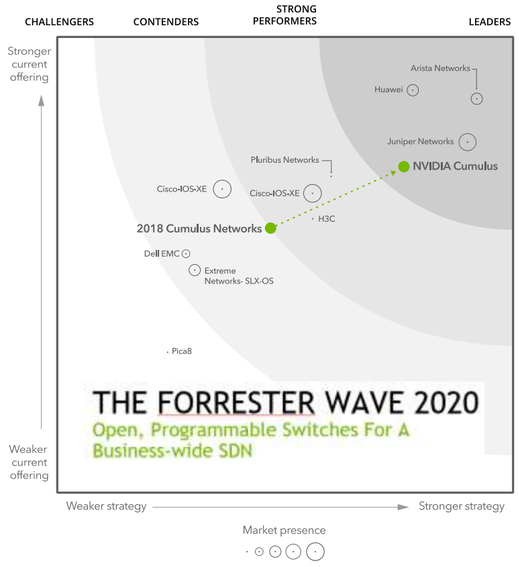
LINKX

World's Most Reliable Optical Transceivers, AOCs & Copper Cables

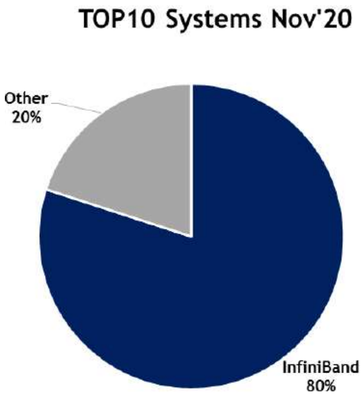
Market Leader in 25G+ Ethernet Adapters[1]



Data Center Infrastructure on a Chip



Top500 Interconnect Trends



1. *3Q20 CREHAN Quarterly Market Shares - Ethernet Server-Class Adapter & LOM_Controllers*

INTELLIGENT NETWORKING SWITCH

DATA CENTER EVOLUTION

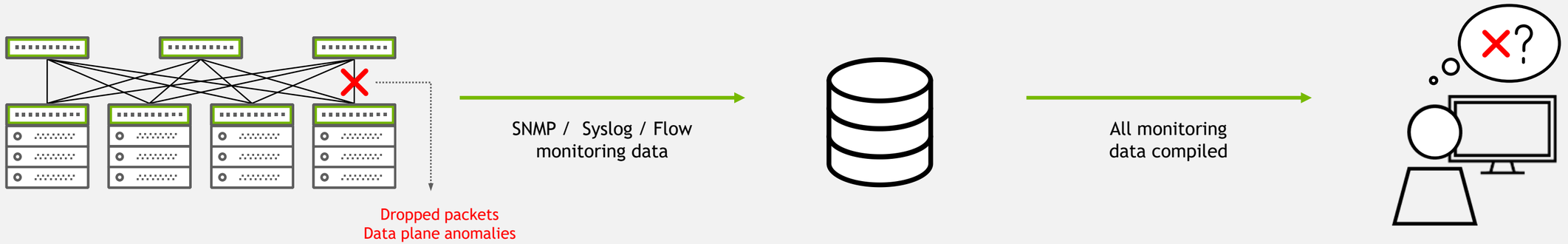
Legacy Mindset

Protocols Telemetry Features

Webscale Mindset

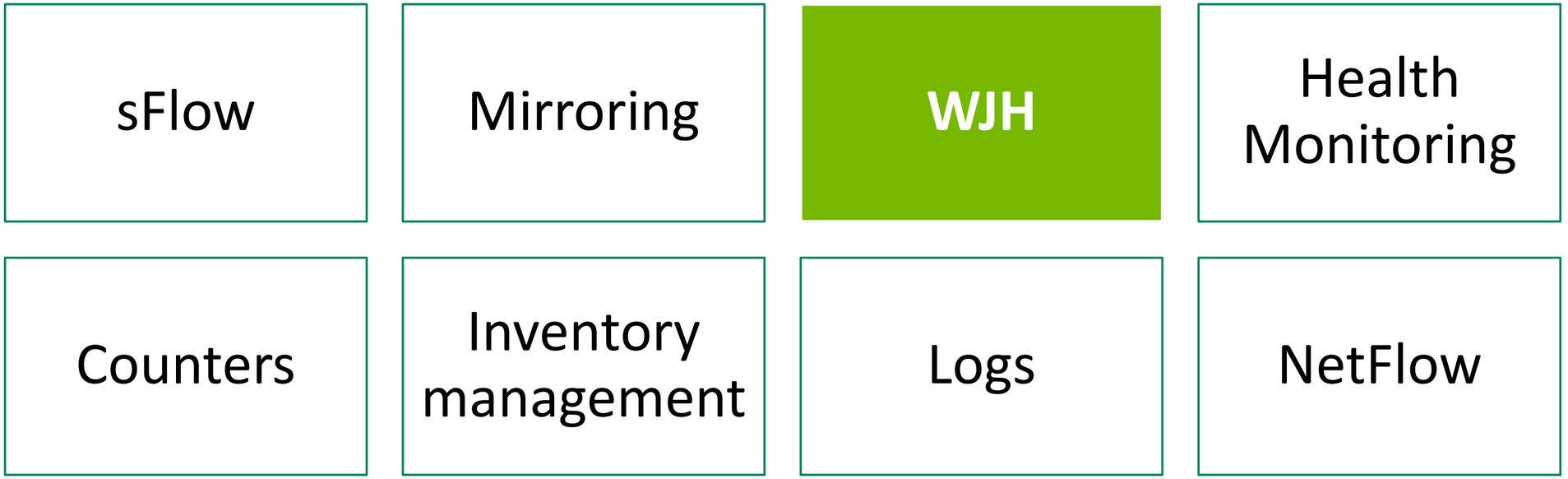
ACCELERATING TIME TO ROOT CAUSE

Legacy Network Monitoring



What Just Happened

WHAT IS TELEMETRY?



Telemetry is an automated communications process by which measurements and other data are collected at remote or inaccessible points and transmitted to receiving equipment for monitoring.

