TABLE OF CONTENTS

SOLUTION BRIEF

POWER-ON AI

Supermicro NGC-Ready Systems

EXECUTIVE SUMMARY

AI is helping to solve some of the world's most complex problems. Solving these enormous challenges require the computation of large amounts of data and highly optimized AI models running at scale. NVIDIA GPU Cloud (NGC) is the GPU-accelerated software hub for optimized AI and HPC. Supermicro's NGC-Ready systems make it easy and efficient to run large workloads with a complete end-to-end NVIDIA Tensor Core GPU-accelerated hardware and software stack:

- Immediate AI software development and deployment when the systems are powered on. Operating System, CUDA, CUDA-X, NVIDIA drivers, container infrastructure are preloaded. Full support is available.
- Access to the latest, cutting-edge deep learning AI models. Using pre-trained models, new AI systems can be constructed quickly with additional training.
- The NGC containers run anywhere, whether they are on the Supermicro systems in the data center, in edge micro datacenters, on edge servers, and in the cloud if Cloud Bursting is needed. Kubernetes can orchestrate the containers in an extended set of systems to scale the processing for large enterprises.

Super Micro Computer, Inc.
980 Rock Avenue
San Jose, CA 95131 USA
www.supermicro.com

SUPERMICRO VALIDATED NGC-READY SYSTEMS

- Performance-Validated: "Out-of-the-box" systems accelerate time to solution
- Purpose Built for AI: Choose the right systems for the appropriate AI workload
- Enterprise-Grade Support: Resolve issues during deployment to ensure minimal disruption

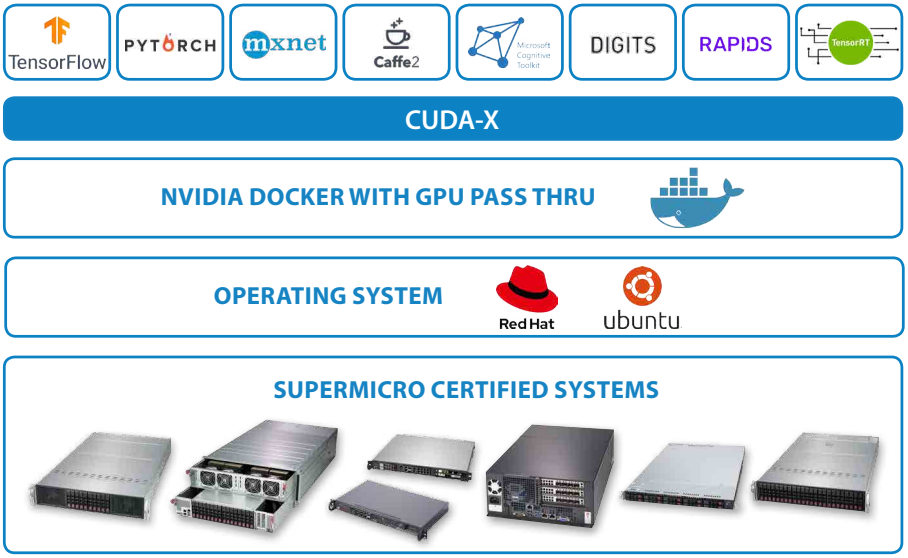

Figure 2. *NGC Ready Systems hardware and software offering.*

CUTTING-EDGE AI

The Supermicro NGC-Ready Systems run any of the NGC software, which is updated monthly with the latest deep learning models across multiple frameworks, including Tensorflow, PyTorch, MXNet. The models come pre-trained, allowing for faster training on new data. Some example models are available as shown in figure 2. NGC support service option removes road blocks in the development

MACHINE LEARNING TOOLS	RAPIDS, DIGITS, TensorRT	Tensorflow	PyTorch
RECOMMENDATION ENGINE		VAE, NCF, BERT	NCF
OBJECT RECOGNITION		SSD	Mask R-CNN
IMAGE RECOGNITION		ResNet50	
VIDEO PROCESSING	DeepStream		
TEXT, TRANSLATION		BERT, GNMTv2	
SPEECH			NeMO ASR, Jasper
MEDICAL	CLARA	V-Net, Unet	

Figure 2. *Example NGC AI models and frameworks available from ngc.nvidia.com.*

SUPERMICRO VALIDATED NVIDIA GPU CLOUD (NGC) SERVERS		
SYSTEM	**CONFIGURATION**	**LOCATION**
SYS-5019D-FN8TP with NVIDIA T4 GPU	Xeon-D, Up to 512GB memory, 1 x PCIe x8 slot for GPU, 1 to 4 internal drives. 9.8" depth	Edge
SYS-1019D-FHN13TP with NVIDIA T4 GPUs	Xeon-D, max 512GB memory 2x PCI-E 3.0 x16 slots for GPU and I/O, 2 SATA, 15" depth	Edge
SYS-1019P-FHN2T with NVIDIA T4 GPUs	Single Xeon Scalable Gen 2, max 1.5TB memory 2x PCI-E 3.0 x16 slots for GPU, 2 SATA, 15" depth	Edge
SYS-1019P-WTR with NVIDIA T4 GPUs	Single Xeon Scalable Gen 2, max 1.5TB memory 2x PCI-E 3.0 x16 slots for GPU, 1 PCI-E 3.0 x8 for I/O, 10 SAS/SATA or 2 NVMe	Edge
SYS-2029GP-TR with NVIDIA V100, T4 GPUs	Dual Xeon Scalable Gen 2, max 4TB memory 6x PCI-E 3.0 x16 slots for GPU and I/O, 8 SAS/SATA or 2 NVMe	Edge
SYS-5039MD18-H8TNR with NVIDIA T4 GPUs	8 Modules in 3U. Each Module has Xeon-D, max 512GB memory 1 PCIe x16 slot for GPU and I/O, 2 SATA with optional NVMe, 23" depth	Micro Data Center, Data Center
SYS-1029U-TRT with NVIDIA T4 GPUs	Dual Xeon Scalable Gen 2, max 6TB memory 2x PCI-E 3.0 x16 slots for GPU, 2 PCI-E 3.0 x8 for I/O, 10 SAS/SATA or 2 NVMe	Micro Data Center, Data Center
SYS-2029GP-TR with NVIDIA V100 GPUs	Dual Xeon Scalable Gen 2, max 4TB memory 6 PCIe x16 slots for GPU and I/O, 8 SAS/SATA or 2 NVMe	Micro Data Center, Data Center
SYS-4029GP-TRT2 with NVIDIA V100 GPUs	Dual Xeon Scalable Gen 2, max 6TB memory 11x PCI-E 3.0 x16 slots for GPU and I/O, 16 SAS/SATA or 8 NVMe	Data Center
SYS-4029GP-TVRT with NVIDIA V100 SXM2 GPUs	Dual Xeon Scalable Gen 2, max 3TB memory 6x PCI-E 3.0 x16 slots for I/O, Single Root, 8 SAS/SATA/NVMe	Data Center

Figure 3. *Supermicro NGC Ready Systems. Processor, memory, disk, network could be adjusted to reflect customer needs. Please consult with your Supermicro representative to build larger systems.*

FULL SUPPORT

The NGC-Ready Systems can be configured to come with pre-installed operating systems and NGC support software. Full support for the hardware system, the Red Hat or Ubuntu operating system, and the NGC software is available.

ABOUT SUPER MICRO COMPUTER, INC.

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.

www.supermicro.com

Printed in USA ♻ Please Recycle 14_Power-on-AI_2020_01-1