



# NVIDIA Blackwell Architecture Solutions

End-to-End AI Data Center Building Block Solutions



✓ 5,000+ Racks Per Month Capacity with Worldwide Manufacturing

✓ Onsite Deployment Services and Management Software

✓ Unmatched Time-to-Online with Extensive Experience

February 2025



## NVIDIA Blackwell Journey Starts Here



The complete multi-GPU scalable compute units built for trillion parameter AI models, directly available from Supermicro.

### The Most Powerful and Efficient NVIDIA Blackwell Architecture Solutions

In this transformative moment of AI, where the evolving scaling laws continue to push the limits of data center capabilities, our latest NVIDIA Blackwell-powered solutions, developed through close collaboration with NVIDIA, offer unprecedented computational performance, density, and efficiency with the next generation air-cooled and liquid-cooled architecture. With our readily deployable AI Data Center Building Block solutions, Supermicro is your premier partner to start your NVIDIA Blackwell journey, providing sustainable, cutting-edge solutions that accelerate AI innovations.



# End-to-End AI Data Center Building Block Solutions

Making direct liquid-cooled AI infrastructure easy for customers to deploy and maintain, including the facility-side cooling tower.

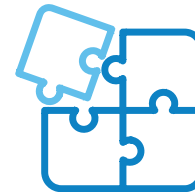
## End-to-End AI Data Center Building Block Solutions Advantage

Select from a broad range of air-cooled and liquid-cooled systems with multiple CPU options, featuring a full data center management software suite, turn-key rack level integration with full networking, cabling, and cluster level L12 validation, global delivery, support, and service.



### Vast Experience

Supermicro's AI Data Center Building Block Solutions power the largest liquid-cooled-AI Data Center deployment in the World.



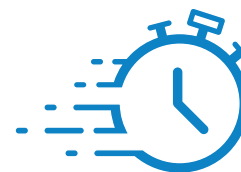
### Flexible Offerings

Air or liquid-cooled, GPU-optimized, multiple system and rack form factors, CPUs, storage, and networking options, optimized for your needs.



### Liquid-Cooling Pioneer

Proven, scalable, and plug-and-play liquid-cooling solutions to sustain the AI revolution. Designed specifically for NVIDIA Blackwell.

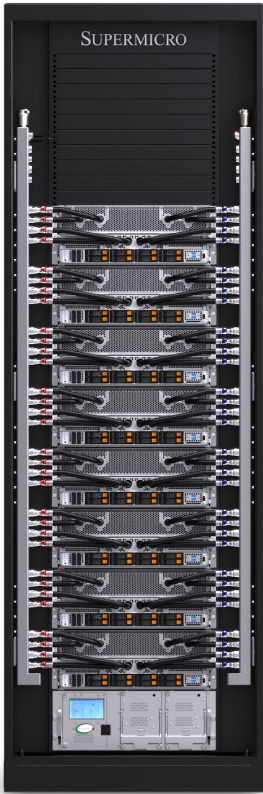


### Fast Time to Online

Accelerated delivery with global capacity, world class deployment expertise, one-site services, to bring your AI to production, fast.



## 4U Liquid-Cooled Rack Configuration



### Networking

- In-band management switch
- Out-of-band IPMI management switch
- Non-blocking network
- Leaf switches in a centralized networking rack

### Compute

- 8x SYS-422GA-NBRT-LCC, AS-4126GS-NBR-LCC, or SYS-421GE-NBRT-LCC per rack
- 8x NVIDIA HGX B200 8-GPU per rack
- 64x NVIDIA B200 Tensor Core GPUs
- 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 Coolant Distribution Unit (CDU) with redundant PSU and dual hot-swap pumps
- Supermicro vertical Coolant Distribution Manifolds (CDM)

## Next-Gen Liquid-Cooled System

The new liquid-cooled 4U NVIDIA HGX B200 8-GPU system features newly developed cold plates and advanced tubing design paired with the new 250kW coolant distribution unit (CDU) more than doubling the cooling capacity of the previous generation in the same 4U form factor. The new architecture further enhances efficiency and serviceability of the predecessor that are designed for NVIDIA HGX H100/H200 8-GPU. Available in 42U, 48U or 52U configuration, the rack scale design with the new vertical coolant distribution manifolds (CDM) means that horizontal manifolds no longer occupy valuable rack units. This enables 8 systems, 64 NVIDIA Blackwell GPUs in a 42U rack and all the way up to 12 systems with 96 NVIDIA GPUs in a 52U rack.



