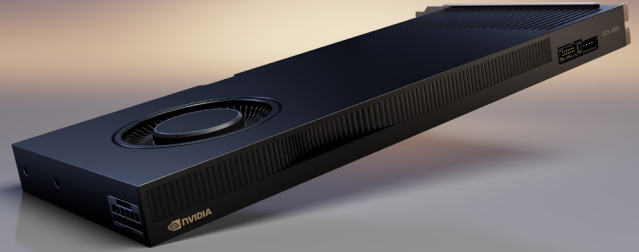Datasheet

NVIDIA

NVIDIA RTX 4000 Ada Generation

Performance for endless possibilities.
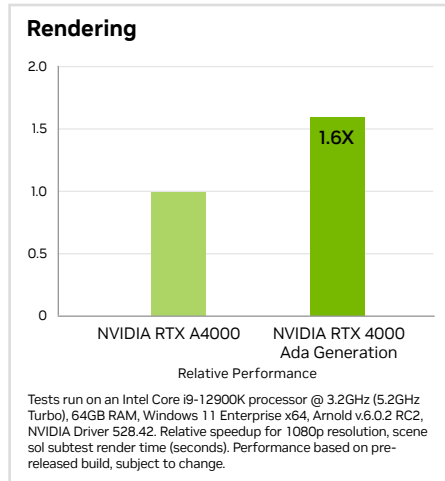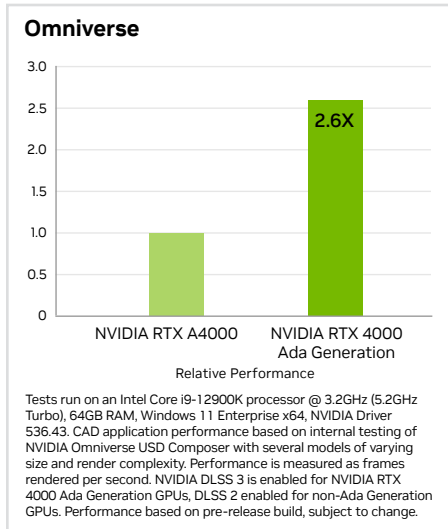
Powering the Next Era of Innovation

Industries are embracing accelerated computing and AI to tackle powerful dynamics and unlock transformative possibilities. Generative AI is reshaping the way professionals create and innovate across various domains, from design and engineering to entertainment and healthcare.

The NVIDIA RTX™ 4000 Ada Generation is the most powerful single-slot GPU for professionals, delivering remarkable acceleration for AI, real-time rendering, graphics, and compute workloads to your desktop. Built on the NVIDIA Ada Lovelace architecture, RTX 4000 combines 48 third-generation RT Cores, 192 fourth-generation Tensor Cores, and 6,144 CUDA® cores with 20GB of graphics memory to effortlessly handle large datasets and complex visual workloads.
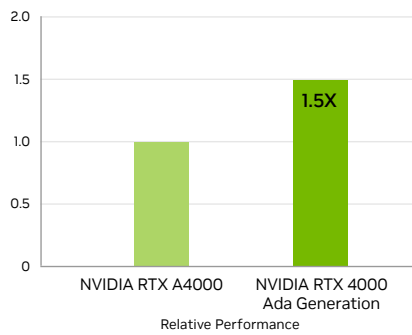
NVIDIA RTX professional graphics cards are certified with a broad range of professional applications, tested by leading independent software vendors (ISVs) and workstation manufacturers, and backed by a global team of support specialists. Get the peace of mind to focus on what matters with the premier visual computing solution for mission-critical business.

Key Features

> PCIe Gen4

> Four DisplayPort 1.4a connectors

> AV1 encode and decode support

> DisplayPort with audio

> 3D stereo support with stereo connector

> NVIDIA® GPUDirect® for Video support

> NVIDIA GPUDirect remote direct memory access (RDMA) support

> NVIDIA Quadro® Sync II[1] compatibility

> NVIDIA RTX Experience

> NVIDIA RTX Desktop Manager software

> NVIDIA RTX IO support

> HDCP 2.2 support

> NVIDIA Mosaic[2] technology

Omniverse

Relative Performance

Tests run on an Intel Core i9-12900K processor @ 3.2GHz (5.2GHz Turbo), 64GB RAM, Windows 11 Enterprise x64, NVIDIA Driver 536.43. CAD application performance based on internal testing of NVIDIA Omniverse USD Composer with several models of varying size and render complexity. Performance is measured as frames rendered per second. NVIDIA DLSS 3 is enabled for NVIDIA RTX 4000 Ada Generation GPUs, DLSS 2 enabled for non-Ada Generation GPUs. Performance based on pre-release build, subject to change.

