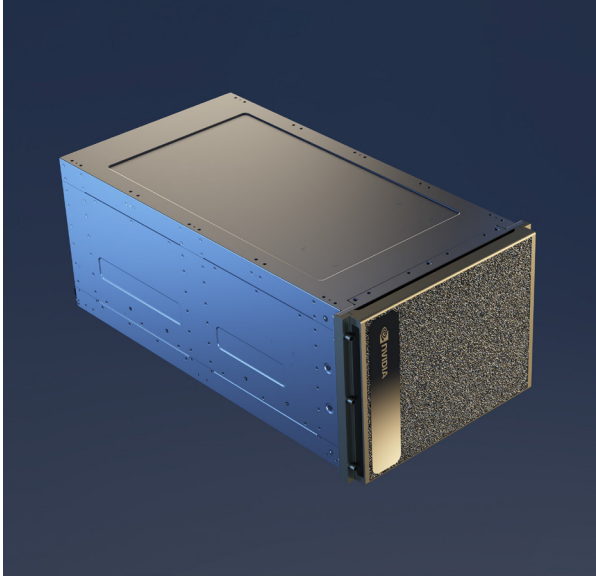
データシート



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となりました。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、ジェネレーティブ AI やその他のTransformer モデルを活用した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™ は、このような重要な概念を実現する最も強力で完結した AI 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。DGX プラットフォームの一部であり、NVIDIA の伝説的な DGX システムの最新モデルである DGX H100 は、NVIDIA DGX SuperPOD™ の基礎をなす AI の原動力であり、NVIDIA H100 Tensor コア GPU の画期的なパフォーマンスにより高速化されます。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、システム化された拡張性の優れたプラットフォームで、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、企業のあらゆる革新的な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ギャップを埋めました。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 プラットフォームは、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。

