

# HP ZGX Fury G1n AI Station

KI vor Ort, für Skalierbarkeit konzipiert. Keine nutzungsbasierten Token-Kosten.

Flexible Softwareumgebung und lokale Rechenleistung unterstützen die Entwicklung, Bereitstellung und Ausführung von KI-Anwendungen und KI-Agenten. Die HP ZGX Fury AI Station hostet große Modelle mit bis zu 1 Billion Parameter<sup>1</sup> lokal, damit die Daten sicher und die Kosten vorhersehbar bleiben<sup>2</sup>.

**Lokale Entwicklung von KI-Anwendungen**

HP ZGX Fury kombiniert den NVIDIA® GB300 Grace Blackwell Ultra Desktop Superchip mit 748 GB gemeinsamem Speicher. Unterstützt die Feinabstimmung von KI-Modellen mit bis zu 100 Milliarden Parametern<sup>1</sup> auf lokalen Systemen und führt zu besser planbaren Kosten.

**Inferenz für den produktiven Einsatz. Planbare Kosten.<sup>2</sup>**

Lokale Inferenz für Unternehmensanwendungen ermöglicht die Ausführung von KI-Modellen vor Ort und kann nutzungsabhängige Token-Gebühren reduzieren. HP ZGX Fury kann als gemeinsam genutztes KI-System bereitgestellt werden und unterstützt die lokale Ausführung von KI-Modellen mit bis zu 1 Billion Parametern<sup>1</sup> sowie proprietären Anwendungen und KI-Agenten. Die Plattform bietet hohe Datenraten und geringe Latenzen für die Zusammenarbeit im Team.

**Offene, flexible Tools. Schnelle Einrichtung. Keine Bindung.**

HP ZGX-Toolkit vereinfacht die Geräteerkennung, Kopplung und Installation von Open-Source-Tools innerhalb der integrierten NVIDIA KI-Softwareumgebung.<sup>3</sup> Die vorinstallierte HP Z Runtime CLI-App unterstützt die Bereitstellung und Ausführung von Hugging Face Modellen.



\*Das Produktbild kann vom tatsächlichen Produkt abweichen

**Nachhaltigkeit in Aktion**

**Gemeinsam nutzbares KI-Computing. Für den Einsatz im Büro entwickelt.**

Ein einzelnes HP ZGX Fury System unterstützt die gleichzeitige Inferenz für mehrere Benutzer:innen sowie die parallele Entwicklung und Feinabstimmung von KI-Modellen durch mehrere Teams. Das System kann im Rack oder im Büro eingesetzt werden und lässt sich in bestehende Unternehmensumgebungen integrieren.



# HP ZGX Fury G1n AI Station

## Funktionsumfang

### Lokale Entwicklung von KI-Anwendungen

HP ZGX Fury kombiniert den NVIDIA® GB300 Grace Blackwell Ultra Desktop Superchip mit 748 GB gemeinsamem Speicher. Unterstützt die Feinabstimmung von KI-Modellen mit bis zu 100 Milliarden Parametern<sup>1</sup> auf lokalen Systemen und führt zu besser planbaren Kosten.

### Inferenz für den produktiven Einsatz. Vorhersehbare Kosten <sup>2</sup>.

Lokale Inferenz für Unternehmensanwendungen ermöglicht die Ausführung von KI-Modellen vor Ort und kann nutzungsabhängige Token-Gebühren reduzieren. HP ZGX Fury kann als gemeinsam genutztes KI-System bereitgestellt werden und unterstützt die lokale Ausführung von KI-Modellen mit bis zu 1 Billion Parametern<sup>1</sup> sowie proprietären Anwendungen und KI-Agenten. Die Plattform bietet hohe Datenraten und geringe Latenzen für die Zusammenarbeit im Team.

### Offene, flexible Tools. Schnelle Einrichtung. Keine Bindung.

HP ZGX-Toolkit vereinfacht die Geräteerkennung, Kopplung und Installation von Open-Source-Tools innerhalb der integrierten NVIDIA KI-Softwareumgebung.<sup>3</sup> Die vorinstallierte HP Z Runtime CLI-App unterstützt die Bereitstellung und Ausführung von Hugging Face Modellen.

