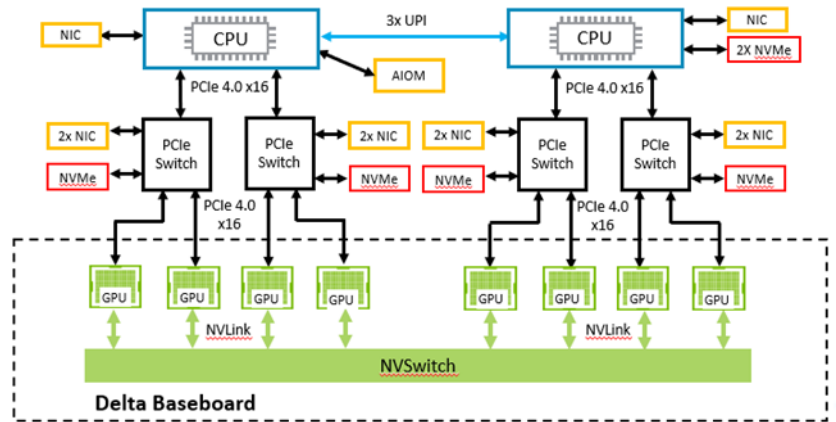


# AI GPU Accelerated Server with 8 NVIDIA A100 SXM4

Part Number: OSS-AMPERE8

## FEATURES

- 4U Chassis
- Dual Future Generation Intel® Xeon® Scalable processors
- Up to 6TB DDR4 LRDIMM System Memory
- Six 2.5" NVMe SSD drive bays
- Eight Ampere GPU SXM4 with 600GB/s NVLink 3.0
- Ten x16 PCIe 4.0 slots
- Four 2200W Titanium Power Supplies
- GPU Management and Monitoring pre-installed
- Software bundle pre-installed



The OSS-AMPERE8 is the newest and most powerful PCIe 4.0 AI training and inference platform supporting NVIDIA’s HGX-A100 baseboard with 600GB/s Bi-Directional NVLink 3.0 peer-to-peer GPU connections designed to tackle the largest AI models. Supporting eight of the latest A100 NVIDIA GPUs, the OSS-AMPERE8 provides 2.5 PetaFLOPS TensorFloat 32 for AI training and 10 PetaOPS INT4 AI inference of sparse neural networks for the most demanding HPC applications. The OSS-AMPERE8 comes completely NVIDIA GPU Cloud certified with deep learning frameworks, popular pre-trained models, GPU management and monitoring software preinstalled to provide data scientists the fastest time to actionable intelligence. The GPU accelerated server also includes dual high-performance Future Intel® Xeon® Scalable processors and DDR4 memory scalable to 6TB. Ten PCIe Gen4 x16 slots are available for additional expansion and for scale out creating a balanced 1:1 GPU Ratio when using IB or highspeed Ethernet networking. The appliance includes six 2.5” removable NVMe drive bays.



## SPECIFICATIONS

Dimensions:	4U Rack Units 7” H x 17.2” W x 35.2” D (39.3” with rails)
CPUs:	Dual Future Generation Intel® Xeon® Scalable processors
System Memory	Memory Capacity <ul style="list-style-type: none"> <li>o 32x 288-pin DDR4 DIMM slots</li> </ul> Memory Type <ul style="list-style-type: none"> <li>o 3200 MHz ECC DDR4 RDIMM/LRDIMM</li> </ul>
GPUs:	8x Tesla A100 SXM4 <ul style="list-style-type: none"> <li>o 10 PetaTOPS INT4<sup>1</sup>; 2.5 PetaFLOPs Tensor Float 32<sup>2</sup>, 156 TeraFLOPS FP64</li> <li>o 320GB HBM2 Memory</li> <li>o 4.8 TB/s Total Aggregate GPU Bandwidth</li> </ul>
Software Bundle	Choice of Operating System <ul style="list-style-type: none"> <li>o CentOS</li> <li>o RHEL</li> <li>o SLES</li> </ul>

SPECIFICATIONS CONTINUED

Part Number: OSS-AMPERE8

Software Bundle (Continued)	<p>Choice of Machine Learning Framework</p> <ul style="list-style-type: none"> <li>o Caffe2</li> <li>o Pytorch</li> <li>o Mxnet</li> <li>o Microsoft Cognitive Toolkit</li> <li>o Tensorflow</li> <li>o Theano</li> <li>o MLPython</li> <li>o ML Dependencies (400MB Python)</li> <li>o cuDNN</li> <li>o DIGITS</li> <li>o Caffe on Spark</li> <li>o CUDA &amp; NVIDIA driver</li> <li>o CUB (CUDA building blocks)</li> <li>o NCCL</li> </ul> <p>GPU Management from Bright Computing</p> <ul style="list-style-type: none"> <li>o Health Management</li> <li>o Workload Integration</li> </ul>
Expansion Slots	<p>PCI-Express</p> <ul style="list-style-type: none"> <li>o Ten PCIe 4.0 x16 LP slots</li> <li>o Six PCIe 4.0 x4 U.3 2.5" drive bays</li> </ul>
On-board devices:	<p>Chipset: Dual Future Intel® Xeon® Scalable Processors</p> <p>SATA: SATA3 (6Gbps) with RAID 0, 1, 5, 10</p> <p>IPMI</p> <ul style="list-style-type: none"> <li>o Support for Intelligent Platform Management Interface v.2.0</li> <li>o IPMI 2.0 with virtual media over LAN and KVM-over-LAN support</li> </ul> <p>Network Controllers</p> <ul style="list-style-type: none"> <li>o Intel® X540 Dual Port 10GBase-T</li> <li>o Virtual Machine Device Queues reduce I/O overhead</li> <li>o Supports 10GBASE-T, 100BASE-TX, and 1000BASE-T, RJ45 output</li> </ul> <p>Graphics: ASPEED AST2400 BMC</p>
Drive Bays:	Hot-swap 6x 2.5" Hot-swap NVMe U.3 drive bays
System BIOS:	BIOS Type: 128Mb SPI Flash EEPROM with AMI® BIOS
Front Panel:	<p>Buttons</p> <ul style="list-style-type: none"> <li>o Power On/Off button</li> <li>o System Reset button</li> </ul> <p>LEDs</p> <ul style="list-style-type: none"> <li>o Power LED</li> <li>o Hard drive activity LED</li> <li>o Two Network activity LEDs</li> <li>o System Overheat LED / Fan fail LED /</li> <li>o UID LED</li> </ul>
Cooling Fans:	8x 92mm + 4x 80mm heavy duty counter-rotating fans with air shroud & optimal fan speed control
Power Supply	<p>4x 2200W Redundant Power Supplies with PMBus</p> <p>Total Output Power: 1200W/1800W/1980W/2200W</p> <p>Dimension (W x H x L): 73.5 x 40 x 265 mm</p> <p>Certification: UL/cUL/CB/BSMI/CE/CCC Titanium Level</p>
Environment:	<p>Operating Temperature: 10°C to 30°C (50°F to 86°F)</p> <p>Non-operating Temperature: -40°C to 70°C (-40°F to 158°F)</p> <p>Operating Relative Humidity: 8% to 90% (non-condensing)</p> <p>Non-operating Relative Humidity: 5% to 95% (non-condensing)</p>
Compliance:	RoHS 6 of 6, WEEE