Rendering

Relative Performance

Tests run on an Intel Core i9-12900K processor @ 3.2GHz (5.2GHz Turbo), 64GB RAM, Windows 11 Enterprise x64, Arnold v.6.0.2 RC2, NVIDIA Driver 528.42. Relative speedup for 1080p resolution, scene sol subtest render time (seconds). Performance based on pre-released build, subject to change.

Graphics

Relative Performance

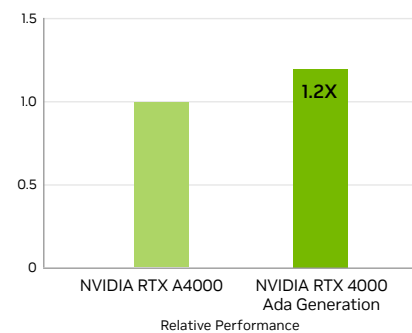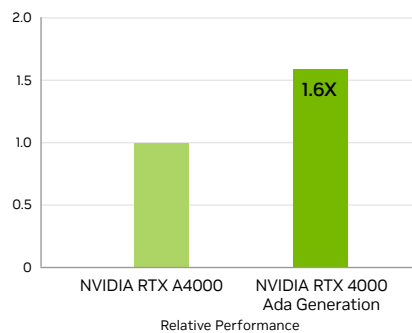| NVIDIA RTX A4000	NVIDIA RTX 4000 Ada Generation
1.0	1.5X

Tests run on an Intel Core i9-12900K processor @ 3.2GHz (5.2GHz Turbo), 64GB RAM, Windows 11 Enterprise x64, NVIDIA Driver 528.49. Relative speedup for 4K geomean score. Performance based on pre-release build, subject to change.

Generative AI

Relative Performance

NVIDIA RTX A4000	NVIDIA RTX 4000 Ada Generation
1.0	1.2X

Tests run on an Intel Core i9-12900K processor @ 3.2GHz (5.2GHz Turbo), 64GB RAM, Windows 11 Enterprise x64, Stable Diffusion WebUI v1.3.1, NVIDIA Driver 536.15. Relative speedup for 512x512 image generation. Performance based on pre-released build, subject to change.

Inference

Relative Performance

NVIDIA RTX A4000	NVIDIA RTX 4000 Ada Generation
1.0	1.6X

Tests run on an Intel Core i9-12900K processor @ 3.2GHz (5.2GHz Turbo), 64GB RAM, Windows 11 Enterprise x64, TensorRT v8.5.1, NVIDIA Driver 525.85. Relative speedup for ResNet-50 V1.5 Inference, mixed precision. Performance based on pre-release build, subject to change.

Specifications

GPU Memory	20GB GDDR6
Memory Interface	160 bit
Memory Bandwidth	360GB/s
Error Correcting Code (ECC)	Yes
NVIDIA Ada Lovelace Architecture-Based CUDA Cores	6,144
NVIDIA Fourth-Generation Tensor Cores	192
NVIDIA Third-Generation RT Cores	48
Single-Precision Performance	26.7 TFLOPS[3]
RT Core Performance	61.8 TFLOPS[3]
Tensor Performance	327.6 TFLOPS[4]
System Interface	PCIe 4.0 x16
Power Consumption	Total board power: 130W
Thermal Solution	Active
Form Factor	4.4" H x 9.5" L, single slot
Display Connectors	4x DisplayPort 1.4a[5]
Max Simultaneous Displays	4x 4096 x 2160 @ 120Hz 4x 5120 x 2880 @ 60Hz 2x 7680 x 4320 @ 60Hz
Encode/Decode Engines	2x encode, 2x decode (+AV1 encode and decode)
VR Ready	Yes
Graphics APIs	DirectX 12, Shader Model 6.7, OpenGL 4.6[6], Vulkan 1.3[6]
Compute APIs	CUDA 12.2, OpenCL 3.0, DirectCompute
NVIDIA NVLink ®	No

Ready to Get Started?

To learn more about NVIDIA RTX 4000, visit:
www.nvidia.com/rtx-4000