

NVIDIA DGX Station A100

AI時代のワークグループ アプライアンス



データサイエンスチームは、AIイノベーションの最先端に立ち、企業や私たちの世界を変革できるプロジェクトを構築しています。しかしながら、最も複雑なモデルをトレーニングするのに役立つ予備のコンピューティング環境を探し続けることも少なくありません。そのようなデータサイエンスチームは、どこにでも接続可能で、ハードウェアとソフトウェア全体に対して完全に最適化され、世界中にいる大勢の同時接続ユーザーに革新的な性能を提供できる専用のAIプラットフォームを必要としています

NVIDIA DGX Station™ A100の概要

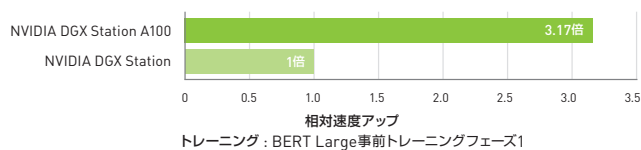
- お客様のチームが無制限に使用できる2.5petaFLOPSの性能を提供可能なAIワークグループサーバー：トレーニング、推論、およびデータ分析用
- サーバーグレードのプラグアンドゴー方式。データセンターの電力と冷却が不要。
- 世界最高峰のAIプラットフォーム。複雑なインストールやITのサポートが不要。
- 完全に相互接続された4基のNVIDIA A100 TensorコアGPUと、最大320ギガバイト (GB) のGPUメモリを搭載した世界で唯一のワークステーション型システム。
- NVIDIAのノウハウと経験を活かしてAIトランスフォーメーションを速やかに実現

NVIDIA DGX Station A100は、AIスーパーコンピューティングをデータサイエンスチームにもたらし、データセンターやITインフラストラクチャの増設なしにデータセンターテクノロジーを提供します。複数ユーザーが同時に接続できるように設計されており、DGX Station A100はオフィスに適したフォームファクタでサーバーグレードコンポーネントを活用します。完全に相互接続されたマルチインスタンスGPU (MIG) 対応の4基のNVIDIA A100 TensorコアGPUと全体で最大320GBのGPUメモリを備えた唯一のシステムであり、標準的な電源コンセントにプラグ接続できるため、どこにでも配置できる強力なAIアプライアンスです。

データサイエンスチーム向けの AIスーパーコンピューティング

DGX Station A100は、複数の同時接続ユーザーに集中型AIリソースを手軽に提供できるAI時代のワークグループアプライアンスです。トレーニング、推論、分析のワークロードをMIGで並列実行可能。最大28台の独立したGPUデバイスが個々のユーザーとジョブに割り当てられるので、システム全体の性能に影響を与えません。DGX Station A100は、すべての

DGX Station A100のトレーニング性能は3倍以上高速



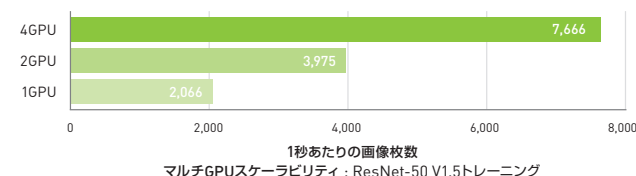
DGX Station A100 320GB; バッチサイズ=64; 混合精度; AMP使用; 実データ; シーケンス長=128

DGX Station A100の推論性能は4倍以上高速



DGX Station A100 320GB; バッチサイズ=256; INT8精度; 合成データ; シーケンス長=128; cuDNN 8.0.4

DGX Station A100はリニアスケーラビリティを提供



DGX Station A100 320GB; バッチサイズ=192; 混合精度; 実データ; cuDNNバージョン=8.0.4; NCCLバージョン=2.7.8; NGC MXNet20.10コンテナ

DGXシステムと同じく完全に最適化されたNVIDIA DGX™ソフトウェアスタックを備えており、個々のシステムからNVIDIA DGX POD™やNVIDIA DGX SuperPOD™に至るまで、最高の性能とDGXベースインフラストラクチャとの完全相互運用性を提供し、規模の大小を問わず、あらゆる組織のチームにとって理想的なプラットフォームとなります。

