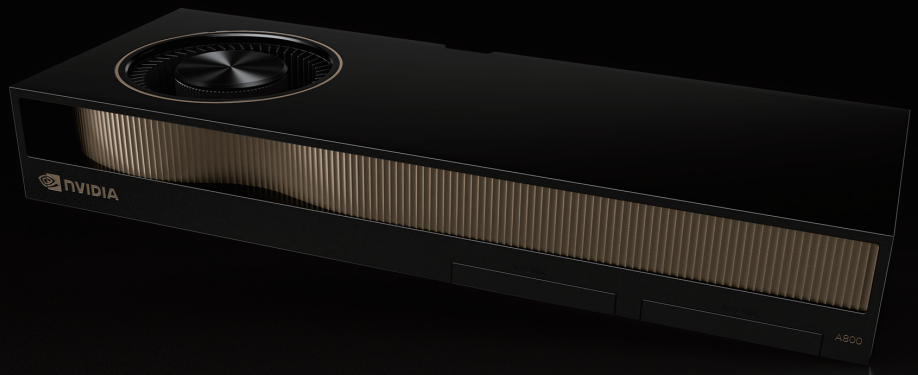
Datasheet

NVIDIA A800 40GB Active



AI、データ サイエンス、ハイ パフォーマンス コンピューティング (HPC) のための究極のワークステーション開発プラットフォーム

高度なコンピューティング ワークロード向けの画期的なパフォーマンス

スーパーコンピューターのパワーをワークステーションにもたらす。NVIDIA A800 40GB Active GPU はエンドツーエンドのデータ サイエンス ワークフローを加速します。NVIDIA Ampere アーキテクチャを搭載した A800 40GB Active の提供する強力なコンピューティング、高速メモリ、拡張性により、プロフェッショナルは最も困難なデータ サイエンス、AI、HPC ワークロードに取り組むことができます。

データサイエンスと解析

複雑なデータ サイエンス ワークフローを強化し、データの読み込みやデータ操作から機械学習や視覚化に至るまで、エンドツーエンドのデータ サイエンス パイプラインを高速化します。

AI トレーニングと推論

データの準備と処理、モデルの最適化とチューニング、初期段階のトレーニングなど、要求の厳しい AI 開発、トレーニング、推論ワークフローに対応します。

HPC

大規模なシミュレーションを 完全なFP64 精度で驚異的な速度で実行し、開発タイムラインを短縮し、価値実現までの時間を短縮します。

すぐに使える AI 開発を強化する NVIDIA AI Enterprise

NVIDIA A800 40GB Active GPU には、エンタープライズ セキュリティ、安定性、管理性、サポートを備えたエンドツーエンド ソフトウェア プラットフォームである NVIDIA AI Enterprise の 3 年間のサブスクリプションが付属しています。NVIDIA AI Enterprise には、実稼働対応の AI およびデータ サイエンスを迅速に開発および展開するための 100 以上の AI フレームワーク、ライブラリ、事前トレーニング済みモデル、ツールが含まれています。NVIDIA A800 40GB と組み合わせることで、NVIDIA AI Enterprise は AI の導入を簡素化し、最高のパフォーマンスでより迅速にビジネスの洞察を実現します。NVIDIA AI Enterprise ソフトウェア サブスクリプションにアクセスして、その利点について詳しく学びましょう。

主な機能

NVIDIA Ampere アーキテクチャ

第3世代 Tensor コア

- 強力な倍精度 (FP64) 機能
- トレーニングと推論パフォーマンスの高速化

第3世代 NVIDIA® NVLink™

- 2 つの A800 GPU を接続してメモリを 80 ギガバイト (GB) まで拡張
- 400 ギガバイト/秒 (GB/s) の双方向帯域幅

超高速 HBM2 メモリ

- 40GBの高速HBM2メモリ
- 1.5 TB/s のメモリー帯域幅

マルチインスタンス GPU (MIG)

- 完全に分離された安全なマルチテナント
- 最大 7 つのインスタンスまでパーティション化

仕様	
GPU メモリー	40GB HBM2
メモリーインタフェース	5,120-bit
メモリー帯域幅	1.5 TB/s
CUDA® コア	6,912
Tensor コア	432
倍精度演算性能	9.7 TFLOPS
単精度演算性能	19.5 TFLOPS
ピーク Tensor 性能	623.8 TFLOPS
マルチインスタンス GPU	最大 7 MIG インスタンス @ 5GB
NVIDIA NVLink	Yes
NVLink 帯域幅	400GB/s
グラフィックスバス	PCIe 4.0 x 16
消費電力	240W
サーマル	Active
フォームファクター	4.4” H x 10.5” L, デュアルスロット

