

DELLTechnologies

GPU 가속 AI로 더 빠르게 혁신 실현

Dell Technologies 및 NVIDIA의 솔루션으로 인공
지능의 잠재력을 최대한 활용



 **NVIDIA**

목차

꿈임없이 변화하는 세상을 이끄는 AI	3
AI의 진입 장벽 해소	4
'AI 가능'에서 'AI 검증'으로 도약	5
AI 통찰력을 빠르게 확보하도록 설계	6
Dell PowerEdge 서버로 AI 이점 활용	6
Dell PowerEdge XE 서버	7
Dell PowerEdge 랙 서버	8
NVIDIA GPU로 AI 활용 극대화	9
NVIDIA 기술 내장	11
권장 구성	12
고객 성공 사례	13
대규모로 콘텐츠를 추천하는 Taboola	13
기차가 최고 속도를 유지하도록 하는 Duos Technologies	13
과학적 발견을 가속화한 케임브리지대학교	14
AI의 잠재력을 확장한 피사대학교	14
Dell Technologies를 선택해야 하는 이유	15
지능적인 성과 가속화	16

꿈임없이 변화하는 세상을 이끄는 AI

AI 시대에 데이터에 대한 액세스 증가와 새로운 데이터 관리 기법은 유형과 규모에 관계없이 모든 기업과 조직에 AI 기반 통찰력을 제공하는 동력이 되고 있습니다. 처리 성능의 발전에서 엔터프라이즈 멀티클라우드의 부상에 이르기까지 AI의 보편성은 기업이 온프레미스, 프라이빗 클라우드 및 퍼블릭 클라우드, 엣지에서 새롭게 떠오르는 다양한 워크로드에 대해 AI의 이점을 누릴 수 있게 합니다.



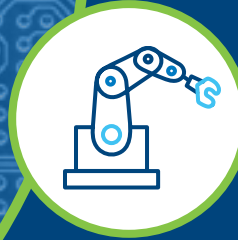
디지털 트윈

- 가상 오브젝트, 시스템 또는 프로세스를 시뮬레이션 하여 실제 동작을 예측합니다.
- 더 우수하고 빠르게 경제적인 R&D(Research and Development) 주기를 지원합니다.
- 비용과 시간이 많이 드는 물리적 프로토타입 제작에 투자하기 전에 제품을 개선합니다.



생성형 AI/NLP(Natural Language Processing)/LLM(Large Language Model)

- 기계가 인간의 언어를 이해할 수 있도록 합니다.
- 인간이 컴퓨터 시스템과 보다 개인 맞춤형 방식으로 상호 작용하도록 돕습니다.
- GPT/Transformer 모델, 챗봇, 디지털 어시스턴트, 감성 분석, 부정 행위 탐지 등에 사용합니다.



CAD(Computer-Aided Design), CAM(Computer-Aided Manufacturing) 및 CAE(Computer-Aided Engineering)

- 급진적이고 새로운 제품 설계 및 생산 방법에 대한 통찰력을 얻습니다.
- 더욱 혁신적이고 품질 높은 제품으로 출시 기간을 단축합니다.
- 설계와 엔지니어링을 혁신하여 미래형 공장을 지원합니다.



엣지 추론

- 클라우드 또는 코어에서 데이터를 전송하는 레이턴시와 연결 문제를 해결합니다.
- 의료 영상 분석에 사용하여 응급 의료 대응을 지원합니다.
- 컴퓨터 비전을 활용하여 현장 장비를 분석하고 자율 주행 차량을 구동합니다.

AI의 진입 장벽 해소

AI를 위한 데이터 사이언스 전문 지식 및 IT 인프라스트럭처 기술 역량 부족.
기술 역량 부족은 AI의 도입 또는 확장을 가로막는 가장 큰 장애물 중 하나입니다.

데이터 작업의 규모 및 복잡성 증가.
기존의 분석 방식으로는 데이터의 양과 복잡성을 따라잡을 수 없습니다.