WJH – DROP REASONS AND TYPES

Packet Drop

L1
Bad CRC
Flaky cable

L2/L3/Overlay
BGP
VLAN

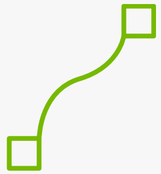
Buffer
Incast
Rate limit

ACLs
Deny based on LP
Deny based on VLAN

No Packet Drop

Congestion
Incast
Busy storage device

Latency
Pause frames
Congestion -> latency

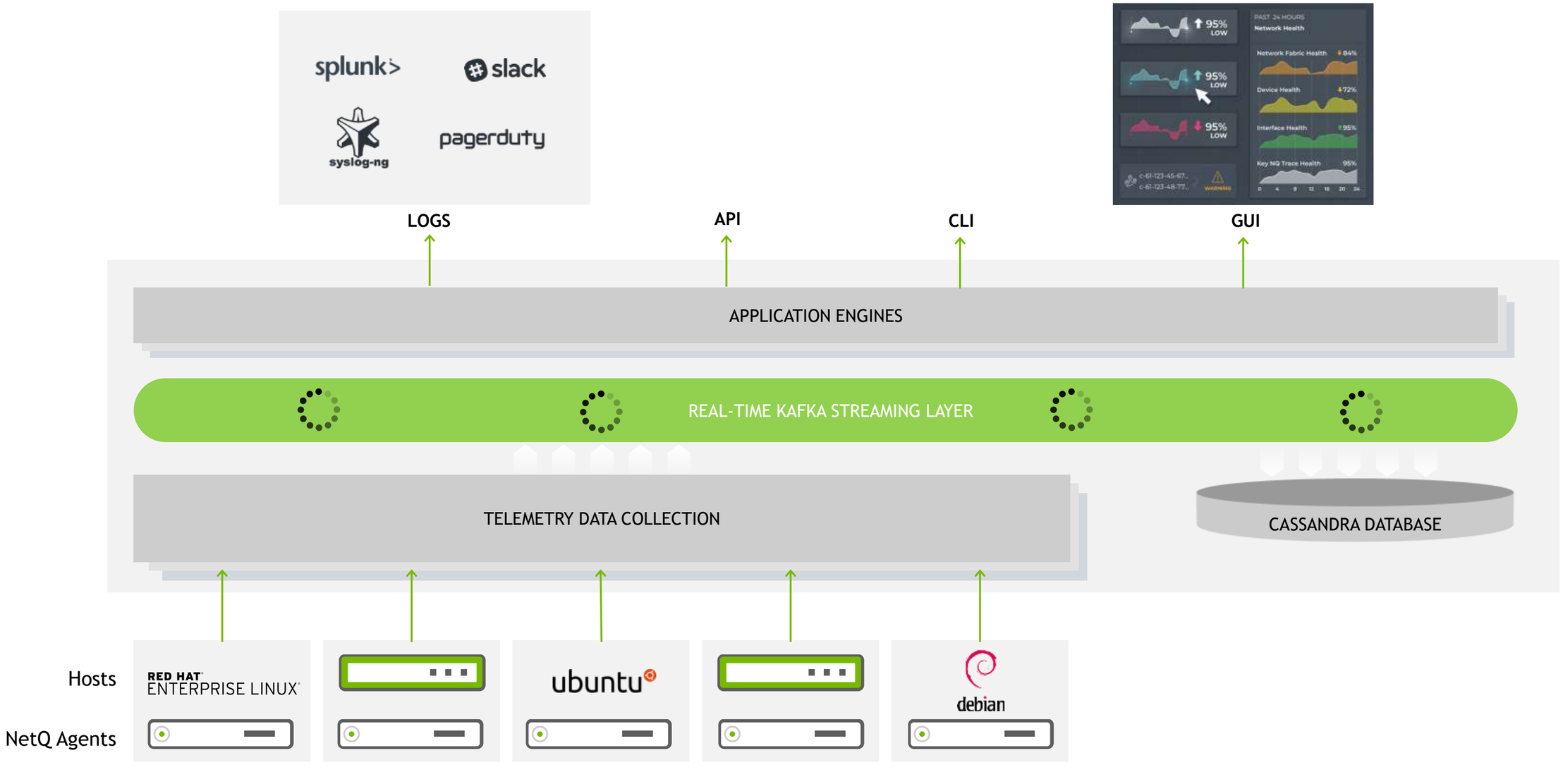


Route Validation
Packet doesn’t reach the firewall
Packet go through a sub-optimized path

Load Balance Validation
Suboptimal ECMP
Suboptimal LAG

NETQ

Real-Time Data Collection and Analysis



TRANSFORMING NETWORK OPERATIONS

CI/CD WITH CUMULUS NETQ

CONFIG CHANGE

Push to repo

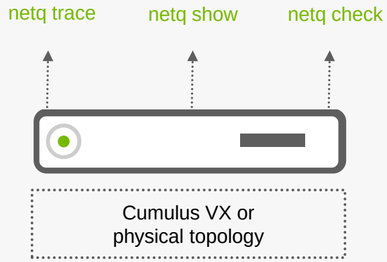
INFRASTRUCTURE AS CODE (IaC) REPO

GitLab

GitHub

Deploy on simulation or staging network

VALIDATE WITH NETQ



Deploy with confidence

PRODUCTION

NetQ eases unit test development for more robust testing and accelerated adoption

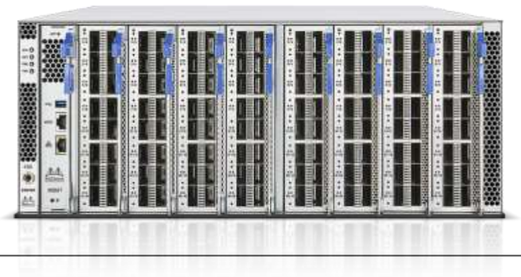
OPEN ETHERNET SN4000 SERIES SWITCH FAMILY

SN4700C - 32x400GbE
400G Spine/Super-spine Switch

SN4600C - 64x100G
High density 100G Spine Switch

SN4600 - 64x200GbE
2U, 200G Spine/Super-spine Switch

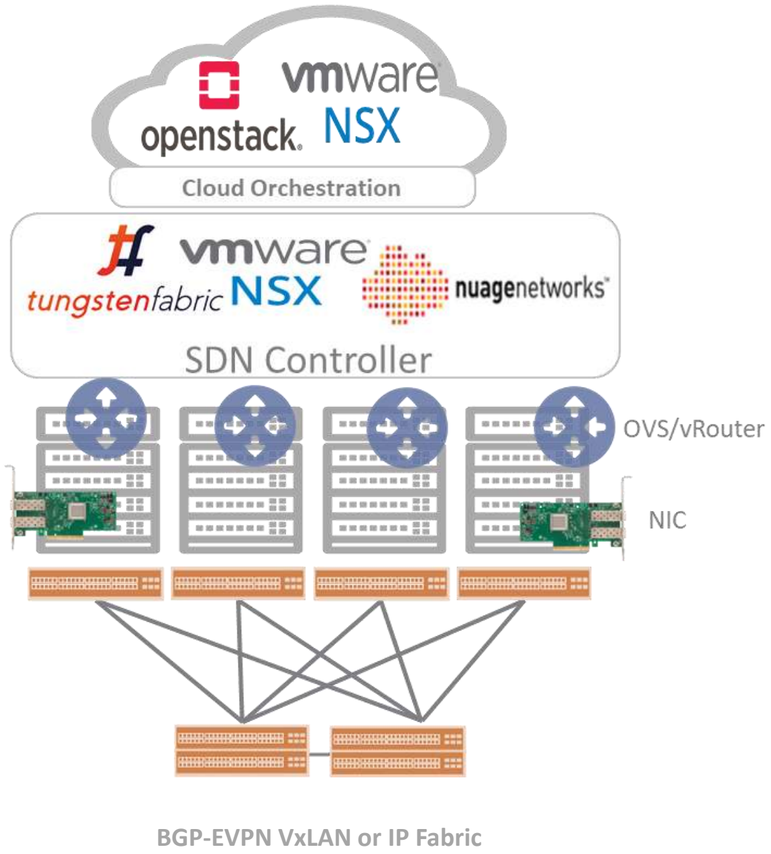
SN4800 - 128x100GbE
4U, Modular platform with 8 line card slots
Line cards – 16x100GbE, 8x200GbE, 4x400BgE



NVIDIA Mellanox Spectrum-3

- 25.6Tbps throughput
- Predictable performance
- Fair traffic distribution for cloud
- Advanced telemetry with WJH
- Best-in-class throughput, latency, power consumption and functionality

CLOUD/SDN AND UNDERLAY



Cloud Orchestration

- Docker/VM life cycle management
- Control OVS on Hypervisor, or Nutron to SDN controller
- Can talking to Switch by ML2 interface directly

Overlay SDN Controller

- Control OVS/vRouter on hypervisor,
- OVSDB to switch for Bare metal Service
- Micro-segmentation integration with service chain VNF

Underlay IP/EVPN Fabric

- Scale out Leaf/Spine Architectures
- Very good Automation, Very good Telemetry
- No Hidden License for BGP/VXLAN/RoCE/Telemetry

cumulus

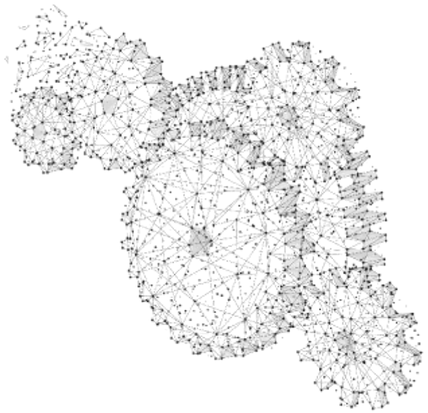
Mellanox TECHNOLOGIES

Open Standards	Scale Out Architecture	Physical Layer Abstraction	Total Host & IP Mobility

