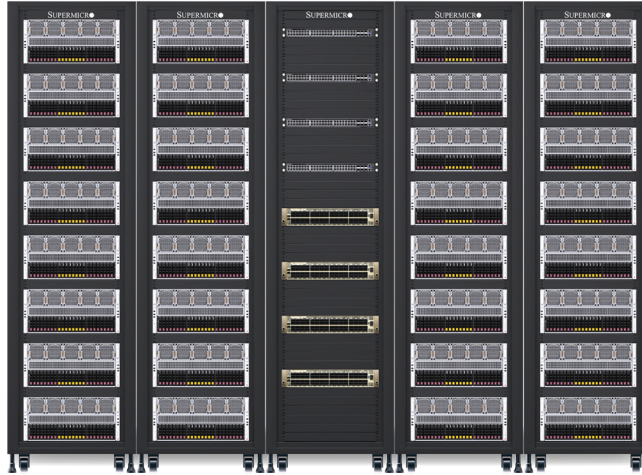


AI Factory Solutions with NVIDIA RTX PRO™ 6000 Blackwell Server Edition

Flexible, proven end-to-end solutions to accelerate at-scale AI factory deployments



Why Choose Supermicro & NVIDIA for AI Factories?

AI factories from Supermicro and NVIDIA are complete, turnkey solutions simplifying the deployment of AI at scale for faster time-to-online and time-to-revenue, with full-stack solutions including compute, software, networking, and storage. Supermicro delivers AI infrastructure optimized for performance and efficiency, with fully-integrated solutions based on NVIDIA Enterprise Reference Architecture Designs and NVIDIA-Certified Systems™ for guaranteed full-stack performance and compatibility. Supermicro's rack-level testing and validation goes beyond industry standards, ensuring quality and seamless plug-and-play deployment for complete AI confidence.

Industry-leading Time-to-Online for the Latest AI Technologies

Supermicro has a proven track record of rapidly bringing new acceleration technologies to market, with a flexible building block approach enabling faster adoption cycles for new NVIDIA GPU platforms. Supermicro can help enterprises bring latest-generation AI infrastructure online faster, accelerating time-to-revenue and maximizing AI-powered competitive advantage. With production capacity in the USA of over 5,000 racks per month, Supermicro is able to build, test, and validate cluster-scale deployments faster, ensuring solutions are delivered ready to begin generating revenue from day one. Additionally, Supermicro's Data Center Building Block Solutions® (DCBBS) can further facilitate the build-out of enterprise AI factories, providing everything needed to develop or refurbish a data center to become an AI factory, reducing lead times and eliminating coordination between multiple vendors.

Flexible, End-to-End AI Solutions Tailored to Your Enterprise

Supermicro offers a broad portfolio of accelerated systems supporting the NVIDIA RTX PRO 6000 Blackwell Server Edition GPU, enabling customers to create AI solutions that deliver maximum performance, efficiency, and value. At scale, Supermicro provides complete AI cluster solutions, backed by deep expertise in networking, topology design, deployment, and cabling. A comprehensive storage portfolio is also available, supporting every stage of the AI data pipeline and integrating the NVIDIA AI Data Platform to simplify data workflows and accelerate innovation. Supermicro AI factory solutions are endorsed by NVIDIA for Infrastructure Configuration, NVIDIA Spectrum™-X Ethernet, and Software Reference Stack and based on the NVIDIA Enterprise Reference Architecture for RTX PRO 6000 Blackwell Server Edition.

Turnkey solutions with Enterprise-grade support from Supermicro & NVIDIA

Working in close cooperation, Supermicro and NVIDIA ensure performance-optimized AI hardware integrates easily into full-stack AI solutions. Supermicro's range of NVIDIA-Certified Systems™ is fully tested and validated for performance, reliability, and compatibility with the NVIDIA software stack (NVIDIA AI Enterprise, NVIDIA Omniverse™, and NVIDIA Run:ai), NVIDIA Spectrum-X Ethernet networking, and forms the building blocks for scaling AI factories seamlessly. As a single-vendor provider, Supermicro supplies everything for a complete AI factory while controlling quality, integrity, and compatibility across the supply chain. Complete L12 system and cluster-level validation before shipment ensure seamless plug-and-play deployment at any scale.

Do More with the NVIDIA RTX PRO 6000 Blackwell Server Edition GPU

The NVIDIA RTX PRO 6000 Blackwell Server Edition is the ultimate universal data center GPU, delivering a powerful combination of AI and visual computing capabilities to accelerate enterprise data center workloads. Equipped with 96GB of ultra-fast GDDR7 memory, the NVIDIA RTX PRO 6000 Blackwell Server Edition GPU provides unparalleled performance and flexibility to accelerate a broad range of use cases - from agentic AI, physical AI, and scientific computing to rendering, 3D graphics, and video. Supermicro AI factories based on the NVIDIA RTX PRO 6000 Blackwell Server Edition GPU offer unmatched performance-per-dollar and flexibility for enterprises looking to deploy multi-workload accelerated infrastructure.



AI Factory Solutions	SRS-48UAC-4N-RTXPRO	SRS-48UAC-8N-RTXPRO	SRS-48UAC-32N-RTXPRO
Nodes per Cluster	4	8	32
GPUs per Node/Cluster	8/32	8/64	8/256
System SKUs	SYS-522GA-NRT, AS-5126GS-TNRT2 (5U) SYS-422GL-NR (4U)		
Networking	NVIDIA Spectrum-X Ethernet	NVIDIA Spectrum-X Ethernet	NVIDIA Spectrum-X Ethernet
Node Pattern (CPU-GPU-NIC-Bandwidth)	2-8-5-200	2-8-5-200	2-8-5-200
Power per Rack (4-node)	36.6kW (SYS-522GA-TNR) 33.3W (SYS-422GL-NR) 34.2kW (AS-5126GS-TNRT2)		
Target Deployment Use Cases	AI inference / Retrieval Augmented Generation (RAG), HPC, visual computing		



NVIDIA-Certified Systems	SYS-522GA-NRT	SYS-422GL-NR	AS-5126GS-TNRT2
Form Factor	5U	4U	5U
GPU	8x NVIDIA RTX PRO 6000 Blackwell Server Edition GPU	8x NVIDIA RTX PRO 6000 Blackwell Server Edition GPU	8x NVIDIA RTX PRO 6000 Blackwell Server Edition GPU
CPU	2x Intel® Xeon® 6972P Processor 96-Core 2.40GHz 480MB Cache (400W)	2x Intel® Xeon® 6972P Processor 96-Core 2.40GHz 480MB Cache (400W)	2x AMD EPYC™ 9575F Processor 64-Core 3.36GHz 256MB Cache (400W)
Memory	24x 64GB DDR5 6400MHz ECC RDIMM	24x 64GB DDR5 6400MHz ECC RDIMM	24x 64GB DDR5 6400MHz ECC RDIMM
Local Node Storage	Micron 7450 PRO 960GB NVMe	Micron 7450 PRO 960GB NVMe	Samsung PM9A3 960GB NVMe PCIe
Networking	1x NVIDIA BlueField®-3 (B3220) 4x NVIDIA BlueField-3 (B3140H)	1x NVIDIA BlueField-3 (B3240) 4x NVIDIA BlueField-3 (B3140H)	1x NVIDIA BlueField-3 (B3220) 4x NVIDIA BlueField-3 (B3140H)
Node Max Power Draw (Full Load)	9.1kW	8.3kW	8.6kW
Node Max Heat (Full Load)	31,214 BTU/h	28,399 BTU/h	29,198 BTU/h