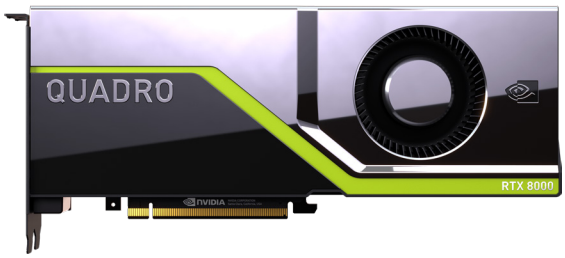
NVIDIA RTX SERVERS

BUILT TO AN NVIDIA REFERENCE DESIGN BY SELECT PNY PARTNERS

NVIDIA QUADRO AUTHORIZED PARTNER

PNY



NVIDIA Quadro RTX features such as RT Cores for cinematic quality production ray-tracing, multi-precision Tensor Cores that accelerate AI/DL/ML/MV and big data analytics, or AI enhanced design and visualization tools, support for virtual GPU (vGPU) that brings Quadro performance with data center security to devices ranging from tablets to non-Quadro equipped mobile or desktop PCs, and ideal attributes for at the edge eXtended Reality (XR) AR and VR deployments in 5G and Wi-Fi 6 environments, are compelling and beneficial reasons for institutions to purchase NVIDIA RTX Servers.

NVIDIA authorizes and supports the utilization of NVIDIA Quadro RTX 8000 and RTX 6000 (actively or passively cooled) professional graphics boards in server enclosures for data center deployment to realize these use cases. All offer similar GPU performance, but the RTX 8000 option offers an unprecedented 48 GB of GPU memory, while the NVIDIA Quadro RTX 6000 provides 24 GB. All utilize ultra-fast GDDR6 with optional ECC, and NVLink offers GPU memory pooling for two cards, providing 96 GB or 48 GB respectively, along with performance scaling since GPU core counts are effectively doubled. Passively cooled versions of the RTX 8000 and RTX 6000 are available across a much wider range of server chassis than their actively cooled counterparts, making Quadro RTX in the data center more accessible than ever before to offer previously unrealizable levels of performance and paradigm shifting capabilities – all with Quadro IT manageability.

SERVERS COMPATIBLE WITH ACTIVELY COOLED QUADRO RTX 8000 OR RTX 6000 BOARDS

VENDOR	SERVER CHASSIS SUPPORTED
ASUS	ESC8000 G4 | ESC4000 G4/G4S/G4X | E900 G4 | RS720-E9-RS8-G
Quanta	D528V-2U | D52G 4U | D43J-3U | D43KQ-2U | D43N-5U
Supermicro	4029GP-TRT2 | 7049GP-TR
TYAN	FT77D-B7109 | T48T-B7105

SERVERS FOR PASSIVELY COOLED QUADRO RTX 8000 OR RTX 6000 BOARDS

The introduction of passively cooled Quadro RTX 8000 and RTX 6000 products enables the integration of RTX-powered solutions into existing (or future) server enclosures suitable for NVIDIA Tesla full-height and dual-slot width passively cooled boards – a change that significantly widens systems compatibility – and a first for any ultra-high-end Quadro product.

FOR MORE INFORMATION, CONTACT YOUR PNY ACCOUNT MANAGER OR EMAIL GOPNY@PNY.COM
WWW.PNY.COM/QUADRO

PNY Technologies, Inc. 100 Jefferson Road, Parsippany, NJ 07054 | Tel 973-515-9700 | Fax 973-560-5590 | WWW.PNY.COM/PNYPRO

KEY SPECIFICATIONS FOR NVIDIA QUADRO RTX 8000 AND RTX 6000 (ACTVE OR PASSIVE) IN MULTI-GPU CONFIGURATIONS:

	1X GPU	2X GPU	4X GPU	8X GPU
CUDA Cores	4608	9216	18432	36864
RT Cores	72	144	288	576
Tensor Cores	576	1152	2304	4608
RTX-OPS	84T	168T	336T	672T
Rays Cast	10 Giga Rays/Sec	20 Giga Rays/Sec	40 Giga Rays/Sec	80 Giga Rays/Sec
Peak FP32 Performance	16.3 TFLOPS	32.6 TFLOPS	65.2 TFLOPS	130.4 TFLOPS
Peak FP16 Performance	32.6 TFLOPS	65.2 TFLOPS	130.4 TFLOPS	260.8 TFLOPS
Peak INT8 Performance	206.1 TOPS	412.2 TOPS	824.4 TOPS	1684.8 TOPS
Deep Learning TFLOPS	130.5 Tensor TFLOPS	261.0 Tensor TFLOPS	522.0 Tensor TFLOPS	1044.0 Tensor TFLOPS
Board Power Consumption	295 W	590 W	1180 W	2360 W
RTX 8000 GPU Memory	48 GB	96 GB	192 GB	384 GB
RTX 6000 GPU Memory	24 GB	48 GB	96 GB	192 GB
NVLink Bandwidth	100 GB/sec Bidirectional | Between 2x GPUs			

NVIDIA QUADRO RTX 8000 AND RTX 6000 (ACTIVE OR PASSIVE) SERVERS SUPPORT A WIDE ARRAY OF MARKETS AND SOLUTIONS:

	VIRTUAL WORKSTATIONS	RENDERING	DATA SCIENCE	HPC AND SIMULATION	XR/AR/VR AT THE EDGE
Workload	Workstations for Design and Visualization	Offline Rendering, On-Demand Viewport Rendering, Workstations and Render Nodes	Workstations for Data Science R&D	Workstations for HPC Compute and Visualization	Development Platforms for AR/VR over 5G
ISV Software	Hypervisor, ISV Applications	Renderer, ISV Applications, Hypervisor	Data Science Software, Hypervisor	HPC Applications, Hypervisor	XR/AR/VR/Applications, Development Tools, Hypervisor
NVIDIA Software	Quadro vDWS, CUDA-X AI, OptiX	Quadro vDWS, CUDA-X AI, OptiX	Quadro vDWS, CUDA-X AI, NGC Containers	Quadro vDWS, NGC Containers	Quadro vDWS, Development Tools
NVLink Bandwidth	100 GB/sec Bidirectional | Between 2x GPUs				
Server Enclosure	Any Server Compatible with NVIDIA Tesla Full-height, Dual Slot Width Passively Cooled Boards				

NVIDIA Quadro RTX 8000 and RTX 6000 (actively or passively cooled) servers deliver exponential power at a fraction of the cost of CPU-based alternatives. For rendering the RTX solution is typically 1/4th the cost, for AI 1/5th the cost, and for HPC 1/7th the cost. To learn more about how NVIDIA RTX Servers can enhance innovation, boost productivity, offer data center security for sensitive IP, and realize significant operational efficiencies, please email **GOPNY@PNY.COM**.

FOR MORE INFORMATION, CONTACT YOUR PNY ACCOUNT MANAGER OR EMAIL GOPNY@PNY.COM
WWW.PNY.COM/QUADRO

NVIDIA QUADRO AUTHORIZED PARTNER
PNY

PNY Technologies, Inc. 100 Jefferson Road, Parsippany, NJ 07054 | Tel 973-515-9700 | Fax 973-560-5590 | WWW.PNY.COM/PNYPRO