# Supermicro Boosts Performance for HPC and AI Applications with Optimized Servers Featuring New NVIDIA A100 80GB PCIe GPUs

*High Performance 200G Networking Accelerates HPC and AI Applications*

**ISC Digital, June 28, 2021 — Super Micro Computer, Inc. (Nasdaq: SMCI)**, a global leader in enterprise computing, storage, networking, and green computing technology, announces a wide range of Supermicro servers will support the latest NVIDIA technologies. Both 3rd Gen Intel Xeon Scalable processors or 3rd Gen AMD EPYC™ processor-based servers will be available that incorporate the new NVIDIA® A100 80GB PCIe Tensor Core GPUs. With these new technologies, Supermicro systems will speed up HPC simulations and analytics applications. Moreover, AI applications, including training, inference, and recommendation engines, will run faster than previous generations of systems.

Supermicro servers, designed with a Building Block Solutions® approach, give Supermicro a first to market advantage when incorporating new technologies. State-of-the-art accelerated computing options can be quickly integrated and tested, bringing faster and more powerful systems to demanding customers quicker than other vendors.

"Supermicro has a long history of bringing systems supporting the latest NVIDIA products to market faster than other server manufacturers," said Charles Liang, president, and CEO, Supermicro. "The new products from NVIDIA will give our customers the ability to solve complex HPC and AI challenges much faster than before, decreasing their TCO. We continually bring application-optimized servers with the latest technologies available, such as our GPU and SuperBlade® systems, to help accelerate customer's HPC and AI workloads."


The new NVIDIA A100 PCIe 80GB GPU will enable faster execution of AI and HPC applications, as bigger models can be stored in the GPU memory. In addition, future systems will include networking between high-performance servers with current and future InfiniBand capabilities.

The GPU SuperServer SYS-420GP-TNR with dual 3rd Gen Intel Xeon Scalable processors supports up to ten single or double width GPUs, including the new NVIDIA A100 80GB PCIe GPUs. Also, the AS -4124GS-TNR server contains dual AMD EPYC 7003 processors with up to eight single or double width GPUs, with the new NVIDIA A100 80GB PCIe GPUs. These new 80GB GPUs from NVIDIA will enable more extensive data sets to be held in memory, accelerating computationally-intensive HPC and AI applications.

"Supercomputing has the potential to transform entire industries and help solve some of the toughest problems they face," said Dion Harris, lead technical product marketing manager, accelerated computing at NVIDIA. "The extreme performance of the NVIDIA A100 80GB PCIe Tensor Core GPUs and NVIDIA NDR InfiniBand supported by Supermicro servers provides researchers with the unparalleled acceleration required to pursue the toughest industrial HPC challenges in the world."

Multiple NVIDIA A100 80GB PCIe devices can be installed in a range of Supermicro systems, allowing for flexible AI computing. By implementing NVIDIA's Multi-Instance GPU (MIG) technology, organizations can guarantee a high quality of service for single applications and as a service for multiple users concurrently. In addition, with the larger memory available, data will not have to be moved between CPU memory and GPU memory, allowing for higher GPU utilization.

Leveraging NVIDIA NDR 400Gb/s InfiniBand technology with In-Networking Computing, networking tasks are offloaded, enabling a dramatic leap in performance is enabled, which is required for HPC, AI, and hyperscale cloud infrastructures. The new networking performance will allow applications designed to run on multiple systems, such as the Supermicro SuperBlade and BigTwin® servers, to produce results faster, enabling new insights into complex problems. Integrated systems combining industry-leading cards, cables, and switches on a complex HPC application will soon be able to share data and communicate twice as fast as before.

Customers worldwide use Supermicro systems powered by NVIDIA GPUs to tackle some of the world's most challenging problems. For example, SZTAKI, the Institute for Computer Science and Control, uses NVIDIA HGX™ A100-powered servers to perform AI research in medical processing, machine learning, and language processing.

Learn more about how Supermicro can accelerate AI and HPC applications by visiting the Supermicro booth at https://www.supermicro.com/en/solutions/high-performance-computing

**About Super Micro Computer, Inc.**

Supermicro (SMCI), the leading innovator in high-performance, high-efficiency server technology, is a premier provider of advanced Server Building Block Solutions® for Enterprise Data Center, Cloud Computing, Artificial Intelligence, and Edge Computing Systems worldwide. Supermicro is committed to protecting the environment through its "We Keep IT Green®" initiative and provides customers with the most energy-efficient, environmentally-friendly solutions available on the market.

Media Contact:
Greg Kaufman
Super Micro Computer, Inc.
PR@supermicro.com