

NVIDIA DGX Station A100

AI시대를 위한 워크그룹 어플라이언스



데이터 사이언스 팀은 AI 혁신의 선두에 서서 기업과 세상을 변화시킬 수 있는 프로젝트를 개발하고 있습니다. 그러나 가장 복잡한 모델을 교육하는 데 도움이 될 수 있는 여러분의 컴퓨팅 자원을 찾는 경우가 많습니다. 이러한 팀에는 전 세계 어디에서나 여러 동시 사용자에게 획기적인 성능을 제공하기 위해 하드웨어 및 소프트웨어 전반에 걸쳐 완벽하게 최적화되고 어디서나 연결할 수 있는 전용 AI 플랫폼이 필요합니다.

NVIDIA DGX Station™ A100 개요

- > 2.5 peta FLOPS의 성능을 제공하는 AI 워크그룹 서버 교육, 추론 및 데이터 분석을 위해 팀이 제한 없이 사용할 수 있는 성능
- > 서버급, 플러그 앤 고, 데이터 센터 전력 및 냉각이 필요하지 않음
- > 복잡한 설치나 IT 지원 필요 없음
- > 4개의 완전히 상호 연결된 NVIDIA A100 Tensor Core GPU와 최대 320GB의 GPU 메모리를 갖춘 세계 유일의 워크스테이션 스타일 시스템
- > NVIDIA의 지식을 바탕으로 AI 혁신에 대한 빠른 경로 제안과 경험 제공

NVIDIA DGX Station A100은 AI 슈퍼컴퓨팅을 데이터 사이언스팀에 제공하여 데이터 센터 또는 추가 IT 인프라 없이 데이터 센터 기술을 제공 합니다.

여러 동시 사용자를 위해 설계된 DGX Station A100은 사무실 친화적인 폼 팩터에서 서버급 구성 요소를 활용합니다.

표준 전원 콘센트에 연결할 수 있는 최대 320GB의 총 GPU 메모리를 갖춘 4개의 완전히 상호 연결된 다중 인스턴스 GPU(MIG) 지원.

NVIDIA A100 Tensor Core GPU가 있는 유일한 시스템으로, 어디에나 두십시오.

데이터 사이언스팀을 위한 AI 슈퍼컴퓨팅 다수의 동시 사용자에게 중앙 집중식 AI 리소스를 손쉽게 제공하는

DGX Station A100은 AI 시대를 위한 워크그룹 어플라이언스 입니다 교육, 추론 및 분석 워크로드를 병렬로 실행할 수 있으며 MIG를 사용하여 개별 GPU 장치를 최대 28개까지 제공할 수 있습니다.

DGX Station A100은 3배 이상 빠른 교육 성능을 제공합니다.



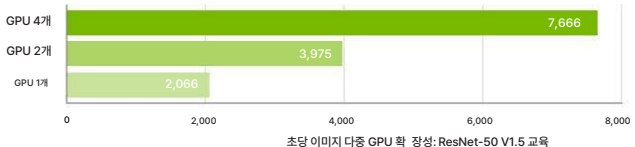
DGX 스테이션 A100 320GB; 배치 크기=64; 혼합 정밀도; AMP; 실제 데이터; 시퀀스 길이=128

DGX Station A100은 4배 이상 빠른 추론 성능을 제공합니다.



DGX 스테이션 A100 320GB; 배치 크기=256; INT8 정밀도; 합성 데이터; 시퀀스 길이=128, cuDNN 8.0.4

선형 확장성을 제공하는 DGX Station A100



DGX 스테이션 A100 320GB; 배치 크기=192; 혼합 정밀도; 실제 데이터; cuDNN 버전=8.0.4; NCCL 버전=2.7.8; NGC MXNet 20.10 컨테이너

활동이 포함되고 시스템 전체의 성능에 영향을 미치지 않도록 사용자 및 작업. DGX Station A100은 모든 DGX 시스템과 마찬가지로 완벽하게 최적화된 NVIDIA DGX™ 소프트웨어 스택을 갖추고 있어 개별 시스템에서 NVIDIA DGX POD™ 및 NVIDIA DGX SuperPOD™에 이르기까지 DGX 기반 인프라와의 완벽한 상호 운용성과 최대 성능을 제공하여 DGX Station A100을 이상적으로 만듭니다. 크고 작은 모든 조직의 팀을 위한 플랫폼입니다.

데이터센터가 없어도 데이터센터성능 구현
NVIDIA DGX Station A100은 워크스테이션 폼 팩터에서 데이터 센터급 AI 서버를 제공하며 특수 전력 및 냉각 없 이 표준 사무실 환경에서 사용하기에 적합합니다. 그 설계에는 4개의 초강력 NVIDIA A100 Tensor Core GPU, 최고급 서버급 CPU, 초고속 NVMe 스토리지 및 첨단 PCIe Gen4 버스가 포함됩니다. DGX Station A100에는 NVIDIA DGX A100과 동일한 BMC(Baseboard Management Controller)가 포함되어 있어 시스템 관리자가 원격 연결을 통해 필요한 모든 작업을 수행할 수 있습니다. DGX Station A100은 데이터센터 없이도 데이터센터 기술을 제공하는 가장 강력한 오피스 환경 AI 시스템입니다.

