DATA SHEET

GigaIO™ SuperNODE™ - d-Matrix Corsair

IMMENSE INFERENCING PERFORMANCE

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

INNOVATIVE AI MEMORY FABRIC

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

EMPOWERING SCALABLE INFERENCING

Delivers unprecedented scale and efficiency for next-generation AI inference workloads

SIMPLIFIED AI DEPLOYMENT

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

READY TO GET STARTED?

Contact a GigaIO authorized representative today.

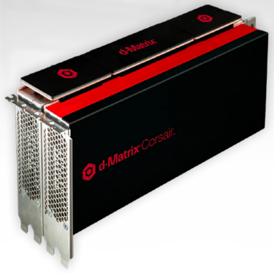
info@gigaio.com

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 d-Matrix Corsair Cards
Data rate (each direction)	512Gb/s (Accelerator-to-Accelerator)
MXINT8	76.8 PFLOPS
MXINT4	307.2 PFLOPS
Accelerator memory	8.19TB DDR5
Failover[1]	Primary and secondary servers
CPU cores[2]	64 AMD EPYC™ "Genoa" 9534
System memory[2]	3.0 TB
Storage[2]	2x 15.3TB NVMe-U.2 (30TB total)
Boot drive[2]	2x 960GB NVMe-M.2
Network[2]	1x 400G QSFP112-DD, 2x 25G/10G SFP28
Rack Units[3]	26U
Rack Power[3]	30.0kW
Weight[3]	711 lbs (322.5 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)



d-Matrix®

[1] Dual head node servers configured indentically 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)