Al supercomputing goes Nano

Accelerate AI workflows with HP ZGX Nano and ZGX Toolkit¹, high-performance local compute combined with a curated open-source stack for prototyping, fine-tuning and inferencing. Built-in IP discovery, model export and local serving reduce friction, ensuring repeatable, deployment-ready results.

Big Al performance. Tiny footprint.

Prototype, fine-tune, and inference models of up to 200B parameters locally. Powered by NVIDIA® GB10 Grace Blackwell Superchip² and 128 GB of coherent unified memory, the HP ZGX Nano delivers 1,000 TOPS of FP4 Al performance in a compact desktop.

HP ZGX Toolkit: Accelerate time to results

Move from idea to deployment faster. The ZGX Toolkit¹ streamlines Al workflows by pairing powerful local compute with open-source tools, built-in discovery, and easy exporthelping teams cut friction, boost productivity, and scale results anywhere.

Offload Al workloads without the cloud

Pair existing laptop and desktop systems—Windows, Mac, or Linux—with a network-connected HP ZGX Nano. Get developer-grade performance without data center queues and costly cloud instances, minimizing latency and keeping sensitive information local.



*Product image may differ from actual product

Sustainability in action

Advance edge Al solutions

Realize the true potential of edge Al solutions. From real-time computer vision to running agentic and domain-specific models locally, get the performance needed to deploy intelligent, responsive solutions directly where data is generated.



Featuring

Big Al performance. Tiny footprint.

Prototype, fine-tune, and inference models of up to 200B parameters locally. Powered by NVIDIA® GB10 Grace Blackwell Superchip² and 128 GB of coherent unified memory, the HP ZGX Nano delivers 1,000 TOPS of FP4 AI performance in a compact desktop.

HP ZGX Toolkit: Accelerate time to results

Move from idea to deployment faster. The ZGX Toolkit¹ streamlines AI workflows by pairing powerful local compute with open-source tools, built-in discovery, and easy export—helping teams cut friction, boost productivity, and scale results anywhere.

Offload Al workloads without the cloud

Pair existing laptop and desktop systems—Windows, Mac, or Linux—with a network-connected HP ZGX Nano. Get developer-grade performance without data center queues and costly cloud instances, minimizing latency and keeping sensitive information local.

Advance edge Al solutions

Realize the true potential of edge Al solutions. From real-time computer vision to running agentic and domain-specific models locally, get the performance needed to deploy intelligent, responsive solutions directly where data is generated.

NVIDIA® DGX™ OS and AI software stack

Hit the ground prototyping, fine-tuning, and inferencing with familiar NVIDIA® DGX™ OS and integrated NVIDIA® AI software stack purpose-built for modern AI development.

NVIDIA® GB10 Grace Blackwell Superchip

Get up to 1000 TOPS of AI compute at FP4 precision with an NVIDIA® Blackwell GPU. Supercharge data preprocessing and orchestration with a Grace 20-core Arm CPU.2

Unified system memory

Run Al development and testing workloads with Al models of up to 200 billion parameters at your desk with 128 GB of coherent unified system memory.

Fast, secure storage.

Choose between 1 or 4 TB of NVMe M.2 self-encrypted storage to work efficiently with large files and keep more of your data local and secure.³

NVIDIA® ConnectX™ Networking

Work with even larger Al models locally—up to 405 billion parameters—by connecting two HP ZGX Nano systems together to scale local compute resources.⁴

Tiny Al powerhouse

Measuring at 150mm L x 150mm W x 51mm H, this new class of desktop is purpose-built for Al development yet fits in the palm of your hand.⁵

HP ZGX Toolkit for Al

Get open-source frameworks, MLflow tracking, and Ollama testing. With instant discovery, sync, and export, train locally, track results, and deploy with cloud, data center, or workstation.¹



