



NVIDIA DGX B300

훈련부터 추론까지, AI 팩토리 성능의 새로운 기준을 제시



모든 기업에 전례 없는 성능 제공

산업 전반에 걸쳐 단기간에 AI 도입이 폭발적으로 증가함에 따라, 기업의 AI 전환 방식에 근본적인 변화를 가져오고 있습니다. 진보하는 기술의 도입에 회의적이었던 조직들도 이제는 앞다투어 데이터 센터에 적합한 인프라를 구축하고, AI를 활용하기 위한 인재 확보에 나서고 있습니다. 그러나 많은 기업들이 실제 AI를 배포하는 것이 계획만큼 간단하지 않다는 사실을 깨닫고 있습니다.

생성형 AI의 혁신적인 잠재력에도 불구하고, 기업들은 이들 기술을 도입 및 확장하는 데 있어 여러 공통된 과제에 직면하고 있습니다. 통합 복잡성, 전문성 부족, 에너지 소비와 비용 문제 등에 적합한 솔루션을 필요로 합니다. 기존의 하이퍼스케일러처럼 AI를 확장하고 운영할 역량을 갖추지 못한 기업들이 대부분입니다.

NVIDIA DGX SuperPOD의 빌딩 블록인 NVIDIA DGX™ B300은 이러한 문제를 해결하기 위해 설계된 목적 기반의 AI 인프라 솔루션으로, AI 추론의 연산 수요를 충족하고, 풀스택 소프트웨어를 활용하여 AI를 효율적으로 배포 가능하게 하여 기업의 부담을 덜어줍니다. NVIDIA Blackwell Ultra GPU 기반의 DGX B300은 144 petaFLOPS의 추론 성능과 72 petaFLOPS의 훈련 성능을 제공하며, 현대의 데이터 센터와 끊임 없이 통합될 수 있도록 새로운 폼팩터로 설계되었습니다. 이는 NVIDIA MGX™뿐만 아니라 기존의 엔터프라이즈 랙과도 호환 가능합니다. 이제는 어떤 기업이든 DGX B300을 사용하여 전례 없는 효율성으로 다양한 AI 워크로드를 훈련하고 추론할 수 있습니다.

실시간 AI의 최강자

DGX B300이 실시간 추론 능력에서 큰 도약을 이루면서, 기존에는 하이퍼스케일러로만 가능했던 AI 성능들을 이제는 소규모 기업에서도 활용 가능하게 되었습니다.

DGX B300은 NVIDIA Blackwell Ultra GPU와 NVIDIA ConnectX-8 네트워킹과 완전 통합된 세계 최초 시스템으로, 최적화된 NVIDIA Mission Control 소프트웨어를 기반으로 이전 세대 대비 추론 성능은 11배, 훈련 성능은 4배 향상되었습니다.

이제는 DGX B300으로 어떤 조직에서도 AI 추론 시대의 새로운 가능성들을 실현할 수 있습니다.

주요 특징

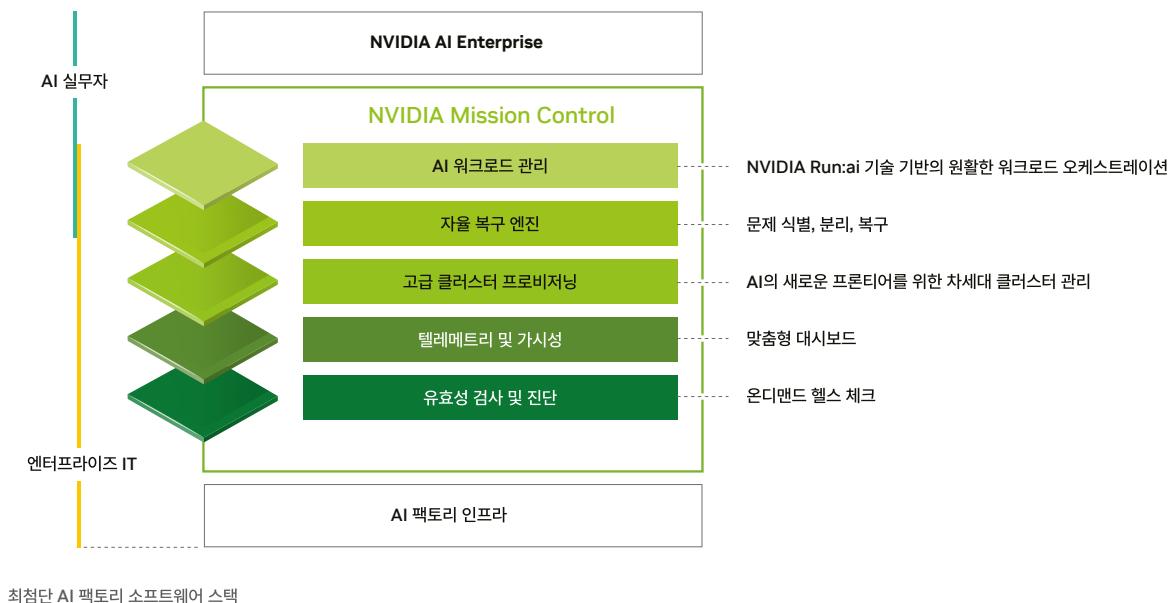
- > NVIDIA Blackwell Ultra GPU 탑재
- > 2.3TB GPU 메모리 공간
- > 72 petaFLOPS의 훈련 성능
- > 144 petaFLOPS의 추론 성능
- > NVIDIA 네트워킹
- > 듀얼 Intel® Xeon® 프로세서
- > NVIDIA DGX BasePOD™ 및 DGX SuperPOD™ 기반
- > NVIDIA AI Enterprise 및 NVIDIA Mission Control 소프트웨어 활용

