

NVIDIA Blackwell Ultra DGX SuperPOD Delivers Out-of-the-Box AI Supercomputer for Enterprises to Build AI Factories

- ***NVIDIA Blackwell Ultra-Powered DGX Systems Supercharge AI Reasoning for Real-Time AI Agent Responses***
- ***Equinix First to Offer NVIDIA Instant AI Factory Service, With Preconfigured Space in Blackwell-Ready Facilities for DGX GB300 and DGX B300 Systems to Meet Global Demand for AI Infrastructure***

GTC—NVIDIA today announced the world’s most advanced enterprise AI infrastructure — [NVIDIA DGX SuperPOD™](#) built with [NVIDIA Blackwell Ultra GPUs](#) — which provides enterprises across industries with AI factory supercomputing for state-of-the-art agentic AI reasoning.

Enterprises can use new [NVIDIA DGX™ GB300](#) and [NVIDIA DGX B300](#) systems, integrated with NVIDIA networking, to deliver out-of-the-box DGX SuperPOD AI supercomputers that offer FP4 precision and faster AI reasoning to supercharge [token](#) generation for AI applications.

AI factories provide purpose-built infrastructure for agentic, generative and physical AI workloads, which can require significant computing resources for AI [pretraining, post-training and test-time scaling](#) for applications running in production.

“AI is advancing at light speed, and companies are racing to build AI factories that can scale to meet the processing demands of reasoning AI and inference time scaling,” said Jensen Huang, founder and CEO of NVIDIA. “The NVIDIA Blackwell Ultra DGX SuperPOD provides out-of-the-box AI supercomputing for the age of agentic and physical AI.”

DGX GB300 systems feature NVIDIA Grace Blackwell Ultra Superchips — which include 36 [NVIDIA Grace™ CPUs](#) and 72 [NVIDIA Blackwell Ultra GPUs](#) — and a rack-scale, liquid-cooled architecture designed for real-time agent responses on advanced reasoning models.

Air-cooled NVIDIA DGX B300 systems harness the NVIDIA B300 NVL16 architecture to help data centers everywhere meet the computational demands of generative and agentic AI applications.

To meet growing demand for advanced accelerated infrastructure, NVIDIA also unveiled NVIDIA Instant AI Factory, a managed service featuring the Blackwell Ultra-powered NVIDIA DGX SuperPOD. [Equinix](#) will be first to offer the new DGX GB300 and DGX B300 systems in its preconfigured liquid- or air-cooled AI-ready data centers located in 45 markets around the world.

NVIDIA DGX SuperPOD With DGX GB300 Powers Age of AI Reasoning

DGX SuperPOD with DGX GB300 systems can scale up to tens of thousands of NVIDIA Grace Blackwell Ultra Superchips — connected via [NVIDIA NVLink™](#), [NVIDIA Quantum-X800 InfiniBand](#) and [NVIDIA Spectrum-X™ Ethernet](#) networking — to supercharge training and inference for the most compute-intensive workloads.

DGX GB300 systems deliver up to 70x more AI performance than AI factories built with [NVIDIA Hopper™](#) systems and 38TB of fast memory to offer unmatched performance at scale for multistep reasoning on agentic AI and reasoning applications.

The 72 Grace Blackwell Ultra GPUs in each DGX GB300 system are connected by fifth-generation NVLink technology to become one massive, shared memory space through the NVLink Switch system.

Each DGX GB300 system features 72 [NVIDIA ConnectX®-8](#) SuperNICs, delivering accelerated networking speeds of up to 800Gb/s — double the performance of the previous generation. Eighteen [NVIDIA BlueField®-3 DPUs](#) pair with NVIDIA Quantum-X800 InfiniBand or NVIDIA Spectrum-X Ethernet to accelerate performance, efficiency and security in massive-scale AI data centers.

DGX B300 Systems Accelerate AI for Every Data Center

The NVIDIA DGX B300 system is an AI infrastructure platform designed to bring energy-efficient generative AI and AI reasoning to every data center.

Accelerated by NVIDIA Blackwell Ultra GPUs, DGX B300 systems deliver 11x faster AI performance for inference and a 4x speedup for training compared with the Hopper generation.

Each system provides 2.3TB of HBM3e memory and includes advanced networking with eight NVIDIA ConnectX-8

SuperNICs and two BlueField-3 DPUs.

NVIDIA Software Accelerates AI Development and Deployment

To enable enterprises to automate the management and operations of their infrastructure, NVIDIA also announced [NVIDIA Mission Control™](#) — AI data center operation and orchestration software for Blackwell-based DGX systems.

NVIDIA DGX systems support the [NVIDIA AI Enterprise](#) software platform for building and deploying enterprise-grade AI agents. This includes [NVIDIA NIM™](#) microservices, such as the new [NVIDIA Llama Nemotron](#) open reasoning model family announced today, and [NVIDIA AI Blueprints](#), frameworks, libraries and tools used to orchestrate and optimize performance of AI agents.

NVIDIA Instant AI Factory to Meet Infrastructure Demand

NVIDIA Instant AI Factory offers enterprises an Equinix managed service featuring the Blackwell Ultra-powered NVIDIA DGX SuperPOD with NVIDIA Mission Control software.

With dedicated Equinix facilities around the globe, the service will provide businesses with fully provisioned, intelligence-generating AI factories optimized for state-of-the-art model training and real-time reasoning workloads — eliminating months of pre-deployment infrastructure planning.

Availability

NVIDIA DGX SuperPOD with DGX GB300 or DGX B300 systems are expected to be available from partners later this year.

NVIDIA Instant AI Factory is planned to be available starting later this year.

Learn more by watching the [NVIDIA GTC keynote](#) and [register to attend sessions](#) from NVIDIA and industry leaders at the show, which runs through March 21.

About NVIDIA

[NVIDIA](#) (NASDAQ: NVDA) is the world leader in accelerated computing.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, availability, and performance of NVIDIA's products, services, and technologies; third parties adopting NVIDIA's products and technologies and the benefits and impact thereof; and AI advancing at light speed, and companies racing to build AI factories that can scale to meet the processing demands of reasoning AI and inference time scaling 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.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

© 2025 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, BlueField, ConnectX, DGX, NVIDIA DGX SuperPOD, NVIDIA Grace, NVIDIA Hopper, NVIDIA Mission Control, NVIDIA NIM, NVIDIA Spectrum-X and NVLink 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.

A photo accompanying this announcement is available at <https://www.globenewswire.com/NewsRoom/AttachmentNg/4f5747e8-5b3d-4764-9d34-6d63cbfb18c2>

Allie Courtney
NVIDIA Corporation
+1-408-706-8995
acourtney@nvidia.com