Dell Pro Max with GB10 Product Introduction

May 2025
Dell Pro Max Product Management

DELL Technologies

NVIDIA

Welcome to Dell Pro Max with GB10

Future of desk-side AI development

Dell Pro Max with GB10 AI computer enables developers to prototype, fine-tune and inference large AI models locally and seamlessly deploy to the data center or cloud.

Dell Pro Max with GB10 Specs

- Grace CPU (20 core ARM-based, 10 Cortex-X925 + 10 Cortex-A725)
  Blackwell GPU featuring NVLink-C2C ultra-low latency interconnection
- 128GB LPDDR5x unified system memory
- (1) M.2 NVMe 2242 Gen4 SSD (1TB & 4TB options)
- 240W Power Adapter
- NVIDIA Connect-X 7 + Wi-Fi + BT5.1
- DGX OS 7 (Linux based) + NVIDIA AI SW Stack
- 150mm x 150mm x 50.5 mm (~1.2L)
- TPM 2.0

Dell Pro Max with GB10 Ports & Connections

Power connector

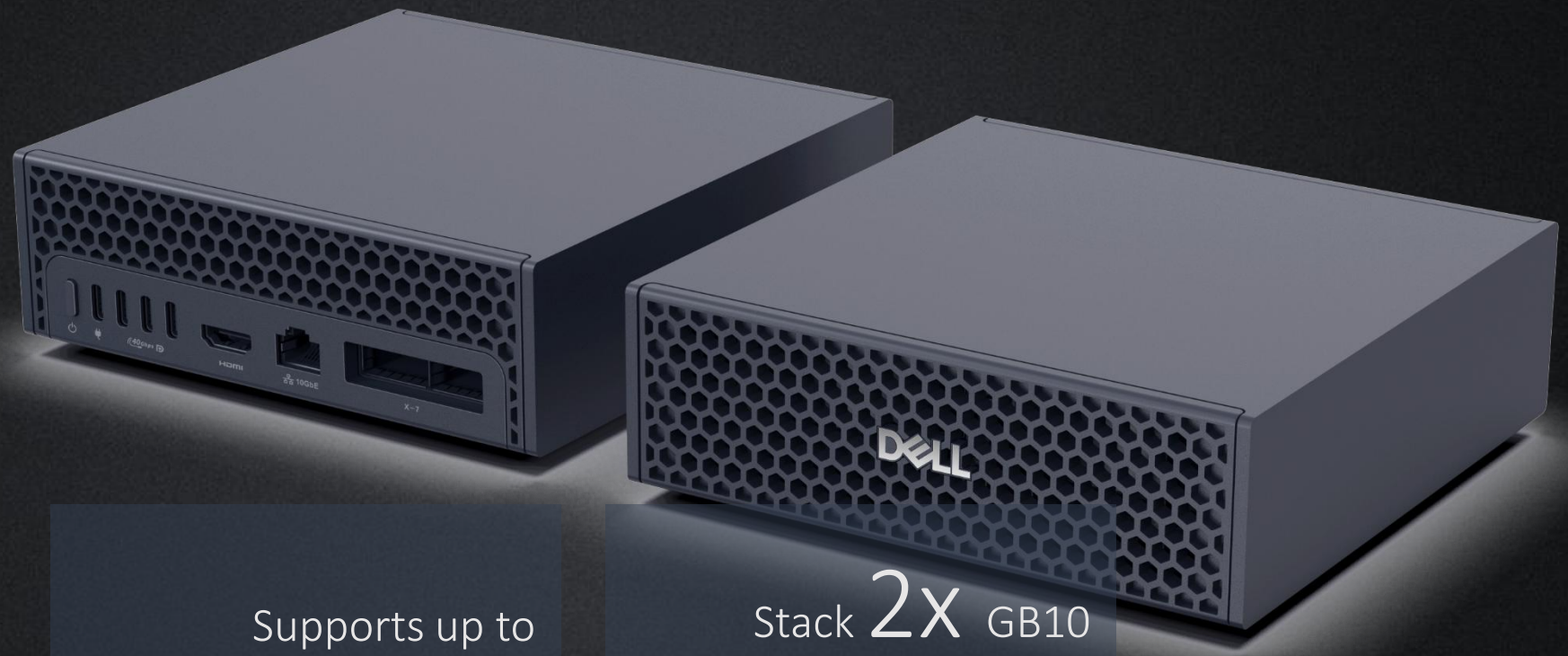
(3) USB Type-C

HDMI 2.1b

RJ45, 10GbE

(2) QSFP, Copper, 200G

Dell Pro Max with GB10 Key Features



1000 TFLOPS of FP4 computing power*

Supports up to **200Bn** parameter models*

Stack **2x** GB10 through CX7 to intake 400Bn parameter models (Llama 4 Maverick level) *

*Preliminary performance, subject to change

NVIDIA. DELL Technologies

Dell Pro Max with GB10 is designed for AI Developers

- Enables developers to run large models, up to 200B parameters, and large AI workloads locally
- NVIDIA AI software stack provides developers with tools to build & run AI
- Provides additional computation – keep your favorite laptop or desktop system (for productivity and other work) and send AI workloads to GB10
- Data stays local, no need to transfer data to external locations, or no contention for data center compute or expense of cloud instances
- Connect two Dell Pro Max with GB10 to work with larger models & workloads

NVIDIA DELL Technologies

Bring AI workloads to the next level

Perform next-level AI with cost-efficiency	• No monthly cloud GPU fees • Predictable one-time CapEx • Avoid usage-based token pricing
Secure local model without reliance on cloud	• Full control over data & model locally • Works on the edge • Runs GenAI in regulated or sensitive sectors
Accelerate workloads with specialized AI hardware	• Runs larger models • More efficient AI inference • Faster CPU – memory – GPU communication
Compact design to deploy AI on edge	• Easy to install at edge or field • Space saving • Power efficient
Fast interconnection to scale AI	• Easily scaling for larger models • Faster than traditional multi-GPU setups • Futureproof flexibility
AI optimized development environment	• Ready out-of-the-box drivers/toolkits/libraries… • Stable, fully tested and validated by NVIDIA • System-specific performance optimizations

R&D
AI prototype

Edge
Applications

AI
Development

Private Data
Deployment

Data
Science

Education

NVIDIA DELL Technologies