NVIDIA IS REINVENTING THE DATA CENTER

The Data Center is the New Unit of Computing

AI & ML

GPU

DPU

Data Processing Unit

Accelerated Computing

Accelerating SDx, security, telemetry, and virt

CPU

Just Announced at GTC2021:

NVIDIA Grace CPU

Learn More at:

www.NVIDIA.com/GTC
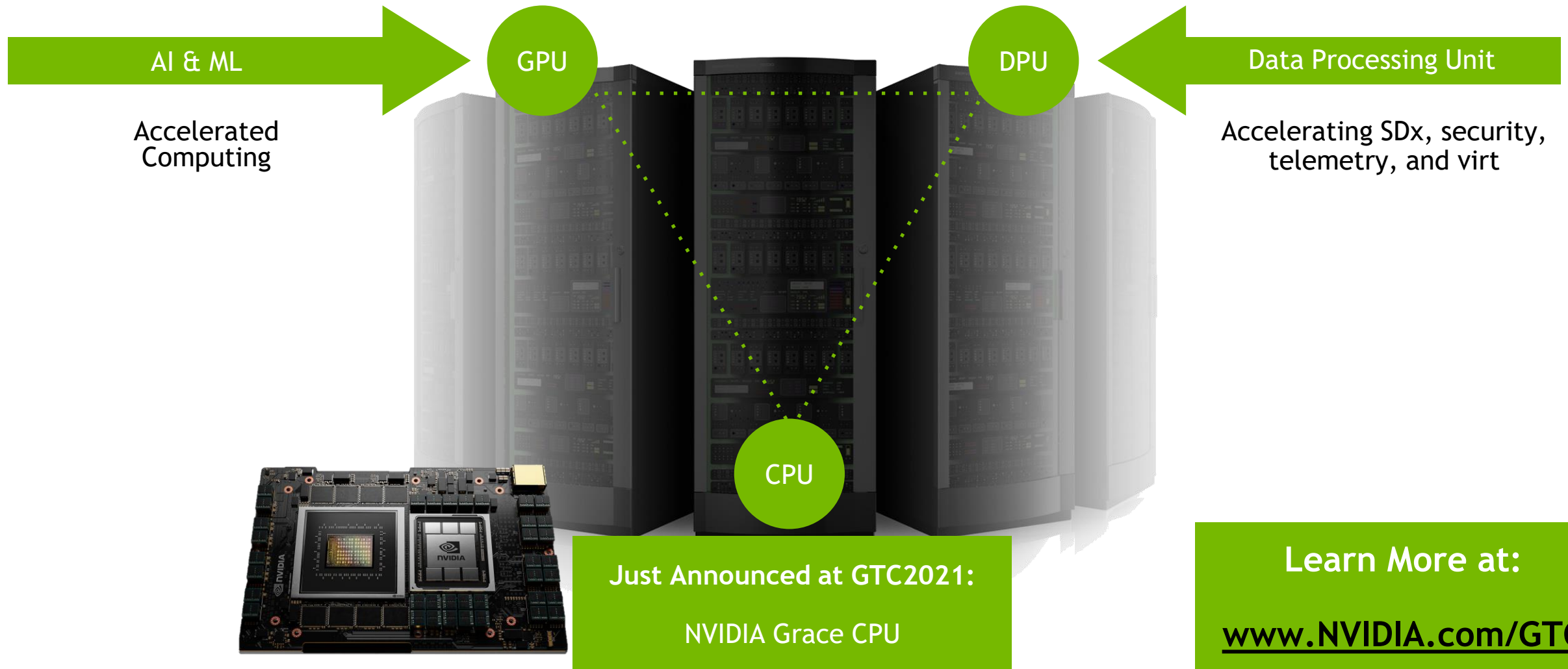
CHALLENGES FOR HIGH FREQUENCY TRADERS

Increasing Market Data rates and bursts

- Market data feeds peaking above 10Gbps
- 25G and 40G handoffs available already

No innovation above 10G

- Existing Low Latency solutions capped at 10G

Obscured Network

- Lack of visibility for congestion
- Lack of real time buffer utilization

Merchant Network Silicon is not aligned with HFT Market

NVIDIA.

