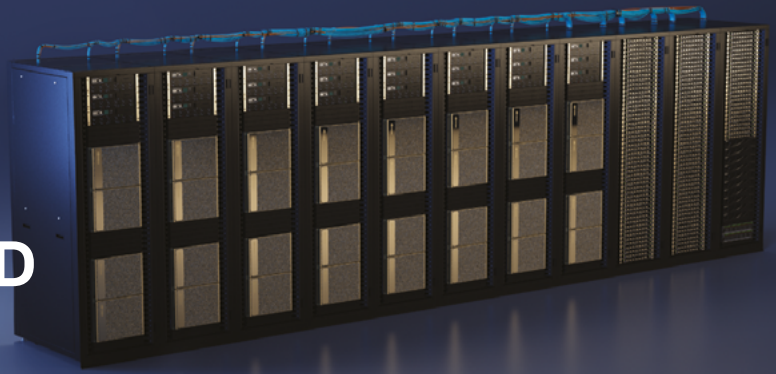




# NVIDIA DGX SuperPOD

오늘날 엔터프라이즈를 위한 AI 팩토리 인프라



NVIDIA DGX SuperPOD™는 어떤 사용자와 워크로드에도 동일한 성능을 제공하는 AI 팩토리 솔루션입니다. DGX SuperPOD는 NVIDIA의 독보적인 AI 슈퍼컴퓨터 설계 및 운영 경험을 바탕으로 하여, 기업에서 몇 개월이 아닌 단 몇 주 만에 AI 인프라를 가동 가능하게 하는 풀스택 AI 인프라 솔루션입니다. DGX SuperPOD는 조직 내 IT 부문에 선도적인 가속 인프라 및 확장성과 함께 업계에서 입증된 결과를 제공하여, 가장 까다로운 AI 훈련 및 추론 워크로드를 처리 가능하게 합니다.

## 풀스택 AI 팩토리 인프라

NVIDIA DGX SuperPOD는 AI 컴퓨팅, 네트워크, 스토리지, 소프트웨어를 설계에 최적화되도록 조합하였습니다. 고밀도, 고성능, 유연성을 제공하는 컴퓨팅 기반에 초고속 대역폭과 저지연시간의 NVIDIA 네트워킹이 상호 연결되어 있습니다. 또한, 인증된 스토리지 솔루션을 통해, 고성능 데이터 패브릭을 통합하여 AI 워크로드를 처리할 수 있습니다. 전체 스택에 최적화된 소프트웨어를 통해 개발자 워크플로우에 맞는 세계적 수준의 도구와 완전한 클러스터 관리를 위한 솔루션을 제공하여, 조직의 운영 부담은 줄이고 자원 활용은 극대화합니다.



NVIDIA DGX SuperPOD architecture

## DGX SuperPOD

### 하드웨어

- > NVIDIA DGX™ GB300, DGX GB200, DGX B300, 또는 DGX B200 시스템
- > NVIDIA 네트워킹
- > 고성능 스토리지

### 소프트웨어

- > NVIDIA AI Enterprise
- > NVIDIA Mission Control
- > NVIDIA DGX 운영 체제

### 라이프사이클 서비스\*

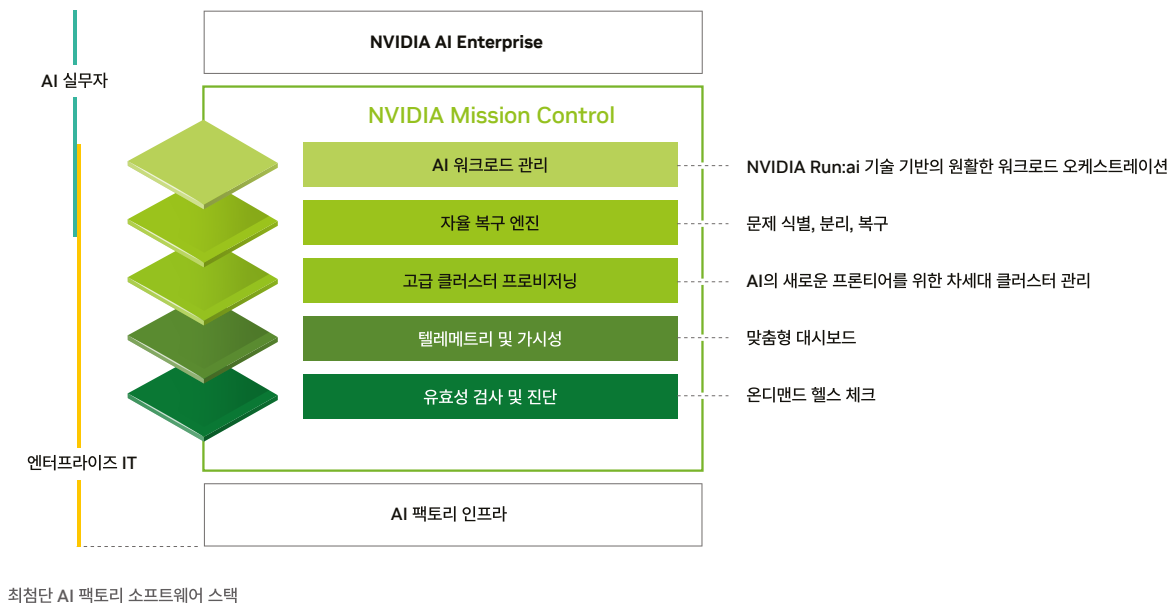
- > 기획 및 배포\*\*
  - 데이터 센터 및 시스템 기획/설계
  - 사전 설치 준비
- > 교육 및 최적화
  - 현장의 물리적 및 논리적 배포
  - 테스트 및 검증
  - 애플리케이션 성능 테스트
  - 현장 문서화 패키지
  - 사용자/DevOps 교육
  - 워크로드 기반 NVIDIA 딥러닝 연구소(DLI) 교육
  - 24/7 전문가 지원

