



NVIDIA DGX H200

AI インフラストラクチャの金字塔。

AIは、困難なビジネス課題を解決するための最善のアプローチとなっています。顧客サービスの向上、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、生成 AI やその他のトランスフォーマーモデルを使用した最先端の製品やサービスの設計など、AI はほぼすべての業界の組織にイノベーションを実現するメカニズムを提供します。

そして、AI インフラストラクチャのパイオニアである NVIDIA DGX™ は、これらの重要なアイデアを実現するための最も強力な完全な AI プラットフォームを提供します。NVIDIA DGX H200 は、ビジネスのイノベーションと最適化を推進します。

NVIDIA の定評ある DGX platform の一部であり、NVIDIA DGX SuperPOD™ と DGX BasePOD™ の基盤である DGX H200 は、画期的な NVIDIA H200 Tensor Core GPU を搭載した AI の強力なツールです。このシステムは 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、体系化されたスケーラブルなプラットフォームを企業に提供し、自然言語処理、レコメンデーション システム、データ分析などの分野での画期的な進歩を支援します。

オンプレミスで利用できる DGX H200 は、さまざまなアクセスおよび展開オプションを通じて、企業が AI で最大の課題を解決するために必要なパフォーマンスを提供します。

AI センターオブエクセレンスの第一歩

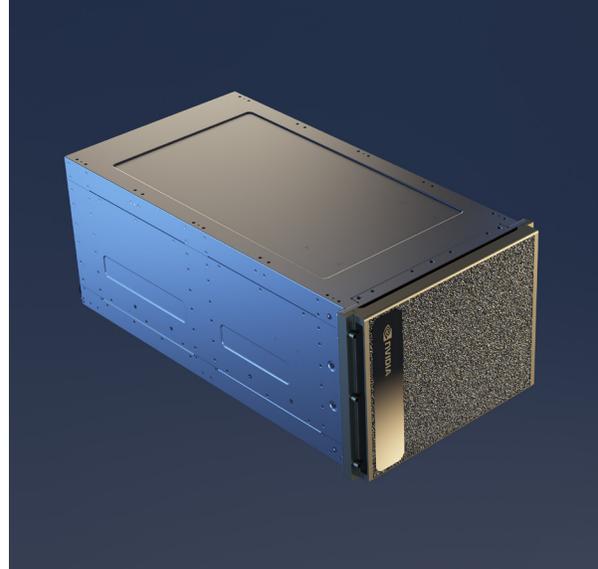
AI は科学とビジネスの間の溝を埋めました。もはや実験の領域ではなく、AI は大小さまざまな企業で日々使用されており、イノベーションを促進し、ビジネスを最適化しています。世界初の専用 AI インフラストラクチャポートフォリオの一部である DGX H200 は、エンタープライズ AI センター オブ エクセレンスの中心となるように設計されています。

これは完全に最適化されたハードウェアおよびソフトウェア プラットフォームであり、新しい一連の NVIDIA AI ソフトウェア ソリューション向けの NVIDIA Enterprise Support、サードパーティ サポートの充実したエコシステム、NVIDIA スペシャリストからの専門的なアドバイスへのアクセスが含まれており、組織は AI によって最大かつ最も複雑なビジネス上の問題を解決できます。

DGX H200 は信頼性が実証されており、DGX プラットフォームは世界中のほぼすべての業界の何千もの顧客に使用されています。

AI スケーリングの障壁を打ち破る

NVIDIA DGX H200 は、AI のスケールとパフォーマンスの限界を打ち破ります。32 ペタフロップスの AI 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) により DGX A100 より 2 倍高速なネットワーク、NVIDIA DGX SuperPOD および DGX BasePOD の高速スケーラビリティを実現します。DGX H200 は、生成 AI、自然言語処理、ディープラーニング推奨モデルなど、最大規模で最も複雑な AI トレーニングおよび推論ジョブに対応するために、1,128 GB の GPU メモリを搭載しています。

