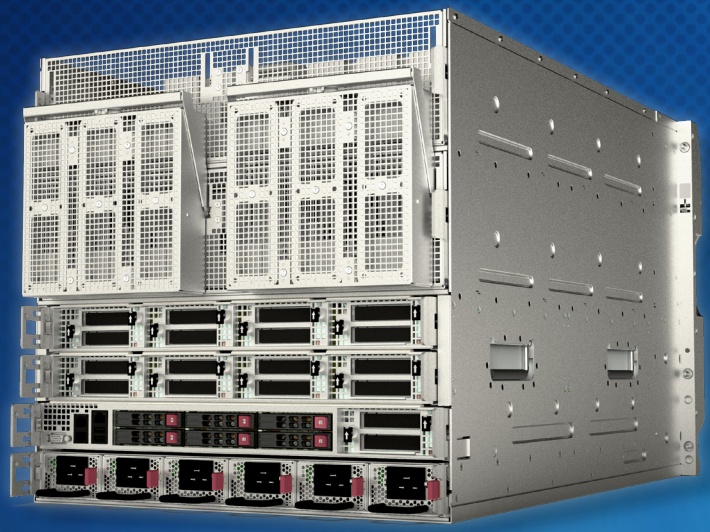
SuperServers Optimized For NVIDIA® Tesla® GPUs

With the Fastest GPU to GPU Interconnections



Most Powerful Solution for *Deep Learning Training*

- Supermicro NVIDIA® **HGX-2** Based Platform
- **16 V100 SXM3** GPUs (512GB GPU Memory)
- NVLink™ and NVSwitch™
- 16 NICs for GPUDirect RDMA
- 16 Hot-swap NVMe Drive Bays
- Fully configurable to order

SYS-9029GP-TNVRT

SYS-1029GQ-TVRT

4 NVIDIA® Tesla® V100 GPUs
with NVIDIA® NVLink™

SYS-4029GP-TVRT

8 NVIDIA® Tesla® V100 GPUs
with NVIDIA® NVLink™

www.supermicro.com/GPU

September 2018

Maximum Acceleration for AI/DL Training Workloads

PERFORMANCE: Highest Parallel peak performance with NVIDIA Tesla V100 GPUs
THROUGHPUT: Best in class GPU-to-GPU bandwidth with a maximum speed of 300GB/s
SCALABILITY: Designed for direct interconections between multiple GPU nodes
FLEXIBILITY: PCI-E 3.0 x16 for low latency I/O expansion capacity & GPU Direct RDMA support
DESIGN: Optimized GPU cooling for highest sustained parallel computing performance
EFFICIENCY: Redundant Titanium Level power supplies & intelligent cooling control

Model	SYS-1029GQ-TVRT	SYS-4029GP-TVRT
CPU Support	• Dual Intel® Xeon® Scalable processors with 3 UPI up to 10.4GT/s • Supports up to 205W TDP CPU	• Dual Intel® Xeon® Scalable processors with 3 UPI up to 10.4GT/s • Supports up to 205W TDP CPU
GPU Support	• 4 NVIDIA Tesla V100 GPUs • NVIDIA® NVLink™ GPU Interconnect up to 300GB/s • Optimized for GPUDirect RDMA	• 8 NVIDIA® Tesla® V100 GPUs • NVIDIA® NVLink™ GPU Interconnect up to 300GB/s • Optimized for GPUDirect RDMA • Independant CPU and GPU thermal zones
Serverboard	SUPER® X11DGQ	SUPER® X11DGO-T
Chipset	Intel® C621	Intel® C621
Memory Support	Up to 1.5TB DDR4-2666MHz in 12 DIMM slots	Up to 3TB DDR4-2666MHz in 24 DIMM slots
Storage Controller	Intel® PCH for 4 SATA3 (6Gbps) ports	Intel® PCH for 8 SATA3 (6Gbps) ports
Drive Bays	• 2 x 2.5" hot-swap SAS/SATA/NVME drive bay • 2x 2.5" Internal Fixed drive bay	• 16x 2.5" hot-swap SATA3 drive bays • Supports up to 8 *NVMe* SSDs
Expansion Slots	• 2 PCI-E 3.0 x16 (FHHL/LP) slots (From PLX for GPUDirect RDMA) • 2 PCI-E 3.0 x16 (FHHL/LP) slots (From CPU)	• 4 PCI-E 3.0 x16 (LP) slots (GPU tray for GPUDirect RDMA) • 2 PCI-E 3.0 x16 (LP) slots (CPU tray)
Networking	Dual 10GBase-T Ethernet ports	Dual 10GBase-T Ethernet ports
Onboard VGA	1 VGA port via ASPEED AST2500 BMC	1 VGA port via ASPEED AST2500 BMC
Management	IPMI 2.0 with virtual media over LAN and KVM-over-LAN support	IPMI 2.0 + KVM with dedicated LAN, Intel® Node Manager, SUM, SPM, SSM, SuperDoctor® 5, Watchdog, Supermicro RSD
Power Supply	2000W Titanium Level (96%+) efficiency redundant power supplies	4x 2000W Redundant Titanium Level (96%+) high-efficiency power supplies with I²C & PMBus
Cooling System	7x 40mm heavy duty counter-rotating PWM fans with air shroud & optimal fan speed control	8x 92mm heavy duty counter-rotating PWM fans with air shroud & optimal fan speed control
Form Factor	1U Rackmount: 437 x 43 x 894mm (17.2" x 1.7" x 35.2")	4U Rackmount: 447 x 178 x 805mm (17.6" x 7.0" x 31.7")

Printed on Recycled Paper

01_NVLink-Family_180914_Rev2