

All-in-One Multi-Model Servers

SuperX's All-in-One Multi-Model Servers ("MMS") is the first enterprise-grade AI infrastructure to support the dynamic collaboration of multiple models, this MMS is centered on being out-of-the-box ready, multi-model fused, and deeply integrated into application scenarios. It provides enterprises with a secure and efficient full-stack AI solutions, with customized specifications for enterprises of various scales, marking a new era where large model applications evolve towards the collaboration of multi-model intelligent agents.

Four Key Functions

Multi-Model Architecture



- Supports seven model types including inference, text-to-image, etc, with presets and collaboration.
- Model Reconfiguration, Invocation, Acceleration, Iteration, and Management
- Cloud-integrated with Local Model Library

RAG Knowledge Base



- Enterprise-grade knowledge base management system, running locally to ensure data security.
- Enables multi-source data fusion and retrieval from documents, databases, and APIs

Intelligent Agent Platform



- Flexible construction and management: Zero-code to build custom agents.

Intelligent Dialogue



- 60+ pre-built smart agents (e.g., document drafting, legal consultation, policy comparison)

Features

Localized Operation



- All data remains on-device, preserving privacy and safeguarding security

Minimal Operation



- Plug-and-play, deploy within minutes

Supports All Platform



- Supports Windows, MacOS, and Linux systems

Open Ecosystem



- Supports integration with external knowledge bases, agents, and MCP services to meet diverse requirements.

Knowledge Empowerment



- Local deployment of user knowledge bases
- Multi-source data search from documents, databases, and APIs

Enhanced Security



- NVIDIA's confidential computing on B200 Series



Workstation Standard

- 1-2x NVIDIA GeForce RTX 5090 Graphics Card, 32GB GDDR7 VRAM
- Intel® Core™ i9-14900KF 24-Core, 32-thread
- 2x 32GB 5200 MT/s DDR5
- 1TB M.2 NVMe SSD
- SuperX Multi-Model AI Platform

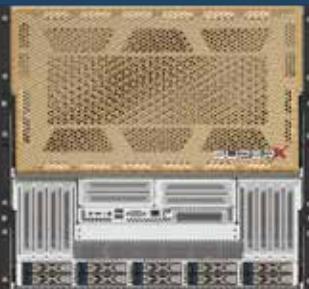
Starting at
US\$50,000



Workstation Ultra

- 4x NVIDIA GeForce RTX PRO 6000 Graphics Card, 32GB GDDR7 VRAM
- AMD Ryzen Threadripper, 24-Core, 48-thread
- 128GB 4800MHz DDR5
- 2TB NVMe SSD
- SuperX Multi-Model AI Platform

Starting at
US\$250,000



B200 Standard

- 8x NVIDIA Blackwell B200, 192GB HBM3E
- 2x intel® Xeon 6767P 64-Core
- 32x 64 GB DDR5 RDIMM
- 8x 3.84TB NVMe SSD
- SuperX Multi-Model AI Platform

Starting at
US\$340,000



B300 Standard

- 8x NVIDIA Blackwell B300, 288GB HBM3E
- 2x intel® Xeon 6767P 64-Core
- 32x 96 GB DDR5 RDIMM
- 8x 3.84TB NVMe SSD
- SuperX Multi-Model AI Platform

Starting at
US\$380,000



NVIDIA GB200/GB300 NVL72 Rack

- 72x NVIDIA Blackwell B200/Ultra GPUs
- 36x NVIDIA Grace™ Arm Neoverse V2 CPUs
- 18x Compute Trays
- 72x ConnectX-7/ConnectX-8 SuperNICs
- 9x 5th Gen NVIDIA NVLink Switches Total NVLink Bandwidth 130TB/s
- 3x NVIDIA SN2201 Management Switches
- 6x 1U 33kW(6x 5.5KW PSUs)
- 48U NVIDIA MGXRack

Starting at
US\$3,200,000/US\$3,800,000