






















































NVIDIA Data Center Platform




Artificial intelligence is driving a new era of advanced reasoning, revolutionizing workflows and transforming every industry. Accelerated computing allows organizations to address AI, scientific computing, digital twin, and graphics-based workloads. Meeting these challenges requires attention to the three scaling laws: pretraining scaling (initial large dataset training), post-training scaling (enhancing models through fine-tuning), and test-time scaling (allocating more compute during inference to enable reasoning). Each scaling law places unique demands on data center infrastructure.


The NVIDIA accelerated computing platform provides a unified solution, delivering breakthrough performance for AI reasoning, simulation, digital twins, and advanced graphics. As AI workloads grow in complexity and diversity, selecting the right GPU—matched to your needs and specific use case—is critical for maximizing performance and innovation. This guide will help you navigate the NVIDIA solution portfolio, ensuring your infrastructure is ready for the next generation of AI reasoning, simulation, or graphics-driven applications.

NVIDIA Data Center Compute Portfolio

							
Solution	Networking	Pretraining	Post-Training	Test-Time/Inference	Data Analytics	HPC/Scientific	AI Video Generation
GB300 NVL72	QTM-X800/SPTMX						
GB200 NVL72	QTM-X800/SPTMX						
GB200 NVL4	QTM-2/SPTMX						
GH200 NVL2 / NVL4	QTM-2/SPTMX						
HGX B300	QTM-X800/SPTMX						
HGX B200	QTM-X800/SPTMX						
HGX H200	QTM-2/SPTMX						
H200 NVL	QTM-2/SPTMX						

Price-performance comparisons are relative within each column. This chart should be used in conjunction with measured data for targeted workloads.

 Best










 Good

QTM-X800 NVIDIA Quantum-X800 InfiniBand switch plus BlueField®-3 DPUs and ConnectX®-8 SuperNICs

QTM-2 NVIDIA Quantum-2 InfiniBand switch plus BlueField-3 DPUs and ConnectX-7 NICs

SPTMX NVIDIA Spectrum-X™ Ethernet switch and SuperNICs plus BlueField-3 DPUs

NVIDIA Data Center Universal Graphics and Compute Portfolio

Solution	 Networking	 Pretraining	 Post-Training	 Test-Time/ Inference	 Data Analytics	 HPC/Scientific	 Omniverse™/ Render	 Virtualization¹ vWS/VD	 Far Edge/AI Video
RTX PRO 6000 Blackwell Server Edition	QTM-2/SPTMX	■	■	■	■	■	■	■	
L40S	QTM-2/SPTM3	■	■	■	■	■	■	■	
L4	QTM-2/SPTM3		■	■	■	■	■	■	■

1. vGPU is supported across all products shown

Price-performance comparisons are relative within each column. This chart should be used in conjunction with measured data for targeted workloads.

■ Best

■

■

■ Good

QTM-2 NVIDIA Quantum-2 InfiniBand switch plus BlueField-3 DPUs and ConnectX-7 NICs

SPTMX NVIDIA Spectrum-X Ethernet switch and SuperNICs plus BlueField-3 DPU

SPTM3 NVIDIA Spectrum-3 Ethernet switch plus ConnectX-6 NICs and BlueField-2 DPUs

Ready to Get Started?

For more information on NVIDIA data center GPUs, visit: nvidia.com/data-center-gpus