가치 창출 지연.
처리 성능과 기술 역량이 부족하면 데이터에서 가치를 창출하는 데 지연이 발생합니다.



AI 관련 리더십 확보

62%

인력, 프로세스, 기술을 비롯하여 AI에 대한 지출을 늘릴 계획인 기업과 조직의 비율.¹

2x

적절한 기술 역량을 갖춘 경우 운영 환경에서 AI를 활용할 가능성 증가.²

86%

성공적인 AI 활용을 가로막는 기술 장애물을 하나 이상 식별한 기업과 조직의 비율.¹

69.3%

평균 8개의 GPU가 탑재된 서버에서 AI 워크로드를 실행하고 있는 AI 리더 그룹의 증가율.³

'AI 가능'에서 'AI 검증'으로 도약

Dell Technologies는 AI 여정의 어떤 단계에서든 요구 사항을 충족할 수 있도록 준비되어 있습니다. AI 활용을 이제 막 시작한 경우든, DL(Deep Learning) 클러스터를 배포할 준비가 된 경우든, Dell Technologies는 미개척 시장 기회를 인지하고 활용하는 데 도움이 되는 완벽한 솔루션 포트폴리오를 보유하고 있습니다.

Dell PowerEdge 서버는 AI 솔루션의 기본 구성 요소로, AI를 시작하고 필요에 따라 확장하는 데 필요한 성능, 집적도 및 효율성을 제공합니다. 또한 Dell PowerEdge 서버는 최대 12개의 NVIDIA® GPU(Graphics Processing Unit)를 지원하여 AI 워크로드 및 결과를 가속화할 수 있습니다.

선도적인 AI 소프트웨어 회사와의 파트너십을 통해 데이터 및 AI 포트폴리오에서 지원이 필요한 위치에 관계없이 적합한 솔루션을 제공할 수 있습니다. 워크스테이션부터 데이터 센터, 엣지 및 클라우드에 이르는 기술 혁신이 통합된 생태계를 활용하여 비즈니스를 성공으로 이끄는 포괄적인 AI 접근 방식을 구현할 수 있습니다.

지능형 비즈니스를 위해 설계된 솔루션으로 가치 실현 시간 단축

AI 여정을 가속화하고 간소화하기 위해 Dell은 Validated Designs for AI 포트폴리오를 제공합니다. 이러한 솔루션으로 얻을 수 있는 이점은 다음과 같습니다.



AI 간소화

Dell Validated Designs for AI는 공동으로 엔지니어링되고 NVIDIA GPU, NVIDIA AI Enterprise 제품군 및 기타 NVIDIA 기술로 검증되어 AI 이니셔티브를 가속화하는 데 최적화된 솔루션 스택을 쉽고 빠르게 배포할 수 있습니다.



AI 통찰력을 더 빠르게 확보

NVIDIA GPU 가속 구성은 최적화된 인프라스트럭처에서 AI 툴 및 프레임워크와 함께 제공되므로 개발 및 IT 팀의 운영 준비 시간을 단축할 수 있습니다.



검증된 AI 전문 지식

세계적 수준의 Dell Technologies 서비스 및 지원을 기반으로 엔지니어링 테스트를 거친 AI 솔루션을 자신 있게 배포할 수 있습니다. ProSupport Plus를 선택하여 소프트웨어 및 하드웨어 지원을 위한 일원화된 연락 창구를 활용하십시오.

1일 차

AI 모델에 대한 작업
준비 완료 상태⁴

10x

더 빠른 모델 생성⁴

60% 감소

AI 인프라스트럭처 관리에
소요되는 시간⁵

20% 단축

맞춤형 시스템을
사용하는 AI 프로젝트의
가치 실현 시간⁵

50% 단축

AI 개발 시간⁶

AI 통찰력을 빠르게 확보하도록 설계

Dell PowerEdge 서버로 AI 이점 활용

Dell Technologies는 더욱 빠르게 혁신을 이루기 위해 어디서든 원하는 방식으로 AI를 활용하도록 지원하며, NVIDIA GPU로 가속화된 새로운 PowerEdge 서버에서 빠르게 확보한 통찰력으로 AI 워크로드를 강화합니다. Dell PowerEdge는 우수한 통찰력, 추론 및 시각화를 통해 향상된 AI 워크로드 성과를 얻을 수 있도록 가속화된 컴퓨팅을 더욱 발전시킵니다.