현대의 데이터 센터를 위한 청사진

DGX B300은 NVIDIA MGX 랙에 맞춰 재설계된 새시를 도입하여, 현대의 데이터 센터에 맞는 호환성과 확장성을 보장합니다. 공냉식으로 설계되어 기존 인프라에 쉽게 통합이 가능합니다. 또한, 업계 최초로 버스바 또는 PSU 전원 방식 중 선택이 가능한 유연한 전력 아키텍처를 제공함으로써, 기업에서 기존의 인프라와 지속가능성 목표에 맞춰 선택이 가능합니다. 이러한 새로운 DGX 설계는 가속 컴퓨팅 인프라를 위한 최적화된 설계 청사진을 제공하여, 규모에 맞춰 AI 인프라를 구축, 배포 할 수 있도록 유연한 기반을 제공합니다. DGX B300은 최첨단 설계에 실용적인 유지보수성을 결합하고 적응력과 효율성을 높여 미래를 대비한 AI 인프라의 새로운 표준을 제시합니다.

NVIDIA Mission Control로 모델 실행과 핵심 업무 자동화

NVIDIA Mission Control은 세계 최고 수준의 운영 역량을 소프트웨어 형태로 조직에 제공하여, 개발자 워크로드부터 인프라, 시설 관리까지 AI 팩토리 운영의 모든 측면을 제어하는 플랫폼입니다. 이를 통해 훈련과 추론의 즉각적인 민첩성을 높이는 동시에, 인프라 복원력을 위해 풀스택 인텔리전스를 제공합니다. NVIDIA Mission Control은 어떤 기업이든 하이퍼스케일 수준의 효율성으로 AI를 구동하여 AI 실험을 가속화할 수 있도록 합니다. 또한, AI 개발과 배포를 간소화하는 소프트웨어 제품군인 NVIDIA AI Enterprise가 NVIDIA DGX 시스템에 최적화되어 있으며, NVIDIA NIM™ 마이크로서비스를 사용하여 속도, 사용 편의성, 관리 용이성, 보안성과 함께 최적의 모델을 배포할 수 있습니다.



DGX B300 Technical Specifications

	DGX B300
GPU	NVIDIA Blackwell Ultra GPUs
Total GPU Memory	2.3TB
Performance	144 PFLOPS FP4 inference* 72 PFLOPS FP8 training*
NVIDIA NVLink Switch System	2x
NVIDIA NVLink™ Bandwidth	14.4 TB/s aggregate bandwidth
Power Consumption	~14kW
CPU	Dual Intel® Xeon® Processors
Networking	8x OSFP ports serving 8x single-port NVIDIA ConnectX-8 VPI >Up to 800Gb/s NVIDIA InfiniBand/Ethernet 2x dual-port QSFP112 NVIDIA BlueField-3 DPU >Up to 800Gb/s NVIDIA InfiniBand/Ethernet
Management Network	1GbE onboard NIC with RJ45 1GbE RJ45 Host baseboard management controller (BMC)
Storage	OS: 2x 1.9TB NVMe M.2 Internal storage: 8x 3.84TB NVMe E1.S
Software	NVIDIA DGX OS / NVIDIA Mission Control / NVIDIA Base Command Manager / NVIDIA AI Enterprise Supports Red Hat Enterprise Linux / Rocky / Ubuntu
Rack Units	10RU
Support	Three-year business-standard hardware and software support

*Shown with sparsity.

Ready to Get Started?

To learn more about DGX B300, visit nvidia.com/dgx-b300

© 2025 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, ConnectX, DGX, DGX BasePOD, DGX SuperPOD, Mission Control, MGX, NIM, and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective owners with which they are associated. 3725750. MAR25

