Front I/O Air & Liquid-Cooled B200 Systems
Supermicro Front I/O NVIDIA HGX™ B200 Systems, 4U Liquid-Cooled and 8U Air-Cooled

Liquid Cooling

Air Cooling

Accelerate NVIDIA Blackwell Deployment for a Diverse Range of AI Factory Environments

Supermicro's Front I/O NVIDIA HGX B200 systems simplify the deployment, management, and maintenance of air or liquid-cooled AI infrastructure, allowing easy front I/O access, simplifying cabling, improving thermal efficiency and compute density, and reducing operational expenses (OPEX).

Supermicro designed the liquid-cooled 4U Front I/O B200 system as the building block for densely populated AI factories that can reach cluster sizes beyond thousands of system nodes. This system features DLC-2 technology with support for higher inlet temperatures, reduced noise levels, additional cold plate coverage, delivering benefits of up to 40% data center power savings versus air cooling.

Supermicro offers the air-cooled 8U Front I/O system as a streamlined solution for AI factories without liquid-cooling infrastructure, while still retaining enhanced serviceability, increased memory capacity, and effortless scalability.

Liquid Cooled 4U
NVIDIA HGX B200 8-GPU Front I/O DLC-2 System

System Info	SYS-422GS-NBRT-LCC
Overview	4U DLC-2 liquid-cooled system with front I/O NICs, DPUs, storage, and management
CPU	Dual-Socket Intel® Xeon® 6700 Series Processors with P-cores, up to 350W
GPU	NVIDIA HGX B200 8-GPU with 5th Generation NVLink® 1.8TB/s, 1.4TB of HBM3e GPU memory per system
Memory	32 DIMMs with up to 8TB at 5200MT/s or up to 4TB at 6400MT/s DDR5 RDIMM
Storage	8 Hot-swap E1.S NVMe storage drive bays and 2 M.2 NVMe boot drives
Networking	Up to 8 single-port NVIDIA ConnectX®-7 NICs or NVIDIA BlueField®-3 SuperNICs Up to two dual-port NVIDIA BlueField®-3 DPUs
Power	4x Redundant 6600W Titanium Level power supplies
Liquid Cooling with DLC-2	Up to 98% system heat capture by liquid-cooling CPUs, GPUs, DIMMs, PCIe switches, VRMs, power supplies, and more
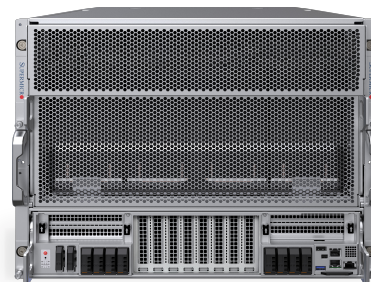
Air-Cooled 8U
NVIDIA HGX B200 8-GPU Front I/O System

System Info	SYS-822GS-NBRT
Overview	8U air-cooled system with front I/O NICs, DPUs, storage, and management
CPU	Dual-Socket Intel® Xeon® 6700 Series Processors with P-cores, up to 350W
GPU	NVIDIA HGX B200 8-GPU with 5th Generation NVLink® 1.8TB/s, 1.4TB of HBM3e GPU memory per system
Memory	32 DIMMs with up to 8TB at 5200MT/s or up to 4TB at 6400MT/s DDR5 RDIMM
Storage	8 Hot-swap E1.S NVMe storage drive bays and 2 M.2 NVMe boot drives
Networking	Up to 8 single-port NVIDIA ConnectX®-7 NICs or NVIDIA BlueField®-3 SuperNICs Up to two dual-port NVIDIA BlueField®-3 DPUs
Power	6 redundant (3+3) 6600W Titanium level power supplies

SUPERMICRO

Liquid-Cooled Rack

Networking
- NVIDIA Quantum-2 400G InfiniBand switches or NVIDIA Spectrum-4 400GbE Ethernet switches dedicated for compute and storage
- Ethernet leaf switches for in-band management
- Out-of-band 1G/10G IPMI switch
- Non-blocking network

Compute
- 8x SYS-422GS-NBRT-LCC, 64x NVIDIA HGX B200 GPUs per rack
- 11.5TB of HBM3e per rack
- Flexible storage options with local or dedicated storage fabric with full NVIDIA GPUDirect RDMA and Storage or RoCE support

Liquid-Cooling
- Supermicro 250kW capacity Cooling Distribution Unit (CDU) with redundant PSU and dual hot-swap pumps
- Vertical Cooling Distribution Manifold (CDM)

Air-Cooled Rack

Networking
- NVIDIA Quantum-2 400G InfiniBand switches or NVIDIA Spectrum-4 400GbE Ethernet switches dedicated for compute and storage
- Ethernet leaf switches for in-band management
- Out-of-band 1G/10G IPMI switch
- Non-blocking network

Compute
- 4x SYS-822GS-NBRT, 32x NVIDIA HGX B200 GPUs per rack
- 5.76TB of HBM3e per rack
- Flexible storage options with local or dedicated storage fabric with full NVIDIA GPUDirect RDMA and Storage or RoCE support

---

Liquid Cooling

32-Node Scalable Unit — Cluster with 256 NVIDIA B200 GPUs in 5 Racks

Overview	Fully integrated liquid-cooled 32-node cluster with 256 NVIDIA B200 GPUs
Compute Fabric Leaf	8x NVIDIA Quantum-2 400G InfiniBand Switch or 8x NVIDIA Spectrum-4 400GbE Ethernet Switch
Compute Fabric Spine	4x NVIDIA Quantum-2 400G InfiniBand Switch or 4x NVIDIA Spectrum-4 400GbE Ethernet Switch
In-band Management Switch	3x NVIDIA Spectrum SN4600 100GbE Ethernet Switch
Out-of-band Management Switch	2x SSE-G3748R-SMIS, 48-port 1Gbps Ethernet ToR management switch 1x SSE-F3548SR, 48-port 10Gbps Ethernet ToR management switch
Rack and PDU	5x 48U x 800mm x 1400mm PDU: 18x 415V 60A/100A 3Ph
Liquid Cooling	4x Supermicro 250kW capacity CDU with redundant PSU and dual hot-swap pumps

*Recommended configuration. Other network switch options and rack dimension and layouts are available.

Air Cooling

32-Node Scalable Unit — Cluster with 256 NVIDIA B200 GPUs in 9 Racks

Overview	Fully integrated air-cooled 32-node cluster with 256 NVIDIA B200 GPUs
Compute Fabric Leaf	8x NVIDIA Quantum-2 400G InfiniBand Switch or 8x NVIDIA Spectrum-4 400GbE Ethernet Switch
Compute Fabric Spine	4x NVIDIA Quantum-2 400G InfiniBand Switch or 4x NVIDIA Spectrum-4 400GbE Ethernet Switch
In-band Management Switch	3x NVIDIA Spectrum SN4600 100GbE Ethernet Switch
Out-of-band Management Switch	2x SSE-G3748R-SMIS, 48-port 1Gbps Ethernet ToR management switch 1x SSE-F3548SR, 48-port 10Gbps Ethernet ToR management switch
Rack	9x 48U 750mm x 1200mm
PDU	34x 208V 60A 3Ph

*Recommended configuration. Other network switch options and rack dimension and layouts are available.

SUPERMICRO