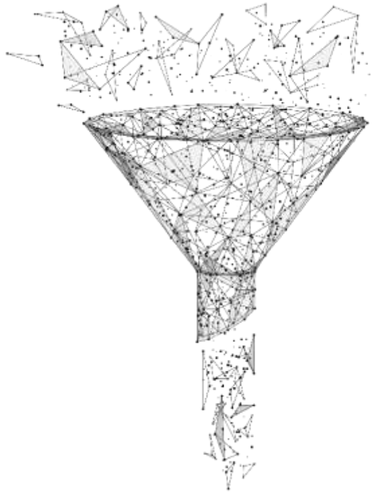
DPU - DATA CENTER INFRASTRUCTURE ON A CHIP

SDN PERFORMANCE CHALLENGES

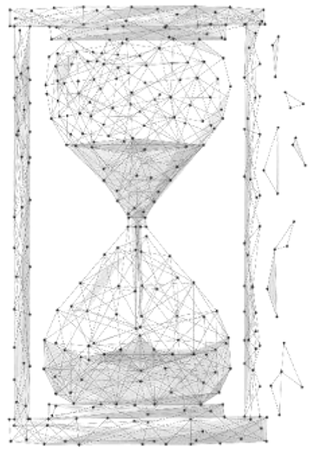
Software-defined networking (SDN) has major performance burdens



High CPU Utilization



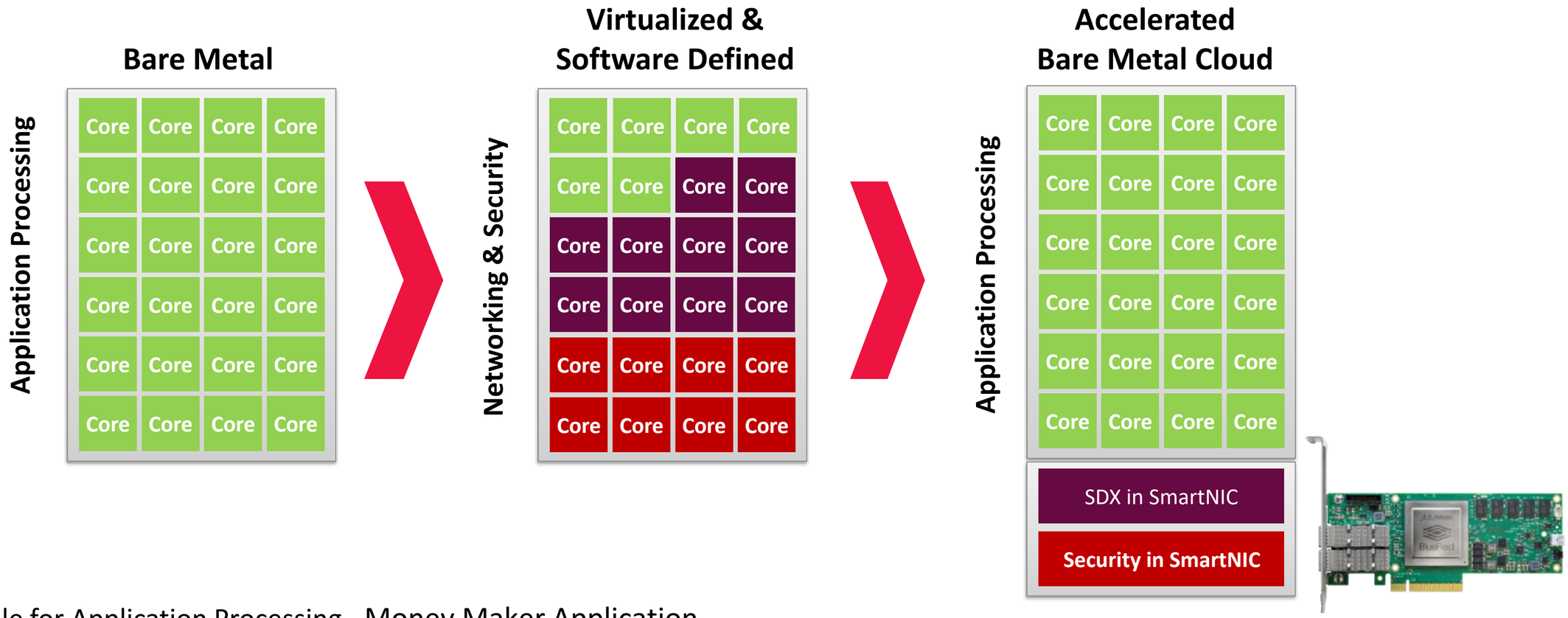
Limited Throughput



Higher Latency

SOFTWARE DEFINED EVERYTHING

Creates Bottlenecks



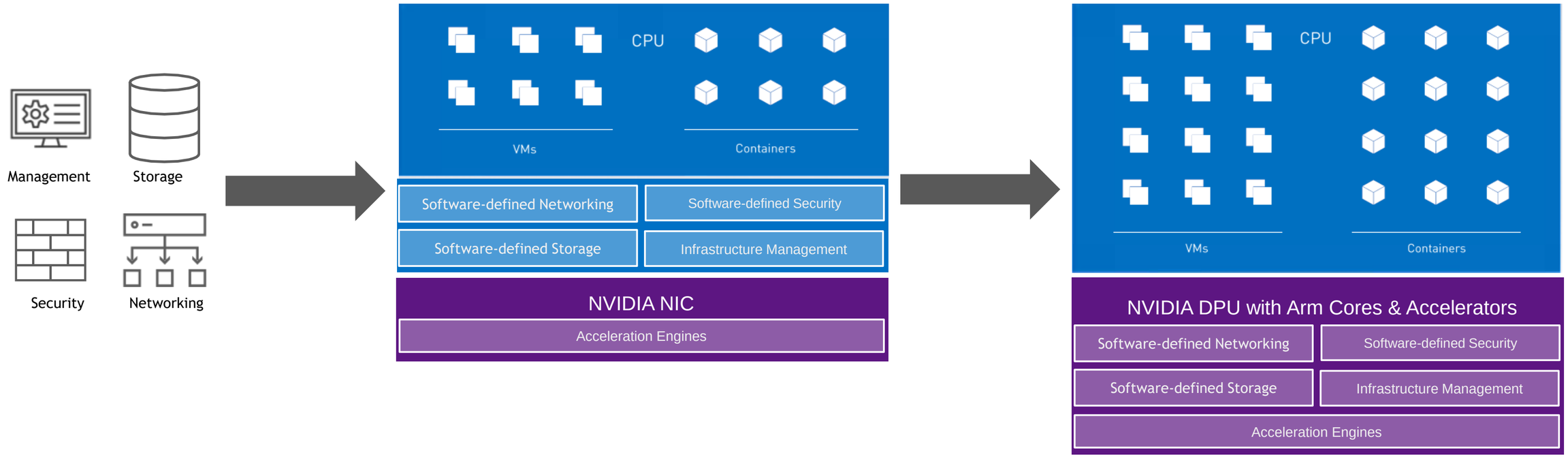
Core Available for Application Processing - Money Maker Application

Core Software Defined Everything (SDX) Consumes CPU cores for Packet Processing
- Virtualization, Storage, Switching, Routing, Load Balancing

Core Security: Consumes CPU cores for Security Processing
- Layer 4 Firewall, encryption, host introspection
- Intrusion detection & prevention

