



NVIDIA DGX STATION A100

WORKGROUP APPLIANCE FOR THE AGE OF AI



Data science teams are at the leading edge of AI innovation, developing projects that can transform enterprises and our world. But they're often left searching for spare compute cycles that can help train the most complex models. These teams need a dedicated AI platform that can plug in anywhere and is fully optimized across hardware and software to deliver groundbreaking performance for multiple, simultaneous users anywhere in the world.

NVIDIA DGX Station™ A100 at a Glance

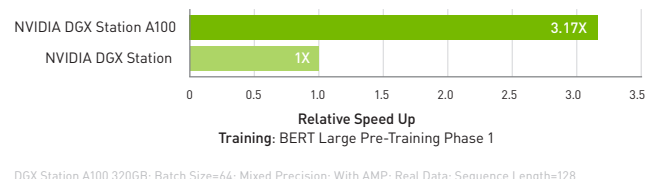
- > AI workgroup server delivering 2.5 petaFLOPS of performance that your team can use without limits—for training, inference, and data analytics
- > Server-grade, plug-and-go, and doesn't require data center power and cooling
- > World-class AI platform, with no complicated installation or IT help needed
- > The world's only workstation-style system with four fully interconnected NVIDIA A100 Tensor Core GPUs and up to 320 gigabytes (GB) of GPU memory
- > Delivers a fast-track to AI transformation with NVIDIA know-how and experience

NVIDIA DGX Station A100 brings AI supercomputing to data science teams, offering data center technology without a data center or additional IT infrastructure. Designed for multiple, simultaneous users, DGX Station A100 leverages server-grade components in an office-friendly form factor. It's the only system with four fully interconnected and Multi-Instance GPU (MIG)-capable NVIDIA A100 Tensor Core GPUs with up to 320 GB of total GPU memory that can plug into a standard power outlet, resulting in a powerful AI appliance that you can place anywhere.

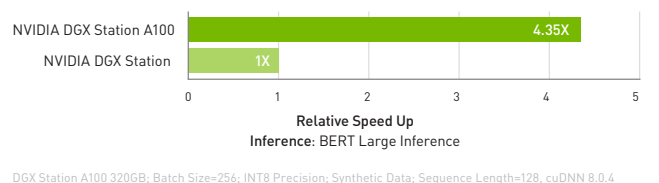
AI Supercomputing for Data Science Teams

Effortlessly providing multiple, simultaneous users with a centralized AI resource, DGX Station A100 is the workgroup appliance for the age of AI. It's capable of running training, inference, and analytics workloads in parallel, and with MIG, it can provide up to 28 separate GPU devices to individual

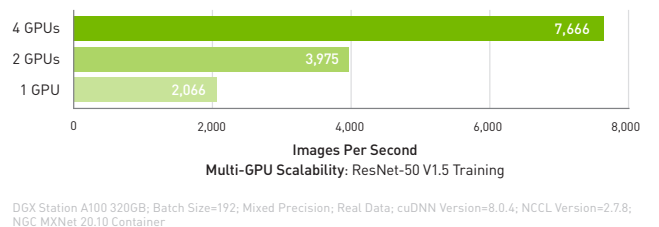
DGX Station A100 Delivers Over 3X Faster The Training Performance



DGX Station A100 Delivers Over 4X Faster The Inference Performance



DGX Station A100 Delivers Linear Scalability



users and jobs so that activity is contained and doesn't impact performance across the system. DGX Station A100 features the same fully optimized NVIDIA DGX™ software stack as all DGX systems, delivering maximum performance and complete interoperability with DGX-based infrastructure, from individual systems to NVIDIA DGX POD™ and NVIDIA DGX SuperPOD™, making DGX Station A100 an ideal platform for teams from all organizations, large and small.

Data Center Performance Without the Data Center

NVIDIA DGX Station A100 provides a data center-class AI server in a workstation form factor, suitable for use in a standard office environment without specialized power and cooling. Its design includes four ultra-powerful NVIDIA A100 Tensor Core GPUs, a top-of-the-line, server-grade CPU, super-fast NVMe storage, and leading-edge PCIe Gen4 buses. DGX Station A100 also includes the same Baseboard Management Controller (BMC) as NVIDIA DGX A100, allowing system administrators to perform any required tasks over a remote connection. DGX Station A100 is the most powerful AI system for an office environment, providing data center technology without the data center.

An AI Appliance You Can Place Anywhere

NVIDIA DGX Station A100 is designed for today's agile data science teams working in corporate offices, labs, research facilities, or even from home. Whereas installing large-scale AI infrastructure requires significant IT investment and large data centers with industrial-strength power and cooling, DGX Station A100 simply plugs into any standard wall outlet, wherever your team's workspace may be. And its innovative, refrigeration-based design means that it stays cool to the touch. With a simple, one person set up, you can be up and running in minutes with a world-class AI platform that needs just two cables to operate.

Bigger Models, Faster Answers

NVIDIA DGX Station A100 isn't a workstation. It's an AI workgroup server that can sit under your desk. In addition to its 64-core, data center-grade CPU, it features the same NVIDIA A100 Tensor Core GPUs as the NVIDIA DGX A100 server, with either 40 or 80 GB of GPU memory each, connected via high-speed SXM4. NVIDIA DGX Station A100 is the only office-friendly system that has four fully interconnected GPUs, leveraging NVIDIA® NVLink®, and that supports MIG, delivering up to 28 separate GPU devices for parallel jobs and multiple users—without impacting system performance.

Integrated Access to Unmatched AI Expertise

NVIDIA DGX Station A100 is a complete hardware and software platform backed by thousands of AI experts at NVIDIA and built upon the knowledge gained from the world's largest DGX proving ground, NVIDIA DGX SATURNV. Owning a DGX Station A100 gives you direct access to NVIDIA DGXperts, a global team of AI-fluent practitioners who offer prescriptive guidance and design expertise to help fast-track AI transformation with know-how and experience from NVIDIA's decade plus of AI leadership. This ensures mission-critical applications get up and running quickly and stay running smoothly, dramatically improving time to insights.

To learn more about NVIDIA DGX Station A100, visit www.nvidia.com/DGXStationA100

SYSTEM SPECIFICATIONS

	NVIDIA DGX Station A100 320GB	NVIDIA DGX Station A100 160GB
GPUs	4x NVIDIA A100 80 GB GPUs	4x NVIDIA A100 40 GB GPUs
GPU Memory	320 GB total	160 GB total
Performance	2.5 petaFLOPS AI 5 petaOPS INT8	
System Power Usage	1.5 kW at 100–120 Vac	
CPU	Single AMD 7742, 64 cores, 2.25 GHz (base)–3.4 GHz (max boost)	
System Memory	512 GB DDR4	
Networking	Dual-port 10Gbase-T Ethernet LAN Single-port 1Gbase-T Ethernet BMC management port	
Storage	OS: 1x 1.92 TB NVME drive Internal storage: 7.68 TB U.2 NVME drive	
DGX Display Adapter	4 GB GPU memory, 4x Mini DisplayPort	
System Acoustics	<37 dB	
Software	Ubuntu Linux OS	
System Weight	91.0 lbs (43.1 kgs)	
Packaged System Weight	127.7 lbs (57.93 kgs)	
System Dimensions	Height: 25.1 in (639 mm) Width: 10.1 in (256 mm) Length: 20.4 in (518 mm)	
Operating Temperature Range	5–35 °C (41–95 °F)	