LOWEST LATENCY SWITCHES FOR HFT

	ToR & Edge	ToR & Spine	
Lowest Latency + Features Spectrum	SN2010: 18x1/10/25G + 4x100G SN2410: 48x1/10/25G + 8x100G	SN2100: 16x100G SN2700: 32x100G	✓ 300ns Latency * ✓ PTP Boundary Clock ✓ Egress SPAN Timestamping
Low Latency + Advanced Features Spectrum-2 Spectrum-3	SN3420: 48x1/10/25G + 12x100G **AI Switches** SN3700: 32x200G SN4600: 64x200G	SN3700C: 32x100G SN4600C: 64x100G SN4800: 128x100G, 64x200, 32x400G	✓ **NAT in Hardware** ✓ 400 or 500ns Latency * ✓ PTP Boundary Clock ✓ Stratum3 Oscillator ✓ In-Band and SPAN Timestamping

3 GENERATIONS OF ASICS

Lowest Latency, High Performance, HFT Feature Set

|	Spectrum	Spectrum-2	Spectrum-3
Buffer Architecture	Fully Shared	Fully Shared	Fully Shared
Buffer Size	16MB	42MB	64MB
Latency	300ns *	400ns *	500ns *
Low Jitter	Yes	Yes	Yes
Forwarding Table	256k	512k	512k
Max. Multicast Scale	13k	100k+	100k+
PTP Boundary Clock	Yes	Yes	Yes
Stratum3 Oscillator	No	Yes	Yes
SPAN Timestamping	Yes	Yes	Yes
In-Band Timestamping	No	Yes	Yes
NAT in Hardware	No	Yes	Yes

* Not STAC Benchmarks

NETWORK PACKET BROKERING

cVu® 32100/32100E Series

- Data Center
- HPC Cluster
- HFT / FinServ
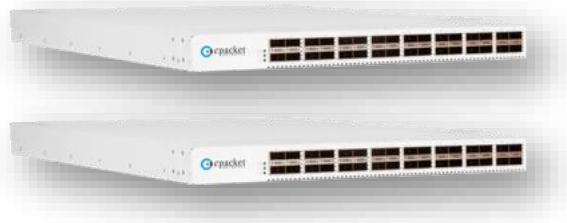
▸ Up to 100G real-time intelligence