INTRODUCING THE DATA PROCESSING UNIT

Software Defined Data Center Infrastructure-on-a-Chip



From Hardware Appliances

To Software Defined Infrastructure on CPU

To Software Defined Infrastructure on DPU

BLUEFIELD-2
DATA PROCESSING UNIT

Data Center Infrastructure-on-a-Chip

6.9B Transistors

8 64-bit Arm CPUs Cores

Dual 16-way VLIW Engine

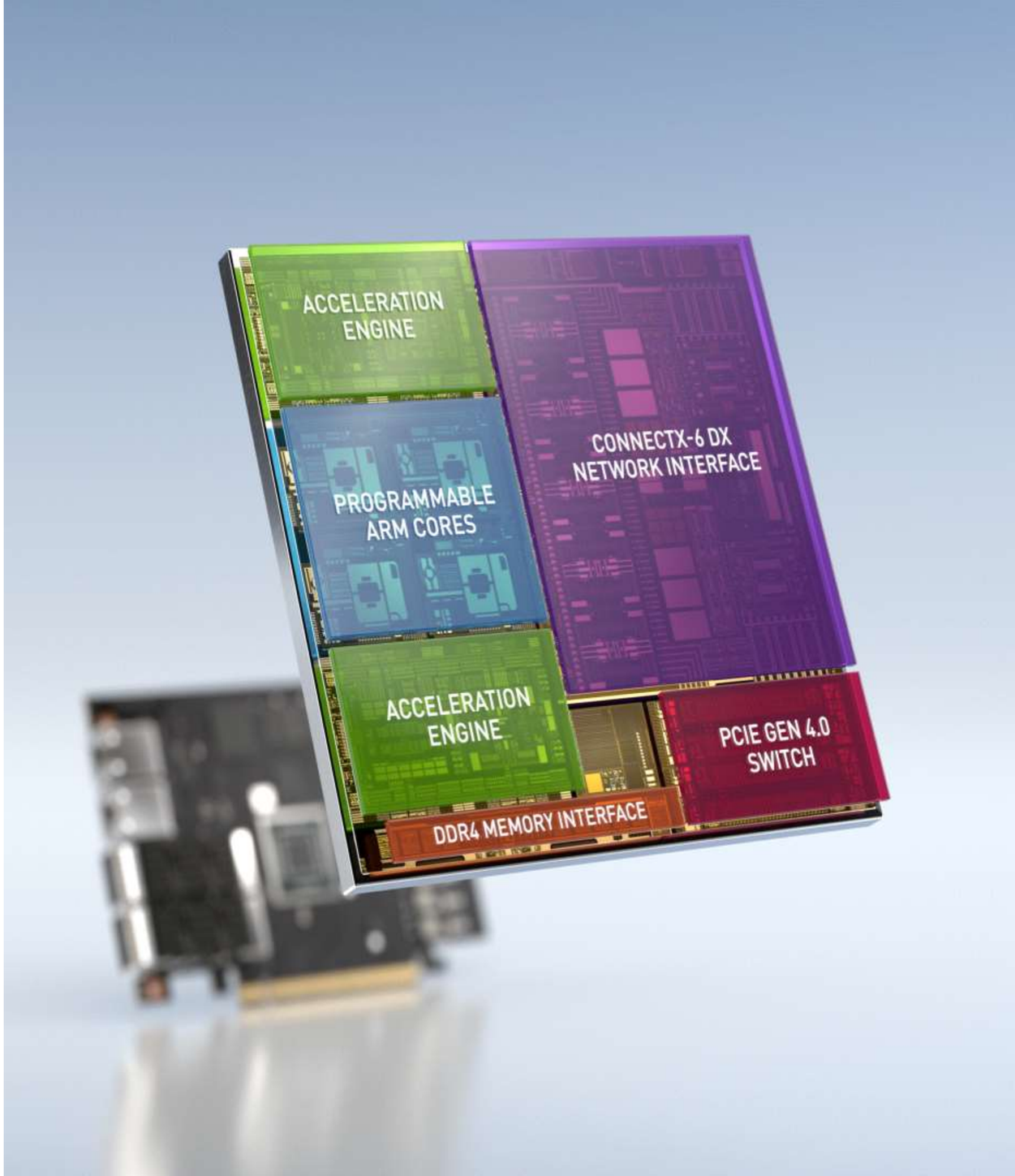
100 Gbps IPsec

50 Gbps RegEx

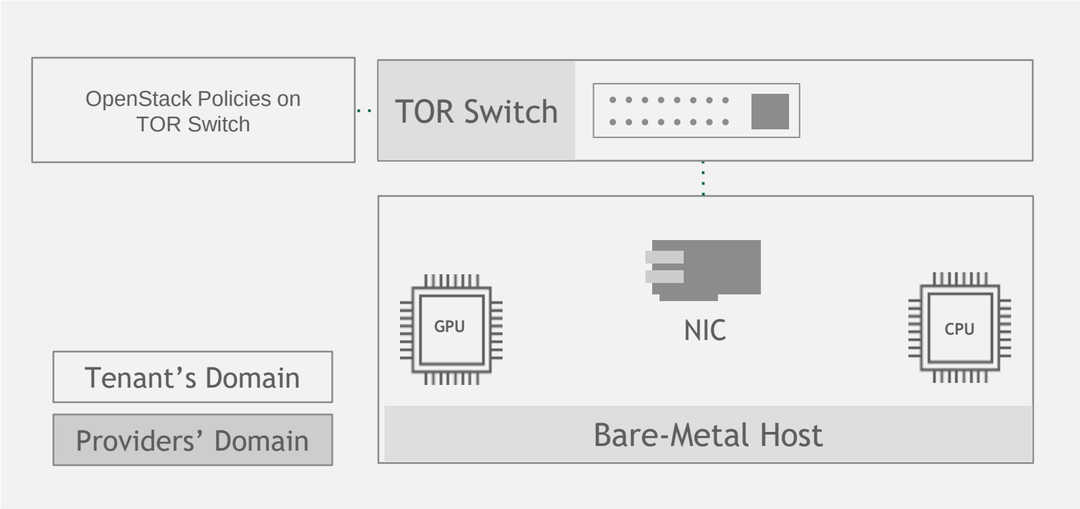
100 Gbps Video Streaming

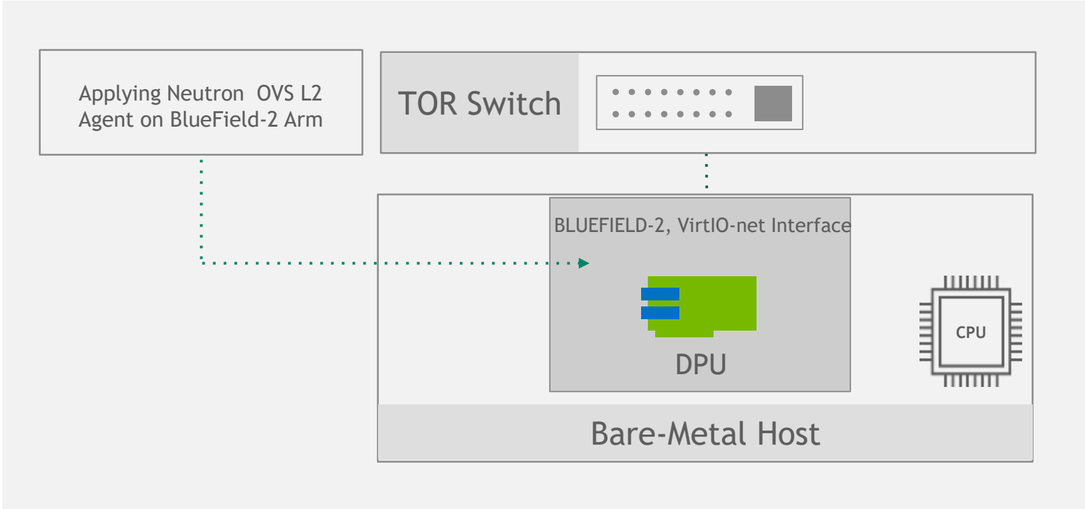
5M NVMe IOPs

Replaces 125 x86 CPU Cores



BLUEFIELD ENABLES SDN IN BARE-METAL CLOUDS





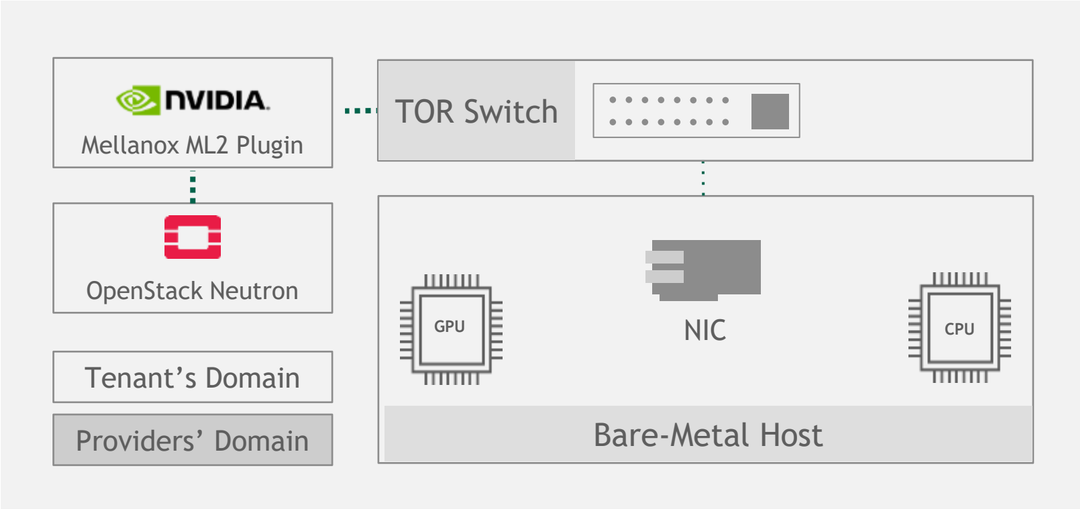
Networking in TOR Switch

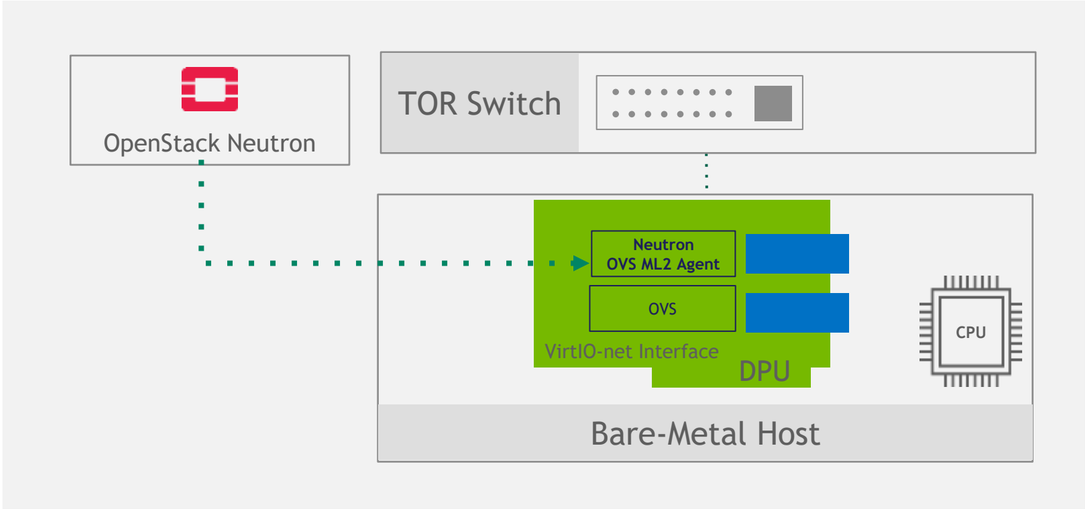
- Limited to no SDN capabilities
