

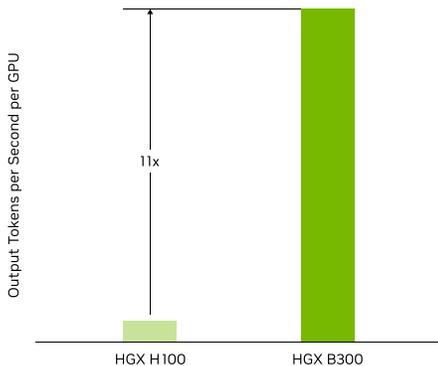
AMAX GPU Solutions with NVIDIA HGX™ B300

Accelerating Advanced AI in Every Data Center

AMAX ENGINEERING

AMAX is a global leader in advanced computing solutions for AI, HPC, edge, and data center applications. Backed by decades of engineering expertise, AMAX partners with organizations worldwide to design, build, and deploy scalable systems.

Llama 3.1 405B Real-Time Throughput



*Source: NVIDIA

Infrastructure Built for Next-Generation AI

Accelerated by the NVIDIA Blackwell architecture, the AMAX RackScale 32 with NVIDIA HGX B300 delivers a high-density, rack-scale solution engineered for large-scale AI, HPC, and enterprise workloads. Designed for training, inference, and advanced LLM reasoning, the AMAX RackScale 32 combines compute performance with efficient scalability for industries including healthcare, finance, public sectors, manufacturing, neoclouds, and research institutions.

AMAX RackScale 32 with NVIDIA HGX B300 Features

- **Tensor/Transformer Cores:** Up to 576 specialized cores that accelerate transformer operations (e.g., attention layers) critical to LLM performance.
- **High Memory Capacity:** Up to 8.4TB of HBM3e memory, enabling efficient handling of massive models and long context windows without bottlenecks.
- **In-Node Bandwidth:** NVIDIA NVLink/NVSwitch interconnect keeps GPUs fully utilized across large-scale training workloads.
- **High-Speed Networking:** 800Gbps NVIDIA InfiniBand for rapid multi-node synchronization across distributed AI clusters.

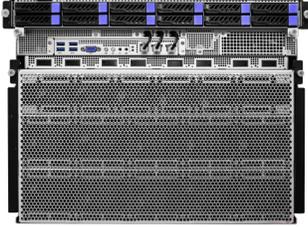
Accelerated Performance with NVIDIA Blackwell Ultra

NVIDIA HGX B300 delivers up to 11x faster real-time throughput for Llama 3.1 405B models compared to HGX H100 systems, accelerating development cycles, shortening time-to-insight, and enabling organizations to bring next-generation AI innovations to market faster.

AMAX GPU Solutions with NVIDIA HGX™ B300

AMAX AceleMax® AXG-828U with HGX B300

8U rackmount server with high GPU density for large-scale AI training and inference.



| AMAX AceleMax® AXG-828U with HGX B300 | |
|---------------------------------------|--|
| CPU | Dual Socket Intel® Xeon® 6700E/6700P series processors |
| GPU | NVIDIA HGX™ B300 8-GPU with NVSwitch |
| Cooling | High-efficiency air cooling |
| System Memory | Up to 32 DDR5 DIMM slots, up to 6400 MT/s |
| Networking | 8× OSFP 800 Gbps InfiniBand ports |
| Storage | Up to 12× 2.5" hot-swap bays, plus 1× M.2 |

AMAX RackScale 32 with NVIDIA HGX™ B300

Air-cooled, rack-scale solution built on the NVIDIA reference design platform, delivering high-performance AI computing with NVIDIA HGX B300 GPUs, high-speed interconnects, and offering up to 8.4TB HBM3e memory per rack.



| AMAX RackScale 32 with NVIDIA HGX™ B300 | |
|---|--|
| CPU | Dual Socket Intel® Xeon® 6700E/6700P series processors |
| GPU | 32x NVIDIA Blackwell Ultra GPUs |
| Cooling | High-efficiency air cooling |
| Architecture | 4× 8U compute servers per rack |
| GPU Memory | Up to 8.4TB total HBM3e GPU memory per rack |
| Networking | NVIDIA NDR 800Gbps InfiniBand switches |
| Storage | High Performance Storage Appliance |
| Total FP4 Tensor Core | 576 PFLOPS |
| Total FP8 Tensor Core | 288 PFLOPS |

Design to Deployment

AMAX delivers systems that are fully engineered, tested, and ready for operation on day one. Our process includes:

- Site survey and infrastructure assessment
- Burn-in, performance, and environmental testing
- GPU benchmarking and validation
- Remote monitoring
- Ongoing support
- Rack & stack deployment

For customers awaiting permanent data center space, AMAX offers HostMax™, an interim hosting service that supports both air-cooled and liquid-cooled systems. HostMax enables newly built solutions to go live immediately after assembly, with secure remote access and dedicated AMAX engineering support.