64-GPU Scalable Unit	SRS-48UDLC-4U8N-L1
GPUs	8x NVIDIA HGX B200 8-GPU (64 GPUs)
CPUs	16x Intel® Xeon® or AMD EPYC™ processors
GPU Systems	8x SYS-422GA-NBRT-LCC / AS-4126GS-NBR-LCC / SYS-421GE-NBRT-LCC
NVLink	5th Generation NVIDIA NVLink at 1.8TB/s
Networking*	NVIDIA Quantum-2 InfiniBand 400G NDR or NVIDIA Spectrum-X Ethernet 400Gb/s Ethernet ToR management switches
Rack Dimension*	48U x 800mm x 1400mm
Liquid Cooling Options*	1 in-rack Supermicro 4U 250kW capacity CDU with redundant PSU and dual hot-swap pumps Optional 1.3MW capacity in-row CDU

\*Recommended configuration. Other network switch options and rack dimensions and layouts are available. Login node may be required. NVIDIA Unified Fabric Manager (UFM) node optional.

4U 8-GPU System	SYS-422GA-NBRT-LCC / AS-4126GS-NBR-LCC / SYS-421GE-NBRT-LCC
Overview	4U Liquid-cooled System with NVIDIA HGX B200 8-GPU
CPU	Dual Intel® Xeon® 6900 series processors with P-cores (SYS-422GA-NBRT-LCC) Dual AMD EPYC™ 9005/9004 Series Processors (AS-4126GS-NBR-LCC) Dual 5th/4th Gen Intel® Xeon® Scalable processors (SYS-421GE-NBRT-LCC)
Memory	24 DIMMs, up to DDR5-6400 (SYS-422GA-NBRT-LCC) 24 DIMMs, up to DDR5-6000 (AS-4126GS-NBR-LCC) 32 DIMMs, up to DDR5-5600 (SYS-421GE-NBRT-LCC)
GPU	NVIDIA HGX B200 8-GPU (180GB HBM3e per GPU) 1.8TB/s NVLink GPU-GPU interconnect with NVSwitch
Networking*	8 single-port NVIDIA ConnectX®-7 NICs or NVIDIA BlueField®-3 SuperNICs Up to 400Gbps 2 dual-port NVIDIA BlueField®-3 DPUs
Storage	8 front hot-swap 2.5" NVMe drive bays 2 M.2 NVMe slots
Power Supply	4x 6.6kW redundant Titanium Level power supplies

\*Recommended configuration, other system memory, networking, storage options are available.

# 10U Air-Cooled Rack Configuration



## Networking

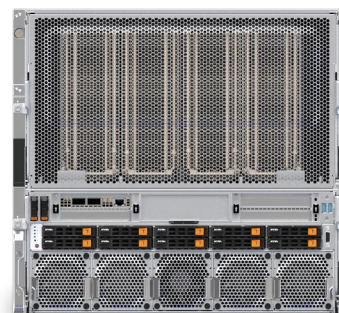
- In-band management switch
- Out-of-band IPMI management switch
- Non-blocking network
- Leaf switches in a centralized networking rack

## Compute

- 4x SYS-A22GA-NBRT, AS -A126GS-TNBR, or SYS-A21GE-NBRT per rack
- 4x NVIDIA HGX B200 8-GPU per rack
- 32x NVIDIA B200 Tensor Core GPUs
- 5.76TB HBM3e per rack
- Flexible storage options with local or dedicated storage fabric with full NVIDIA GPUDirect RDMA and Storage or RoCE support

## Air-Cooled System, Evolved

The new air-cooled 10U NVIDIA HGX B200 system features a redesigned chassis with expanded thermal headroom to accommodate eight 1000W TDP Blackwell GPUs. Up to 4 of the new 10U air-cooled systems can be installed and fully integrated in a rack, the same density as the previous generation, while providing up to 15x inference and 3x training performance. All Supermicro NVIDIA HGX B200 systems are equipped with a 1:1 GPU-to-NIC ratio supporting NVIDIA BlueField<sup>®</sup>-3 or NVIDIA ConnectX<sup>®</sup>-7 for scaling across a high-performance compute fabric.