\* NVIDIA와 파트너사가 공동으로 제공

\*\* 온프레미스 또는 DGX-Ready 데이터 센터에서 배포

# 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.



## AI의 새로운 프론티어를 위한 대규모 슈퍼컴퓨팅

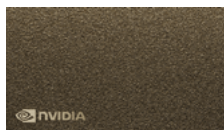
DGX SuperPOD는 수만 개의 NVIDIA GPU로 확장 가능하며 최첨단 생성형 AI 모델의 훈련과 추론을 처리하는 슈퍼컴퓨팅 플랫폼입니다. DGX SuperPOD는 업계 최고의 성능과 확장성을 제공하여 최대 규모의 AI 워크로드를 처리하는 동시에, 운영 비용과 인프라 복잡성을 최소화합니다. DGX SuperPOD는 엔터프라이즈 AI 추론을 위한 세계 최고의 플랫폼으로, 업계 최대 규모의 NVIDIA NVLink 도메인, 독보적인 총 GPU 메모리 용량과 GPU 간 메모리 대역폭, 선도적인 컴퓨팅 성능을 제공합니다. AI 훈련 측면에서는 DGX SuperPOD 아키텍처에 기반한 슈퍼컴퓨팅 클러스터를 통해 기업의 AI 운영 속도를 획기적으로 향상시키고 있으며, 매년 TOP500 및 Green500의 상위권에 오르며 MLPerf 벤치마크 기록을 갱신하고 있습니다.<sup>2</sup>



NVIDIA DGX GB300



NVIDIA DGX GB200



NVIDIA DGX B300



NVIDIA DGX B200

## NVIDIA가 보장하는 완전한 전문 서비스 라이프사이클

기업에서는 아키텍처 설계 이상의, 가속 컴퓨팅 인프라를 실질적인 비즈니스 가치로 신속하게 전환하는 방법을 필요로 합니다. 이를 위해서는 IT 환경에 최적화된 신속한 턴키 방식의 구축 경험이 중요합니다. 이러한 경험을 바탕으로 데이터 과학자가 1일 차부터 바로 사용할 수 있고, 시간이 지남에 따라 지속적으로 향상시킬 수 있는 환경이 필요합니다. NVIDIA DGX SuperPOD는 시작부터 끝까지 전체 라이프사이클을 아우르는 완전한 서비스를 제공합니다. 기획부터 설치, 검증, 인프라 관리, 운영에 이르는 전 과정에서 전담 가이드 및 지원 서비스를 제공합니다. DGX SuperPOD 전체 스택에 걸쳐서 NVIDIA 전문가들의 지원을 받을 수 있어 기업이 AI 워크로드에서 항상 최상의 성능을 유지할 수 있도록 돕습니다.

## AI 성공을 이끄는 경험

DGX SuperPOD는 매일 DGX SuperPOD 플랫폼을 활용해 새로운 혁신을 만들어내는 수천 명의 NVIDIA 연구원과 엔지니어의 전문성이 집약된 결과물입니다. 턴키 AI 팩토리 솔루션을 제공하여, 고객들은 자체 데이터 센터나 다양한 관리형 서비스 제공업체, 코로케이션 환경, 프라이빗 클라우드 등을 통해 이를 배포할 수 있습니다.



NVIDIA's DGX Supercomputer Eos is used for NVIDIA's research and development.

## 시작할 준비가 되셨나요?

NVIDIA DGX SuperPOD에 대한 자세한 정보는 [nvidia.com/dgx-superpod](https://nvidia.com/dgx-superpod)에서 확인해 보세요.

1. 자세한 내용은 [top500.org](https://top500.org)를 참조하세요.
2. 자세한 내용은 [mlperf.org](https://mlperf.org)를 참조하세요.