Dell PowerEdge 서버로 어디서든 혁신 가속화

AI 기반 혁신 가속화



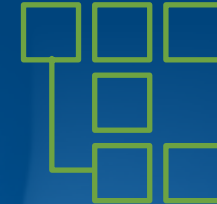
엣지부터 클라우드까지
현대화

제로 트러스트 도입 가속화



보안 강화

더 빠르게 자동화 구현



운영 효율성 향상

환경 보호 및 지속 가능한 발전

Dell PowerEdge XE 서버

복잡한 컴퓨팅, AI/ML/DL 및 HPC(High Performance Computing) 집약적 워크로드를 위해 특별히 설계되고 가속 최적화됨

	PowerEdge XE9680 성능 저하 없이 가속화된 AI를 지원하는 강력하고 유연한 서버	PowerEdge XE9640* 실시간 AI 통찰력을 제공하는 고집적 스마트 냉각 서버	PowerEdge XE8640* GPU 최적화 설계로 탁월한 성능 제공	PowerEdge XE8545 AI, ML(Machine Learning) 및 HPC를 위한 올인원 서버
활용 분야 및 활용 사례	<ul style="list-style-type: none"> AI/ML/DL 훈련, HPC, CRISP 생성형 AI 의료, CSP(Cloud Service Provider), 재무, 학계 	<ul style="list-style-type: none"> AI/ML/DL 훈련, HPC 모델링 및 시뮬레이션 	<ul style="list-style-type: none"> 중간 규모의 데이터 세트 언어 모델, NLP, 모델링 및 시뮬레이션 AI/ML/DL 훈련 및 추론, 이미지 인식 	<ul style="list-style-type: none"> AI/ML 훈련 및 추론, 소규모/중간 규모의 데이터 세트 언어 모델
프로세서	<ul style="list-style-type: none"> 4세대 인텔® 제온® 스케일러블 프로세서 2개 	<ul style="list-style-type: none"> 4세대 인텔 제온 스케일러블 프로세서 2개 	<ul style="list-style-type: none"> 4세대 인텔 제온 스케일러블 프로세서 2개 	<ul style="list-style-type: none"> 3세대 AMD® EPYC™ 프로세서 2개
GPU 지원	<ul style="list-style-type: none"> 전체 NVLink™ 연결을 지원하는 NVIDIA H100 SXM5 또는 NVIDIA A100 SXM4 GPU 최대 8개 	<ul style="list-style-type: none"> 인텔 GPU 최대 4개 	<ul style="list-style-type: none"> 전체 NVLink 연결을 지원하는 NVIDIA H100 SXM5 GPU 최대 4개 	<ul style="list-style-type: none"> NVLink 연결을 지원하는 NVIDIA A100 SXM4 GPU 최대 4개
기능	<ul style="list-style-type: none"> 6U 랙 높이 공랭식(최고 35°C) DDR5 DIMM 32개 x16 PCIe Gen5 슬롯 최대 10개 	<ul style="list-style-type: none"> 2U 랙 높이 수랭식 CPU 및 GPU 작동 DDR5 DIMM 32개 PCIe Gen5 슬롯 최대 2개 	<ul style="list-style-type: none"> 4U 랙 높이 공랭식(최고 35°C) DDR5 DIMM 32개 PCIe Gen5 슬롯 최대 4개 	<ul style="list-style-type: none"> 4U 랙 높이 공랭식(최고 35°C) DDR4 DIMM 32개 x16 PCIe Gen4 슬롯 최대 4개