データセンターなしでデータセンターの性能を実現

NVIDIA DGX Station A100は、ワークステーションのフォームファクタでデータセンター級のAIサーバーを実現し、専用の電源と冷却なしに標準的なオフィス環境で使用するのに適しています。4つの超強力なNVIDIA A100 TensorコアGPU、最上位のサーバーグレードCPU、超高速NVMeストレージ、および最先端のPCIe Gen4バスを備える設計になっています。DGX Station A100にもNVIDIA DGX A100と同じベースボードマネジメントコントローラ(BMC)が搭載されており、システム管理者はリモート接続によって要求タスクを実行できます。DGX Station A100は、オフィス環境向けの最も強力なAIシステムであり、データセンターなしにデータセンターテクノロジーを提供します。

どこにでも配置できるAIアブライアンス

NVIDIA DGX Station A100は、企業のオフィス、実験室、研究施設、さらには自宅から作業する昨今のアジャイルデータサイエンスチーム向けに設計されています。大規模なAIインフラストラクチャを設置するには多大なIT投資と工業用の強力な電源供給と冷却機能を備えた大型データセンターが必要ですが、DGX Station A100なら、お客様のチームの作業スペースがどこにあっても標準的な壁のコンセントに接続するだけです。さらに、その革新的な冷却ベースのデザインにより、手で触れても熱くありません。1人で簡単にセットアップできるので、たった2本のケーブルで動作する世界最高峰のAIプラットフォームを数分で立ち上げて稼働させることができます。

モデルの大規模化、応答の高速化

NVIDIA DGX Station A100はワークステーションではありません。これは、デスクの下に設置できるAIワークグループサーバーです。64コアのデータセンターグレードCPUに加えて、NVIDIA DGX A100サーバーと同じNVIDIA A100 TensorコアGPUを備え、それぞれ40GBまたは80GBのGPUメモリが高速なSXM4を介して接続されています。NVIDIA DGX Station A100は、NVIDIA®NVLink®を活用して4つのGPUを完全相互接続し、かつMIG対応でシステム性能に影響を与えずに並列ジョブと複数のユーザー向けに最大28台の独立したGPUデバイスを提供する唯一のオフィスに適したシステムです。

比類なきAIノウハウへの統合アクセス

NVIDIA DGX Station A100は、NVIDIAの数千人のAI専門家に支えられた完全なハードウェア&ソフトウェアプラットフォームであり、世界最大のDGX実験場であるNVIDIA DGX SATURNVから得られた知識に基づいて構築されています。DGX Station A100を所有すれば、NVIDIA DGXpertsに直接アクセスできます。NVIDIA DGXpertsは、AIに精通した専門家の国際チームであり、NVIDIAの10年を超えるAIリーダーシップのノウハウと経験に基づき、AI変革の急速な発展に貢献する規範的なガイダンスと設計の専門知識を提供します。これにより、ミッションクリティカルなアプリケーションが確実に素早く立ち上がってスムーズに稼働し続けることになり、インサイトを得るまでの時間が大幅に改善されます。

システム仕様

	NVIDIA DGX Station A100 320GB	NVIDIA DGX Station A100 160GB
GPU	4x NVIDIA A100 80GB GPU	4x NVIDIA A100 40GB GPUs
GPUメモリ	総計320GB	総計160GB
性能	2.5petaFLOPS AI 5petaOPS INT8	
システム消費電力	1.5kW (100~120Vac時)	
CPU	Single AMD 7742、64コア、 2.25GHz (ベース) ~3.4GHz (最大ブースト)	
システムメモリ	512GB DDR4	
ネットワーク	デュアルポート10Gbase-T イーサネットLAN シングルポート1Gbase-T イーサネットBMCマネジメントポート	
ストレージ	OS:1x 1.92TB NVMEドライブ 内部ストレージ:7.68TB U.2 NVMEドライブ	
DGXディスプレイアダプタ	4GB GPUメモリ、 4x Mini DisplayPort	
システム音響特性	37dB未満	
ソフトウェア	Ubuntu Linux OS	
システム重量	43.1kg (91.0lbs)	
システム梱包重量	57.93kg (127.7lbs)	
システムサイズ	高さ:639mm (25.1in) 幅:256mm (10.1in) 長さ:518mm (20.4in)	
運用温度範囲	5~35 °C (41~95 °F)	

NVIDIA DGX Station A100について詳しくは www.nvidia.com/DGXStationA100 をご覧ください。