### 32-GPU Scalable Unit

SRS-48UAC-10U4N-A1

GPUs	4x NVIDIA HGX B200 8-GPU (32 GPUs)
CPUs	8x Intel <sup>®</sup> Xeon <sup>®</sup> or AMD EPYC <sup>™</sup> processors
GPU Systems	4x SYS-A22GA-NBRT / AS -A126GS-TNBR / SYS-A21GE-NBRT
NVLink	5th Generation NVIDIA NVLink at 1.8TB/s
Networking*	NVIDIA Quantum-2 InfiniBand 400G NDR NVIDIA Spectrum-X Ethernet 400Gb/s Ethernet ToR management switch
Rack Dimension*	48U x 750mm x 1295mm

\*Recommended configuration. Other network switch options and rack dimensions and layouts are available. Login node may be required. NVIDIA Unified Fabric Manager (UFM) node optional.

### 10U 8-GPU System

SYS-A22GA-NBRT / AS-A126GS-TNBR /  
SYS-A21GE-NBRT

Overview	10U Air-cooled System with NVIDIA HGX B200 8-GPU
CPU	Dual Intel <sup>®</sup> Xeon <sup>®</sup> 6900 series processors with P-cores (SYS-A22GA-NBRT) Dual AMD EPYC <sup>™</sup> 9005/9004 Series Processors (AS -A126GS-TNBR) Dual 5th/4th Gen Intel <sup>®</sup> Xeon <sup>®</sup> Scalable processors (SYS-A21GE-NBRT)
Memory	24 DIMMs, up to DDR5-6400 (SYS-A22GA-NBRT) 24 DIMMs, up to DDR5-6000 (AS -A126GS-TNBR) 32 DIMMs, up to DDR5-5600 (SYS-A21GE-NBRT)
GPU	NVIDIA HGX B200 8-GPU (180GB HBM3e per GPU) 1.8TB/s NVLink GPU-GPU interconnect with NVSwitch
Networking*	8 single-port NVIDIA ConnectX <sup>®</sup> -7 NICs or NVIDIA BlueField <sup>®</sup> -3 SuperNICs Up to 400Gbps 2 dual-port NVIDIA BlueField <sup>®</sup> -3 DPUs
Storage	10 front hot-swap 2.5" NVMe drive bays 2 M.2 NVMe slots
Power Supply	6x 5250W redundant Titanium Level power supplies

\*Recommended configuration, other system memory, networking, storage options are available.



## NVIDIA GB200 NVL72 Rack Configuration



- Management Networking**
  - In-band management switch
  - Out-of-band management switch
- 10 Compute Trays**
  - 4x NVIDIA Blackwell GPUs per tray
  - 2x NVIDIA Grace CPUs per tray
- Compute Interconnect**
  - 9x NVLink Switches
  - 72 GPUs and 36 CPUs interconnected at 1.8TB/s
- 8 Compute Trays**
  - 4x NVIDIA Blackwell GPUs per tray
  - 2x NVIDIA Grace CPUs per tray
- Liquid-Cooling Options**
  - Supermicro 250kW capacity coolant distribution unit (CDU) with redundant PSU and dual hot-swap pumps
  - 240kW or 180kW capacity Liquid-to-air solution (no facility water required)

### Powered by Supermicro End-to-End Liquid-Cooling Solution

Supermicro NVIDIA GB200 NVL72 SuperCluster features the new advanced in-rack coolant distribution unit (CDU) and custom coldplates designed for the compute trays housing the NVIDIA GB200 Grace™ Blackwell Superchips. The NVIDIA GB200 NVL72 delivers exascale computing capabilities in a single rack with fully integrated Liquid-Cooling. It incorporates 72 NVIDIA Blackwell GPUs and 36 Grace CPUs interconnected by NVIDIA's largest NVLink™ network to date. The NVLink Switch System facilitates 130 terabytes per second (TB/s) of total GPU communications with low latency, enhancing performance for AI and high-performance computing (HPC) workloads.



72-GPU Scalable Unit		SRS-GB200-NVL72-M1
GPUs	72x NVIDIA Blackwell B200 GPUs	
CPUs	36x NVIDIA 72-core Grace Arm Neoverse V2	
Compute Trays	18x 1U ARS-121GL-NB0	
NVLink Switch Trays	9x NVLink Switch, 4-ports per compute tray connecting 72 GPUs to provide 1.8TB/s GPU-to-GPU interconnect	
Power Shelves	8x 1U 33kW (6x 5.5kW PSUs), total power 132kW	
Rack Dimensions (mm)	2236mm x 600 mm x 1068mm	
Liquid Cooling Options	1x in-rack Supermicro 4U 250kW capacity CDU with redundant PSU and dual hot-swap pumps 1.3MW capacity in-row CDU Optional 180kW/240kW capacity liquid-to-air solutions for facilities without cooling tower and water supply	