Dell PowerEdge 랙 서버

광범위한 활용 분야, 활용 사례 및 워크로드를 위한 유연한 메인스트림 컴퓨팅 기반

	PowerEdge R760xa* GPU 기반 워크로드를 위한 주력 서버	PowerEdge R750xa 유연성을 위해 특별히 설계	PowerEdge R750/7525/7515 R650/6525/6515 메인스트림 성능	PowerEdge XR12 엣지 성능
활용 분야 및 활용 사례	<ul style="list-style-type: none"> · AI/ML/DL 훈련 및 추론, 분석, HPC · 생성형 AI 및 고집적 추론 · VDI 및 고성능 그래픽 	<ul style="list-style-type: none"> · AI/ML/DL 훈련 및 추론, 분석, HPC · VDI 및 고성능 그래픽 	<ul style="list-style-type: none"> · 경량 AI/ML/DL 훈련 및 추론 · VDI, 고성능 그래픽 · 엣지 	<ul style="list-style-type: none"> · 엣지 AI 훈련 및 추론 · 통신 · 렌더링/모델링
프로세서	<ul style="list-style-type: none"> · 4세대 인텔 제온 스케일러블 프로세서 2개 	<ul style="list-style-type: none"> · 3세대 인텔 제온 스케일러블 프로세서 2개 	<ul style="list-style-type: none"> · 3세대 인텔 제온 스케일러블 또는 3세대 AMD EPYC 프로세서 최대 2개 	<ul style="list-style-type: none"> · 3세대 인텔 제온 스케일러블 프로세서 1개
GPU 지원	<ul style="list-style-type: none"> · 더블 와이드 NVIDIA PCIe GPU 최대 4개 또는 싱글 와이드 NVIDIA PCIe GPU 최대 12개 	<ul style="list-style-type: none"> · 더블 와이드 NVIDIA PCIe GPU 최대 4개 또는 싱글 와이드 NVIDIA PCIe GPU 최대 6개 	<ul style="list-style-type: none"> · 더블 와이드 NVIDIA PCIe GPU 최대 3개 또는 싱글 와이드 NVIDIA PCIe GPU 최대 6개 	<ul style="list-style-type: none"> · 더블 와이드 또는 싱글 와이드 NVIDIA PCIe GPU 최대 2개
기능	<ul style="list-style-type: none"> · 2U 랙 높이 · 공랭식(최고 35°C) · DDR5 DIMM 32개 · PCIe Gen5 슬롯 최대 4개 	<ul style="list-style-type: none"> · 2U 랙 높이 · 공랭식(최고 35°C) · DDR5 DIMM 32개 · PCIe Gen4 슬롯 최대 4개 	<ul style="list-style-type: none"> · 1U 또는 2U 랙 높이 · 공랭식(최고 35°C) · DDR4 DIMM 32개 · PCIe 4 Gen4 슬롯 최대 8개 	<ul style="list-style-type: none"> · 2U 랙 높이 · 작동 허용 범위: -5°C ~ 55°C · PCIe 4 Gen4 슬롯 최대 4개

* 2023년 상반기에 제공

베어 메탈에
근접한 성능 구현

97.5%

VMware를 사용한 베어
메탈 대비 성능⁷

66%

와트당 성능 향상⁸

66%

HPL(High-Performance Linpack)
성능 향상⁹

NVIDIA GPU로 AI 활용 극대화

Dell Technologies는 엔트리 레벨부터 메인스트림, 최고 수준의 성능에 이르기까지 Hopper 및 Ampere GPU의 완벽한 포트폴리오를 제공하는 유일한 공급업체인 NVIDIA와 긴밀히 협력하고 있습니다. 각각은 엣지, 클라우드, 온프레미스 등 어디에서든 가장 광범위한 AI 애플리케이션을 가속화할 수 있는 다양한 기능을 제공합니다.

