



Supermicro New NVIDIA HGX-2 based SuperServer is World's Most Powerful Cloud Server Platform for AI and HPC

September 13, 2018

Designed for the Next-Generation of AI, New HGX-2 System with 16 Tesla V100 GPUs and NVSwitch leverages over 80,000 Cuda Cores to deliver unmatched performance for deep learning and compute workloads

TOKYO, Sept. 13, 2018 /PRNewswire/ -- **Super Micro Computer, Inc. (NASDAQ: SMCI)**, a global leader in enterprise computing, storage, networking solutions and green computing technology, today announced that the company's upcoming NVIDIA® HGX-2 cloud server platform will be the world's most powerful system for artificial intelligence (AI) and high-performance computing (HPC) capable of performing at 2 PetaFLOPS.



**World's Most Powerful
AI Training Machine**
Based on NVIDIA HGX-2 Platform



Supermicro is hosting a Platinum Sponsor Booth at the GPU Technology Conference (GTC) Japan 2018 in Tokyo on September 13-14.

"Supermicro's new SuperServer based on the HGX-2 platform will deliver more than double the performance of current systems, which will help enterprises address the rapidly expanding size of AI models that sometimes require weeks to train," said Charles Liang, president and CEO of Supermicro. "Our new HGX-2 system will enable efficient training of complex models. It combines sixteen Tesla V100 32GB SXM3 GPUs connected via NVLink and NVSwitch to work as a unified 2 PetaFlop accelerator with half a terabyte of aggregate GPU memory to deliver unmatched compute power."

From natural speech by computers to autonomous vehicles, rapid progress in AI has transformed entire industries. To enable these capabilities, AI models are exploding in size. HPC applications are similarly growing in complexity as they unlock new scientific insights. Supermicro's HGX-2 based SuperServer (SYS-9029GP-TNVRT) will provide a superset design for datacenters accelerating AI and HPC in the cloud. With fine-tuned optimizations, this SuperServer will deliver the highest compute performance and memory for rapid model training.

Supermicro GPU systems also support the ultra-efficient Tesla T4 that is designed to accelerate inference workloads in any scale-out server. The hardware accelerated transcode engine in Tesla T4 delivers multiple HD video streams in real-time and allows integrating deep learning into the video transcoding pipeline to enable a new class of smart video applications. As deep learning shapes our world like no other computing model in history, deeper and more complex neural networks are trained on exponentially larger volumes of data. To achieve responsiveness, these models are deployed on powerful Supermicro GPU servers to deliver maximum throughput for inference workloads.

With the convergence of big data analytics and machine learning, the latest NVIDIA GPU architectures, and improved machine learning algorithms, deep learning applications require the processing power of multiple GPUs that must communicate efficiently and effectively to expand the GPU network. Supermicro's single-root GPU system allows multiple NVIDIA GPUs to communicate efficiently to minimize latency and maximize throughput as measured by the NCCL P2PBandwidthTest.

For comprehensive information on Supermicro NVIDIA GPU system product lines, please go to <https://www.supermicro.com/products/nfo/gpu.cfm>.

Follow Supermicro on [Facebook](#) and [Twitter](#) to receive the latest news and announcements.

About Super Micro Computer, Inc. (NASDAQ: SMCI)

Supermicro® (NASDAQ: SMCI), the leading innovator in high-performance, high-efficiency server technology is a premier provider of advanced Server Building Block Solutions® for Data Center, Cloud Computing, Enterprise IT, Hadoop/Big Data, HPC and Embedded 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.

Supermicro, Server Building Block Solutions, and We Keep IT Green are trademarks and/or registered trademarks of Super Micro Computer, Inc.

All other brands, names and trademarks are the property of their respective owners.

SMCI-F

 View original content to download multimedia: <http://www.prnewswire.com/news-releases/supermicro-new-nvidia-hgx-2-based-superserver-is-worlds-most-powerful-cloud-server-platform-for-ai-and-hpc-300712489.html>

SOURCE Super Micro Computer, Inc.

Michael Kalodrich, Super Micro Computer, Inc., michaelk@supermicro.com