Compute Tray		ARS-121GL-NB0
Overview	1U Liquid-cooled System with 2x NVIDIA GB200 Grace Blackwell Superchips	
CPU and GPU	2 72-core NVIDIA Grace Arm Neoverse V2 CPUs 4 NVIDIA Blackwell Tensor Core GPUs	
GPU Memory	Up to 384GB HBM3e per Superchip (768GB per tray)	
CPU Memory	Up to 480GB LPDDR5X per Superchip (960GB per tray)	
Networking	4 NVIDIA NVLink Switch ports (rear) 4 single-port NVIDIA ConnectX®-7 NICs (front) Up to 2 NVIDIA BlueField®-3 DPUs (front)	
Storage	Up to 8 E1.S PCIe 5.0 drives	
Power Supply	Shared power through 4+4 rack power shelves	

# AI Data Center End-to-End Liquid-Cooling

## Total Liquid-Cooling Offerings for a Wide Range of AI Data Center Environments



**Coolant Distribution Unit**



**Liquid-to-Air Cooling Rack**  
(no facility water required)

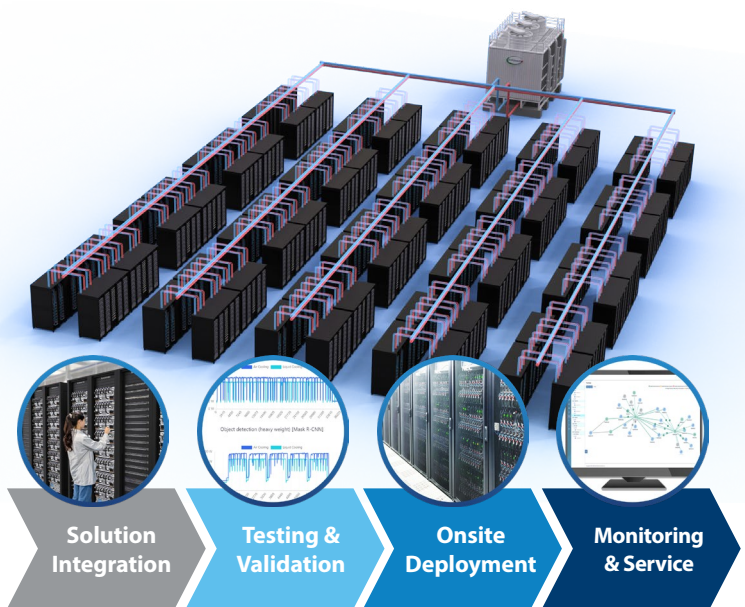


**Cooling Tower**

## End-to-end Data Center Solution and Deployment Services for NVIDIA Blackwell

Supermicro serves as a comprehensive one-stop solution provider with global manufacturing scale, delivering data center-level solution design, liquid-cooling technologies, switching, cabling, a full data center management software suite, L11 and L12 solution validation, onsite installation, and professional support and service. With production facilities across San Jose, Europe, and Asia, Supermicro offers unmatched manufacturing capacity for liquid-cooled or air-cooled rack systems, ensuring timely delivery, reduced total cost of ownership (TCO), and consistent quality.

Supermicro's comprehensive datacenter management platform, SuperCloud Composer software, provides powerful tools to monitor vital information on liquid-cooled systems and racks, coolant distribution units, and cooling towers, including pressure, humidity, pump and valve conditions, and more. SuperCloud Composer's Liquid-Cooling Consult Module (LCCM) optimizes the operational cost and manages the integrity of liquid-cooled data centers.



SUPERMICRO

SUPERMICRO

SUPERMICRO

# Supermicro AI Data Center Building Block Solutions



## Worldwide Headquarters

**Super Micro Computer, Inc.**  
980 Rock Ave.  
San Jose, CA 95131, USA  
Tel: +1-408-503-8000

## EMEA Headquarters

**Super Micro Computer, B.V.**  
Het Sterrenbeeld 12, 5215 ML,  
's-Hertogenbosch, The Netherlands  
Tel: +31-73-640-0390

## APAC Headquarters

**Super Micro Computer, Taiwan Inc.**  
3F, No. 150, Jian 1st Rd., Zhonghe Dist.,  
New Taipei City 235, Taiwan  
Tel: +886-2-8226-3990

[www.supermicro.com](http://www.supermicro.com)

© Super Micro Computer, Inc. Specifications subject to change without notice. All other brands and names are the property of their respective owners.

MKT-0002\_2025-NV-Blackwell\_R14



Please Recycle