H100 SXM

최고 수준 성능의 AI, ML 훈련 및 엑사스케일 HPC

- 3,958TFLOPS FP8 Tensor 코어*
- NVLink: 900GB/s PCIe Gen5
- MIG 최대 7개, 각각 10GB

H100 PCIe

최고 수준 성능의 AI, ML 훈련 및 엑사스케일 HPC

- 3,026TFLOPS FP8 Tensor 코어*
- NVLink: 600GB/s PCIe Gen5
- MIG 최대 7개, 각각 10GB
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU
- 소프트웨어 지원

A100

고성능 AI, ML 훈련 및 추론

- 312TFLOPS FP16 Tensor 코어*
- 최대 2개의 GPU를 지원하는 NVLink Bridge: 600GB/s
- MIG 최대 7개, 각각 10GB
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

A30

메인스트림 그래픽 및 AI 추론

- 165TFLOPS TF32 Tensor 코어*
- 최대 2개의 GPU를 지원하는 NVLink Bridge: 200GB/s
- 최대 4개의 GPU 인스턴스, 각각 6GB
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

A10

메인스트림 엔터프라이즈 서버용 AI로 그래픽 및 비디오 가속

- 250TFLOPS FP16*
- PCIe Gen4 x16
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

비약적인 도약: NVIDIA H100 Tensor 코어 GPU

데이터 센터 규모로 H100 GPU를 배포하면 탁월한 성능을 제공하고 차세대 엑사스케일 HPC와 조 단위 매개변수를 지원하는 AI를 경험할 수 있습니다.

*구조적 희소성(structural sparsity) 지원.

9X

초거대 모델에서 AI 훈련 속도 향상¹⁰

30X

초거대 모델에서 AI 추론 성능의 속도 향상¹¹

3,958

TFLOPS FP8 Tensor 코어¹²

L40

최고 수준 성능의
그래픽 및 렌더링

- 90.5 FP32 TFLOPS(비Tensor)
- 724.1 FP8 Tensor TFLOPS(FP32 누적 포함)*
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원
- NVIDIA Omniverse를 위한 OVX 지원

A40

고성능 그래픽 및
렌더링

- 299.4 BF16 Tensor TFLOPS(FP32 누적 포함)*
- NVLink 112.5GB/s(양방향)
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

A16

CAD/CAM/CAE를 비롯한
원격 작업이 가능하도록 다양한
멀티미디어를 지원하는 VDI

- 35.9TFLOPS FP16 4개*
- PCI Express Gen4 x16
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

L4

효율적인 비디오, 그래픽 및 AI를
위한 획기적인 범용 가속기

- 485TFLOPS FP8*
- PCIe Gen4 x16
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

A2

엣지에서의 AI 추론을 위한
엔트리 레벨 GPU

- 36TFLOPS FP16 Tensor 코어*
- PCIe Gen4 x8
- NVIDIA AI Enterprise 소프트웨어 포함
- NVIDIA vGPU 소프트웨어 지원

각 Dell PowerEdge 서버가 지원하는 NVIDIA GPU에 대한 자세한 내용은 [GPU 매트릭스](#)를 참조하십시오.

*구조적 회소성 지원.

NVIDIA 기술 내장

솔루션의 핵심인 Dell PowerEdge 서버는 AI 워크로드 및 결과를 가속화하는 데 도움이 되는 통합 NVIDIA 기술과 함께 제공됩니다.

NVIDIA vGPU(Virtual GPU)

NVIDIA vGPU 소프트웨어를 사용하면 어디서든 원하는 디바이스에서 액세스할 수 있도록 여러 VM에 걸쳐 GPU 리소스를 공유할 수 있습니다.

NVIDIA MIG(Multi-Instance GPU)

NVIDIA MIG는 GPU를 최대 7개의 인스턴스로 파티셔닝하여 모든 워크로드를 지원하고 가속화된 리소스를 더 많은 사용자와 공유함으로써 GPU의 성능과 가치를 확대합니다.