Technical specifications

Available Operating Systems	NVIDIA DGX ™ OS
Processor family	NVIDIA Grace Blackwell
Available Processors ¹	NVIDIA® GB10 Grace Blackwell Superchip (20-core Arm, 10 Cortex-X925 and 10 Cortex-A725) with NVIDIA® Blackwell GPU Architecture
Form factor	Mini
Maximum memory	128 GB LPDDR5x (unified, onboard) Memory bandwidth up to 273 GB/s.
Internal storage	1TB PCIe® NVMe™ OPAL M.2 SSD ² 4TB PCIe® NVMe™ OPAL M.2 SSD ²
Available Graphics	Integrated: NVIDIA® Blackwell
Audio	DP/HDMI Audio Output
Ports and connectors	Rear: 1 USB Type-C® power connector; 3 USB Type-C® 20Gbps signaling rate; 1 RJ-45 (10Gbps); 2 QSFP 200 Gbps signaling rate; 1 HDMI 2.1a
Communications	LAN: Realtek RTL8127 10 GbE Ethernet Controller; NVIDIA ConnectX-7 200 GbE Ethernet Controller ; WLAN: MediaTek Wi-Fi 7 MT7925 (2x2) and Bluetooth® 5.4 wireless card ⁵
Software	NVIDIA AI software stack



Technical specifications

Power	240W external USB Type-C power adapter, 89% efficiency, active PFC.
Dimensions	$5.9 \times 5.9 \times 2.01$ in (without feet); $5.9 \times 5.9 \times 2.1$ in (with feet); $15 \times 15 \times 5.1$ cm (without feet); $15 \times 15 \times 5.45$ cm (with feet); (Standard desktop orientation.); $8.5 \times 5.6 \times 10.7$ in $21.6 \times 14.2 \times 27.2$ cm (Package)
Weight	Starting at 2.76 lb; Starting at 1.25 kg; (Exact weights depend upon configuration (System weight only))
Sustainable impact specifications	40% post-consumer recycled plastic; 60% post-consumer recycled plastic; 100% sustainably sourced packaging; Contains at least 20% post-industrial recycled steel; 100% sustainably sourced packaging or recycled; 75% recycled aluminum ^{5,67,8}



Footnotes

Messaging Footnotes

¹The HP ZGX Toolkit is provided free of charge. Use requires a client device running Windows 11 or Ubuntu 24.04 (or later) with Visual Studio Code installed and host device ZGX Nano. The client device must be x86-based, but otherwise there are no restrictions regarding hardware specifications or device manufacturer. Availability may vary by region and is subject to applicable local laws, regulations, and restrictions.

² Multi-core is designed to improve performance of certain software products. Not all customers or software applications will necessarily benefit from use of this technology. Performance and clock frequency will vary depending on application workload and your hardware and software configurations.

 $^3\,{\rm 1TB}$ or ${\rm 4TB}$ storage configuration must be selected at time of purchase.

⁴ Requires compatible QSFP cable. Sold separately.

⁵ Height excludes feet.

Specification Footnotes

¹ Multi-core is designed to improve performance of certain software products. Not all customers or software applications will necessarily benefit from use of this technology. Performance and clock frequency will vary depending on application workload and your hardware and software configurations.

 $^{2}\,\mbox{1TB}$ or 4TB storage configuration must be selected at time of purchase.

⁵ Wireless access point and Internet service required and sold separately. Availability of public wireless access points limited. Wi-Fi 7 (802.11BE) functionality requires compatible Windows 11 24H2 OS, compatible processor, and separately purchased Wi-Fi 7 router to support backward compatibility with prior 802.11 specs. Available in countries where Wi-Fi 7 is supported.

⁶ Recycled metal is expressed as a percentage of the total weight of the metal according to ISO 14021 definitions for metal parts over 25 grams.

⁷ Recycled plastic is expressed as a percentage of the total weight plastic. Post-consumer recycled is based on the definition set in the EPEAT standard for computers, IEEE 1680.1-2018 standard.

⁸ HP paper and fiber based packaging for PCs, displays, home and office print, and supplies is reported by suppliers as recycled or certified, with a minimum of 97% by volume verified by HP. Packaging is the box that comes with the product and all paper-based materials inside the box. Packaging for commercial, personal systems accessories, and spare parts is not included

