

# NVIDIA DGX Station A100 Offers Researchers AI Data-Center-in-a-Box

## World's Only Petascale Integrated AI Workgroup Server, Second-Gen DGX Station Packs Four NVIDIA A100 GPUs, Debuts with up to 320GB of GPU Memory to Bring AI into Offices and Labs

**SC20**—NVIDIA today announced the [NVIDIA DGX Station™ A100](#)— the world's only petascale workgroup server. The second generation of the groundbreaking AI system, DGX Station A100 accelerates demanding machine learning and data science workloads for teams working in corporate offices, research facilities, labs or home offices everywhere.

Delivering 2.5 petaflops of AI performance, DGX Station A100 is the only workgroup server with four of the latest [NVIDIA® A100 Tensor Core GPUs](#) fully interconnected with [NVIDIA NVLink®](#), providing up to 320GB of GPU memory to speed breakthroughs in enterprise data science and AI.

DGX Station A100 is also the only workgroup server that supports NVIDIA's [multi-instance GPU \(MIG\)](#) technology. With MIG, a single DGX Station A100 provides up to 28 separate GPU instances to run parallel jobs and support multiple users without impacting system performance.

"DGX Station A100 brings AI out of the data center with a server-class system that can plug in anywhere," said Charlie Boyle, vice president and general manager of DGX systems at NVIDIA. "Teams of data science and AI researchers can accelerate their work using the same software stack as NVIDIA DGX A100 systems, enabling them to easily scale from development to deployment."

### DGX Station Powers AI Innovation

Organizations around the world have adopted DGX Station to power AI and data science across industries such as education, financial services, government, healthcare and retail. These AI leaders include:

- **BMW Group Production** is using NVIDIA DGX Station to explore insights faster as they develop and deploy AI models that improve operations.
- **DFKI**, the German Research Center for Artificial Intelligence, is using DGX Station to build models that tackle critical challenges for society and industry, including computer vision systems that help emergency services respond rapidly to natural disasters.
- **Lockheed Martin** is using DGX Station to develop AI models that use sensor data and service logs to predict the need for maintenance to improve manufacturing uptime, increase safety for workers, and reduce operational costs.
- **Pacific Northwest National Laboratory** is using NVIDIA DGX Station to conduct federally funded research in support of national security. Focused on technological innovation in energy resiliency and national security, PNNL is a leading U.S. HPC center for scientific discovery, energy resilience, chemistry, Earth science and data analytics.
- **NTT Docomo**, Japan's leading mobile operator with over 79 million subscribers, uses DGX Station to develop innovative AI-driven services such as its image recognition solution.

### An AI Supercomputer Anywhere

While DGX Station A100 does not require data-center-grade power or cooling, it is a server-class system that features the same remote management capabilities as [NVIDIA DGX A100](#) data center systems. System administrators can easily perform any management tasks over a remote connection when data scientists and researchers are working at home or in labs.

DGX Station A100 is available with four 80GB or 40GB [NVIDIA A100 Tensor Core GPUs](#), providing options for data science and AI research teams to select a system according to their unique workloads and budgets.

To power complex conversational AI models like BERT Large inference, DGX Station A100 is more than 4x faster than the previous generation DGX Station. It delivers nearly a 3x performance boost for BERT Large AI training.

### Doubling Down on GPU Memory for AI, HPC

For advanced data center workloads, DGX A100 systems will be available with the new NVIDIA A100 80GB GPUs, doubling GPU memory capacity to 640GB per system to enable AI teams to boost accuracy with larger datasets and models.

The new NVIDIA DGX A100 640GB systems can also be integrated into the [NVIDIA DGX SuperPOD™ Solution for Enterprise](#), allowing organizations to build, train and deploy massive AI models on turnkey AI supercomputers available in units of 20 DGX A100 systems.

The first installments of NVIDIA DGX SuperPOD systems with DGX A100 640GB will include the [Cambridge-1](#) supercomputer being installed in the U.K. to supercharge healthcare research, as well as the new [University of Florida HiPerGator](#) AI supercomputer that will power AI-infused discovery across the Sunshine State.

## Availability

NVIDIA DGX Station A100 and NVIDIA DGX A100 640GB systems will be available this quarter through [NVIDIA Partner Network](#) resellers worldwide. An upgrade option is available for NVIDIA DGX A100 320GB customers.

Learn more about DGX Station A100 in the live [NVIDIA SC20 Special Address](#) at 3 p.m. PT today, and register for the [Learn to Build Bigger AI, Faster](#) with DGX Station webinar, on Wednesday, Dec. 2.

Watch the [introductory video of the NVIDIA DGX Station A100](#).

## About NVIDIA

NVIDIA's (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI — the next era of computing — with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at <https://nvidianews.nvidia.com/>.

Certain statements in this press release including, but not limited to, statements as to: the benefits, performance, features and abilities of the NVIDIA DGX Station A100; DGX Station A100 bringing AI out of the data center and under desks everywhere, its installation rate, it accelerating the work of millions of data scientists and AI researchers, and its scalability; the companies using NVIDIA DGX Stations and how they are using them; the size and growth of AI models and DGX Station A100 systems helping to address this; DGX A100 systems being integrated into the NVIDIA DGX SuperPOD Solution for Enterprise and what it enables; the first installments of NVIDIA DGX SuperPOD systems; and the price and availability of NVIDIA DGX Station A100 systems 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.

© 2020 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, NVIDIA DGX Station, NVIDIA DGX SuperPOD and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. Features, pricing, availability and specifications are subject to change without notice.

Shannon McPhee  
+1-310-920-9642  
[smcphee@nvidia.com](mailto:smcphee@nvidia.com)