



NVIDIA DGX H200

AI 인프라의 표준

인공 지능은 이제 까다로운 비즈니스 문제를 해결할 때 가장 많이 찾는 솔루션으로 자리를 잡았습니다. 고객 서비스를 개선하거나, 공급망을 최적화하거나, 비즈니스 인텔리전스를 추출하거나, 생성형 AI와 기타 트랜스포머 모델을 사용해 최첨단 제품 및 서비스를 설계할 때도 AI는 산업 전반에 걸쳐 기업들에게 새로운 혁신 메커니즘을 제공하기 때문입니다. AI 인프라의 선구자인 NVIDIA DGX™는 가장 강력하고 완전한 성능을 바탕으로 중요한 아이디어를 실현할 수 있는 AI 플랫폼입니다.

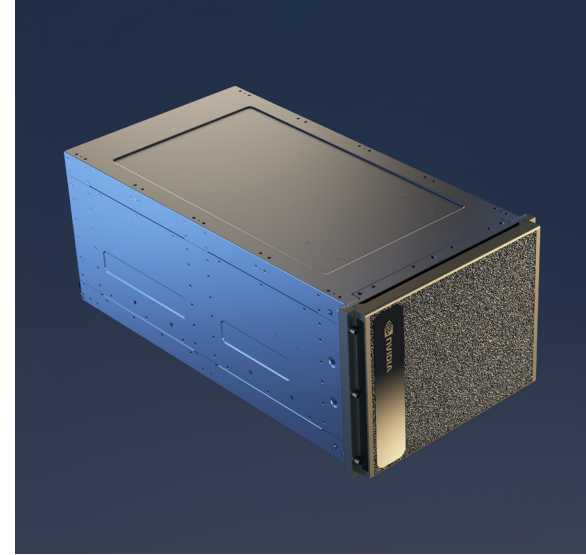
NVIDIA DGX H200은 비즈니스 혁신과 최적화를 지원합니다. DGX H200은 NVIDIA가 자랑하는 **DGX 플랫폼**을 구성할 뿐만 아니라 **NVIDIA DGX SuperPOD™**와 **DGX BasePOD™**를 기반으로 개발 되었으며, 획기적인 **NVIDIA H200 Tensor 코어 GPU**와 Intel® Xeon® 플래티넘 프로세서까지 탑재되어 AI의 성장 동력이라고 해도 과언이 아닙니다. DGX H200은 AI 처리량을 극대화하도록 설계되어 기업에게 매우 정교하고, 체계적이고, 확장 가능한 플랫폼을 제공하기 때문에 자연어 처리, 추천 시스템, 데이터 분석 등에서 문제를 해결하는 데 효과적입니다. 그 밖에도 온프레미스 환경에서 사용할 수 있고, 광범위한 액세스 및 배포 옵션도 제공하여 AI로 까다로운 문제를 해결하기 위해 필요한 성능을 제공합니다.

AI Center of Excellence의 초석

AI는 과학과 비즈니스를 연결하는 가교입니다. 이제는 실험의 단계에서 벗어나 대기업이든, 중소기업이든 상관없이 일상적으로 혁신을 앞당기고 비즈니스를 최적화할 수 있습니다. 세계 최초의 특수 목적 AI 인프라 포트폴리오를 구성하는 DGX H200은 기업의 AI Center of Excellence에서 핵심 역할을 합니다. 또한 완전히 최적화된 하드웨어/소프트웨어 플랫폼으로 새로운 NVIDIA AI 소프트웨어 솔루션을 위한 **NVIDIA Enterprise Support**, 풍부한 서드파티 지원 에코시스템, NVIDIA 전문 서비스를 통한 전문가 자문 등이 포함되어 아무리 크고 복잡한 비즈니스 문제라고 해도 AI를 활용해 해결할 수 있습니다. 산업 전반에 걸쳐 전 세계 수많은 고객들이 DGX 플랫폼을 사용하고 있기 때문에 DGX H200의 안정성 또한 이미 검증된 셈입니다.

대규모 AI를 가로막는 장애물 해결

NVIDIA DGX H200은 NVIDIA H200 Tensor 코어 GPU 8개와 Intel Xeon 프로세서 2개를 탑재하여 AI 규모 및 성능의 한계를 허물었습니다. 또한 32 petaFLOPS의 AI 성능, NVIDIA ConnectX®-7 스마트 네트워크 인터페이스 카드(SmartNICs)가 탑재되어 DGX A100과 비교했을 때 2배 더 빨라진 네트워킹 속도, 그리고