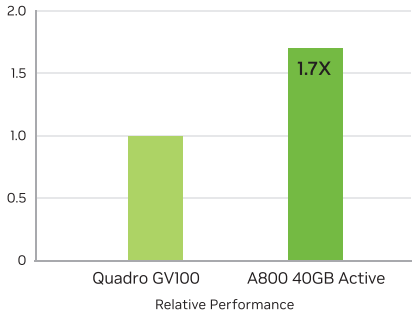
NVIDIA AI Enterprise Software

エンドツーエンドの AI ソフトウェア プラットフォーム

- A800 40GB アクティブ GPU には 3 年間の NVIDIA AI Enterprise ライセンスを含む
- フレームワーク、ライブラリ、ツールにアクセス、AI の運用までの時間を短縮
- エンタープライズセキュリティ、安定性、管理性、サポート
- ソフトウェアのアクティベーションが必要

HPC

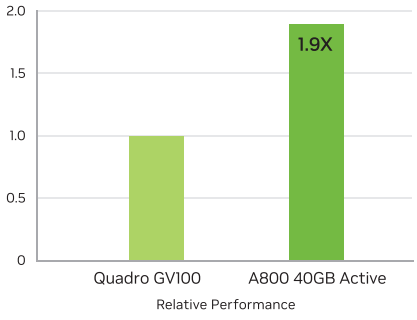
LAMMPS



Tests run on an Intel Xeon Gold 6126 processor, NVIDIA Driver 535.104. Relative speedup for LAMMPS patch_8Feb2023, Atomic Fluid Lennard-Jones 2.5 (cutoff); Precision=FP64;

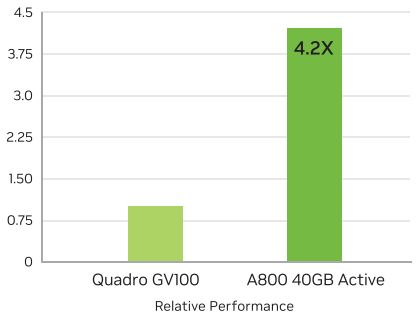
HPC

GTC



Tests run on an Intel Xeon Gold 6126 processor, NVIDIA Driver 535.104. Relative speedup for GTC Version 4.5, TAE, Precision=FP32

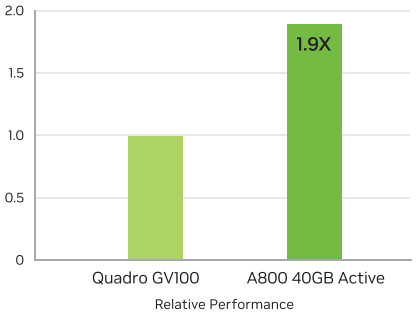
AI 推論



Tests run on an Intel Xeon Gold 6126 processor, NVIDIA Driver 535.104. Relative speedup for GPT2 Inference . Batch Size=32; Precision=Mixed; Data=Synthetic; cuDNN Version=8.9.3.22;

AI トレーニング

BERT - Large



Tests run on an Intel Xeon Gold 6126 processor, NVIDIA Driver 535.104. Relative speedup for BERT Large Pre-Training Phase 2 Batch Size=8; Precision=Mixed; AMP=Yes; Data=Real; Sequence Length=512; Gradient Accumulation Steps=_SEE_OUTPUTS_; cuDNN Version=8.9.3.28; NCCL Version=2.18.3

始める準備はできましたか？

NVIDIA A800 40GB Active のさらに詳しい情報は www.nvidia.com/ja-jp/design-visualization/a800

NVIDIA AI Enterprise の 3 年間のサブスクリプションの有効化は www.nvidia.com/activate-license

1. A800 40GB Active にはディスプレイ ポートが装備されていません。NVIDIA RTX 4000 Ada 世代、NVIDIA RTX A4000、および NVIDIA T1000 は、ディスプレイ出力機能をサポートすることが認定されています。