- Orchestration through proprietary TOR switch vendor plugins
- Mandates proprietary network driver installation in bare-metal host

SDN Integration with BlueField-2 DPU

- ✔ Full-featured SDN hardware-accelerated capabilities
- ✔ Full orchestration through upstream OpenStack
- ✔ No installation of network driver in bare-metal host
- ✔ Unified and consistent architecture across virtualized and bare-metal cloud environment

BLUEFIELD ENABLES SDN IN BARE-METAL CLOUDS





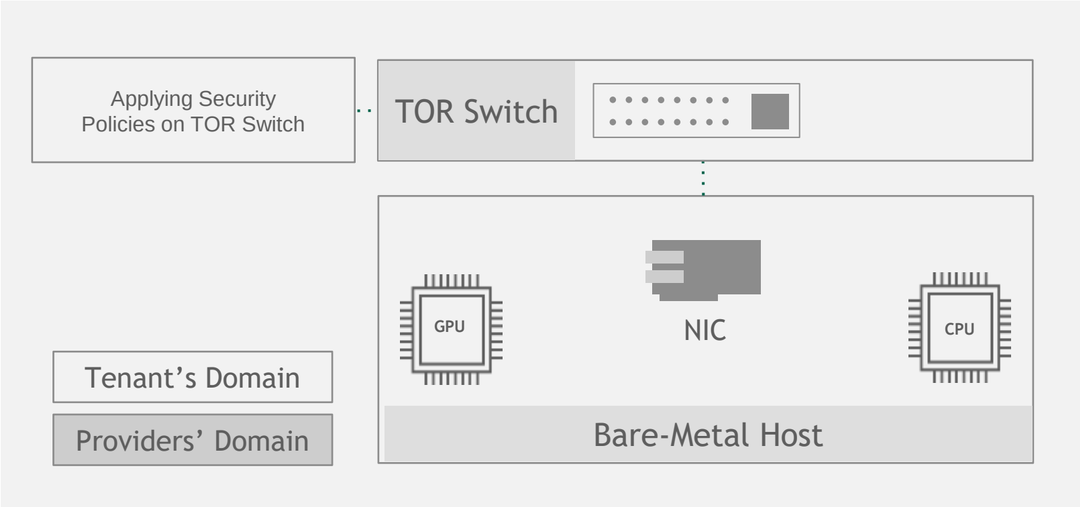
Networking in TOR Switch

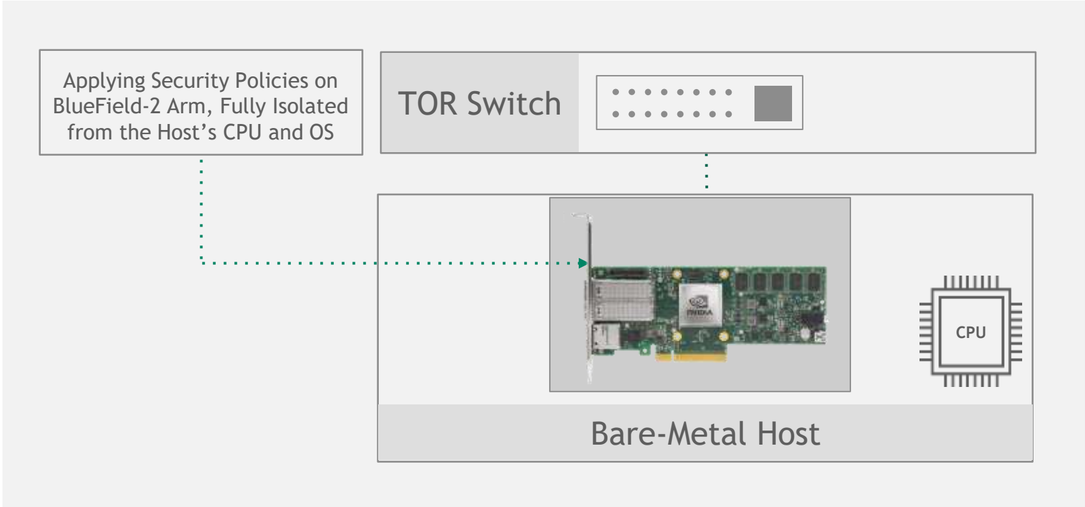
- Limited to no SDN capabilities
- Orchestration through proprietary TOR switch vendor plugins
- Mandates proprietary network driver installation in bare-metal host

