

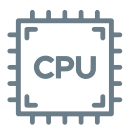


**NVIDIA MGX™ Server  
AEX-2UA1**

## Aetina Edge Server with NVIDIA MGX™ Accelerates On-Premises Enterprise AI Deployment

Aetina SuperEdge AEX-2UA1, an accelerated edge server built on the NVIDIA MGX™ modular design, integrates a single socket Intel® Xeon® 6 processor and dual double-width NVIDIA data center GPUs with NVLink™, delivering maximum AI performance in a space-efficient 2U form factor, while offering greater scalability and flexibility for enterprises to achieve faster time to value. Built for both performance and scalability, the edge server accelerates generative AI workloads, including large language model (LLM) training and inference, retrieval-augmented generation (RAG) and cutting-edge enterprise AI applications.

Optimized for on-premises generative AI, the AEX-2UA1 enables secure processing of sensitive data with local LLMs and enterprise AI applications. It's ideal for privacy-critical sectors like finance, healthcare, and space-constrained 5G telco environments, offering an ideal entry-level solution for LLM deployment compared to standard GPU servers on the market.



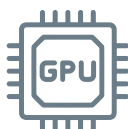
### Single Intel Xeon 6700-series with P-cores

- High performance data processing



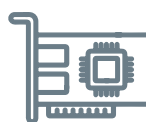
### NVIDIA MGX™ Modular Server Design

- Scalable, future-ready modular architecture



### Dual Double-Width NVIDIA GPUs with NVLink

- Optimized for high performance AI workloads



### High-Speed Networking

- NVIDIA Bluefield®-3 DPUs and NVIDIA ConnectX®-7 NICs



### 2U Short-Depth Footprint

- Fits standard, short-depth, or non-rack setups



### Enterprise AI Training and Inference

- Ideal for private LLMs, RAG, data analytics, simulation, rendering

# Short-Depth Server



## AEX-2UA1 \*Preliminary

- Unmatched edge AI performance: Built on NVIDIA MGX™ server architecture and Intel® Xeon® 6 with P-cores.
- Enhanced AI Training and Inferencing: Dual Double-width GPUs with NVIDIA® NVLink™ for seamless AI training and inferencing.
- Optimized Network transfer for demanding workloads: Minimal latency with NVIDIA® BlueField®-3 DPU and ConnectX®-7 NIC.
- Deployments in space-constrained locations: Compact 2U short-depth (420mm) rackmount server.



**READ MORE**  
Aetina Short-Depth Server

Model Number		AEX-2UA1
Chassis	Form Factor	Front Access 2U Rackmount
	Dimensions (W x D x H)	438mm x 420mm x 88mm
Processor		Single Intel® Xeon® 6700-series with P-cores
Chipset		N/A
GPU	Max Count	Up to 2 Double-width PCIe Passive GPUs
	Qualified	L40S, RTX PRO™ 6000 Blackwell Server Edition, H200NVL (with NVLink bridge)
System Memory		8 x Channel Memory
Storage	Drive Bays	Rear: 4x Hot-swappable E1.S, up to PCIe Gen5, up to 15mm E1.S
	M.2	2 x M.2, up to PCIe Gen5 x4, M-Key 2280/22110
	RAID Support	NVMe RAID 0/1/5/10 support (Intel® VROC RAID key required)
Input/Output		1 x RJ45 1Gbe Dedicated BMC LAN port 1 x MiniDP port 2 x USB 2.0 ports
Front Panel	Buttons	System Power UID/BMC reset
	LED	Power status UID Storage Activity Network Activity Information
PCIe Expansion slots		2 x PCIe Gen5 x16 link (PCIe x16 slot, FHFL) for 2 Double-width GPUs with NVLink bridge 1 x PCIe Gen5 x8 link (PCIe x16 slot, FHFL)* 1 x PCIe Gen5 x16 link (PCIe x16 slot, FHHL) * Slot not available when 2 Double-width AOC is installed.
Power	Power Supply	Default: 2 x 2,000W Redundant (1+1) Titanium Level (96%) power supplies
	Power Input	Default 2,000W power supply: 1,000W: 100-127V~, 12A, 47-63Hz 2,000W: 220-240V~, 10A, 47-63Hz
	Power Connectors	IEC 60320 C13
Package		TBD
Weight		TBD
MISC. Function		Trusted Platform Module (TPM) 2.0 Intel® Platform Firmware Resilience (Intel® PFR)
Temperature		Operating Temperature: 10°C ~ 35°C (50°F ~ 95°F) Non-operating Temperature: -30°C to 60°C (-22°F to 140°F)
Humidity		Operating Relative Humidity: 8% to 80% (non-condensing) Non-operating Relative Humidity: 8% to 90% (non-condensing)
OS Support		TBD
Certification		TBD

## Key Applications

Enterprise / Finance / 5G Telco

Healthcare

Retails

Automotive

Manufacture

