AMAX RackScale GPU Solutions for Cloud Service Providers

Built for Multi-Tenant AI and HPC at Scale

AMAX ENGINEERING

AMAX provides RackScale GPU solutions purpose-built for cloud service providers, combining intelligent engineering design with supply chain logistics, sizing, validation, and global deployment support. Our expertise ensures CSPs can scale capacity efficiently while maintaining performance, reliability, and cost control.

Supercharge Your Cloud

AMAX RackScale solutions are built for cloud service providers (CSPs) that demand dense GPU capacity, efficient cooling, and predictable uptime to support multi-tenant customers at scale. AMAX solutions deliver maximum compute per square foot, with secure workload isolation and flexible allocation across tenants. Advanced liquid-cooling sustains high utilization without inflating operating costs, while modular rack designs enable providers to scale capacity quickly as demand grows.

Core Features for CSP Deployments

- **Up to 128 GPUs Per Rack:** Maximize compute density for multi-tenant cloud environments.
- **800G Networking:** High-bandwidth connectivity to support diverse, simultaneous customer workloads.
- **Liquid Cooling Efficiency:** Maintain stable performance at scale while reducing PUE and operating costs.
- **Workload Partitioning:** Hardware-level isolation for secure, predictable multi-tenant performance.

High-Density Liquid-Cooled Rack Designs

AMAX RackScale solutions maximize GPU compute within compact rack footprints, helping CSPs increase capacity without expanding facility space. Liquid cooling sustains high-density performance, while modular designs allow smooth scaling as demand grows. Built-in telemetry supports workload isolation and utilization tracking, ensuring multi-tenant environments stay efficient and predictable.

AMAX | 1565 Reliance Way, Fremont, CA 94539 | 1 (800) 800-6328 | www.amax.com | info@amax.com |

AMAX LiquidMax® RackScale 128

High-density, liquid-cooled 51OU rack for CSPs, delivering maximum GPU throughput, scalable multi-tenant performance, and low PUE efficiency.

LiquidMax® RackScale 128	
GPU	128x NVIDIA HGX™ B300 or AMD Instinct™ MI355X/ MI325X
Cooling	Direct liquid cooling with rack-level CDUs
Power	8 × 33kW shelves
Memory	Up to 36 TB HBM3e GPU memory per rack[1]
Networking	800G InfiniBand or Ethernet options
Storage	High Performance Storage Appliance
Total FP4 Tensor Core	2,304 PFLOPS[1]
Total FP8 Tensor Core	1,152 PFLOPS[1]

[1] Specifications for memory and Tensor Core performance are based on configurations with NVIDIA HGX B300.

AMAX LiquidMax® ALC-B4872U GB300 NVL72 AI POD

48RU liquid-to-air cooled rack powered by NVIDIA Grace Blackwell Ultra Superchips, delivering high-performance AI compute with rapid deployment without requiring facility liquid cooling upgrades.

LiquidMax® ALC-B4872U GB300 NVL72 AI POD	
GPU	NVIDIA GB300 Grace Blackwell Superchip (72 per rack)
Cooling	Direct liquid cooling with rack-integrated CDU and manifolds
Power	6 × 1U 33 kW (6 × 5.5 kW PSUs)
Memory	Up to 21TB of HBM3e GPU memory per rack
Networking	800G InfiniBand or Ethernet options
Storage	High Performance Storage Appliance
Total FP4 Tensor Core	1,400 PFLOPS
Total FP8 Tensor Core	720 PFLOPS

End-to-End Deployment Services

AMAX delivers both liquid-cooled and air-cooled rack-scale solutions engineered, assembled, and validated before shipment. Each rack undergoes a site survey, full burn-in, performance benchmarking, and environmental testing to ensure readiness for multi-tenant AI and HPC workloads in cloud environments. Remote monitoring and ongoing support options help maintain peak performance and uptime. For customers awaiting permanent data center space, HostMax™ provides temporary hosting so systems can go live immediately after build.

Visit www.amax.com/contact to get started today

AMAX | 1565 Reliance Way, Fremont, CA 94539 | 1 (800) 800-6328 | www.amax.com | info@amax.com |