SDN Integration with BlueField-2 DPU

- ✔ Full-featured SDN hardware-accelerated capabilities
- ✔ Full orchestration through upstream OpenStack
- ✔ No installation of network driver in bare-metal host
- ✔ Unified and consistent architecture across virtualized and bare-metal cloud environment

BLUEFIELD EMPOWERS SECURE BARE-METAL KUBERNETES





Security Policy in TOR Switch

- Limited to network security with ACLs
- No visibility to the host's workloads, failing to implement effective security strategies
- Increased surface for east-west attacks

Security Policy BlueField-2 DPU

- ✔ Complete isolation of the security enforcement from the tenant's workload
- ✔ Enabling diverse cyber security solutions, enhancing data-center security
- ✔ No need to install agents on servers
- ✔ No impact on server performance

BLUEFIELD-2 VIRTIO-NET ACCELERATION

Bare Metal Use Case – Full Emulation

Use case: Any host with zero changes; Unified configuration

Serving networking and storage to bare-metal host

Isolation of Control plane

Provision rules in ARM/OVS instead of TOR – allowing unified configuration for both Virtualized and Bare Metal Servers

Supports NIC Hot Plug/Un-Plug

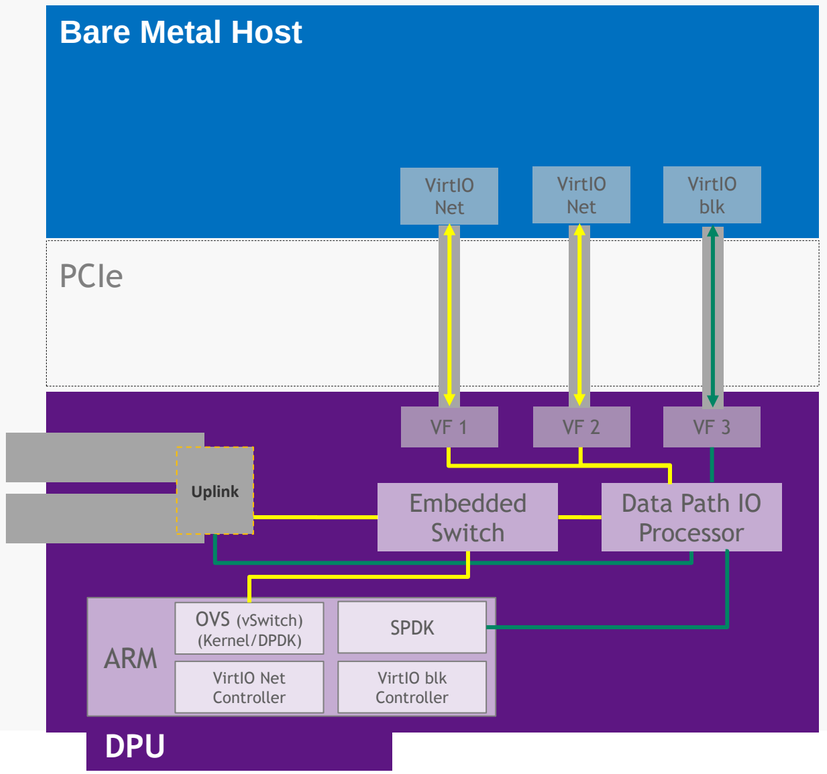
No changes to host drivers

Hardware implements VirtIO spec version 0.95, 1.0 & 1.1

Performance:

VirtIO-Net : 20 Mpps Rx + 20 Mpps Tx

VirtIO-blk : 5.4M IOPs (4K blocks)



ASAP² FEATURES AND CAPABILITIES

Flow based Classification and action

Hierarchal multiple layer tables

Table consists of classification and action

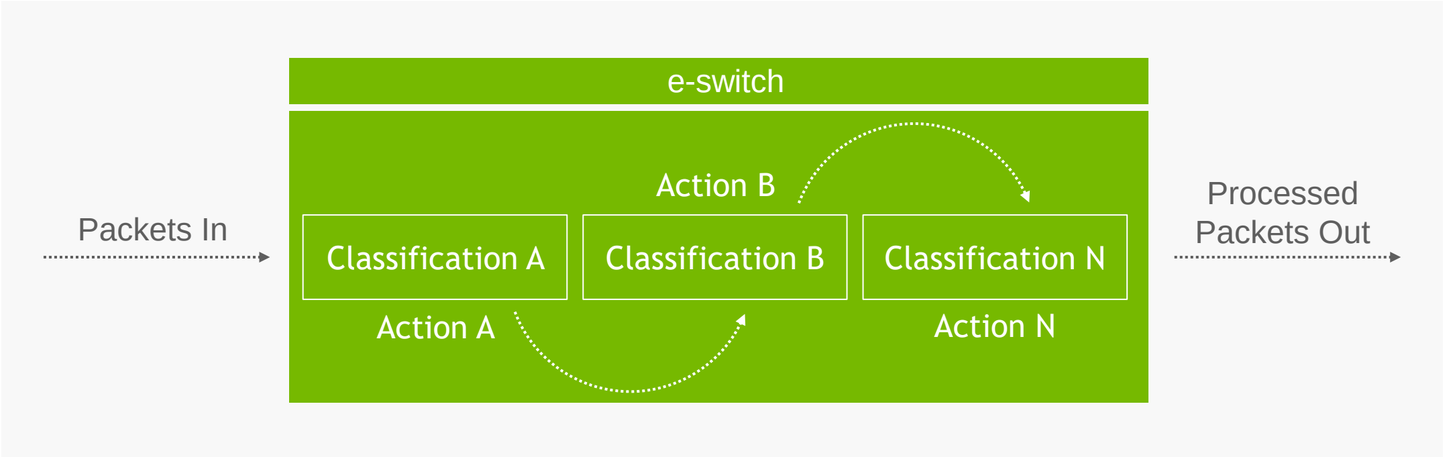
Builds a tree of classification and action

Key fields include: Ethernet L2/IPv4/IPv6/TCP/UDP/Inner Packet (VXLAN/GENEVE/etc.)

Can store and match Meta Data

Flexible fields extraction by “Flexparser”

Action: Allow / Deny, Re-write (Route/NAT), Encap/Decap of headers, Meta Data set, Hairpin, Sample, Counter, etc.