NVIDIA H100 GPU

NVIDIA H100 Tensor 코어 GPU는 모든 데이터 센터에 전례 없는 성능, 확장성 및 보안을 제공합니다. NVIDIA H100 PCIe GPU에는 AI 개발 및 배포를 간소화할 수 있는 NVIDIA AI Enterprise 소프트웨어 제품군이 포함되어 있습니다. 초거대 모델에서 AI를 9배 더 빠르게 훈련시키고¹⁰ AI 추론 성능의 속도를 30배 더 높입니다.¹¹

NVIDIA A100 GPU

이전 세대보다 최대 20배 더 높은 성능으로 AI 워크로드를 가속화합니다. A100은 기존의 PCIe 기반 솔루션보다 훨씬 더 빠른 멀티 GPU 시스템에 대한 대안을 제공하는 세계 최초의 고속 GPU 상호 연결인 NVLink Bridge를 지원합니다.¹⁴

VMware vSphere 기반 NVIDIA AI Enterprise

NVIDIA AI Enterprise는 AI에 대한 전문 지식 없이도 지원되는 컨테이너, 프레임워크 및 워크플로를 통해 AI 여정을 시작할 수 있도록 지원하는 포괄적인 클라우드 네이티브 제품군입니다. Dell Technologies의 NVIDIA Certified Systems™에서 실행하도록 인증되었으며 AI 개발 및 배포 툴, 인프라스트럭처 최적화 소프트웨어, AI 프로젝트의 원활한 추진을 위한 글로벌 엔터프라이즈 지원을 포함합니다. 따라서 인프라스트럭처 배포가 아닌 AI의 비즈니스 가치 활용에 집중할 수 있습니다.



NVIDIA-Certified Systems

NVIDIA Certified Systems™인 Dell VxRail HCI와 Dell PowerEdge는 NVIDIA GPU, NVIDIA ConnectX® SmartNIC(Smart Network Interface Card), NVIDIA BlueField® DPU를 최적화된 구성으로 통합합니다. 이러한 기술은 성능, 관리 용이성, 보안 및 확장성에 대한 검증을 거쳤으며 NVIDIA 및 Dell Technologies의 엔터프라이즈급 지원으로 뒷받침됩니다.

NVIDIA LaunchPad

무료로 제공되는 선별된 랩 경험을 통해 AI, 데이터 사이언스, 3D 디자인 협업 및 시뮬레이션 등 다양한 분야의 포괄적인 솔루션 워크플로를 경험하는 데 필요한 하드웨어 및 소프트웨어 스택에 즉시 단기적으로 액세스할 수 있습니다. NVIDIA LaunchPad는 Dell PowerEdge 서버를 기반으로 구축됩니다. 자세한 정보: [nvidia.com/dell-launchpad](https://www.nvidia.com/dell-launchpad).

NVIDIA BlueField DPU(Data Processing Unit)

BlueField DPU는 다양한 고급 네트워킹, 스토리지 및 보안 서비스를 오프로드, 가속화, 격리하여 클라우드부터 데이터 센터, 엣지에 이르기까지 모든 환경에서 모든 워크로드에 대해 안전하고 가속화된 인프라스트럭처를 제공합니다.

