



NVIDIA DGX H100

The Gold Standard for AI Infrastructure

人工知能(以下、AI)は、ビジネス上の難題を解決するための有力なアプローチとなりました。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におけるデータ抽出、最先端の製品・サービスの設計など、ほぼすべての産業・分野でイノベーションを実現するため仕組みを提供しています。そしてAIインフラのパイオニアであるNVIDIA DGX™システムは、そういった不可欠な要望を実現するための、最も強力で完全なAIプラットフォームとなります。

NVIDIA DGX H100は、ビジネスの革新と最適化を促進させます。NVIDIAの伝説的なDGXシステムの最新版であるNVIDIA DGX SuperPOD™の基盤となるDGX H100は、画期的なNVIDIA H100 Tensor Core GPUを搭載したAIパワーハウスです。AIのスループットを最大化するように設計されており、システム化され、高度に洗練された拡張可能なプラットフォームを企業に提供することで、自然言語処理・レコメンドシステム・データ分析など様々な分野におけるブレークスルーの達成を支援します。DGX H100は、オンプレミスはもちろんのこと、多様なアクセスや導入オプションを通じて利用でき、AIによる課題解決に必要な性能を企業に提供します。

卓越したAIセンターの基礎となるもの

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はもはや実験の領域ではなく、大企業から中小企業までイノベーションを促進しビジネスを最適化するために日々活用されています。世界初の専用AIインフラの第4世代であるDGX H100は、企業の卓越したAIセンターの中心的存在となるように設計されています。また、完全に最適化されたハードウェアとソフトウェアのプラットフォームとして、新しく広げた領域に対するフルサポートを行います。そこには、NVIDIAのAIソフトウェアソリューション、サードパーティ・サポートの豊富なエコシステム、NVIDIAプロフェッショナルサービスによる専門家のアドバイスなどが含まれます。ほぼすべての業界にわたって世界中の何千もの顧客がDGXシステムを使用していることが物語るように、DGX H100は実証済みの信頼性を提供します。

AI拡張の障壁を打破する

NVIDIA H100 Tensor Core GPUを搭載した世界初のシステムとして、NVIDIA DGX H100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。6倍のパフォーマンス向上、NVIDIA ConnectX®-7 スマートネットワークインターフェースカード (SmartNICs) と NVIDIA BlueField®-3 データ処理ユニット (DPUs) による2倍のネットワーク高速化、NVIDIA DGX SuperPODの高速スケーラビリティを特徴としています。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ドモデルなど、膨大かつ最も複雑なAIジョブのために超高速化されています。

SPECIFICATIONS

GPU	8x NVIDIA H100 Tensor Core GPUs
GPU memory	640GB total
Performance	32 petaFLOPS FP8
NVIDIA® NVSwitch™	4x
System power usage	~10.2kW max
CPU	Dual x86
System memory	2TB
Networking	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI 400 Gb/s InfiniBand or 200 Gb/s Ethernet 2x dual-port NVIDIA ConnectX-7 VPI 1x 400 Gb/s InfiniBand 1x 200 Gb/s Ethernet
Management network	10 Gb/s onboard NIC with RJ45 50 Gb/s Ethernet optional NIC Host baseboard management controller (BMC) with RJ45
Storage	OS: 2x 1.9TB 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 Ubuntu / Red Hat Enterprise Linux / CentOS – Operating system
Support	Comes with 3-year business-standard hardware and software support
Operating temperature range	5–30°C (41–86°F)

お客様のご要望に応じたリーダーシップクラスのインフラストラクチャ

ビジネス向けのAIは、パフォーマンスや機能だけではありません。それは、組織のIT環境と慣習にきちんと適合することでもあります。DGX H100は、オンプレミスで設置して直接管理することも、NVIDIA DGXの受け入れ態勢が整ったデータセンターに併置することも、NVIDIA DGX 製造工場からレンタルすることも、NVIDIA認定のサービスプロバイダーを経由してアクセスすることも可能です。また、DGX-Ready Lifecycle Managementプログラムにより、企業はその導入を最新に維持するための財務モデルをあらかじめ把握できます。これにより、DGX H100は従来のITインフラと同様に取得および使用が容易となり、多忙なITスタッフに新たな負担をかけることなく、企業は今日からビジネスにAIを活用できるようになります。

Ordering Information

NVIDIA DGX H100 is coming late 2022. Contact your NVIDIA sales representative for more details.

Learn More

To learn more about NVIDIA DGX H100, visit
[nvidia.com/DGXH100](https://www.nvidia.com/DGXH100)

© 2022 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, Base Command, ConnectX, DGX, 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. 2372184. AUG22



株式会社 HPC テック
<https://www.hpctech.co.jp>
mail : info@hpctech.co.jp

本社 : 〒103-0006 東京都中央区日本橋富沢町 7-13
TEL: 03-5643-2681 FAX: 03-5643-2682
大阪営業所 : 〒532-0011 大阪市淀川区西中島4丁目5-1
TEL 06-6195-6464 FAX 06-6195-6468