어디에나 놓을 수 있는 AI 어플라이언스
NVIDIA DGX Station A100은 기업 사무실, 연구실, 연구 시설 또는 심지어 집에서 작업하는 오늘날의 민첩한 데이터 사이언스 팀을 위해 설계되었습니다. 대규모 AI 인프라를 설치하려면 막대한 IT 투자와 강력한 전력 및 냉각 기능을 갖춘 대규모 데이터 센터가 필요한 반면, DGX Station A100은 팀의 작업 공간이 어디에 있든 표준 벽면 콘센트에 연결하기만 하면 됩니다. 또한 혁신적인 냉각 기반 설계로 만져도 시원함을 유지합니다. 간단한 1인 설정을 통해 단 2개의 케이블만 있으면 작동하는 세계적 수준의 AI 플랫폼을 몇 분 만에 가동할 수 있습니다.

더 큰 모델, 더 빠른 답변
NVIDIA DGX Station A100은 워크스테이션이 아닙니다. 책상 아래에 놓을 있는 AI 작업 그룹 서버입니다. 64코 어 데이터 센터급 CPU 외에도 NVIDIA DGX A100 서버와 동일한 NVIDIA A100 Tensor Core GPU를 갖추고 있으며 각각 40GB 또는 80GB의 GPU 메모리를 고속 SXM4를 통해 연결합니다. NVIDIA DGX Station A100은 완전히 상호 연결된 4개의 GPU가 있고 NVIDIA® NVLink®를 활용 하고 MIG를 지원하여 시스템 성능에 영향을 주지 않으면 서 병렬 작업과 여러 사용자를 위해 최대 28개의 개별 GPU 장치를 제공하는 유일한 사무용 시스템입니다.

타의 추종을 불허하는 AI 전문성에 대한 통합 액세스
NVIDIA DGX Station A100은 NVIDIA의 수천 명의 AI 전문가가 지원하고 세계 최대의 DGX 시험장인 NVIDIA DGX SATURNV에서 얻은 지식을 기반으로 구축된 완전한 하드웨어 및 소프트웨어 플랫폼입니다. DGX Station A100을 소유하면 NVIDIA DGXperts에 직접 액세스할 수 있습니다. NVIDIA DGXperts는 규범적 지침과 설계 전문 지식을 제공 하는 AI 유창한 실무자로 구성된 글로벌 팀으로 NVIDIA의 10년 이상의 AI 리더십에서 얻은 노하우와 경험으로 AI 혁신 을 빠르게 추적할 수 있도록 지원합니다. 이를 통해 미션 크리티컬 애플리케이션을 빠르게 시작하고 원활하게 실행하여 인사이트를 얻는 시간을 크게 단축할 수 있습니다.

시스템 사양

	NVIDIA DGX Station A100 320GB	NVIDIA DGX Station A100 160GB
GPU	4x NVIDIA A100 80GB GPU	4x NVIDIA A100 40GB GPU
GPU	메모리 총 320GB	총 160GB
성능	2.5 petaFLOPS AI 5 petaOPS INT8	
체계 전력 사용량	1.5 kW at 100~120 Vac	
CPU	단일 AMD 7742, 64코어, 2.25GHz(기본)~3.4GHz(최대 부스트)	
체계 메모리	512GB DDR4	
네트워킹	듀얼 포트 10Gbase-T 이더넷 LAN 단일 포트 1Gbase-T 이더넷 BMC 관리 포트	
스토리지	OS: 1.92TB NVME 드라이브 1개 내장 스토리지: 7.68TB U.2 NVME 드라이브	
DGX 디스플레이 어댑터	4GB GPU 메모리, 4x Mini DisplayPort	
시스템 음향	<37dB	
소프트웨어	우분투 리눅스 OS	
시스템무게	91.0파운드(43.1kg)	
포장된 시스템무게	127.7파운드(57.93kg)	
체계 치수	높이: 639mm(25.1인치) 너비: 10.1인치(256mm) 길이: 518mm(20.4인치)	
운영 온도 범위	5-35°C(41-95°F)	

제품 문의

☎ 02-6206-5024

✉ jysuh@krinfra.co.kr



06288) 서울특별시 강남구 삼성로 150(대치동, 극동교회빌딩 3층) | Tel. 02-6204-5000 | Fax. 02-6204-5099
51139) 경남 창원시 의창구 용동로 85 한마음빌딩 805호 | Tel. 055-262-8086 | Fax. 055-722-7218