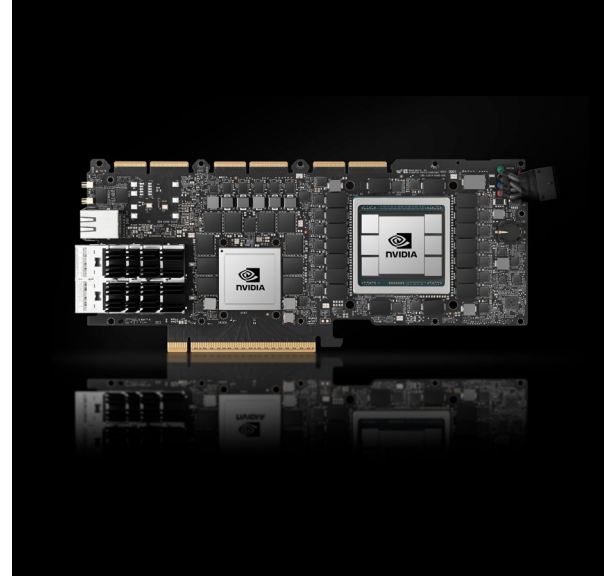




Converged Accelerators

Networking and compute, unified.



A More Powerful, Secure Enterprise

NVIDIA converged accelerators combine the power of the **NVIDIA Ampere architecture** with the enhanced security and networking of the **NVIDIA® BlueField® data processing unit (DPU)**, all in one high-performance package.

Better Performance

Because the GPU architecture and DPU are connected via an integrated PCIe switch, there's a dedicated path for data transfer between the GPU and network. This eliminates performance bottlenecks and enables much more predictable performance—important for time-sensitive applications.

Enhanced Security

The convergence of GPU and DPU creates a more secure AI processing engine, where data generated at the edge can be sent across the network fully encrypted without traveling over the server PCIe bus. This boosts protection from network-based threats.

Smarter Networking

The architecture of converged cards allows GPU processing to be applied directly to traffic flowing to and from the DPU. This enables a new class of AI-based networking and security applications, such as data leak detection, network optimization, and more.

Cost Savings

Because the GPU, DPU, and PCIe switch are combined on a single card, mainstream servers can perform tasks previously only possible with high-end or purpose-built systems. Even edge servers can benefit from the same performance boost.

Key Features

- > NVIDIA Ampere architecture
- > NVIDIA BlueField DPU
- > Integrated PCIe switch

Top Use Cases

- > 5G vRAN
- > AI-based cybersecurity
- > AI on 5G

NVIDIA AX800

- > BlueField-3 DPU
- > Ampere architecture GPU
- > PCIe Gen5

NVIDIA A100X

- > BlueField-2 DPU
- > NVIDIA A100 GPU
- > PCIe Gen4

NVIDIA A30X

- > BlueField-2 DPU
- > NVIDIA A30 GPU
- > PCIe Gen4

Products

NVIDIA converged accelerators are available in three form factors.

A30X

The A30X combines the NVIDIA A30 Tensor Core GPU with the BlueField-2 DPU. It provides a good balance of compute and input/output (IO) performance for use cases such as 5G vRAN and AI-based cybersecurity.

A100X

The A100X brings together the NVIDIA A100 Tensor Core GPU with the BlueField-2 DPU. It's ideal for workloads where compute demands are greater, such as AI-on-5G, signal processing, and multi-node training.

AX800

The AX800 combines an NVIDIA Ampere architecture GPU with a BlueField-3 DPU. It delivers maximum performance on a dense, power efficient form factor when performing AI-driven data processing and analytics on edge and data center traffic.

Developer Ecosystem

NVIDIA converged accelerators expand the capabilities of CUDA® and NVIDIA DOCA™ programming libraries for workload acceleration and offloading. CUDA applications can be run on the x86 host or on the DPU's Arm processor for isolated AI and inferencing applications.

Technical Specifications			
	AX800	A100X	A30X
Arm Cores	16x Arm A78 cores at 2.3GHz	8x Arm A72 cores at 2GHz	8x Arm A72 cores at 2GHz
GPU Memory	80GB HBM2e	80GB HBM2e	24GB HBM2e
Memory Bandwidth	2,039GB/s	2,039GB/s	1,223GB/s
Multi-Instance GPU (MIG) Instances	7	7	4
Interconnect	PCIe Gen5, NVIDIA NVLink® bridge	PCIe Gen4, NVIDIA NVLink® bridge	PCIe Gen4, NVIDIA NVLink® bridge
Networking	2x 200Gb/s ports, Ethernet or InfiniBand	2x 100Gb/s ports, Ethernet or InfiniBand	2x 100Gb/s ports, Ethernet or InfiniBand
Form Factor	Dual-slot FHFL	Dual-slot FHFL	Dual-slot FHFL
Max Power	350W	300W	230W

Ready to Get Started?

To learn more about NVIDIA converged accelerators, visit www.nvidia.com/converged-accelerators/