▸ Non-blocking design with zero packet-loss

▸ Low latency monitoring with millisecond analytics

▸ Nanosecond time-stamping with PTP synchronization

▸ Header stripping for protocols

▸ Open monitoring architecture with RESTful

▸ Flexible aggregation through TAP/SPAN/RSPAN

▸ Centrally manageable through cPacket cClear®

▸ Integrated w/ cPacket cStor® Packet Capture & cProbe® flow devices

"The NVIDIA Spectrum-2 switch provides the lowest latency available with a non-blocking design and zero packet-loss at speeds up to 100GbE. This performance is mandatory to provide crucial network analytics and monitoring in the most demanding trading environments."
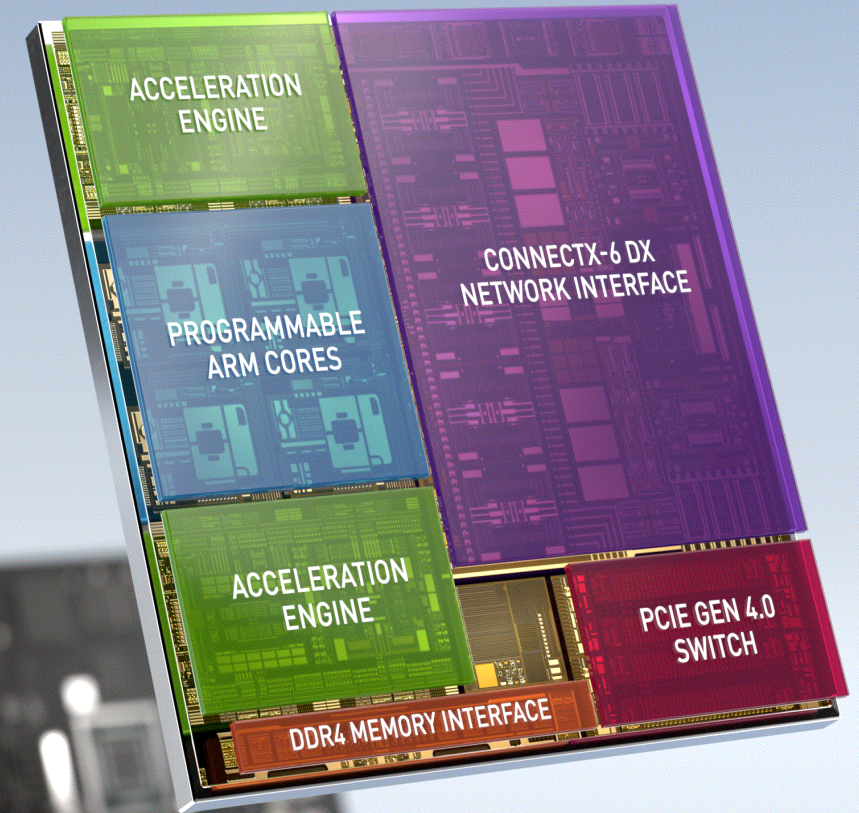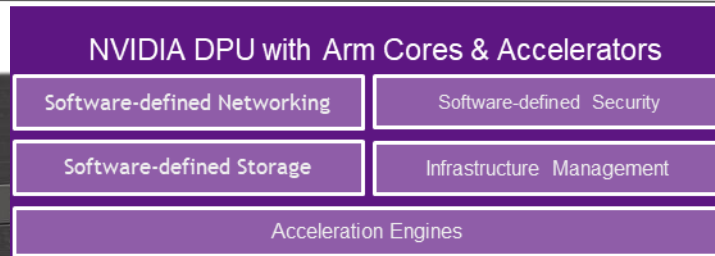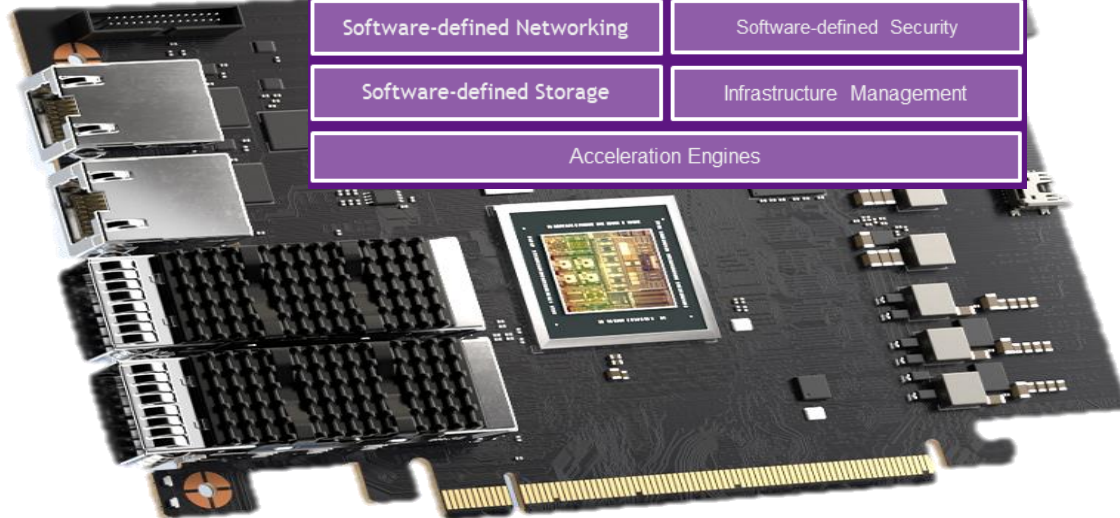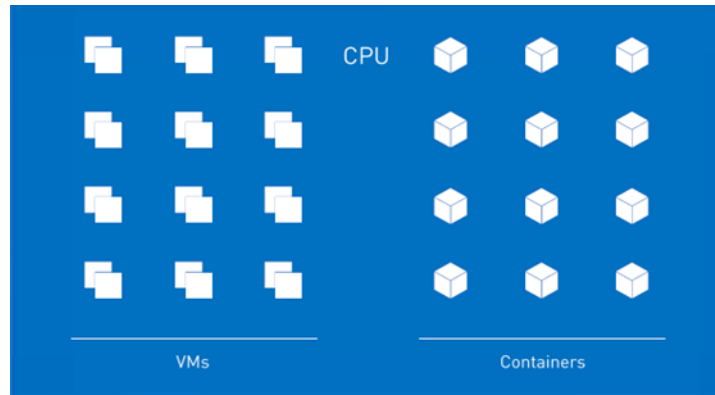
– Daryl Stanbery, Vice President, cPacket

