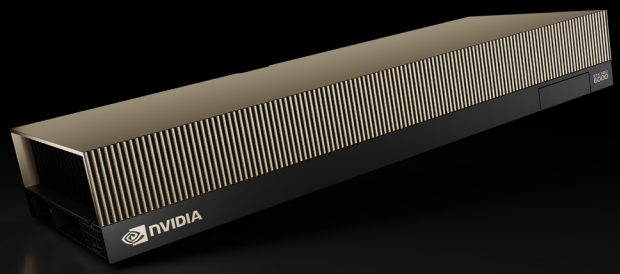
Datasheet

NVIDIA RTX PRO 6000 Blackwell Server Edition



Universal AI and visual computing performance for the data center.

The NVIDIA RTX PRO™ 6000 Blackwell Server Edition is the most powerful Blackwell data center platform for AI and visual computing, delivering breakthrough performance to accelerate a wide array of enterprise data center workloads. Equipped with 96 GB of ultra-fast GDDR7 memory, the NVIDIA RTX PRO 6000 Blackwell Server Edition GPU provides unparalleled performance and flexibility to accelerate a broad range of use cases—from multimodal AI inference, physical AI, and scientific computing to rendering, 3D graphics, video, and more.

Enterprises can configure NVIDIA RTX PRO 6000 Blackwell GPUs in high-density server platforms to deliver unmatched levels of compute power, memory capacity, and throughput to power mission-critical AI-enabled applications and accelerate use cases across industries—from healthcare, manufacturing, and geoscience to retail, media, and live broadcast. The RTX PRO 6000 delivers performance, scalability, and reliability for data center workloads, helping unlock innovation needed to drive breakthroughs and push the boundaries of what's possible.

Blackwell Architecture Innovations

Fifth-Generation Tensor Cores: Deliver up to 5x the performance of the previous generation for LLM inference with support for FP4 precision for faster AI model processing times and reduced memory usage.

Fourth-Generation 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: Accelerate multimodal AI inference, enhance video production, and power streaming workflows with real-time AI processing. The RTX PRO 6000 Blackwell features four (4) ninth-generation NVENC and four (4) sixth-generation NVDEC engines with support for 4:2:2 encoding and decoding.

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 1.6 TB/s bandwidth, enterprises can deliver low-latency inference with larger models and render massive 3D scenes.

Key Features

- Fifth-Generation Tensor Cores, FP4 precision
- Fourth-Generation Ray-Tracing Cores
- 96 GB of GDDR7 memory
- 1597 GB/s of memory bandwidth
- Ninth-Generation NVENC and Sixth-Generation NVDEC with 4:2:2 support
- PCIe Gen 5
- Multi-instance GPU (MIG) support
- Passive thermal design

Target Workloads

- NVIDIA AI Enterprise
- NVIDIA Omniverse™
- Multimodal AI inference
- Generative AI
- Physical AI
- Rendering
- 3D graphics
- Video

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.

Universal MIG: Divide a single RTX PRO 6000 Blackwell into up to four fully isolated instances, each with dedicated resources, allowing for concurrent execution of graphics and AI workloads, optimized GPU utilization, and secure isolation of different applications or users.

Streamline AI Development and Deployment

Accelerate AI development and deployment for production-ready AI agents and generative AI solutions, including multimodal AI, recommenders, computer vision, retrieval-augmented generation (RAG), and more with NVIDIA AI Enterprise. NVIDIA AI Enterprise includes NVIDIA NIM™, a set of easy-to-use microservices to speed up enterprise generative AI deployment, and NVIDIA AI Blueprints, predefined, customizable AI workflows designed to assist developers in creating and deploying generative AI applications. With the RTX PRO 6000 and NVIDIA AI Enterprise, deployments have enterprise-grade security, manageability, stability, and support, resulting in performance-optimized AI solutions that deliver faster business value and actionable insights.[1]

Build Next-Generation Physical AI-Enabled Applications

NVIDIA Omniverse makes it possible to develop generative physical AI-powered applications for industrial and robotics use cases. With powerful AI compute and fourth-generation RTX technology, the RTX PRO 6000 Blackwell Server Edition accelerates physical AI and industrial digitalization workflows powered by NVIDIA Omniverse, including photoreal design and rendering of massive digital twins, physically accurate 3D synthetic data generation, and robotics simulation.

Enterprise-Ready

The RTX PRO 6000 is optimized for 24/7 enterprise data center operations and is designed, built, tested, and supported exclusively by NVIDIA to ensure maximum uptime. The RTX PRO 6000 will be available from leading cloud and data center computing partners, including NVIDIA-Certified servers.

NVIDIA-Certified servers are the essential platform for building high-performance, scalable, and secure data center infrastructure for AI and accelerated computing workloads. NVIDIA-Certified servers are tested and validated to deliver optimal performance for a wide range of workloads, including NVIDIA AI Enterprise, NVIDIA Omniverse, graphics, visualization, and high-performance computing (HPC).

Technical Specifications

	RTX PRO 6000 Blackwell Server Edition
PNY Part Number	TCSRTX6000PROSE-PB
EAN Code	3536403403638
CUDA® parallel processing cores	24,064
NVIDIA Tensor Cores	752 (fifth-generation)
NVIDIA RT Cores	188 (fourth-generation)
Single-precision performance (FP32)	120 TFLOPS
Peak FP4 AI PFLOPS	4 PFLOPS
RT Core performance	355 TFLOPS
GPU memory	96 GB GDDR7 with ECC
Memory interface	512-bit
Memory bandwidth	1597 GB/s
Power consumption	Up to 600W (Configurable)
Multi-instance GPU	Up to 4 MIGs @ 24GB
NVENC | NVDEC | JPEG	4x | 4x | 4x
Confidential compute	Supported
Secure boot with root of trust	Yes
Graphics bus	PCI Express 5.0 x16
Display connectors	4x DisplayPort 2.1
Form factor	4.4" (H) x 10.5" (L), dual slot
Thermal solution	Passive
Power connector	1x PCIe CEM5 16-pin