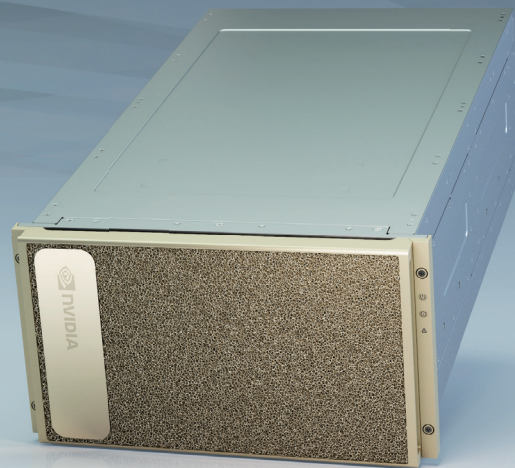
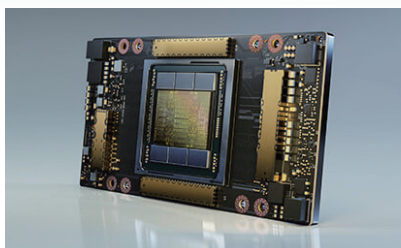


NVIDIA DGX™ A100

第3世代統合 AI システム

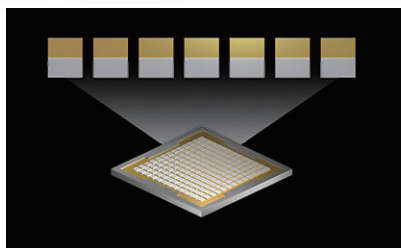


DGX A100はGPUに最新のAmpereアーキテクチャ「NVIDIA A100 Tensor コア GPU」を NVIDIA® NVSwitch™ で結合し8基搭載、CPUはAMDのRome EPYC7742 64コア 3.4GHzを採用しています。新しいGPUであるA100の演算性能は、倍精度で19.5TFLOPSと従来のV100の約2.5倍、単精度は「TF32」と呼ばれるFP32とFP16のハイブリッド方式を採用しソフトウェアの改編の必要なく312TFLOPSのスループットを実現。ディープラーニング推論性能は、1,248TOPSといずれもV100の約20倍の演算性能を誇り、8基のA100を搭載したDGX A100のAI性能は1ノードで5PFLOPSに達します。



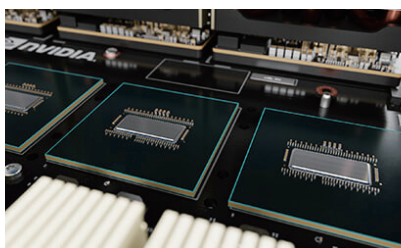
| NVIDIA A100 Tensor コア GPU

第3世代の NVIDIA Tensor コアにより性能を大幅に向上できるようになり、A100 GPUを効率的に数千までスケールアップできます。また、マルチインスタンスGPUを使用すれば、あらゆるサイズのワークフローを加速するために7つの小さな専用インスタンスとして割り当てることができます。



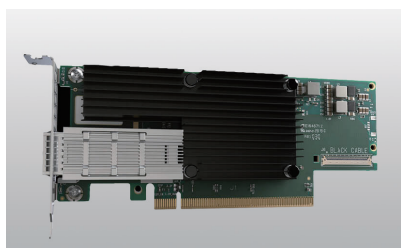
| マルチインスタンス GPU (MIG)

MIGを利用することで、DGX A100の8つのA100 Gpuを最大56個のGPUインスタンスとして構成し、それぞれが独自の高帯域幅メモリ、キャッシュ、コンピューティング コアで完全分離できます。これにより管理者は、複数のワークフローに対してGPUのサイズを適切に設定することができます。



| 次世代 NVLink と NVSwitch

DGX A100の第3世代のNVIDIA® NVLink®は、GPUからGPUへの直接帯域幅を600ギガバイト/秒 (GB/s) まで倍増します。これはPCIe Gen4のほぼ10倍に相当します。DGX A100には、前世代と比較して2倍の速度を可能にする次世代のNVIDIA® NVSwitch™も搭載されています。



| Mellanox ConnectX-6 VPI HDR InfiniBand

DGX A100には最新型のMellanox ConnectX-6 VPI HDR InfiniBand /イーサネット アダプターが搭載されており、それぞれが200ギガバイト/秒 (GB/s) で動作し、大規模AIワークフロー用の高速ファブリックを形成します。

システムスペック

GPU	NVIDIA A100 Tensor Core GPUs × 8
GPUメモリ	Total 320 GB
性能	5 petaFLOPS AI 10 petaOPS INT8
NVIDIA NVSwitches	6
最大消費電力	6.5 kW
CPU	Dual AMD Rome 7742 128 cores total, 2.25 GHz (base), 3.4 GHz (max boost)
システムメモリ	1TB
ネットワーク	8x Single-Port Mellanox ConnectX-6 VPI 200Gb/s HDR InfiniBand 1x Dual-Port Mellanox ConnectX-6 VPI 10/25/50/100/200Gb/s Ethernet
ストレージ	OS : 1.92TB M.2 NVME drives × 2 Internal Storage : 15TB (4x 3.84TB) U.2 NVME drives
ソフトウェア	Ubuntu Linux OS
重量	123 kgs
サイズ	全高 : 264.0 mm 全幅 : 482.3 mm 奥行 : 897.1 mm
運用温度範囲	5 °C - 30 °C