



NVIDIA DGX Spark

A personal AI computer, designed to build and run AI.



Desktop AI Compute Demands

The rise of agentic AI—autonomous agents like OpenClaw that can plan, reason, and execute multiple steps to complete complex tasks—represents a significant technical shift from passive generative AI models. While traditional AI simply answers questions, these agents act as “digital employees,” capable of navigating file systems, managing calendars, and interacting with local software. State-of-the-art mixture-of-experts (MoE) models are the engines driving this evolution, generating tokens significantly faster and enabling agents to “think” through a sequence of actions in near-real time.

Building and deploying autonomous agents—which may coordinate multiple specialized models for reasoning, speech, vision, and tool use—makes local development and deployment challenging, increasing the demand for high-performance AI compute resources and memory to handle this new generation of autonomous AI.

200B Parameter Models on Your Desk

NVIDIA DGX Spark™ is part of a new class of supercomputers designed from the ground up to build and run AI. Powered by the NVIDIA GB10 Grace Blackwell Superchip and based on the NVIDIA Grace Blackwell architecture, NVIDIA DGX Spark delivers up to 1 petaFLOP¹ of AI performance to power large AI workloads. The GB10 Superchip uses NVIDIA NVLink™-C2C technology to deliver a CPU+GPU coherent memory model with 5x the bandwidth of PCIe Gen 5 and combined with 128 GB of unified system memory, developers can experiment, fine-tune, or build agents using the latest generation of reasoning and MoE AI models of up to 200B parameters.² Plus, NVIDIA ConnectX™ networking can connect two NVIDIA DGX Spark systems to enable inference on models up to 405B parameters.²

NVIDIA DGX Spark is powered by NVIDIA DGX OS with Ubuntu Linux and comes preconfigured with the latest NVIDIA AI software stack. With NVIDIA Developer Program access to NVIDIA NIM™ and NVIDIA Blueprints, developers can hit the ground running using common tools such as PyTorch, Jupyter, and Ollama to prototype, fine-tune, and run inference locally. Work can then be easily moved to DGX Cloud or any accelerated data center or cloud infrastructure. Large local memory, strong compute performance, and power efficient design make DGX Spark ideal for desktop autonomous agent development and deployment. NVIDIA OpenShell provides an open-source runtime to build always on autonomous, self-evolving agents more safely on DGX Spark.

Key Features

- > Built on NVIDIA GB10 Grace Blackwell Superchip
- > NVIDIA Blackwell GPU with fifth-generation Tensor Core technology
- > NVIDIA Grace CPU with 20-core high-performance Arm architecture
- > Up to 1 petaFLOP¹ of AI performance using FP4
- > 128 GB of coherent, unified system memory
- > Support for up to 200 billion parameter² models
- > NVIDIA ConnectX™ networking to link two systems to work with models up to 405 billion parameters
- > 4 TB of NVMe storage
- > Compact desktop form factor

An optional NVIDIA AI Enterprise license for NVIDIA DGX Spark provides access to robust AI development and deployment resources—including enterprise-level support, supported container and operator maintenance releases, defect resolutions, security patches, knowledge base access, and phone, email, and online support.

Develop Locally, Deploy Anywhere at Scale

Leveraging the NVIDIA AI platform software architecture makes it possible for NVIDIA DGX Spark users to easily move their work from their desktop to DGX Cloud or any accelerated cloud or data center infrastructure, making it easier than ever to prototype, fine-tune, and iterate.

Technical Specifications*

Architecture	NVIDIA Grace Blackwell
GPU	NVIDIA Blackwell Architecture
CPU	20 core Arm, 10 Cortex-X925 + 10 Cortex-A725 Arm
NVIDIA® CUDA® Cores	NVIDIA Blackwell Generation
Tensor Cores	5th Generation
RT Cores	4th Generation
Tensor Performance ¹	1 PFLOP
System Memory	128 GB LPDDR5x, coherent unified system memory
Memory Interface	256-bit
Memory Bandwidth	Up to 273 GB/s
Storage	4 TB NVME.M2 with self-encryption
USB	4x USB TypeC
Ethernet	1x RJ-45 connector 10 GbE
NIC	ConnectX-7 NIC @ 200 Gbps
Wi-Fi	WiFi 7
Bluetooth	BT 5.4 w/LE
Audio-output	HDMI multichannel audio output
Power Consumption	240 W
Display Connectors	1x HDMI 2.1a
NVENC NVDEC	1x 1x
OS	NVIDIA DGX™ OS
System Dimensions	150 mm L x 150 mm W x 50.5 mm H
System Weight	1.2 kg

* preliminary specifications, subject to change

Ready to Get Started?

To learn more about NVIDIA DGX Spark, visit hypertecsp.com/partners/nvidia/nvidia-dgx-spark/

1. Theoretical FP4 TOPS using the sparsity feature. | 2. Using FP4 precision models.