### Gemeinsam nutzbares KI-Computing. Für den Einsatz im Büro entwickelt.

Ein einzelnes HP ZGX Fury System unterstützt die gleichzeitige Inferenz für mehrere Benutzer:innen sowie die parallele Entwicklung und Feinabstimmung von KI-Modellen durch mehrere Teams. Das System kann im Rack oder im Büro eingesetzt werden und lässt sich in bestehende Unternehmensumgebungen integrieren.

### NVIDIA® DGX™ OS- und KI-Software-Stack

NVIDIA® DGX™ OS unterstützt die Entwicklung, das Testen und die Inferenz großer KI-Modelle in einer vertrauten Softwareumgebung. Der HP ZGX Fury Developer Software Stack unterstützt die Konfiguration von NVIDIA AI Enterprise sowie offenen KI-Umgebungen und wurde für aktuelle KI-Anwendungen entwickelt.

### NVIDIA® GB300 Grace Blackwell Ultra-Superchip

Der NVIDIA® GB300 Grace Blackwell Ultra Superchip bietet bis zu 20 PetaFLOPS FP4 KI-Rechenleistung und 748 GB gemeinsamen Speicher für die Entwicklung und Ausführung großer KI-Modelle vor Ort.

### Einheitlicher Speicher

Mit 748 GB gemeinsamem Speicher lassen sich KI-Modelle mit bis zu 100 Milliarden Parametern entwickeln und feinabstimmen. Darüber hinaus wird die Inferenz für Modelle mit bis zu einer Billion Parametern unterstützt.<sup>1</sup>

### Schnelle und sichere Speicherung

Wählen Sie zwischen 2 oder 4 TB NVMe M.2 selbstverschlüsselndem Speicher, um große Dateien effizient zu verarbeiten und mehr Daten lokal und sicher zu speichern.<sup>4</sup>

### Für den Einsatz im Büro entwickelt

KI-Rechenleistung vor Ort – vom Schreibtisch über den Technikraum bis zum Rechenzentrum. Das für den Einsatz im Büro nutzbare Rack-Einbau lässt sich perfekt in Standard-Unternehmensumgebungen integrieren.

### HP ZGX Toolkit für KI

Sichern Sie sich Open-Source-Frameworks, MLflow-Tracking und Ollama-Tests. Mit sofortiger Erkennung, Synchronisierung und Export, lokalem Training, Nachverfolgung von Ergebnissen und Bereitstellung in der Cloud, im Rechenzentrum oder auf einer Workstation.<sup>5</sup>

### HP Z Runtime CLI-App

Die vorinstallierte HP Z Runtime unterstützt den schnellen Übergang von der Modellentwicklung zur Inferenz. KI-Teams können Modelle lokal bereitstellen, ausführen und verwalten.<sup>6</sup>

### Grafische KI-Unterstützung

Eine optionale NVIDIA RTX™ PRO GPU übernimmt die Videoausgabe. Dadurch stehen GB300 GPU-Ressourcen für anspruchsvolle KI-Anwendungen zur Verfügung.

### NVIDIA ConnectX®-8 SuperNIC™ Networking

Unterstützt die lokale Ausführung größerer KI-Modelle. Mit bis zu 800 Gb/s unterstützt die NVIDIA ConnectX®-8 SuperNIC™ eine leistungsfähige Netzwerkverbindung. Zwei ZGX Fury Systeme lassen sich koppeln, um größere KI-Modelle zu verarbeiten und die verfügbaren Rechenressourcen zu erweitern.<sup>6,7</sup>



