



# UNLEASH THE FUTURE OF INNOVATION WITH HPC & AI

*Transforming businesses with a new class of Supermicro and NVIDIA high-performance solutions*

## TABLE OF CONTENTS

Executive Summary .....	1
Unlocking Business Insights .....	1
The New Wave of HPC and AI .....	2
Success Built on Unparalleled Computing .....	2
World-Class Solutions for HPC and AI .....	3
Conclusion .....	4

## SUPERMICRO

Supermicro (Nasdaq: SMCI), the leading innovator in high-performance, high-efficiency server and storage 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.

## Executive Summary

Data is the driving force for success in the global marketplace. Data volumes are erupting in size and complexity as organizations work to collect, analyze, and derive intelligence from a growing number of sources and devices. These workloads are critical to powering applications that translate insight into business value.

High-performance computing (HPC) and [artificial intelligence \(AI\)](#) are changing the way businesses operate, allowing them to use their data to extract deeper insights, enhance processes, and realize better outcomes. However, executing these tasks is a major concern for today’s organizations. Data analytics, HPC, AI, machine learning, and deep learning workloads require massive levels of power and flexibility, which rapidly exceed the capacity of traditional infrastructure. In order to succeed, businesses must invest in a new breed of technology that can deliver revolutionary performance and breakneck speeds. Cutting-edge solutions from Supermicro and NVIDIA are enabling customers to transform and capitalize on HPC and AI innovation.

## Unlocking Business Insights

The rise of HPC and AI is triggering a fundamental shift in numerous industries. Businesses utilize these capabilities to drive improvements across their operations by boosting productivity and efficiency, as well as accelerating insights to solve their most significant challenges. HPC and AI are the foundation for a variety of applications such as modeling and simulation, genome sequencing, real-time and predictive analytics, autonomous driving, and climate monitoring.

As HPC and AI become increasingly intertwined, businesses will continue to leverage their data in new ways.

Legacy infrastructure is the biggest roadblock to data-intensive applications. To overcome this pain point, businesses adopt the latest technological advances to satisfy escalating demands on compute. Innovation is crucial to unlocking future insights, and the ideal solutions will equip businesses with extreme capacity, agility, and flexibility to optimize any workload.

## The New Wave of HPC and AI

As industries recognize the need for advanced technologies to support widespread HPC and AI usage, the next generation of accelerated computing has emerged, offering unprecedented performance. Purpose-built solutions for HPC and AI are redefining the modern enterprise. Backed by the massively parallel processing power of GPUs, businesses can enhance how they work, collaborate, and solve problems:

- **Faster Data Movement** – Greater bandwidth to move massive volumes of data.
- **Workload Optimization** – Adapt quickly as workloads grow in size, scope, and complexity.
- **Performance at Scale** – High throughput and low latency to ensure peak levels of efficiency.
- **Technology Standardization** – Centralized frameworks to simplify IT and increase cost savings.

As the requirements for HPC and AI continue to evolve, Supermicro is committed to helping businesses transform their operating environments. Our [accelerated computing solutions](#) provide the right balance of density and speed to manage diverse workloads while reducing time-to-insight. These groundbreaking platforms leverage the [third generation NVIDIA® Tensor Core GPUs](#) to enable maximum acceleration at every scale, so customers can harness the full potential of HPC and AI.

## Success Built on Unparalleled Computing

Supermicro has extended its leadership in the GPU server market, with the broadest portfolio of high-performance technologies to keep pace with the increasing demand of HPC and AI applications. From practical and efficient clusters to accelerated computing, our comprehensive offerings empower businesses to build the ideal solution for their needs.

At Supermicro, we deliver significant value with an extensive lineup of highly configurable and sustainable technologies, featuring [fundamental building blocks](#) that help customers succeed.

Supermicro is the foremost developer of new solutions, guaranteeing that customers will benefit competitively from the latest technologies and capabilities through our first-to-market advantage. Supermicro solutions are produced in the U.S., and with total control of manufacturing, we prioritize customer requirements and accelerate time-to-market.

