

NVIDIA, Global Workstation Manufacturers to Launch Powerful Systems for Generative AI and LLM Development, Content Creation, Data Science

Desktops Feature NVIDIA RTX 6000 Ada GPUs, NVIDIA Omniverse and NVIDIA AI Enterprise Software

SIGGRAPH—NVIDIA and global manufacturers today announced powerful new NVIDIA RTX™ workstations designed for development and content creation in the age of generative AI and digitalization.

The systems, including those from BOXX, Dell Technologies, HP and Lenovo, are based on [NVIDIA RTX 6000 Ada Generation GPUs](#) and incorporate [NVIDIA AI Enterprise](#) and [NVIDIA Omniverse™ Enterprise](#) software.

Separately, NVIDIA also released three new desktop workstation Ada Generation GPUs — the [NVIDIA RTX 5000](#), [RTX 4500](#) and [RTX 4000](#) — to deliver the latest AI, graphics and real-time rendering technology to professionals worldwide.

“Few workloads are as challenging as generative AI and digitalization applications, which require a full-stack approach to computing,” said Bob Pette, vice president of professional visualization at NVIDIA. “Professionals can now tackle these on a desktop with the latest NVIDIA-powered RTX workstations, enabling them to build vast, digitalized worlds in the new age of generative AI.”

The new RTX workstations offer up to four NVIDIA RTX 6000 Ada GPUs, each equipped with 48GB of memory, and a single desktop workstation can provide up to 5,828 TFLOPS of AI performance and 192GB of GPU memory. Depending on user needs, systems can be configured with NVIDIA AI Enterprise or Omniverse Enterprise to power a breadth of demanding generative AI and graphics-intensive workloads.

[NVIDIA AI Enterprise 4.0](#), announced separately today, now includes [NVIDIA NeMo™](#), an end-to-end framework for building and customizing foundation models for generative AI, [NVIDIA RAPIDS™](#) libraries for data science, as well as frameworks, pretrained models and tools for building common enterprise AI use cases, including recommenders, virtual assistants and cybersecurity solutions.

Omniverse Enterprise is a platform for industrial digitalization that enables teams to develop interoperable 3D workflows and [OpenUSD](#) applications. As an OpenUSD-native platform, Omniverse enables globally distributed teams to collaborate on full-design-fidelity datasets from hundreds of 3D applications.

“Yurts provides a full-stack generative AI solution aligning with multiple form factors, deployment models and budgets of our customers. We’ve achieved this by leveraging LLMs for various natural language processing tasks and incorporating the RTX 6000 Ada. From private data centers to workstation-sized solutions that fit under a desk, Yurts remains committed to scaling our platform and offering alongside NVIDIA,” said Jason Schnitzer, chief technology officer at Yurts.

Workstation users can also take advantage of the new [NVIDIA AI Workbench](#), available soon in early access, which provides developers with a unified, easy-to-use toolkit for creating, fine-tuning and running generative AI models with just a few clicks. Users of any skill level can quickly create, test and customize pretrained generative AI models on a PC or workstation and then scale them to virtually any data center, public cloud or NVIDIA DGX Cloud.

Next-Generation RTX Technology

The new NVIDIA RTX 5000, RTX 4500 and RTX 4000 desktop GPUs feature the latest NVIDIA Ada Lovelace architecture technologies, including:

- NVIDIA CUDA® cores: Up to 2x the single-precision floating point throughput compared to the previous generation.
- Third-generation RT Cores: Up to 2x the throughput of the previous generation with the ability to concurrently run ray tracing with either shading or denoising capabilities.
- Fourth-generation Tensor Cores: Up to 2x faster AI training performance than the previous generation with expanded support for the FP8 data format.
- DLSS 3: New levels of realism and interactivity for real-time graphics with the power of AI.
- Larger GPU memory: The RTX 4000 provides 20GB of GDDR6 memory; the RTX 4500 offers 24GB of GDDR6 memory; and the RTX 5000 boasts 32GB of GDDR6 memory — all supporting error-code correction for error-free computing with large 3D models, rendered images, simulations and AI datasets.
- Extended-reality capabilities: Support for high-resolution augmented-reality and virtual-reality devices to deliver the high-performance graphics required for creating stunning AR, VR and mixed-reality content.

“The NVIDIA RTX 5000 Ada GPU demonstrates NVIDIA’s impressive generational performance improvements — it has significantly increased our efficiency in creating stereo panoramas using Enscape,” said Dan Stine, director of design technology at architectural firm Lake|Flato. “With the performance boost and large frame buffer of RTX 5000 GPUs, our large, complex models look great in virtual reality, which gives our clients a more comfortable and contextual experience.”

Availability

[RTX workstations](#) featuring up to four RTX 6000 Ada GPUs, NVIDIA AI Enterprise and NVIDIA Omniverse Enterprise will be available from system builders starting in the fall.

The new NVIDIA RTX 5000 GPU is now available and shipping from HP and through global distribution partners such as Leadtek, PNY and Ryoyo Electro starting today. The NVIDIA RTX 4500 and RTX 4000 GPUs will be available in the fall from BOXX, Dell Technologies, HP and Lenovo and through global distribution partners.

About NVIDIA

Since its founding in 1993, [NVIDIA](#) (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company’s invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing company with data-center-scale offerings that are reshaping industry. More information at <https://nvidianews.nvidia.com/>.

Certain statements in this press release, including, but not limited to, statements as to: the benefits, impact, performance, features and availability of our products, services and technologies, including NVIDIA RTX 6000 Ada Generation GPUs, NVIDIA AI Enterprise, NVIDIA Omniverse Enterprise, NVIDIA RTX 5000, RTX 4500 and RTX 4000 desktop GPUs, NVIDIA AI Enterprise 4.0, NVIDIA NeMo, [NVIDIA RAPIDS](#) and [NVIDIA AI Workbench](#); and generative AI and digitalization applications requiring a full-stack approach to computing are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners’ products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company’s website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, NeMo, NVIDIA Omniverse, NVIDIA RTX and RAPIDS are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability, and specifications are subject to change without notice.

Kasia Johnston
+1-415-813-8859
kasiaj@nvidia.com
Cliff Edwards
NVIDIA Corporation
+1-415-699-2755
cliffe@nvidia.com