# $GIG \bigwedge IO$

#### DATA SHEET

## GigalO<sup>™</sup> SuperNODE<sup>™</sup> - H200 NVL

#### IMMENSE INFERENCING 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 inferencing, reducing data transfers and boosting performance

## SIMPLIFIED AI

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

#### **READY TO GET STARTED?**

Contact a GigalO authorized representative today.

info@gigaio.com

The GigalO SuperNODE seamlessly presents all accelerators to a single server in an easy, scaleup 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 NVIDIA H200 NVL GPUs                                      |
|----------------------------|---|
| Data rate (each direction) | 512Gb/s (Accelerator-to-Accelerator)                          |
| FP64 Tensor Core           | 1.92 PFLOPS   |
| FP16 Tensor Core           | 53.5 PFLOPS   |
| GPU memory                 | 4.51TB HBM3   |
| Failover <sup>1</sup>      | Primary and secondary servers                                 |
| CPU cores <sup>2</sup>     | 64 AMD EPYC™ "Genoa" 9534                                     |
| System memory <sup>2</sup> | 3.0 TB  |
| Storage <sup>2</sup>       | 2x 15.3TB NVMe-U.2<br>(30TB total)                            |
| Boot drive <sup>2</sup>    | 2x 960GB NVMe-M.2   |
| Network <sup>2</sup>       | 1x 400G QSFP112-DD,<br>2x 25G/10G SFP28                       |
| Rack Units <sup>3</sup>    | 26U   |
| Rack Power <sup>3</sup>    | 28.6kW  |
| Weight <sup>3</sup>        | 665 lbs (301.5 kgs)   |
| Cooling                    | Air cooled, airflow front-to-rear (rack handle to power side) |
| Environmental              | Operating Temperature: 10°C<br>to 35°C (50°F to 95°F)         |
|                            |   |



<sup>1</sup> Dual head node servers configured indentically provide failover standby to maintain up-time
<sup>2</sup> Each server, based on max spec CS511-NA head nodes with 3.0 TB memory, 64 cores, 30 TB storage

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

© 2025 GigalO, all rights reserved. The information contained herein is subject to change without notice. GigalO shall not be liable for technical or editorial errors or omissions contained herein. DS SuperNODE - H200 NVL - v1.01 06112025