# HP ZGX Fury G1n AI Station

## Technische Daten

|                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Verfügbare Betriebssysteme           | Ubuntu 24.04 LTS mit NVIDIA KI-Entwicklertools <sup>2</sup>                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Prozessorproduktfamilie <sup>1</sup> | NVIDIA® Grace Blackwell Ultra Desktop Superchip                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Verfügbare Prozessoren <sup>1</sup>  | NVIDIA® GB300 Grace Blackwell Ultra Desktop (NVIDIA® Grace-72 Core Neoverse V2 CPU) mit NVIDIA® Blackwell Ultra GPU (252 GB HBM3e)                                                                                                                                                                                                                                                                                                                                           |
| Formfaktor                           | Tower                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Maximaler Hauptspeicher              | 496 GB LPDD5X 6400 MT/s SOCAMM<br>Speicherbandbreite von bis zu 396 GB/s                                                                                                                                                                                                                                                                                                                                                                                                     |
| Interner Speicher                    | 2 TB HP Z Turbo Drive PCIe® M.2-SSD <sup>7</sup><br>4 TB HP Z Turbo Drive PCIe® M.2-SSD <sup>7</sup><br>2 TB Z Turbo PCIe NVMe™ M.2 SED NMIC SSD <sup>7</sup>                                                                                                                                                                                                                                                                                                                |
| Verfügbare Grafikkarten              | Integriert: NVIDIA® Blackwell Ultra-GPU<br>Separat: NVIDIA RTX PRO™ 4000 Blackwell Generation (24 GB GDDR7 dediziert) NVIDIA RTX PRO™ 2000 Blackwell Generation (16 GB GDDR7 dediziert)                                                                                                                                                                                                                                                                                      |
| Erweiterungssteckplätze              | 3 PCIe-Steckplätze: 1 PCIe Gen5 x16; 2 PCIe Gen5 x16 (x8 elektrisch)                                                                                                                                                                                                                                                                                                                                                                                                         |
| Ports und anschlüsse                 | Hinten: 4 USB Type-A 10 Gbit/s Signalarate; 1 Mikrofon; 1 RJ-45 (1 GbE); 1 RJ-45 1 Gbit Ethernet (Systemmanagement ausschließlich über BMC) 1 Micro-USB (Systemmanagement ausschließlich über BMC) 1 Mikrofon-Audioanschluss; 1 Mini DisplayPort™ 2.0 (Systemmanagement ausschließlich über BMC); 2 QSFP112 (400 GB pro Port); 1 Audioeingang; 1 Audioausgang ; Vorderseite: 2 USB Type-C® 10 Gbit/s Signalarate; 2 USB Type-A 5 Gbit/s Signalarate; 1 Universal-Audiobuchse |
| Kommunikation                        | NVIDIA® ConnectX-8, 2 400-Gbit-QSFP112-Anschlüsse (unterstützt optische und DAC-Anschlüsse)                                                                                                                                                                                                                                                                                                                                                                                  |
| Software                             | HP Z Laufzeit; NVIDIA KI-Entwickler-Tools; NVIDIA-geprüftes, von Partnern entwickeltes BIOS; NVIDIA-geprüfte, von Partnern entwickelte BMC-Firmware                                                                                                                                                                                                                                                                                                                          |
| Sicherheitsmanagement                | Trusted Platform Module (TPM) 2.0; UEFI-Version 2.9; Kensington-Kabelschloss; UEFI Secure Boot                                                                                                                                                                                                                                                                                                                                                                               |



# HP ZGX Fury G1n AI Station

## Technische Daten

|                                |                                                                                                                                                                        |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Managementfunktionen           | Out-of-Band mit NVIDIA integriertem BMC; Redfish API-Unterstützung                                                                                                     |
| Stromversorgung                | 1600 W Titan                                                                                                                                                           |
| Abmessungen                    | 44,6 × 22,15 × 66,7 cm; (Standard Tower-Bauform, wird mit Rack-Schienen geliefert, damit die Rack-Konfiguration sofort verfügbar ist.); 80,65 × 58,9 × 43,5 cm (Paket) |
| Gewicht                        | Ab 41 kg;<br>(Exakte Gewichte abhängig von der Konfiguration (reines Systemgewicht))                                                                                   |
| Zertifizierung und Konformität | TAA-konform                                                                                                                                                            |





# HP ZGX Fury G1n AI Station

## Fußnoten

### Fußnoten für Text

<sup>4</sup> Die Speicherkonfiguration mit 2 TB oder 4 TB muss zum Zeitpunkt des Kaufs ausgewählt werden.

