



NVIDIA DGX Spark Arrives for World's AI Developers

News Summary:

- NVIDIA founder and CEO Jensen Huang delivers DGX Spark to Elon Musk at SpaceX.
- This week, NVIDIA and its partners are shipping DGX Spark, the world's smallest AI supercomputer, delivering NVIDIA's AI stack in a compact desktop form factor.
- Acer, ASUS, Dell Technologies, GIGABYTE, HPI, Lenovo and MSI debut DGX Spark systems, expanding access to powerful AI computing.
- Built on the NVIDIA Grace Blackwell architecture, DGX Spark integrates NVIDIA GPUs, CPUs, networking, CUDA libraries and NVIDIA AI software, accelerating agentic and physical AI development.

NVIDIA today announced it will start shipping NVIDIA DGX Spark™, the world's smallest AI supercomputer.

AI workloads are quickly outgrowing the memory and software capabilities of the PCs, workstations and laptops millions of developers rely on today — forcing teams to shift work to the cloud or local data centers.

As a new class of computer, DGX Spark delivers a petaflop of AI performance and 128GB of unified memory in a compact desktop form factor, giving developers the power to run inference on AI models with up to 200 billion parameters and fine-tune models of up to 70 billion parameters locally. In addition, DGX Spark lets developers create AI agents and run advanced software stacks locally.

"In 2016, we built DGX-1 to give AI researchers their own supercomputer. I hand-delivered the first system to Elon at a small startup called OpenAI — and from it came ChatGPT, kickstarting the AI revolution," said Jensen Huang, founder and CEO of NVIDIA. "DGX-1 launched the era of AI supercomputers and unlocked the scaling laws that drive modern AI. With DGX Spark, we return to that mission — placing an AI computer in the hands of every developer to ignite the next wave of breakthroughs."



Evolution of AI Computing: NVIDIA DGX-1 to DGX Spark

	DGX-1 (2016)	DGX Spark (2025)
GPU Architecture	NVIDIA Pascal	NVIDIA Blackwell
GPU Memory	128GB (16GB per GPU)	128GB unified system memory
AI Performance	170 TFLOPS (FP16)	1 PFLOP (FP4)
System Power	3,200 W	240 W
Dimensions	866 mm x 444 mm x 131 mm (34.1 in. x 17.48 in. x 5.16 in.)	150 mm x 150 mm x 50.5 mm (5.91 in. x 5.91 in. x 1.99 in.)
System Weight	60.8 kg (134 lb.)	1.2 kg (2.65 lb.)
Price	\$129,000	\$3,999

DGX Spark brings together the full NVIDIA AI platform — including GPUs, CPUs, networking, CUDA® libraries and the NVIDIA AI software stack — into a system small enough for a lab or an office, yet powerful enough to accelerate agentic and physical AI development. By combining breakthrough performance with the reach of the NVIDIA ecosystem, DGX Spark transforms the desktop into an AI development platform.

DGX Spark systems deliver up to 1 petaflop of AI performance, accelerated by a NVIDIA GB10 Grace Blackwell Superchip, NVIDIA ConnectX®-7 200 Gb/s networking and NVIDIA NVLink™-C2C technology, providing 5x the bandwidth of fifth-generation PCIe with 128GB of CPU-GPU coherent memory.

The NVIDIA AI software stack is preinstalled to enable developers to start working on AI projects out of the box. With DGX Spark, developers can access NVIDIA AI ecosystem tools including models, libraries, and NVIDIA NIM™ microservices, enabling local workflows such as customizing Black Forest Labs' FLUX.1 models to refine image generation, creating a vision search and summarization agent using the NVIDIA Cosmos™ Reason vision language model, or building an AI chatbot using Qwen3 that is optimized for DGX Spark.

To celebrate DGX Spark shipping worldwide, Huang hand-delivered one of the first units of DGX Spark to Elon Musk, chief engineer at SpaceX, today in Starbase, Texas. The exchange was a connection to the supercomputer's origins, as Musk was among the team that received the first NVIDIA DGX™-1 supercomputer from Huang in 2016.

Other early recipients of DGX Spark, including [Anaconda](#), Cadence, [ComfyUI](#), [Docker](#), Google, Hugging Face, JetBrains, [LM Studio](#), Meta, Microsoft, [Ollama](#) and [Roboflow](#), are testing, validating and optimizing their tools, software and models for DGX Spark.

Research organizations around the world, including the NYU Global AI Frontier Lab, previewed DGX Spark to boost their AI development.

“DGX Spark allows us to access peta-scale computing on our desktop,” said Kyunghyun Cho, professor of computer and data science at the NYU Global AI Frontier Lab. “This new way to conduct AI research and development enables us to rapidly prototype and experiment with advanced AI algorithms and models — even for privacy- and security-sensitive applications, such as healthcare.”

NVIDIA DGX Spark Now Shipping

Starting Wednesday, Oct. 15, DGX Spark can be ordered on [NVIDIA.com](#). Partner systems will be available from [Acer](#), [ASUS](#), [Dell Technologies](#), [GIGABYTE](#), [HP](#), [Lenovo](#), [MSI](#) as well as [Micro Center](#) stores in the U.S., and from NVIDIA channel partners worldwide.

About NVIDIA

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

Certain statements in this press release including, but not limited to, statements as to: with DGX Spark, NVIDIA placing an AI computer in the hands of every developer to ignite the next wave of breakthroughs; the benefits, impact, performance, and availability of NVIDIA's products, services, and technologies; expectations with respect to NVIDIA's third party arrangements, including with its collaborators and partners; expectations with respect to technology developments; and other statements that are not historical facts are forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, which are subject to the “safe harbor” created by those sections based on management's beliefs and assumptions and on information currently available to management and 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 and political conditions; NVIDIA's reliance on third parties to manufacture, assemble, package and test NVIDIA's products; the impact of technological development and competition; development of new products and technologies or enhancements to NVIDIA's existing product and technologies; market acceptance of NVIDIA's products or NVIDIA's partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA's products or technologies when integrated into systems; and changes in applicable laws and regulations, 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.

© 2025 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, ConnectX, CUDA, DGX, NVIDIA Cosmos, NVIDIA NIM and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other NVIDIA trademarks and service marks not listed herein are also the property of NVIDIA Corporation. All rights reserved. All other company and product names are trademarks or registered trademarks of their respective owners. Features, pricing, availability and specifications are subject to change without notice.

Alex Shapiro
Enterprise Networking
1-415-608-5044
ashapiro@nvidia.com