Specifications

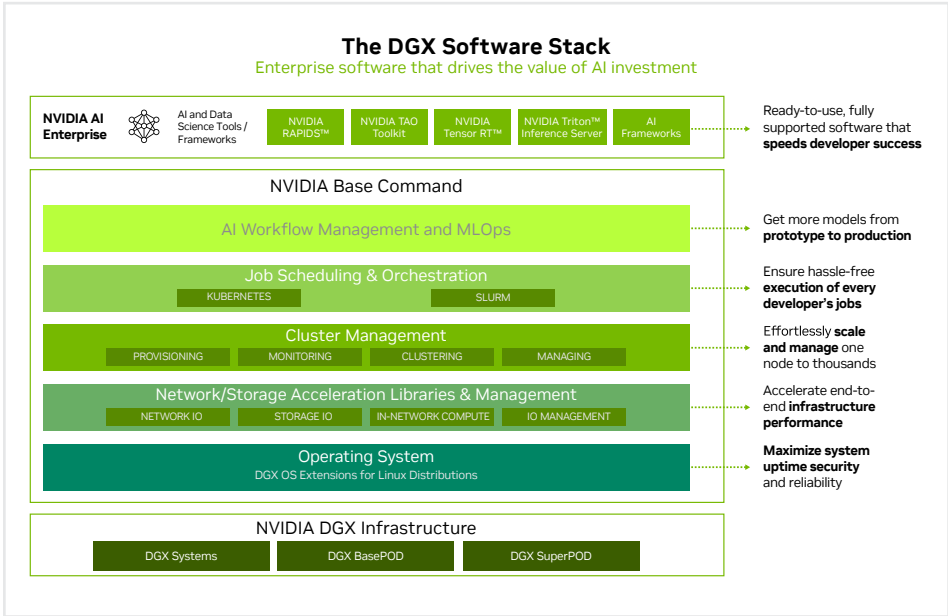
GPU	8x NVIDIA H200 Tensor Core GPUs, with 141GB of GPU memory each
GPU memory	1,128GB total
Performance	32 petaFLOPS FP8
NVIDIA NVSwitch™	4x
System power usage	10.2kW max
CPU	Dual Intel® Xeon® Platinum 8480C Processors 112 Cores total, 2.00 GHz (Base), 3.80 GHz (Max Boost)
System memory	2TB
Networking	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI ➢ Up to 400Gb/s InfiniBand/Ethernet 2x dual-port QSFP112 NVIDIA ConnectX-7 VPI ➢ 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 DGX SuperPOD와 DGX BasePOD를 위한 고속 확장성을 제공합니다. 그 밖에도 1,128GB의 GPU 메모리로 한층 더 강력해져 생성형 AI, 자연어 처리, 딥 러닝 추천 모델 등 아무리 크고 복잡한 AI 훈련 및 추론 작업도 간단히 처리합니다.

NVIDIA Base Command 탑재

NVIDIA Base Command는 DGX 플랫폼에 탑재되어 기업들이 혁신적인 NVIDIA 소프트웨어를 최대한 이용할 수 있도록 지원합니다. 덕분에 기업들은 엔터프라이즈급 오케스트레이션 및 클러스터 관리, 컴퓨팅/스토리지/네트워크 인프라를 가속하는 라이브러리, AI 워크로드에 최적화된 운영 체제가 포함되어 이미 성능이 검증된 플랫폼을 이용해 DGX 인프라의 잠재력을 극대화할 수 있습니다. 또한 AI 개발 및 배포를 간소화할 수 있는 소프트웨어 제품군을 제공하는 NVIDIA AI Enterprise는 DGX 시스템에 최적화되어 있습니다.

NVIDIA NIM™ 추론 마이크로서비스는 속도, 사용 편의성, 관리 용이성, 보안 기능을 제공하여 모델 배포를 최적화합니다.



NVIDIA DGX AI 소프트웨어 스택

기업의 요건을 고려한 최고의 인프라

단순히 성능과 기능만으로는 비즈니스용 AI라고 말할 수 없습니다. 기업의 IT 운영 및 실무에 원활하게 통합되어야 합니다. DGX H200은 직접 관리를 위해 온프레미스 환경에 설치할 수도 있고, **NVIDIA DGX-Ready 데이터 센터**에서 코로케이션 설치도 가능하며, **NVIDIA 인증 서비스 공급업체**를 통한 액세스도 지원합니다.

기업이 **DGX-Ready 라이프사이클 관리 프로그램**에 가입하면 예측 가능한 재정 모델을 통해 최신 플랫폼으로 업그레이드할 수 있습니다. 바쁜 IT 인력에게 가중되는 부담이 없기 때문에 기존 IT 인프라 못지않게 DGX H200을 손쉽게 사용하고 구매할 수 있으며, 기업은 지금 바로 AI를 비즈니스에 활용할 수 있습니다.

시작할 준비가 되었습니까?

이제 NVIDIA H200 Tensor 코어 GPU와 Intel Xeon 프로세서가 탑재된 NVIDIA DGX H200을 사용해 AI 기반 인사이트를 더욱 빠르게 얻으십시오. 자세한 내용은 nvidia.com/DGX-H200에서 확인할 수 있습니다.

© 2024 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, NVIDIA 로고, Base Command, ConnectX, DGX, DGX BasePOD, DGX SuperPOD, NVSwitch는 미국을 비롯한 기타 국가에 위치한 NVIDIA 기업과 계열사의 상표 및/또는 등록 상표입니다. 그 밖의 모든 상표와 저작권은 각 소유권자의 재산입니다. 사양은 사전 공지 없이 변경될 수 있습니다. 3412155. AUG24