INTRODUCING NVIDIA DOCA

Data Center Infrastructure-on-a-Chip Architecture

Multiple OS

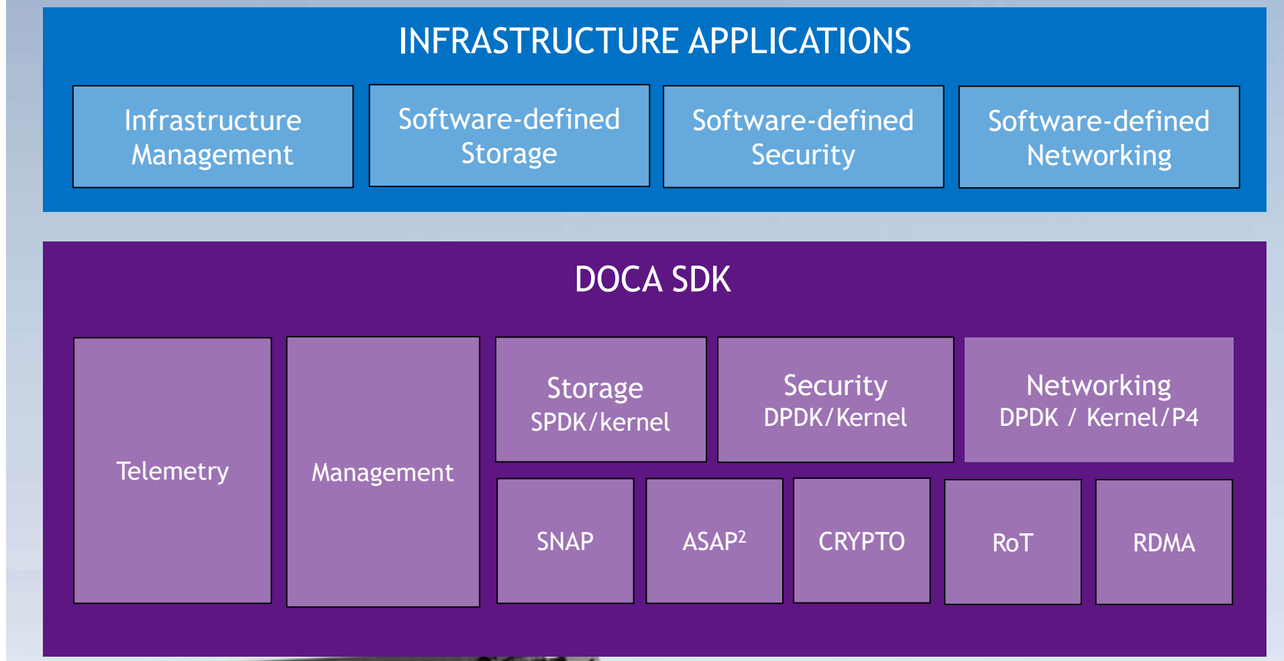
All Programmable Resources

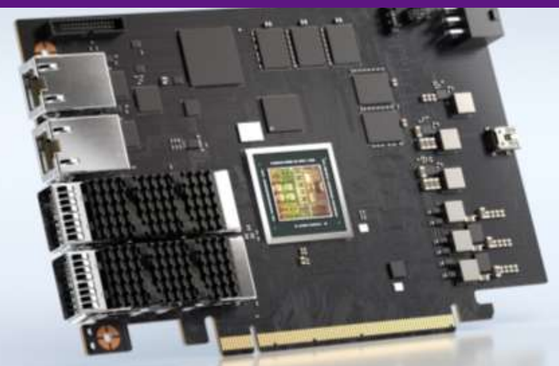
Full Application on DPU

Open Source, Open APIs

Backward and Forward Compatibility

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

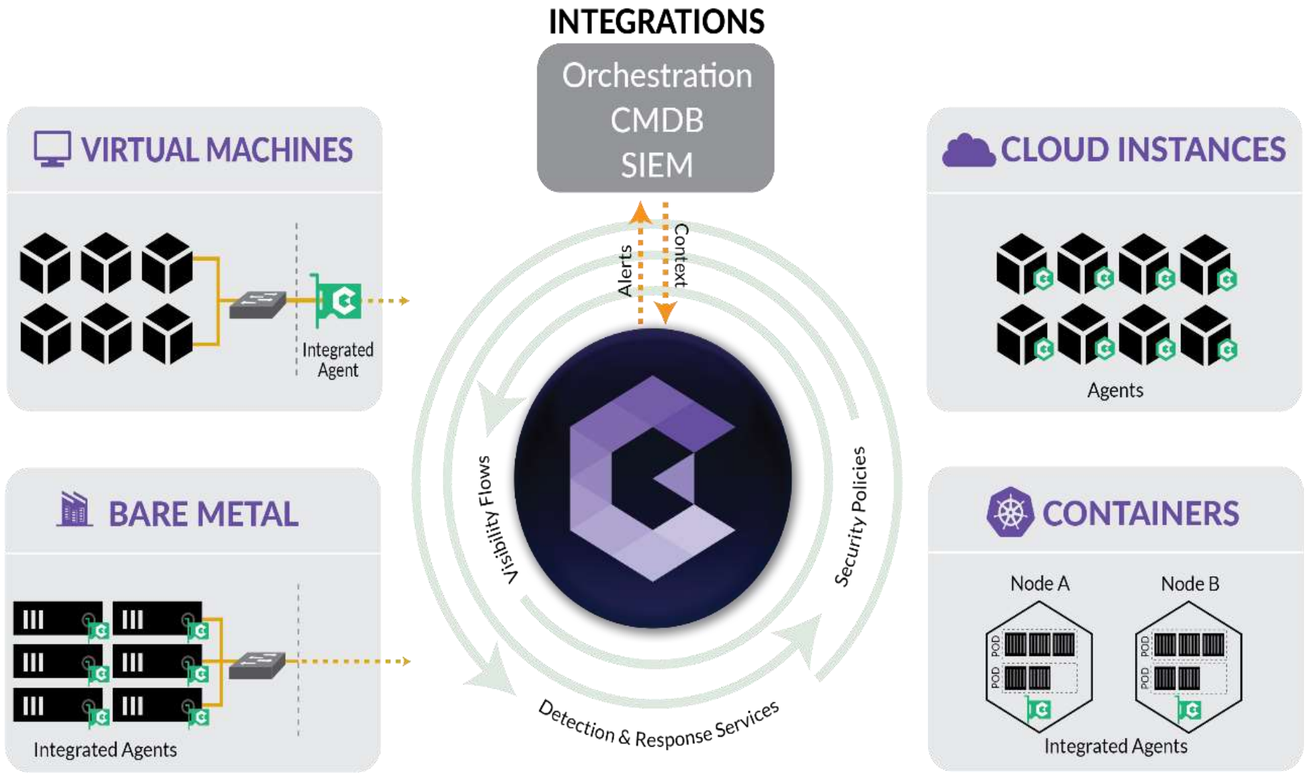




UNLEASHING THE FULL POTENTIAL OF MICRO-SEGMENTATION

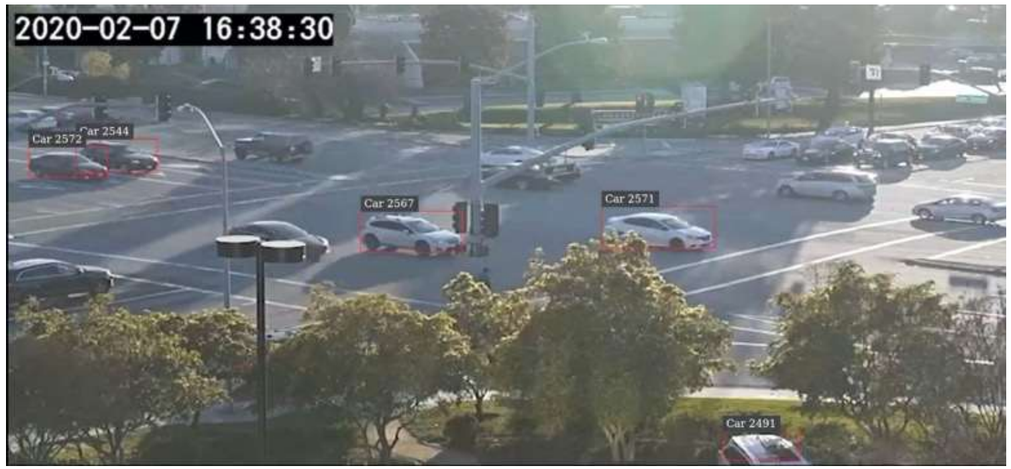
Guardicore and Mellanox Deliver Agentless and High-Performant Micro-Segmentation in Hybrid Cloud

- ✔ Compliance and DevOps Agility
  - No need to install agents on servers
- ✔ Scalability and Performance
  - No impact on server performance
- ✔ Security
  - Full isolation of segmentation enforcement from workload



