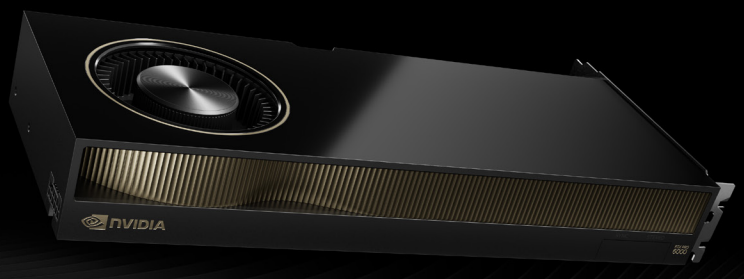
Datasheet

NVIDIA RTX PRO 6000 Blackwell Max-Q Workstation Edition

Powering the next era of AI.



Transform Workflows With Scalable Workstation Performance

As AI continues to advance at an incredible pace, industries face mounting pressure to harness its transformative power and adopt tools capable of handling generative AI, real-time simulation, and hyper-realistic rendering. Enterprises need solutions that combine breakthrough performance, scalability, and versatility to tackle the rise of increasingly complex workloads—from training domain-specific AI models to rendering billion-polygon engineering designs or simulating real-world physics with higher fidelity and precision.

The NVIDIA RTX PRO™ 6000 Blackwell Max-Q Workstation Edition is designed for professionals tackling the most demanding AI, data science, and creative workflows. Built on the NVIDIA Blackwell architecture, this GPU delivers groundbreaking AI and neural rendering capabilities, combining the latest RTX technology with 96 GB of memory to accelerate AI development, data science, and immersive content creation.

With seamless scalability from single to multi-GPU configurations, professionals can tailor performance to their most demanding workloads. Scale up to four RTX PRO 6000 Max-Q GPUs to tackle colossal datasets, fine-tune LLMs, run local AI assistants, and drive virtual production and large-scale simulations. Engineered for multi-GPU flexibility, the RTX PRO 6000 Max-Q enables teams to scale resources as needed, ensuring that today's investments accelerate tomorrow's breakthroughs.

Unmatched desktop performance meets scalable computing, empowering professionals to pioneer new frontiers in AI, graphics, and beyond.

Breakthrough Innovations

The NVIDIA Blackwell architecture combines breakthrough AI, ray tracing, and neural rendering technology, with massive performance and memory improvements to drive cutting-edge professional creative, design, and engineering workflows and power the next decade of innovation.

NVIDIA Blackwell Streaming Multiprocessor: The new SM features increased processing throughput, and new neural shaders that integrate neural networks inside of programmable shaders to drive the next decade of AI-augmented graphics innovations.

Key Features

- Enhanced Streaming Multiprocessors (SMs) built for neural shaders
- 5th Gen Tensor Cores support FP4 precision, DLSS 4 Multi Frame Generation
- 4th Gen Ray Tracing Cores built for detailed geometry
- 96 GB of GDDR7 memory
- 1.8 TB/s of memory bandwidth
- 9th Gen NVENC and 6th Gen NVDEC with 4:2:2 support
- PCIe Gen 5
- Four Display Port 2.1b connectors
- Multi-instance GPU (MIG) support
- AI Management Processor

5th Gen Tensor Cores: Deliver up to 3X the performance of the previous generation and support for FP4 precision for faster AI model processing times with reduced memory usage, enabling local fine-tuning of LLMs and generative AI.

4th Gen Ray Tracing Cores: Double the ray-triangle intersection rate of the previous generation to create photoreal, physically accurate scenes and immersive 3D designs with RTX Mega Geometry, which enables up to 100X more ray-traced triangles.

Next-Gen Video Engines: Enhance video conferencing, video production, and streaming workflows with real-time AI processing. Ninth-generationNVENC and the sixth-generation NVDEC engines provide support for 4:2:2 encoding and decoding to explore a new realm of high-resolution video workflows.

GDDR7 Memory: New and improved GDDR7 memory significantly boosts bandwidth and capacity, empowering your applications to run faster, and work with larger, more complex datasets. With 96 GB of GPU memory and 1.8 TB/s bandwidth, tackle massive 3D and AI projects, fine-tune AI models locally, explore large-scale VR environments, and drive larger multi-app workflows.

DLSS 4: Multi Frame Generation ensures ultra-smooth frame pacing for lifelike simulations. Experience up to 3X faster frame rates and stunning image quality in supported game engines and 3D rendering applications for smoother, more responsive performance.

PCIe Gen 5: Support for PCIe Gen 5 provides double the bandwidth of PCIe Gen 4, improving data-transfer speeds from CPU memory and unlocking faster performance for data-intensive tasks like AI, data science and 3D modeling.

DisplayPort 2.1: Achieve unparalleled visual clarity and performance, driving high-resolution displays at up to 8K at 240 Hz and 16K at 60 Hz. Increased bandwidth enables seamless multi-monitor setups, ideal for multitasking and collaboration, while HDR and higher color depth support ensures superior color accuracy for precision work, such as video editing, 3D design, and live broadcasting.

Universal MIG: Split a single RTX PRO 6000 Max-Q into multiple isolated instances, each with its own resources, allowing for concurrent execution of multiple workloads, optimized GPU utilization, and secure isolation of different applications or users.

Enterprise Reliability

Designed for professionals who demand the best, NVIDIA RTX PRO solutions deliver unparalleled performance, reliability, and support. Every GPU is rigorously tested for a wide range of design, engineering, and AI workflows and continually optimized through enterprise drivers. With extensive ISV certifications, robust IT management tools, and enterprise-grade support, RTX PRO workstations are the trusted choice for enterprise and mission-critical work.

Technical Specifications	
GPU architecture	NVIDIA Blackwell
NVIDIA® CUDA® Cores	24,064
Tensor Cores	5th Generation
Ray Tracing Cores	4th Generation
AI TOPS	3511 AI TOPS[1,2]
Single-precision performance	110 TFLOPS[1]
RT Core performance	333 TFLOPS[1]
GPU memory	96 GB GDDR7 with ECC
Memory interface	512-bit
Memory bandwidth	1792 GB/s
System interface	PCIe 5.0 x16
Display connectors	4x DisplayPort 2.1b
Max simultaneous displays	> 4x 4096 x 2160 @ 120 Hz > 4x 5120 x 2880 @ 60 Hz > 2x 7680 x 4320 @ 60 Hz
Video Engines	> 4x NVENC (9th Gen) > 4x NVDEC (6th Gen)
MIG Instance Types	> Up to 4x 24 GB > Up to 2x 48 GB > Up to 1x 96 GB
Power consumption	Total board power: 300 W
Power connector	1x PCIe CEM5 16-pin
Thermal solution	Active
Form factor	4.4” x 10.5” L, dual slot, full height
Graphics APIs	DirectX 12, Shader Model 6.6, OpenGL 4.63, Vulkan 1.3[3]
Compute APIs	CUDA 12.8, OpenCL 3.0, DirectCompute

Ready to Get Started?

To learn more, visit: nvidia.com/rtx-pro-6000-max-q/

1 Peak rates based on GPU Boost Clock

2 Effective FP4 TOPS with sparsity

3 Product is based on a published Krhonos specification and is expected to pass the Khronos conformation testing process when available. Current conformance status can be found at www.khronos.org/conformance

NVIDIA