NVIDIA BLUEFIELD-2 DPU (DATA PROCESSING UNIT)

Data Center Infrastructure-on-a-Chip

NVIDIA DOCA

Enabling Broad DPU Partner Ecosystem
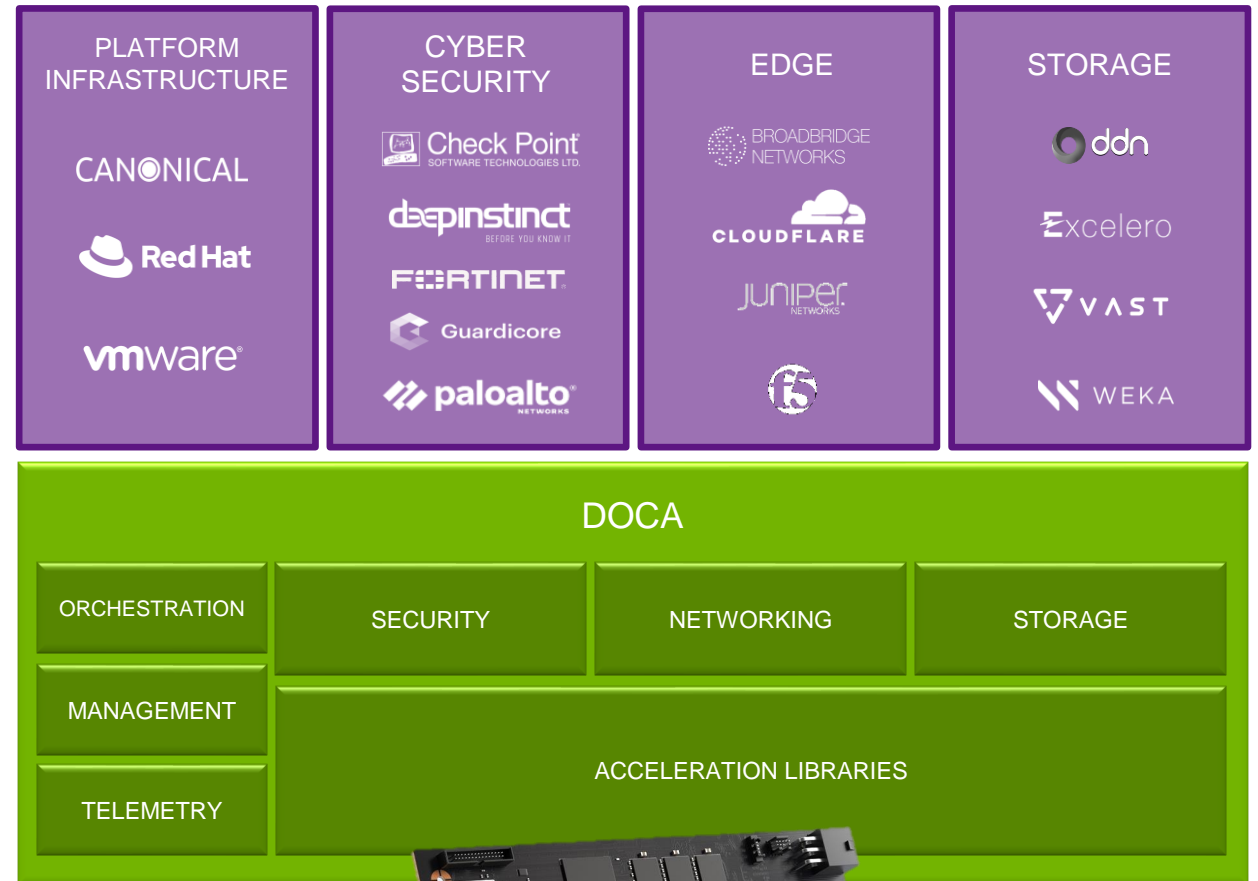
Software application framework for BlueField DPUs