SOLUTION OVERVIEW

Edge Video Analytics with AI



SW on GPU

SW on CPU

Converged AI Edge

DU
Mellanox FH Termination

CU-UP
+UPF
Mellanox Acceleration

DeepStream
(AI)

Container | Container | Container

UBUNTU OS (HOST) | Kubernetes

NVIDIA Mellanox SmartNIC | Tesla GPU | CPU

IP cameras

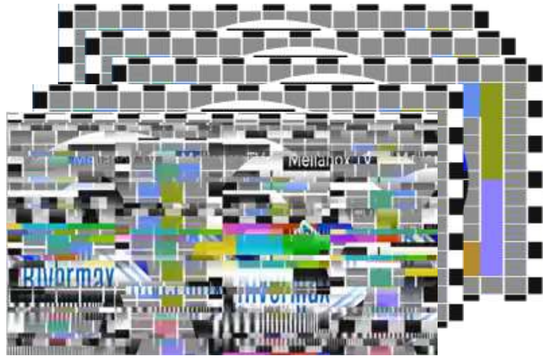
Significantly Reduce Transport Bandwidth Requirement & Cloud Compute + Storage

https://developer.nvidia.com/blog/accelerating-iva-using-ultra-efficient-5g-core-with-mavenir-and-nvidia-edge-ai/

TIMED DATA MOVEMENT – UHD VIDEO (12GBPS) EXAMPLE

Server with standard NIC

High BW timed data works up to a point



Video destroyed
Needs extra hardware to make it function

5 cores ~1 flow

FPGA Needed **~11Gb/s**

Rivermax with NVIDIA NIC

Scales linearly with fewer resources

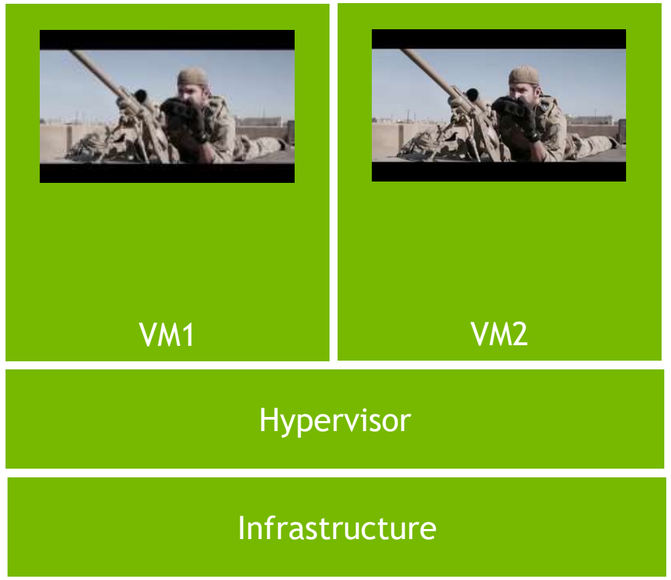
✓ GPU Direct

1 core 8 flows

No FPGA Needed **90Gb/s**

Virtualised Rivermax with NVIDIA NIC

Fully virtualizable for cloud environments



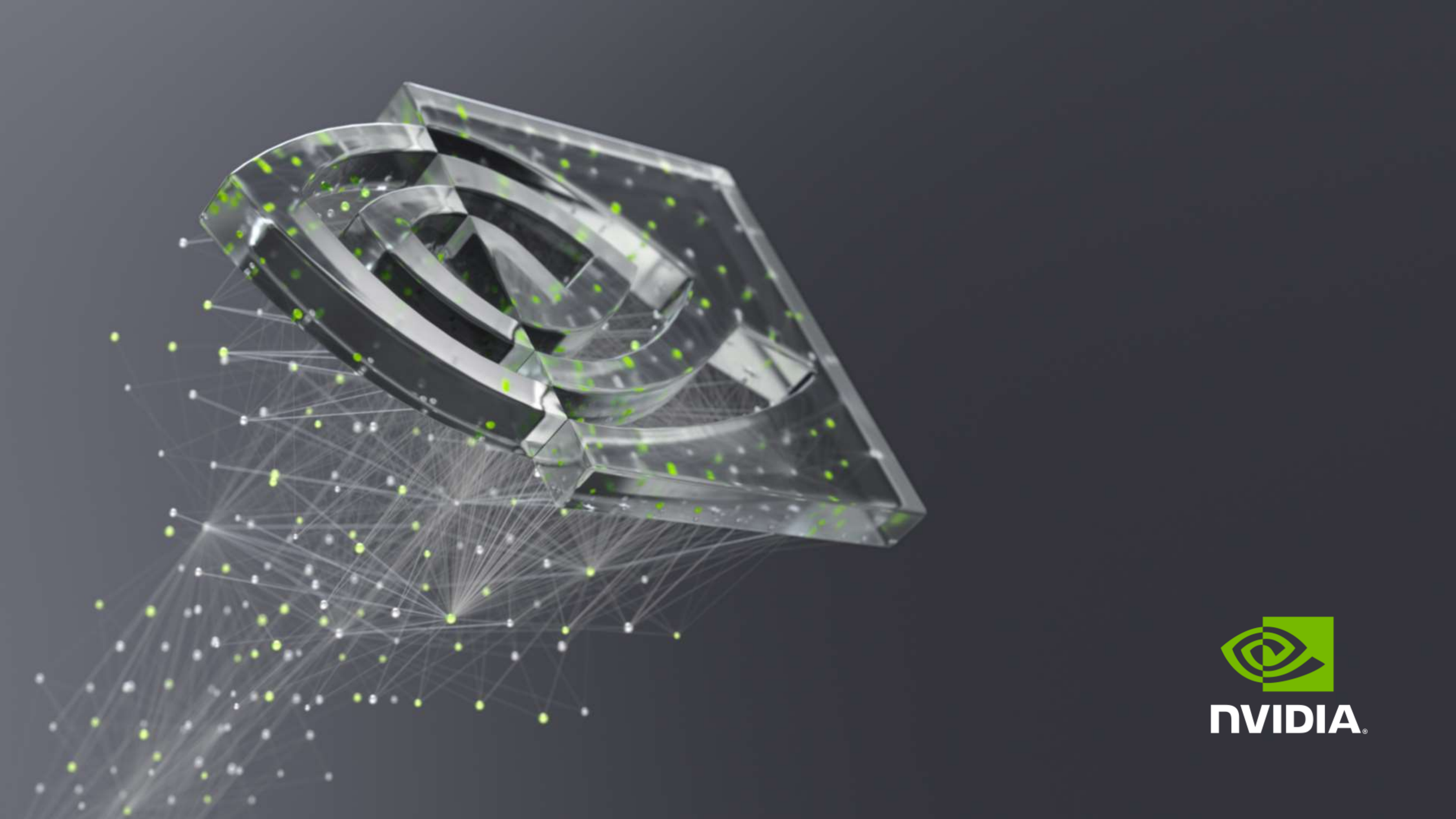
✓ vGPU ✓ vNIC

1 core 8 flows

No FPGA Needed **90Gb/s**

MARKETS AND USE CASES - EXAMPLES

	CSP/Cloud	Telco	Enterprise/FSI	M&E
				
Applications	Cloud Virtualization Bare Metal Cloud Storage AI/HPC Security Video Streaming	5G Transition Cloud RAN	Moving to 25G Cloud Storage	Moving to IP Infra 25G+
Technologies	RoCE ASAP^2, VirtIO, SRIOV Connection Tracking GPU Direct NVMe-OF, NVMe SNAP Security Accelerations Rivermax	ASAP^2, VirtIO, SRIOV 5T for 5G	Ease of use Speeds	Streaming Accelerations - Rivermax GPU Direct Packet Pacing


NVIDIA