NVIDIA DATA CENTER GPU

Linecard

Contact a PNY account manager or email GOPNY@PNY.COM
WWW.PNY.COM/DATA-CENTER-GPUS

PRODUCTS		NVIDIA A100 PCIe 80GB	NVIDIA A100 PCIe 40GB	NVIDIA A30	NVIDIA A40	NVIDIA A10	NVIDIA T4	NVIDIA A16
PNY PART NUMBER		NVA100TCGPU80-KIT	NVA100TCGPU-KIT	NVA30TCGPU-KIT	NVA40TCGPU-KIT	NVA10TCGPU-KIT	TCSCT4-KIT	NVA16TCGPU-KIT
WORKLOAD	DESCRIPTION	Highest Performance Compute	Highest Performance Compute	Mainstream Compute	Highest Performance Graphics	Mainstream Graphics	Small Footprint Low Power	Optimized for VDI
		Recommended Number of GPUs per Server						
Deep Learning (DL) Training and Data Analytics	For the absolute fastest model training and analytics	**4-8 GPUs** 80GB: Bn+ parameter models (DLRM, GPT-2)	**4-8 GPUs** 40GB: Bn+ parameter models (DLRM, GPT-2)					
DL Inference	For batch and real-time inference	**1-2 GPUs** w/multi-instance GPU (MIG) 80GB: large batch size constrained models (RNN-T)	**1-2 GPUs** w/multi-instance GPU (MIG) 40GB: large batch size constrained models (RNN-T)	**2–4 GPUs** with MIG		**4–8 GPUs**	**4–8 GPUs**	
High-Performance Computing (HPC) / AI	For Higher Education Research and scientific computing centers	**1-4 GPUs** with MIG	**1-4 GPUs** with MIG	**2–4 GPUs** with MIG				
Render Farms	For batch and real-time rendering				**4–8 GPUs**	**4–8 GPUs**		
Graphics	For the best graphics performance on professional VDI				**2–4 GPUs** for high-end virtual workstations*	**2-8 GPUs** for mid-range virtual workstations*	**2-8 GPUs** for entry-level virtual workstations*	**2–4 GPUs** for highest virtual desktop user density**
Cloud Gaming	For 4K resolution / Android				**4–8 GPUs** (4K resolution)	**4–8 GPUs** (4K resolution)	**1-2 GPUs** (Android)	
Enterprise Acceleration	For mixed workloads, including graphics, ML, DL, analytics, training, and inference	**1-2 GPUs** with MIG for compute workloads	**1-2 GPUs** with MIG for compute workloads	**1-2 GPUs** with MIG for compute workloads	**1-2 GPUs** for graphics-intensive workloads*	**1-2 GPUs** for graphics-intensive* and compute workloads	**1-4 GPUs** for balanced workloads*	
Edge Acceleration	For differing use cases and deployment locations	**1-2 GPUs** with MIG	**1-2 GPUs** with MIG	**1-2 GPUs** with MIG	**1-4 GPUs** for graphics-intensive workloads & AR / VR*	**1-8 GPUs** for inference and video workloads	**1-8 GPUs** for inference and video workloads	

* NVIDIA RTX Virtual Workstation (vWS) software license required for virtual workstation workloads.
** NVIDIA Virtual PC (vPC) software license required for VDI workloads.

PNY Technologies, Inc. 100 Jefferson Road, Parsippany, NJ 07054 | Tel 973-515-9700 | Fax 973-560-5590 | www.PNY.com

PNY 3 YEAR WARRANTY | PNY PRO | NVIDIA PARTNER