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

Sign up for DOCA access at https://developer.nvidia.com/networking/doca

PLATFORM INFRASTRUCTURE	CYBER SECURITY	EDGE	STORAGE
CANONICAL	Check Point SOFTWARE TECHNOLOGIES LTD.	BROADBRIDGE NETWORKS	ddn
Red Hat	deepinstinct BEFORE YOU KNOW IT	CLOUDFLARE	Excelero
vmware	FORTINET	JUNIPER NETWORKS	VAST
	Guardicore		WEKA
	paloalto NETWORKS	f5	

DOCA

ORCHESTRATION	SECURITY	NETWORKING	STORAGE
MANAGEMENT	ACCELERATION LIBRARIES		
TELEMETRY			

PRECISION TIMING IN DPU & CONNECTX

Ultra-precise Network Time Stamps and Synchronization

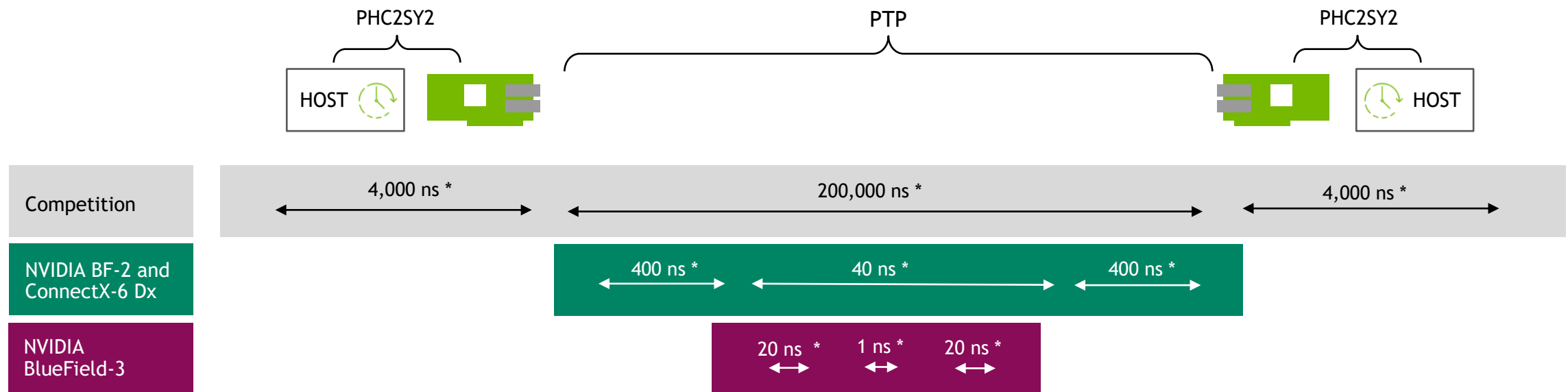| **Transaction Time Stamping**	**Distributed Applications**	**Streaming Data**
Precise transaction records and event logs at every location	Synchronize apps, storage and monitoring across multiple sites	Optimize capture/broadcast of market data

PHC2SY2   PTP   PHC2SY2

HOST   HOST

Competition   4,000 ns *   200,000 ns *   4,000 ns *

NVIDIA BF-2 and ConnectX-6 Dx   400 ns *   40 ns *   400 ns *

NVIDIA BlueField-3   20 ns *   1 ns *   20 ns *

View GTC Session w/ Facebook at:
https://gtc21.event.nvidia.com/media/Time%20Synchronization%20in%20Distributed%20Data%20Centers%20%5BS31889%5D/1_qytjmdhx

* Not STAC Benchmarks