

GigaIO™ SuperNODE™ - MI300X

IMMENSE INFERENCE PERFORMANCE

Breaks traditional server limits by consolidating up to 32 GPUs, simplifying scalability for demanding processing tasks

INNOVATIVE AI MEMORY FABRIC

Optimizes AI applications with a memory-centric infrastructure, aggregating resources for faster, low-latency data transfer

EMPOWERING LARGE MODEL PROCESSING

Vast GPU memory pool handles large model inference, reducing data transfers and boosting performance

SIMPLIFIED AI DEPLOYMENT

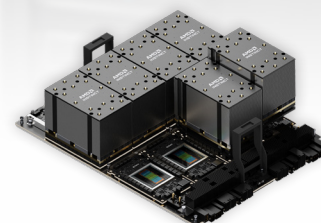
By minimizing the complexity of AI and HPC infrastructure, organizations can quickly get their LLMs and applications up and running, without the usual delays associated with InfiniBand infrastructure setup

The GigaIO SuperNODE seamlessly presents all accelerators to a single server in an easy, scale-up configuration, enabling expansion without infrastructure complexity.

This scale-up architecture minimizes latency, enhances performance, and reduces power consumption by allowing models to efficiently utilize all GPUs as if locally connected, ensuring rapid scaling of AI workloads with minimal operational overhead.

Specifications

Accelerators	32x AMD Instinct™ MI300X OAM GPUs
Data rate (each direction)	512Gb/s (Accelerator-to-Accelerator)
FP64	20.92 PFLOPS
FP8	668.8 PFLOPS
GPU memory	6.14TB HBM3
Failover¹	Primary and secondary servers
CPU cores²	64 AMD EPYC™ “Genoa” 9534
System memory²	3.0 TB
Storage²	2x 15.3TB NVMe-U.2 (30TB total)
Boot drive²	2x 960GB NVMe-M.2
Network²	1x 400G QSFP112-DD 2x 25G/10G SFP28
Rack Units³	42U
Rack Power³	36.5kW
Weight³	1,131 lbs (513.2 kgs)
Cooling	Air cooled, airflow front-to-rear (rack handle to power side)
Environmental	Operating Temperature: 10°C to 35°C (50°F to 95°F)



AMD
INSTINCT
MI300X

READY TO GET STARTED?

Contact a GigaIO authorized representative today.

info@gigaio.com

¹ Dual head node servers configured identically provide failover standby to maintain up-time

² Each server, based on max spec CS511-NA head nodes with 3.0 TB memory, 64 cores, 30 TB storage

³ Does not include available top of the rack Ethernet switches, rack and PDU(s)