NVIDIA Announces New DGX SuperPOD, the First Cloud-Native, Multi-Tenant Supercomputer, Opening World of AI to Enterprise

Customers Worldwide Advance Conversational AI, Drug Discovery, Autonomous Vehicles and More with DGX SuperPOD

GTC — NVIDIA today unveiled the world's first cloud-native, multi-tenant AI supercomputer — the next-generation NVIDIA DGX SuperPOD™ featuring NVIDIA BlueField®-2 DPUs.

Fortifying the DGX SuperPOD with BlueField-2 DPUs — data processing units that offload, accelerate and isolate users' data — provides customers with secure connections to their AI infrastructure.

The company also announced NVIDIA Base Command™, which enables multiple users and IT teams to securely access, share and operate their DGX SuperPOD infrastructure. Base Command coordinates AI training and operations on DGX SuperPOD infrastructure to enable the work of teams of data scientists and developers located around the globe.

"AI is the most powerful technology the world has ever known, and NVIDIA DGX systems are the most effective tool for harnessing it," said Charlie Boyle, vice president and general manager of DGX systems at NVIDIA. "The new DGX SuperPOD, which combines multiple DGX systems, provides a turnkey AI data center that can be securely shared across entire teams of researchers and developers."

Advancing AI with DGX SuperPOD
DGX SuperPODs are AI supercomputers featuring 20 or more NVIDIA DGX A100™ systems and NVIDIA InfiniBand HDR networking. Among the latest to deploy DGX SuperPODs to power new AI solutions and services are:

- **Sony Group Corporation** is using DGX SuperPOD to enable its corporate research and development team to infuse AI across the company.
- **NAVER**, a leading internet technology company in Korea and Japan, is training giant AI language models at scale on DGX SuperPOD to pioneer new services across e-commerce, search, entertainment and payment applications.
- **Recursion**, a digital-biology company working to industrialize drug discovery, is using DGX SuperPOD to accelerate its deep learning models and empower its growing workforce of machine learning experts.
- **MTS**, Russia's largest telecommunications company, is using its DGX SuperPOD as the foundation for its AI-ready development infrastructure and new public cloud service offerings, as well as to integrate AI and data science across its departments.
- **VinAI** is using its DGX SuperPOD, which will rank as Vietnam's fastest AI supercomputer, to accelerate AI initiatives, including autonomous vehicles, healthcare and consumer services, in Southeast Asia and around the world.

Additionally, NVIDIA and Schrödinger today separately announced a strategic partnership designed to harness DGX SuperPODs to further accelerate drug discovery at supercomputing scale. Their jointly developed solution can enable pharmaceutical and biotech companies of all sizes to simulate molecular combinations with physics and AI to identify and optimize the most promising compounds for potential therapeutic use.

Secure, Cloud-Native AI at Scale
NVIDIA designed the new DGX SuperPOD to meet customers' growing security and scalability requirements as AI broadens in adoption. Increasingly, enterprise IT departments need to support the work of multiple teams at different locations, while academic and research institutions often grant outside organizations access to their computing resources.

BlueField-2 DPUs offload, accelerate and isolate users and their data, allowing organizations to safely provide private cloud access to their DGX SuperPOD infrastructure with security that spans user traffic, firewalls and multi-tenant access to storage. DDN is the first NVIDIA storage partner certified to support the next-gen DGX SuperPOD with BlueField-2.

NVIDIA Base Command allows teams of AI developers and data scientists to seamlessly provision and schedule workloads on DGX infrastructure from prototyping to production.

To support the iterative process of building and refining AI models, Base Command provides built-in telemetry for users to validate deep learning techniques, workload settings and resource allocations to constantly improve results.

New DGX Station A100 Subscription Makes AI More Accessible
NVIDIA also introduced a subscription offering available for the NVIDIA DGX Station A100, the world's only workgroup appliance in an office form factor that supports NVIDIA Multi-Instance GPU technology.

The new subscription program makes it easier for companies at every stage of growth to accelerate AI development outside the data center for teams working in corporate offices, research facilities, labs and home offices.

With four [NVIDIA A100](#) 80GB GPUs, DGX Station A100 delivers a 3x speed-up compared to its predecessor for complex AI workloads such as natural language processing. As many as 28 data scientists can share a DGX Station A100 simultaneously, enabling enterprises to accelerate AI development while saving up to $700,000 compared to data center servers.

Availability
Cloud-native, multi-tenant NVIDIA DGX SuperPODs will be available in Q2 through NVIDIA's global partners, which can provide pricing to qualified customers upon request.

NVIDIA Base Command will also be available starting in Q2.

Subscriptions for NVIDIA DGX Station A100 are available starting at a list price of $9,000 per month.

[Register for free](#) to learn more about DGX systems during GTC21, taking place online April 12-16. Tune in to watch NVIDIA founder and CEO Jensen Huang's [GTC21 keynote](#) address streaming live on April 12 starting at 8:30 a.m. PT.

About NVIDIA
[NVIDIA](#)'s (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market and has redefined modern computer graphics, high performance computing and artificial intelligence. The company's pioneering work in accelerated computing and AI is reshaping trillion-dollar industries, such as transportation, healthcare and manufacturing, and fueling the growth of many others. More information at https://nvidianews.nvidia.com/.

Shannon McPhee
+1-310-920-9642
smcphee@nvidia.com