



NVIDIA DATA CENTER PLATFORM

Accelerating Every Workload

Modern applications are transforming every business. From AI for better customer engagement, to data analytics for forecasting, to advanced visualization for product innovation, the need for accelerated computing is rapidly increasing. Because new compute demands are outstripping the capabilities of traditional CPU-only servers, enterprises need to optimize their data centers—making this acceleration a must-have. The NVIDIA data center platform is the world's leading accelerated computing solution, deployed by the largest supercomputing centers and enterprises. It enables breakthrough performance with fewer, more powerful servers, driving faster time to insights, while saving money.

The platform accelerates a broad array of workloads, from AI training and inference to scientific computing and virtual desktop infrastructure (VDI) applications, with a diverse range of GPUs from the highest performance to entry-level, all powered by a single unified architecture. For optimal performance, it's essential to identify the ideal GPU for a specific workload. Use this as a guide to those workloads and the corresponding NVIDIA GPUs that deliver the best results.

CHOOSE THE RIGHT DATA CENTER GPU

Solution Category	GPU	DL Training & Data Analytics	DL Inference	HPC / AI	Omniverse/ Render Farms	Virtual Workstation	Virtual Desktop (VDI)	Mainstream Acceleration	Far Edge Acceleration	AI-on-5G
Compute	H100*	PCIE SXM CNX	PCIE SXM CNX	PCIE SXM CNX				PCIE CNX		CNX
	A100	PCIE SXM A100X	PCIE SXM	PCIE SXM A100X				PCIE A100X		A100X
	A30		PCIE	PCIE				PCIE		A30X
Graphics / Compute	L40									
	A40									
	A10									
	A16									
Small Form Factor Compute/ Graphics	A2									
	T4									

Price-performance comparison within each solution category (Compute, Graphics and Compute, SFF Compute and Graphics) and workload column.

- Best
- Better
- Good

- PCIe Form Factor
- SXM Form Factor

- CNX - H100 + ConnectX-7 Converged Accelerator
- A100X/A30X - A100 or A30 + BlueField-2 Converged Accelerator

* H100 Tensor Core GPUs for mainstream servers include NVIDIA AI Enterprise software suite and support. The NVIDIA AI Enterprise software suite reduces time to development and simplifies deployment of AI workloads and makes H100 the most powerful end-to-end AI and HPC data center platform.

To learn more about NVIDIA data center GPUs, visit [NVIDIA.com/data-center](https://www.nvidia.com/data-center)

