



AIME A4000 - Multi GPU HPC Rack Server

Der AIME A4000 basiert als Enterprise Deep Learning-Server auf dem ASUS ESC4000A-E10 Barebone. Im schlanken 2HE-Formfaktor ist er mit bis zu vier der fortschrittlichsten Deep Learning-Beschleunigern bzw. GPUs konfigurierbar. Treten Sie mit mehr als 4 Peta-TensorOps Deep Learning-Leistung ein in den Peta-FLOPS-HPC-Computerbereich.

Mit dem neuen PCI 4.0-Bus sowie einer 10 GB Netzwerkkonnektivität ist er ausgelegt für verlässliches Hochleistungs-Computing im 24/7-Dauereinsatz in Ihrem hauseigenen Rechenzentrum oder Ihrer Co-Location.

AIMÉ A4000 - Deep Learning Server

Wenn Sie auf der Suche nach einem Server sind, der auf maximales Deep-Learning-Training, Inferenzleistung, maschinelles Lernen und für die höchsten Anforderungen im HPC Computing spezialisiert ist, haben Sie mit dem AIMÉ A4000 Multi-GPU-2HE-Rack-Server die optimale Lösung gefunden.

Der AIMÉ A4000 basiert auf dem neuen ASUS Barebone ESC4000A-E10, welches mit einem AMD EPYC™ 7002-Prozessor mit bis zu 64 Kernen und 128 Threads betrieben werden kann.

Das auf multi-GPU-Betrieb ausgelegte Design, mit seiner effizienten Luftstromkühlung, ermöglicht die Verwendung von vier parallel betriebenen High-End-GPUs in jeweils zwei Doppel-Steckplätzen, wie etwa NVIDIA's A100, Tesla oder Quadro GPUs.

EPYC CPU-Leistung

Die High-End AMD EPYC Server-CPU arbeitet mit bis zu 64 Kernen und insgesamt 128 Threads pro CPU. Die 128 verfügbaren PCI 4.0-Lanes der AMD EPYC-CPU ermöglichen es, dass alle vier GPUs mit der maximalen Anzahl von PCI Lanes angebunden werden. Hierdurch wird die höchste Datenübertragungsrate zwischen der CPU und den GPUs erreicht.

Der Datendurchsatz kann durch die hohe Anzahl verfügbarer CPU-Kerne erheblich verbessert werden, da hierdurch die GPUs optimal mit Aufgaben versorgt werden können, wenn diese für das Preprocessing und die Bereitstellung von Daten genutzt werden.

Bis zu 32 TB High-Speed-SSD-Speicher

Beim Deep Learning müssen hohe Datenmengen verarbeitet und gespeichert werden. Ein hoher Datendurchsatz und schnelle Zugriffszeiten auf die Daten sind daher für schnelle Turnaround-Zeiten unerlässlich.

Der AIMÉ A4000 kann mit bis zu vier wechselbar U.2 NVMe Tripple Level Cell SSDs mit einer Kapazität von jeweils bis zu 8 TB konfiguriert werden. Was sich im Vollausbau zu einer Gesamkapazität von 32 TB schnellsten SSD Speicher addiert.

Da jede der SSDs ist über PCI-4.0-Lanes direkt mit der CPU und dem Hauptspeicher verbunden und erreichen dadurch konstant höchste Lese und Schreibraten von 3000 MB/s.

Wie im Serverbereich üblich haben die SSDs eine MTBF von 2,000,000 Stunden und 5 Jahre Garantie des Herstellers.



Höchste Konnektivität mit Management Interface

Mit den verfügbaren 2x 1 Gbit/s RJ45 und 1x 10 Gbit/s SFP+ LAN Ports sind die schnellsten Verbindungen zu NAS-Ressourcen und Big Data Datenbanken möglich. Auch für den Datenaustausch in einem verteilten Compute Cluster ist die höchste verfügbare LAN-Konnektivität ein Muss.

Der AIME A4000 ist vollständig über ASMB9 (Out-of-Band) verwaltbar, und das ASUS Control Center (In-Band) ermöglicht eine erfolgreiche Integration des AIME A4000 in größere Servercluster.

Wählbare GPU-Konfigurationen

Wählen sie für die Konfiguration aus den leistungsfähigsten NVIDIA Deep-Learning GPUs:

2-4x NVIDIA A100

Nvidia A100 ist das Flaggschiff der Nvidia Ampere-Prozessorgeneration. Mit seinen 6912 CUDA-Kernen, 432 Tensorkernen der dritten Generation und 40 GB HBM2-Speicher mit der höchsten Bandbreite durchbricht eine einzelne A100 GPU schon die Peta-TOPS-Leistungsbarriere. Vier GPUs dieser Art summieren sich somit auf mehr als 1000 teraFLOPS fp32-Leistung.



