

NVIDIA HGX A100 8基搭載 4Uラックマウント GPGPUモデル

4UラックマウントシャーシにNVIDIA A100を8基搭載する超高密度GPGPUモデルです。本製品は、AMD 第3世代 EPYC プロセッサを2基搭載し高いマルチスレッド性能とPCIe 4.0接続による高速なネットワーク帯域幅により高密度なGPUコンピュートノードを構築することが可能でTCOを大幅に削減します。



NVIDIA HGX A100 8基搭載

AMD 第3世代 EPYC プロセッサ 2基搭載

DDR4 3200 ECC メモリ搭載

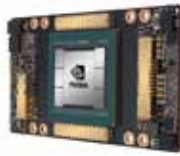
メモリスロット 32 DIMM

NVIDIA CUDA プリインストール

NVIDIA A100 搭載

NVIDIA A100 Tensor Core GPUはAIとHPC用途のためにデータセンター向けGPUとして開発されたVolta世代の後継として製造プロセス7nmのAmpereアーキテクチャを採用。第3世代のTensorコアを実装し、BF16、Tensor Float 32を新たにサポート、前世代のVoltaと比較しFP32演算、INT8のディープラーニング推論処理において20倍の性能を実現し、トレーニングと推論の両方が強化されています。

また、Multi-instance GPU(MIG)機能により単一のA100を最大7個のGPUとして扱うことが可能で複数のユーザーが独立したインスタンスとして利用できます。



■ NVIDIA HGX A100

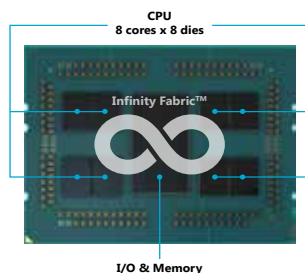
| | |
|----------------|---------------|
| GPU アーキテクチャ | NVIDIA Ampere |
| フォームファクタ | SXM |
| FP64 CUDA コア | 3,456 |
| FP32 CUDA コア | 6,912 |
| Tensorコア | 432 |
| GPUメモリ | 80GB HBM2 |
| GPU間接続帯域 | 最大600 GB/sec |
| マルチインスタンスGPU | 最大7分割 |
| 最大消費電力 | 400 W |
| 主要アプリケーション実行性能 | 100% |
| 冷却方式 | パッシブ(冷却ファンなし) |

AMD 第3世代 EPYC™ プロセッサ搭載

第3世代 EPYCは前世代より製造プロセス、最大コア数などに大きな変化はないものの、新しいアーキテクチャ「ZEN3」に刷新されており、第2世代 EPYCに採用されていた「Zen2」と比較すると、IPC (Instruction Per Clock-cycle) が最大19%ほど改善されています。

アーキテクチャが新しくなったことにより、8つのCPUコアが32MBのL3キャッシュを共有し1コアあたりが使えるL3キャッシュが増加しているほか、6チャンネルのメモリチャネルをサポート。これによりメモリ構成が柔軟に選択でき、導入におけるコスト削減に貢献しています。

このほか、セキュリティ面も強化されており「AMD Infinity Guard」と呼ばれるセキュリティ機能を搭載、新たに「SEV-SNP」「Shadow Stack」機能が追加され仮想マシン環境の安全性を引き上げています。



■仕様

| | |
|------------|--|
| プロセッサ | AMD 第3世代 EPYC™ プロセッサ x2基 |
| チップセット | System on Chip |
| メモリ | DDR4-3200 ECC RDIMM/LRDIMM |
| メモリスロット | 32 DIMM |
| ストレージ | 2.5インチ |
| ドライブベイ | 6 (ホットスワップ対応) ・NVMe x6 (標準) ・SATA/SAS x4 + NVMe x2 (オプション) |
| M.2スロット | インターフェイス: PCIe 4.0 x4 フォームファクタ: 2280/22110 x2 |
| オプティカルドライブ | USB接続 外付けDVDマルチドライブ (オプション) |
| グラフィックス | onboard |
| GPU | NVIDIA HGX A100 80GB x8基 |
| ネットワークI/F | 10GbE 2ポート (RJ45) |
| 管理インターフェース | IPMI 2.0, KVM-over-LAN |
| I/Oポート | VGA x1 (前面) USB 3.0 x2 (前面) |
| 拡張スロット | PCI-Express 4.0 x16 (8スロット LP) PCI-Express 4.0 x16 (1スロット LP) PCI-Express 4.0 x8 (1スロット LP) |
| ケース | 4U ラックマウント (外形寸法: W446mm x H178mm x D900mm) |
| 電源 | 100V 1000W / 200V 1800W リダンダント電源 (80-plus Platinum 高効率電源) |
| 対応OS (別売) | <ul style="list-style-type: none"> ● Ubuntu ● AlmaLinux ● Rocky Linux ● RedHat Enterprise Linux Server ● SUSE Linux Enterprise Server ● Microsoft Windows Server 2022 ● ※その他Linuxディストリビューションについてはご相談ください。 |
| 開発環境 | NVIDIA CUDA (デバイスドライバ、ツールキット、CUDA SDK開発環境、コンパイラ等) Docker |
| フレームワーク | TensorFlow, Pytorch * Deep Learning各種フレームワークのインストールにつきましてはお問い合わせください。 |
| 保守 | 1年間全国出張オンサイトサービス 3年間全国出張オンサイト (オプション) |