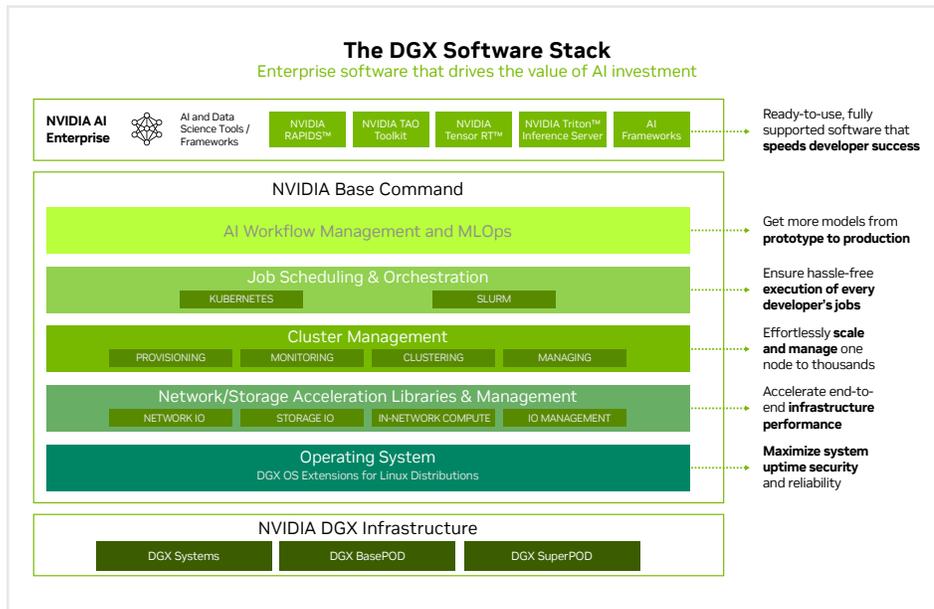


Specifications

GPU	8x NVIDIA H200 Tensor Core GPUs, with 141GB of GPU memory each
GPU memory	1,128GB total
Performance	32 petaFLOPS FP8
NVIDIA NVSwitch™	4x
System power usage	10.2kW max
CPU	Dual Intel® Xeon® Platinum 8480C Processors 112 Cores total, 2.00 GHz (Base), 3.80 GHz (Max Boost)
System memory	2TB
Networking	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI <ul style="list-style-type: none"> Up to 400Gb/s InfiniBand/Ethernet 2x dual-port QSFP112 NVIDIA ConnectX-7 VPI <ul style="list-style-type: none"> Up to 400Gb/s InfiniBand/Ethernet
Management network	10Gb/s onboard NIC with RJ45 100Gb/s Ethernet NIC Host baseboard management controller (BMC) with RJ45
Storage	OS: 2x 1.92TB NVMe M.2
Internal storage:	8x 3.84TB NVMe U.2
Software	NVIDIA AI Enterprise – Optimized AI software NVIDIA Base Command – Orchestration, scheduling, and cluster management DGX OS / Ubuntu / Red Hat Enterprise Linux / Rocky – Operating System
Support	Comes with 3-year business-standard hardware and software support
System weight	287.6lbs (130.45kgs)
Packaged system weight	376lbs (170.45kgs)
System dimensions	Height: 14.0in (356mm) Width: 19.0in (482.2mm) Length: 35.3in (897.1mm)
Operating temperature range	5–30°C (41–86°F)

NVIDIA Base Command を搭載

NVIDIA Base Command は DGX プラットフォームを強化し、組織が NVIDIA ソフトウェア イノベーションの最高の成果を活用できるようにします。企業は、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 インフラストラクチャ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を使用して、DGX インフラストラクチャの潜在能力を最大限に引き出すことができます。さらに、DGX インフラストラクチャには NVIDIA AI Enterprise が含まれており、AI の開発と展開を効率化するために最適化された一連のソフトウェアを提供します。



NVIDIA DGX AI Software Stack

リーダーシップクラスのインフラを思いのままに

ビジネス向け AI は、パフォーマンスや機能だけではありません。組織の IT 環境や業務にうまく適合することも重要です。DGX H200 は、直接管理するためにオンプレミスにインストールすることも、NVIDIA DGX-Ready データ センターに共存させることも、NVIDIA 認定のマネージド サービス プロバイダーを通じてアクセスすることもできます。また、DGX-Ready ライフサイクル管理プログラムにより、組織は予測可能な財務モデルを手に入れ、最先端の展開を維持できます。これにより、DGX H200 は従来の IT インフラストラクチャと同じくらい簡単に使用および取得でき、多忙な IT スタッフに追加の負担がかかりません。これにより、組織は明日まで待つのではなく、今すぐ AI をビジネスに活用できます。

Ready to Get Started?

To learn more about NVIDIA DGX H200, visit [nvidia.com/DGX-H200](https://www.nvidia.com/DGX-H200)

© 2024 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, Base Command, ConnectX, DGX, DGX BasePOD, DGX SuperPOD, and NVSwitch are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. Specifications are subject to change without notice. 3306343. JUN24