<sup>5</sup> Das HP ZGX Toolkit wird kostenlos zur Verfügung gestellt. Die Verwendung erfordert ein Client-Gerät, auf dem Windows 11 oder Ubuntu 24.04 (oder höher) ausgeführt wird und Visual Studio Code installiert ist, sowie ein ZGX Fury-Hostgerät. Das Client-Gerät muss x86-basiert sein, ansonsten bestehen keinerlei Einschränkungen hinsichtlich der Hardwarespezifikationen oder des Geräteherstellers. Die Verfügbarkeit variiert je nach Region und unterliegt den geltenden lokalen Gesetzen, Vorschriften und Einschränkungen.

<sup>6</sup> Die Geschwindigkeit hängt von der Modellgröße und dem Zugriffsstatus des jeweiligen Modells ab.

<sup>7</sup> Kompatibles QSFP-Kabel erforderlich. Separat erhältlich.

<sup>1</sup> Unterstützt die Quantisierung von Modellen mit bis zu FP4.

<sup>2</sup> KI-Workstations unterstützen eine besser planbare Kostenstruktur, indem nutzungsabhängige Gebühren – beispielsweise für Tokens – entfallen. Gleichzeitig verringern sie die Abhängigkeit von variabler Cloud-Infrastruktur und ermöglichen die lokale Ausführung von KI-Anwendungen auf fest bereitgestellten Systemen.

<sup>3</sup> Das HP ZGX Toolkit wird kostenlos zur Verfügung gestellt. Die Verwendung erfordert ein Client-Gerät, auf dem Windows 11 oder Ubuntu 24.04 (oder höher) ausgeführt wird und Visual Studio Code installiert ist, sowie ein ZGX Fury-Hostgerät. Das Client-Gerät muss x86-basiert sein, ansonsten bestehen keinerlei Einschränkungen hinsichtlich der Hardwarespezifikationen oder des Geräteherstellers. Die Verfügbarkeit variiert je nach Region und unterliegt den geltenden lokalen Gesetzen, Vorschriften und Einschränkungen.

### Fußnoten zur Spezifikation

<sup>1</sup> Die Multi-Core-Technologie dient zur Verbesserung der Leistung bestimmter Softwareprodukte. Die Verwendung dieser Technologie bringt nicht zwangsläufig für jeden Kunden und jede Softwareanwendung Vorteile. Leistung und Taktfrequenz variieren je nach Anwendungs-Workloads und Ihrer Hardware- und Softwarekonfiguration.

<sup>2</sup> Es stehen nicht alle Funktionen in allen Editionen oder Versionen von Ubuntu zur Verfügung. Das System erfordert möglicherweise aktualisierte und/oder separat erworbene Hardware, Treiber, Software oder ein BIOS-Update, damit die Funktionen von Ubuntu vollständig genutzt werden können. Ubuntu kann automatisch aktualisiert werden. Es fallen ggf. ISP-Gebühren an und es müssen im Laufe der Zeit ggf. zusätzliche Anforderungen für Updates erfüllt werden.

<sup>3</sup> Wireless Access Point und Internet-Service sind erforderlich und separat zu erwerben. Die Verfügbarkeit öffentlicher Wireless Access Points ist beschränkt. Die Wi-Fi 7 (802.11BE)-Funktionalität erfordert ein kompatibles Windows 11 24H2-Betriebssystem, einen kompatiblen Prozessor und einen separat erworbenen Wi-Fi 7-Router, um die Abwärtskompatibilität mit früheren 802.11-Spezifikationen zu unterstützen. In Ländern verfügbar, in denen Wi-Fi 7 unterstützt wird.

<sup>7</sup> Die Speicherkonfiguration mit 2 TB oder 4 TB muss zum Zeitpunkt des Kaufs ausgewählt werden.

