



AIME A4000 - Multi GPU HPC Rack Server

The AIME A4000 is an enterprise Deep Learning server based on the ASUS ESC4000A-E10, configurable with up to 4 of the most advanced deep learning accelerators and GPUs to enter the Peta FLOPS HPC computing area with more than 4 Peta TensorOps Deep Learning performance. Packed into a dense form factor of 2 height units, the fastest PCI 4.0 bus speeds and 10GB network connectivity.

Built to perform 24/7 at your inhouse data center or co-location for most reliable high performance computing.

AIME A4000 - Deep Learning Server

If you are looking for a server specialized in maximum deep learning training, inference performance and for the highest demands in HPC computing, the AIME A4000 multi-GPU 2U rack server takes on the task of delivering.

The AIME A4000 is based on the new ASUS ESC4000A-E10 barebone which is powered by an AMD EPYC™ 7002 processor with up to 64 cores, 128 threads.

Its GPU-optimized design with high air flow cooling allows the use of four high-end double-slot GPUs like the NVIDIA A100, Tesla or Quadro GPU models.

EPYC CPU Performance

The high-end AMD EPYC CPU designed for servers delivers up to 64 cores with a total of 128 threads per CPU with an unbeaten price performance ratio.

The available 128 PCI 4.0 lanes of the AMD EPYC CPU allow highest interconnect and data transfer rates between the CPU and the GPUs and ensures that all GPUs are connected with full x16 PCI 4.0 bandwidth.

A large amount of available CPU cores can improve the performance dramatically in case the CPU is used for preprocessing and delivering of data to optimal feed the GPUs with workloads.

Up to 32 TB High-Speed SSD Storage

Deep Learning is most often linked to high amount of data to be processed and stored. A high throughput and fast access times to the data are essential for fast turn around times.

The AIME A4000 can be configured with up to four exchangeable U.2 NVMe triple level cell (TLC) SSDs with a capacity of up to 8 TB each. Which adds up to a total capacity of 32 TB of fastest SSD storage.

Since each of the SSDs is directly connected to the CPU and the main memory via PCI 4.0 lanes, they achieve consistently high read and write rates of 3000 MB/s.

As usual in the server sector, the SSDs have an MTBF of 2,000,000 hours and a 5-year manufacturer's guarantee.



High Connectivity and Management Interface

With the available 2x 1 Gbit/s and 1x 10 Gbit/s SFP+ LAN ports fastest connections to NAS resources and big data collections are achievable. Also for data interchange in a distributed compute cluster the highest available LAN connectivity is a must have.

The AIME A4000 is completely manageable with ASMB9 (out-of-band) and ASUS Control Center (in-band) makes a successful integration of the AIME A4000 into larger server clusters possible.

Definable GPU Configuration

Choose the need configuration among the most powerful NVIDIA GPUs for Deep Learning:

2-4x NVIDIA A100

The Nvidia A100 is the flagship of Nvidia Ampere processor generation. With its 6912 CUDA cores, 432 Third-generation Tensor Cores and 40 GB of highest bandwidth HBM2 memory a **single A100 is breaking the Peta TOPS performance barrier. 4 of those add up to more than 1000 teraFLOPS fp32 performance.**



2-4x Nvidia RTX 3090

The GeForce RTX™ 3090 is a big ferocious GPU (BFGPU) with TITAN class performance. It's powered by NVIDIA's Ampere 2nd gen RTX architecture—doubling down AI performance with 10496 CUDA cores, 328 Third-generation Tensor Cores, and new streaming multiprocessors. It features 24 GB of GDDR6X memory.



2-4x Tesla V100

The Tesla V100 is the flagship of the Volta GPU processor series which takes on any Turing based GPU. With its 640 Tensor Core and 32 GB of highest bandwidth HBM2 memory the Tesla V100 was the first beyond 100 teraFLOPS (TFLOPS) GPU available.



2-4x NVIDIA RTX A6000

The NVIDIA RTX A6000 is the Ampere based refresh of the Quadro RTX 6000. It features the same GPU processor (GA-102) as the RTX 3090 but with all processor cores enabled. It even outperforms the RTX 3090 with its 10752 CUDA and 336 Third-generation Tensor Cores. With the double amount of GPU memory compared to the Quadro RTX 6000 and the RTX 3090: 48 GB GDDR6 ECC. The NVIDIA RTX A6000 is currently the 2nd fastest available NVIDIA GPU, only topped by the NVIDIA A100, with the largest available GPU memory, best suited for the most compute and memory demanding tasks.



All NVIDIA GPUs are supported by NVIDIA's CUDA-X AI SDK, including cuDNN, TensorRT which power nearly all popular deep learning frameworks.

Optimized for Multi GPU Server Applications

The AIME A4000 offers energy efficiency with redundant platinum power supplies, which enable long time fail-safe operation.

Its thermal control technology provides more efficient power consumption for large-scale environments.

All setup, configured and tuned for perfect Multi GPU performance by AIME.

The A4000 come with preinstalled Linux OS configured with latest drivers and frameworks like Tensorflow, Keras, PyTorch and Mxnet. Ready after boot up to start right away to accelerate your deep learning application.

Technical Details AIME A4000

Typ	Rack Server 2HE, 80cm depth
CPU Options	EPYC 7232 (8 cores, 3.1 GHz) EPYC 7302 (16 cores, 3.0 GHz) EPYC 7402 (24 cores, 2.8 GHz) EPYC 7502 (32 cores, 2.5 GHz) EPYC 7462 (48 cores, 2. GHz) EPYC 7742 (64 cores, 2.25 GHz)
RAM	64 / 128 / 256 / 512 GB ECC memory
GPU-Options	2 to 4x NVIDIA A100 40GB or 2 to 4x NVIDIA RTX 3090 20GB or 2 to 4x NVIDIA RTX A6000 48GB or 2 to 4x Tesla V100 16GB or 2 to 4x Tesla V100S 32GB
Cooling	CPU and GPUs are cooled with an air stream provided by 7 high performance fans > 100000h MTBF
Storage	Up to 4x 8TB U.2 NVMe SSD Tripple Level Cell (TLC) quality 3000 MB/s read, 3000 MB/s write MTBF of 2,000,000 hours and 5 years manufacturer's warranty
Network	2x 1 GBit LAN RJ45 1x 10 GBit LAN SFP+ 1x IPMI LAN
USB	4x USB 3.0 Ports (front) 2x USB 3.0 Ports (back)
PSU	2x 1600 Watt redundant power 80 PLUS Platinum certified (94% efficiency)
Noise-Level	80dBA
Dimensions (WxHxD)	800mm x 440mm x 88.9mm (2HE) 31.50" x 17.22" x 3.5"
Operating Environment	Operation temperature: 10°C ~ 35°C Non operation temperature: -40°C ~ 70°C