Our customers benefit from customizable solutions that accommodate any budget. In conjunction with Supermicro's [Green IT innovation](#), our solutions dramatically extend sustainability for unbeatable price performance. Supermicro offers the best energy efficiency on the market to fuel HPC and AI platforms of all sizes. Our architectures optimize power consumption, cooling, shared resources, and refresh cycles to ensure top performance per dollar. Liquid cooling options combined with Titanium Level power supplies offer an [efficiency rating of 96%](#) and [reduce TCO by as much as 40%–50%](#).

Now, businesses can operate with confidence, leveraging solutions designed for high configurability, reliability, and flexibility to fit their needs. By integrating compute power, NVIDIA GPUs, optimal storage, and networking options, our robust platforms can reach extremely high rack density with minimal facility footprint. Customers can implement tight configurations with unparalleled I/O capacity to maintain operational consistency for HPC and AI. This infrastructure makes it easy for businesses to carry out today's workloads and adapt for tomorrow's challenges, regardless of their requirements.

## World-Class Solutions for HPC and AI

Supermicro's portfolio enables next-generation performance in any operating environment. We are committed to developing total computing innovations that are customer-centric, providing a one-stop-shop where they can select and deploy a turnkey platform. Now, Supermicro offers an array of [GPU accelerated computing solutions](#) to help customers meet and exceed even the most daunting challenges of HPC and AI.

Supermicro's adaptable, fastest-to-market servers powered by unmatched [NVIDIA® A100® Tensor Core GPUs](#) allow businesses to tackle diverse tasks with ease—from running AI inference on developed models, to HPC, to AI training requests. [Fully integrated solutions](#) deliver a significant performance boost across Supermicro's extensive of 1U, 2U, 4U, and 10U multi-GPU servers, providing leading compute capacity and agility for data analytics, HPC, AI, machine learning, and deep learning.

Servers with NVIDIA A100 power applications from edge to cloud, creating the world's most powerful accelerated platforms for AI featured in the [Supermicro NVIDIA HGX A100 8-GPU Platform](#) and [Supermicro NVIDIA HGX A100 4-GPU Platform](#). Additionally, our [newly unveiled NVIDIA A100 PCI-E form factor](#) offers high versatility across multiple configurations, including the [Supermicro NVIDIA A100 PCI-E AMD Platform](#), [Supermicro NVIDIA A100 Ultra AMD Platform](#), and [Supermicro NVIDIA A100 Blade AMD Platform](#) to expedite data-heavy workloads.

Our offerings utilize [NVIDIA Mellanox end-to-end Ethernet and InfiniBand interconnected solutions](#) to run complex workloads with outstanding levels of speed and functionality in data centers and the cloud. NVIDIA Mellanox network technologies are the best choice to connect the world's top HPC and AI accelerated computing platforms. The addition of [NVIDIA® NVLink™ and NVIDIA NVSwitch™](#) enhances interconnection consistency by providing higher bandwidth, more links, and improved scalability for multi-GPU server configurations to ensure seamless communication between components in addition to the fastest data speed, lowest latency, and resiliency.



2U w/ NVIDIA HGX A100 4-GPU



4U w/ NVIDIA HGX A100 8-GPU



4U w/ 8x NVIDIA A100 PCI-E

## Conclusion

The expansion of HPC and AI poses new and rising challenges for today's businesses. Savvy organizations

are empowering their operating environments with GPU accelerated computing to tap into their data and extract game-changing insights.

Supermicro and NVIDIA are preparing our customers for success, offering the right technologies to make the leap. We prioritize the advancement of HPC and AI by continually testing, validating, and integrating advanced hardware featuring optimized software components to support a rising number of use cases. Backed by an extensive partner ecosystem, Supermicro's unrivaled computing solutions are expertly engineered to achieve faster intelligence, superior outcomes, and proven sustainability for ongoing innovation.

This is your time to transform. [Visit Supermicro online](#) to learn how you can harness the full power HPC and AI.



**2U Ultra w/ 2x NVIDIA A100 GPU**



**8U 20 Node Blade System**