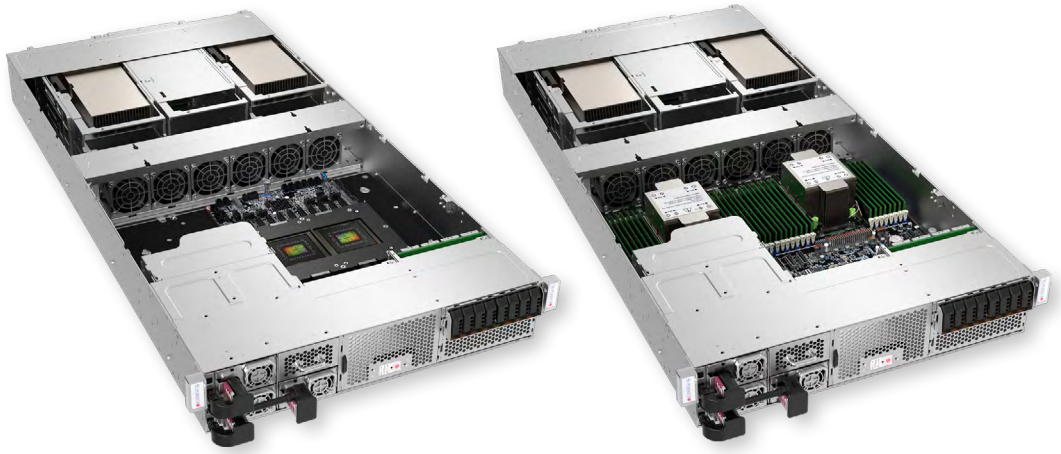
# Supermicro NVIDIA MGX™ Systems

## 1U/2U NVIDIA Grace™ CPU Superchip and x86 Intel® Xeon® Systems



## Ultimate Building Block for Data Center Acceleration

Supermicro NVIDIA MGX™ Systems are designed to standardize AI infrastructure and accelerated computing in compact 1U and 2U form factors while providing ultimate flexibility and expansion ability for current and future GPUs, DPUs, and CPUs. Featuring NVIDIA's new Arm-based Grace™ CPU Superchip as well as x86 processors in the same form factors, these systems support up to 4 double-width GPUs such as the NVIDIA H100 and L40S to enable accelerated computing for hyperscalers, edge, HPC, and cloud.

## Featuring NVIDIA Grace CPU Superchip

### 2x Grace CPUs on one Superchip

The NVIDIA Grace CPU Superchip uses NVLink® Chip-to-Chip (NVLink-C2C) technology to deliver 144 cores (2x 72) and 1TB/s of memory bandwidth with 480GB LPDDR5X on the integrated board. With industry-leading performance-per-watt and memory density, Grace CPUs can enable ground breaking compute densities and energy efficiency in hyperscale and edge data centers excelling at AI, data analytics, digital twins, and hyperscale cloud applications. Supermicro NVIDIA MGX Systems feature up to 2 Grace CPU Superchips, providing up to 288 CPU cores in a system.


## x86-Based Supermicro MGX Systems

Supermicro NVIDIA MGX System featuring 5th/4th Gen Intel® Xeon® Scalable Processors provides the benefits of this modern system architecture while supporting applications that run on x86. Available in 2U with dual Xeon processors with support for up to 4 double-width GPUs and 32 DIMM slots supporting DDR5 memory.

## MGX: A Modern System Architecture

### Optimized for thermals, compatibility, and flexibility.

MGX is a modular and building block platform with support for the leading GPUs, CPUs, and DPUs of today and the future. In addition to enabling GPUs with 400W+ TDP, Supermicro MGX systems also support both air-cooling and liquid-cooling. The modular bays on both sides facilitate constructing finely-tuned systems and can host up to 7 PCIe devices in 2U supporting 4 double-width PCIe GPUs and 3 additional slots for I/O and high-speed networking through NVIDIA BlueField®-3 and ConnectX®-7.

## Cooling + Efficiency + Power Delivery

### Increased Operations-Per-Second. Decreased OPEX.

Due to its mechanical design and component selection, Supermicro MGX Systems optimize cooling, efficiency, and power delivery without sacrifice. Air-cooled systems are designed to optimize cooling in hot/cold aisle configurations. Up to 3x 2000W Redundant Titanium Level power supplies deliver ample power to handle the power requirements of dual CPUs and up to 4 GPUs.

# Accelerate NVIDIA MGX Systems

## 1U/2U NVIDIA Grace™ CPU Superchip and x86 Intel® Xeon® Systems

Supermicro NVIDIA MGX™ 1U/2U Systems with Grace™ CPU Superchip and x86 CPUs are fully optimized to support up to 4 GPUs via PCIe without sacrificing I/O networking, or thermals. The ultimate building block architecture allows you to tailor these systems optimized for a variety of accelerated workloads and fields, including AI training and inference, HPC, data analytics, visualization / NVIDIA Omniverse™, and hyperscale cloud applications.


| | ARS-121L-DNR | ARS-221GL-NR | SYS-221GE-NR |
|---|---|---|---|
| Form Factor | Dual node 1U system with single NVIDIA Grace CPU Superchip per node | 2U system with single NVIDIA Grace CPU Superchip | 2U system with dual x86 processors |
| CPU | 144-core Grace Arm Neoverse V2 CPU in a single chip per node (total of 288 cores in one system) | 144-core Grace Arm Neoverse V2 CPU in a single chip | 5th/4th Gen Intel® Xeon® Scalable processors<br>(Up to 56-core per socket) |
| GPU | Please contact our sales for possible configurations | Up to 4 double-width GPUs including NVIDIA H100 PCIe, H100 NVL, L40S | Up to 4 double-width GPUs including NVIDIA H100 PCIe, H100 NVL, L40S |
| Memory | Up to 480GB of integrated LPDDR5X memory with ECC and up to 1TB/s of bandwidth per node | Up to 480GB of integrated LPDDR5X memory with ECC and up to 1TB/s of bandwidth | Up to 2TB, 32x DIMM slots, ECC DDR5-4800 |
| Drive | 4x Hot-swap E1.S drives and 2x M.2 NVMe drives per node | 8x Hot-swap E1.S drives and 2x M.2 NVMe drives | 8x Hot-swap E1.S drives and 2x M.2 NVMe drives |
| Networking | 2x PCIe 5.0 x16 slots per node supporting NVIDIA BlueField-3 or ConnectX-7 (e.g., 1 GPU and 1 BlueField-3)" | 3x PCIe 5.0 x16 slots supporting NVIDIA BlueField-3 or ConnectX-7 | 3x PCIe 5.0 x16 slots supporting NVIDIA BlueField-3 or ConnectX-7 |
| Interconnect | NVLink-C2C with 900GB/s for CPU-CPU interconnect (within node) | NVLink-C2C with 900GB/s for CPU-CPU interconnect, optional NVLink Bridge GPU-GPU interconnect supported (e.g., H100 NVL) | Optional NVLink Bridge GPU-GPU interconnect supported (e.g., H100 NVL) |
| Cooling | Air-cooling | Air-cooling | Air-cooling |
| Power | 2x 2700W Redundant Titanium Level power supplies | 3x 2000W Redundant Titanium Level power supplies | 3x 2000W Redundant Titanium Level power supplies |

Go to www.supermicro.com/mgx or scan the QR code to learn more.