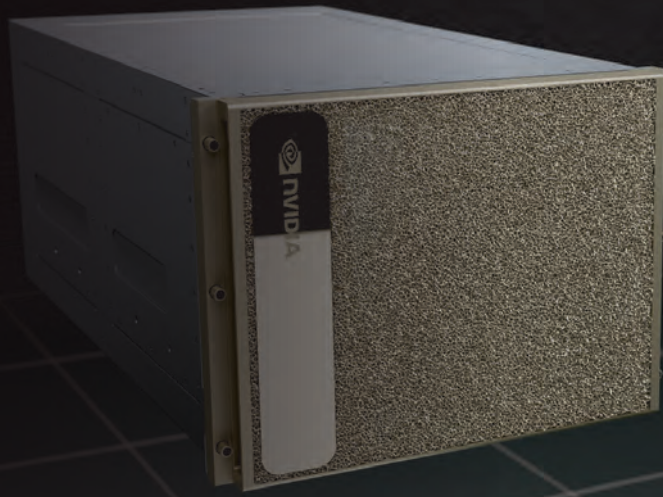




NVIDIA DGX H100

AI インフラの金字塔



人工知能 (AI) はビジネスにおける困難な課題を、解決するために役立つ手段になりました。顧客サービスの改善、サプライチェーンの最適化、ビジネス インテリジェンスの抽出、最先端の製品やサービスの設計など、ほぼすべての業界の組織の課題に対し、AI はイノベーションを実現する包括的なメカニズムを提供します。AI インフラの先駆者となった NVIDIA DGX™ システムは、このような重要な概念を実現するための最も強力なプラットフォームを提供します。

NVIDIA DGX H100 はビジネスのイノベーションと最適化を支援します。NVIDIA における、伝説的な DGX システムと NVIDIA DGX SuperPOD の構築の歴史において、最新モデルである NVIDIA DGX H100 は、画期的な NVIDIA H100 Tensor コア GPU を搭載し、AI の活用を推し進めます。AI スループットを最大化するように設計されており、高度に洗練され、システム化されたスケーラブルなプラットフォームの提供により、自然言語処理、レコメンデーション システム、データ分析など、企業のさまざまな革新的課題解決を支援します。NVIDIA DGX H100 は、オンプレミスに導入できるほか、さまざまなアクセスと展開の選択肢があり、企業の大規模な課題を AI で解決するために必要なパフォーマンスを実現します。

AI のセンター オブ エクセレンスの基礎

AI は、科学とビジネスの間にあった溝を埋めました。AI はすでに実験の域を脱し、大小の企業がイノベーションの推進や事業の最適化のために、日々利用するようになりました。NVIDIA DGX H100 は、AI に特化した世界初のインフラの第 4 世代であり、企業の AI 中核拠点の基礎となるように設計されています。すべてを最適化したハードウェアとソフトウェアのプラットフォームに、新しい各種 NVIDIA AI ソフトウェア ソリューションのフルサポート、充実したサードパーティ サポート エコシステム、NVIDIA プロフェッショナル サービスからの専門的なアドバイスが含まれています。DGX システムは、世界のほぼすべての業界で何千社もの企業が利用しており、NVIDIA DGX H100 には立証された信頼性があります。

AI の大規模活用の壁を打ち破る

NVIDIA DGX H100 は世界で初めて NVIDIA H100 Tensor コア GPU を搭載したシステムであり、AI のスケールとパフォーマンスの限界を突破します。NVIDIA ConnectX®-7 スマート ネットワーク インターフェイス カード (SmartNIC) と NVIDIA BlueField®-3 データ プロセッシング ユニット (DPU) で 6 倍のパフォーマンスと 2 倍高速なネットワーキングを実現し、NVIDIA DGX SuperPOD 向けのハイスピード スケーラビリティを提供します。この次世代アーキテクチャは、自然言語処理やディープ ラーニングによるレコメンデーション モデルなど、大規模で複雑な AI ジョブに対応できるように強化されています。

仕様

GPU	8x NVIDIA H100 Tensor コア GPU
GPU メモリ	合計 640GB
パフォーマンス	32 petaFLOPS FP8
NVIDIA® NVSwitch™	4x
システム消費電力	最大 10.2kW
CPU	デュアル x86
システムメモリ	2TB
ネットワーキング	4 個の OSFP ポートで 8 基のシングルポート ConnectX-7 へ接続 400Gb/s InfiniBand/Ethernet 2 基のデュアルポート NVIDIA BlueField-3 DPU VPI 400Gb/s InfiniBand/Ethernet x1 200Gb/s InfiniBand/Ethernet x1
管理ネットワーク	10Gb/s オンボード NIC (RJ45 付き) 50Gb/s Ethernet オプション NIC ホストベースボード管理コントローラー (BMC) (RJ45 付き) 2x NVIDIA BlueField-3 DPU BMC (それぞれ RJ45 コネクタで接続)
ストレージ	OS : 2x 1.9TB NVMe M.2 内部ストレージ : 8x 3.84TB NVMe U.2
システムソフトウェア	DGX H100 システムにプレインストールされている DGX OS は、Ubuntu Linux をベースとし、DGX ソフトウェア スタック (DGX 用に最適化されている、必要なすべてのパッケージとドライバ) が含まれます。 オプションで、お客様が Ubuntu Linux または Red Hat Enterprise Linux、および必要な DGX ソフトウェア スタックを個別にインストールすることもできます。
動作温度範囲	5°C ~ 30°C (41°F ~ 86°F)

多様な条件に適合するリーダーシップクラスの インフラ

企業で AI を利用する際に重要なのは、パフォーマンスと機能だけではありません。組織の IT 要件と運用要件に適合することも重要です。NVIDIA DGX H100 は、オンプレミスに導入して直接管理できるほか、NVIDIA DGX 対応のデータセンターへのコロケーション、NVIDIA DGX Foundry でのレンタル、NVIDIA 認定のマネージド サービス プロバイダーを介した利用が可能です。また、DGX 対応のライフサイクル管理プログラムにより、最新のシステムの導入を維持するための予測性の高い財務モデルを利用できます。これを利用すると、忙しい IT スタッフの負担を増やすことなく、従来の IT インフラと同じように簡単に NVIDIA DGX H100 を活用、入手できるため、明日を待たずにすぐ AI をビジネスに活用できます。

発注情報

NVIDIA DGX H100 は 2022 年後半に発売する予定です。詳細については、NVIDIA の営業担当者にお問い合わせください。

お問い合わせ先



日本GPUコンピューティングパートナーシップ
<https://www.gdep.jp>