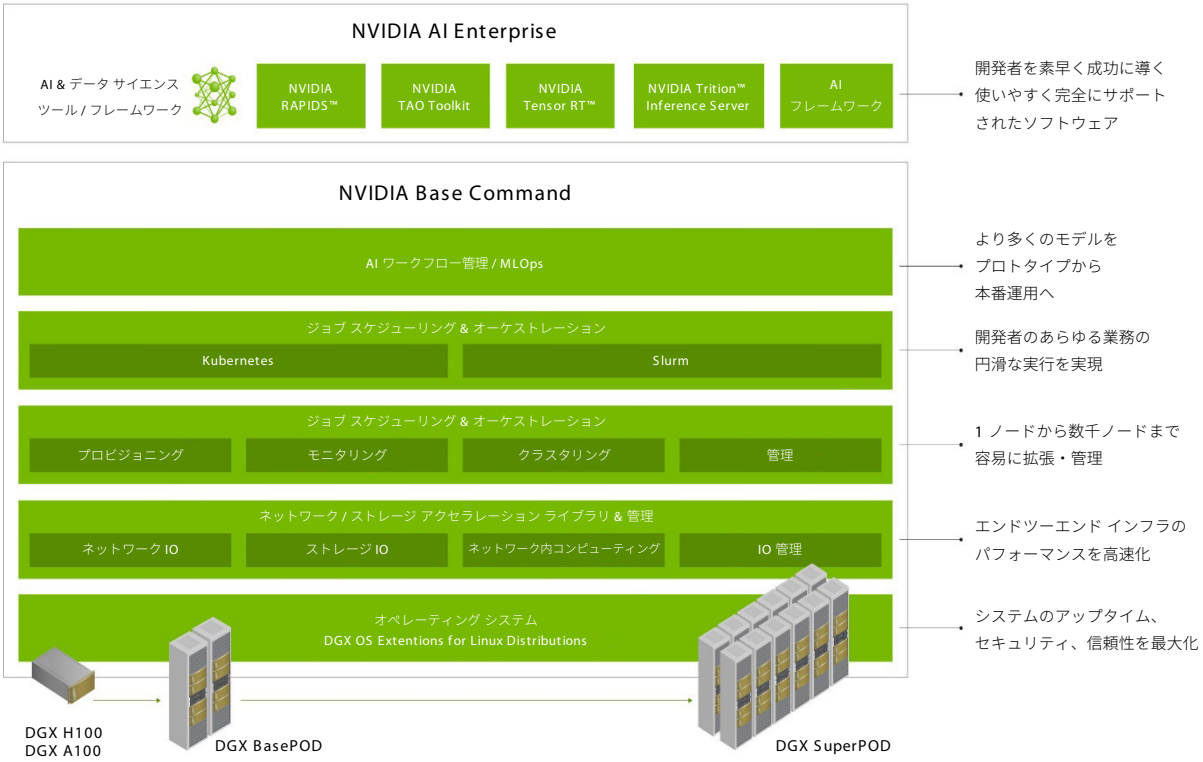
仕様	
GPU	8x NVIDIA H100 Tensor コア GPU
GPU メモリ	合計 640GB
パフォーマンス	32 petaFLOPS FP8
NVIDIA® NVSwitch™	4x
システム消費電力	最大 10.2kW
CPU	デュアル Intel® Xeon® Platinum 8480C プロセッサ（合計 112 コア）、2.00GHz （ベース）、3.80GHz（最大ブースト時）
システム メモリ	2TB
ネットワーキング	4 個の OSFP ポートで 8 基のシングル ポート ConnectX-7 VPI ➢ 400Gb/s InfiniBand/Ethernet 2 基のデュアルポート QSFP112 NVIDIA ConnectX-7 VPI ➢ 400Gb/s InfiniBand / Ethernet
管理ネットワーク	10Gb/s オンボード NIC（RJ 45 付き） 100 Gb/s Ethernet NIC ホスト ベースボード管理コントローラー （BMC）（RJ 45 付き）
ストレージ	OS： 1.92TB NVMe M.2 2枚
内部ストレージ：	3.84TB NVMe U.2 8枚
ソフトウェア	NVIDIA AI Enterprise – AIソフトウェア NVIDIA Base Command - オーケストレーション、スケジューリング、クラスタ管理 DGX OS / Ubuntu
サポート	3 年間のビジネス標準ハードウェアおよび ソフトウェア サポートが付属
システム重量	130.45kg
梱包重量	170.45kg
システム寸法	高さ：356mm 幅：482.2mm 長さ：897.1mm
動作温度範囲	5 °C～ 30 °C（41 °F～ 86 °F）

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 の規模とパフォーマンスの限界を突破します。9倍のパフォーマンス、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）による2倍の高速ネットワークを実現し、NVIDIA DGX SuperPOD向けのハイスピード スケーラビリティを備えています。この次世代アーキテクチャは、ジェネレーティブ AI や自然言語処理、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。

NVIDIA Base Command を搭載

NVIDIA Base Command は DGX プラットフォームを強化し、NVIDIA のソフトウェア イノベーションの最大活用を可能にします。エンタープライズ グレードのオーケストレーションとクラスター管理、コンピューティング / ストレージ / ネットワーク インフラを高速化するライブラリ、AI ワークロードに最適化されたオペレーティング システムを含む実証済みのプラットフォームにより、DGX インフラの可能性を最大限に引き出すことができます。さらに、DGX インフラには AI の開発と展開を効率化するように最適化されたソフトウェア スイート「NVIDIA AI Enterprise」も含まれています。



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できるほか、各種マネージドおよびコロケーション オプションを通じて利用することも可能です。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 インフラと同様に容易に NVIDIA DGX H100 を活用、入手できるため、明日を待つことなくすぐ AI をビジネスに活用できます。

詳細はこちら

NVIDIA DGX H100 の詳細については、nvidia.com/DGX -H100 をご覧ください。

© 2023 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA、NVIDIA ロゴ、BaseCommand,、ConnectX、DGX、DGX SuperPOD、および NVSwitch は NVIDIA Corporation およびその子会社の米国およびその他の国における商標または登録商標です。その他のすべての商標および著作権はそれぞれの所有者に帰属します。仕様は予告なく変更される場合があります。2023 年 3 月