2-4x Nvidia RTX 3090

Die GeForce RTX™ 3090 basiert auf der NVIDIA Ampere RTX-Architektur der 2. Generation und verdoppelt die AI-Leistung mit 10496 CUDA-Kernen, 328 Tensorkernen der 3. Generation und neuen Streaming-Multiprozessoren. Sie bietet die Leistung die vorher nur von GPUs der Titan Klasse erreicht wurden. Die RTX 3090 verfügt über 24 GB GDDR6X-Speicher.



2-4x Tesla V100

Die Tesla V100 ist das Flaggschiff der Volta GPU-Prozessorserie, die jede Turing-basierte GPU im Leistungsvergleich schlägt. Mit seinen 640 Tensor-Kernen und 32 GB HBM2-Speicher mit der höchsten Bandbreite war die Tesla V100 die erste verfügbare GPU mit mehr als 100 TeraFLOPS (TFLOPS).



2-4x NVIDIA RTX A6000

Die NVIDIA RTX A6000 ist das Ampere-basierte Nachfolgemodell der Quadro RTX 6000. Sie verfügt über denselben GPU-Prozessor (GA-102) wie die RTX 3090, allerdings sind alle Kerne des GA-102 Prozessors aktiviert. Mit Ihren 10752 CUDA und 336 Tensor-Kernen der 3. Generation übertreift sie damit die RTX 3090. Ausgestattet mit der doppelten Menge an GPU-Speicher, im Vergleich zur Quadro RTX 6000 und der RTX 3090: 48 GB GDDR6 ECC. Die NVIDIA RTX A6000 ist derzeit zweitschnellste erhältliche NVIDIA-GPU, die nur von der NVIDIA A100 übertroffen wird. Sie verfügt über den größten verfügbaren GPU-Speicher und eignet sich bestens für speicherintensive Anwendungen.



Alle angebotenen NVIDIA GPUs unterstützen NVIDIAs CUDA-X AI SDK inkl. cuDNN und TensorRT und werden damit von allen gängigen Deep-Learning-Frameworks unterstützt.

Optimiert für Multi-GPU-Serveranwendungen

Der AIME A4000 ist energieeffizient und bietet durch seine zwei redundanten Platinum-Netzteile einen ausfallsicheren Langzeitbetrieb. Seine Wärmesteuerungstechnologie ermöglicht einen effizienten Stromverbrauch in Server-Umgebungen.

AIME liefert den A4000 einsatzbereit vorkonfiguriert, optimiert auf Multi-GPU-Dauerleistung wie sie für Deep-Learning Anwendungen benötigt wird. Hierzu gehört auch ein vorinstalliertes Linux-Betriebssystem, mit aktuellsten Treibern und Frameworks wie Tensorflow, Keras, PyTorch und Mxnet, komfortabel verpackt in das AIME ML Container Management Framework.

Beginnen Sie sofort nach dem ersten Booten mit dem Training Ihrer Deep-Learning-Anwendung.

Technische Details AIME A4000

Typ	Rack Server 2HE, 80cm Tiefe
CPU Optionen	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 Speicher
GPU-Optionen	2 bis 4 NVIDIA A100 40GB oder 2 bis 4 NVIDIA RTX 3090 20GB oder 2 bis 4 NVIDIA RTX A6000 48GB oder 2 bis 4 Tesla V100 16GB oder 2 bis 4 Tesla V100S 32GB
Kühlung	CPU und GPUs mit 7 starken Hochleistungslüfter gekühlt > 100000h MTBF
Speicher	Bis zu 4x 8TB U.2 NVMe SSD Tripple Level Cell (TLC) Qualität 3000 MB/s lesen, 3000 MB/s schreiben MTBF von 2,000,000 Stunden und 5 Jahre Herstellergarantie
Netzwerk	2x 1 GBit LAN RJ45 1x 10 GBit LAN SFP+ 1x IPMI LAN
USB	4x USB 3.0 Ports (vorne) 2x USB 3.0 Ports (hinten)
Netzteil	2x 1600 Watt Leistung, redundant 80 PLUS Platinum zertifiziert (94% Effizienz)
Geräuschpegel	80dBA
Abmessungen (WxHxD)	800mm x 440mm x 88.9mm (2HE) 31.50" x 17.22" x 3.5"
Betriebsumgebung	Betriebs-Temperatur: 10°C ~ 35°C Nicht-Betriebs-Temperatur: -40°C ~ 70°C

