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	NVIDIA A30	NVIDIA A2	NVIDIA A40	NVIDIA A16	NVIDIA A100X	NVIDIA A30X
PNY PART NUMBER		NVA100TCGPU80-KIT	NVA30TCGPU-KIT	NVA2TCGPU-KIT	NVA40TCGPU-KIT	NVA16TCGPU-KIT	NVA100XTCGPUCA-KIT	NVA30XTCGPUCA-KIT
WORKLOAD	DESCRIPTION	Highest Performance Compute	Mainstream Compute	Entry-Level Compact AI	Highest Performance Graphics	Optimized for VDI	Highest Performance Converged Accelerator	Mainstream Converged Accelerator
Recommended Number of GPUs or Converged Boards 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-3)					**1-2 cards** for multi-node training	
DL Inference	For batch and real-time inference	**1-2 GPUs** w/ multi-instance GPU (MIG) 80GB: large batch size constrained models (RNN-T)	**2–4 GPUs** with MIG	**1–4 GPUs**				
High-Performance Computing (HPC) / AI	For Higher Education Research and scientific computing centers	**2-4 GPUs** with MIG	**2–4 GPUs** with MIG				**1-2 cards** for multi-node workloads	
Render Farms	For batch and real-time rendering				**4–8 GPUs**			
Graphics	For the best graphics performance on professional VDI			**1–4 GPUs** for entry-level virtual workstations*	**2–4 GPUs** for midrange to high-end virtual workstations*	**2-4 GPUs** for highest virtual desktop and workstation user density**		
Cloud Gaming	For 4K resolution / Android			**1–4 GPUs** for mobile android	**4–8 GPUs** (4K resolution)			
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	**1–4 GPUs** for balanced workloads*	**1-2 GPUs** for graphics-intensive workloads*			**1 card** for compute acceleration with software-defined infrastructure
Edge Acceleration	For differing use cases and deployment locations	**1-2 GPUs** with MIG	**1-2 GPUs** with MIG	**1-4 GPUs** for inference and video workloads	**1-4 GPUs** for graphics-intensive workloads & AR / VR*		**1 card** for AI-on-5G with heavy workloads	**1 card** for AI-on-5G with average workloads
5G vRAN	For low-latency GPU-network communication							**1-2 cards**
AI-Based Security	For GPU-powered network processing							**1 card**

* 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