권장 구성

워크로드	활용 사례	권장 구성	
HPC/AI/ML/DL 훈련 생성형 AI	<ul style="list-style-type: none"> • NLP(Natural Language Processing) • LLM(Large Language Model) • 대규모 추천 엔진 훈련 • HPC, 모델링 및 시뮬레이션 	<ul style="list-style-type: none"> • PowerEdge XE9680 	<ul style="list-style-type: none"> • H100 SXM GPU
HPC/AI/데이터베이스/분석	<ul style="list-style-type: none"> • HPC • AI/ML/DL 훈련 및 추론 • 중간 규모 데이터 세트 언어 모델 • NLP • 이미지 인식 • 모델링 및 시뮬레이션 • 분자동역학 • 유전체 염기 서열 분석 	<ul style="list-style-type: none"> • PowerEdge XE9680 • PowerEdge XE8640 	<ul style="list-style-type: none"> • H100 SXM GPU • A100 SXM GPU
고성능 그래픽/VDI/모델링	<ul style="list-style-type: none"> • 디지털 트윈 및 3D 세계/메타버스 • 고성능 그래픽 • CAD/CAM/CAE • 가상화 • HPC 	<ul style="list-style-type: none"> • PowerEdge R760xa • PowerEdge R750xa • PowerEdge R750 • PowerEdge R7525 	<ul style="list-style-type: none"> • L40 GPU • A40 GPU
메인스트림 AI	<ul style="list-style-type: none"> • HPC • 분석 • GPU 데이터베이스 가속화 • AI/ML 훈련 및 추론 • 경량 AI 훈련 • AI/ML 훈련 및 추론 	<ul style="list-style-type: none"> • PowerEdge R960 • PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7625/7525 • 기타 랙 서버 	<ul style="list-style-type: none"> • A2, A10, A30 또는 A100 GPU • L4 GPU
VDI 및 가상화	<ul style="list-style-type: none"> • 파워 유저를 위한 풍부한 협업 • 지식 근로자를 위한 VDI 	<ul style="list-style-type: none"> • PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7625/R7525 • PowerEdge R960 	<ul style="list-style-type: none"> • A10 또는 A16 GPU • L4 GPU
메인스트림 그래픽 및 VDI	<ul style="list-style-type: none"> • 그래픽 렌더링 	<ul style="list-style-type: none"> • PowerEdge R760/R750 • PowerEdge R7625/R7525 • 기타 랙 서버 	<ul style="list-style-type: none"> • A10 GPU • L4 GPU
추론/엡지/VDI	<ul style="list-style-type: none"> • 엡지 추론 	<ul style="list-style-type: none"> • PowerEdge XR12 • PowerEdge R760xa/R750xa • PowerEdge R760/R750 • PowerEdge R7626/R7525 • 기타 랙 서버 	<ul style="list-style-type: none"> • A2 GPU • L4 GPU

고객 성공 사례

1 대규모 콘텐츠를 추천하는 Taboola.

Taboola®는 뛰어난 컴퓨팅 성능과 간소화된 관리 기능을 활용하여 달성한 탁월한 성능, 확장성 및 자동화를 기반으로 매일 수십억 개의 관련 콘텐츠를 추천하는 AI 모델을 훈련시키고 실행합니다.

150,000

초당 처리되는 AI 기반
요청 수

6x

AI 기반 추론 성능 향상

50밀리초

실시간 추천에 소요되는 시간

"AI 기반 추론을 활용하여 성능을 최대 6배 향상시켰습니다... 그 덕분에 비용이 절감되었습니다."

- Ariel Pisetzky, VP of IT and Cybersecurity, Taboola

[사례 연구 보기.](#)

2 기차가 최고 속도를 유지하도록 하는 Duos Technologies.

Duos Technologies®는 NVIDIA GPU 가속 Dell PowerEdge 서버가 지원하는 엣지에서 AI를 활용해 실시간으로 데이터를 처리하고 분석하여 기차가 검사를 위해 멈출 필요가 없도록 신속하고 실행 가능한 통찰력을 제공합니다.

120:1

검사 시간 단축 비율

1.3TB

사이트당 매일 처리 및
분석되는 데이터의 양

\$3,000 USD

서버 복구를 위한 인스턴스당 절감액

"우리는 PowerEdge 서버에서 AI 모델을 사용하여 연중무휴 24시간 카메라와 센서의 이미지 및 기타 데이터를 처리하고 분석합니다."

- David Ponevac, Chief Technology Officer, Duos Technologies

[사례 연구 보기.](#) [비디오 보기.](#)

3 과학적 발견을 가속화한 케임브리지대학교
Dell Technologies는 케임브리지대학교가 HPC 및 데이터 스토리지 시스템을 구축하여 오늘날 가장 까다로운 데이터 기반 시뮬레이션 및 AI 과제를 해결하도록 지원합니다.

3.8

페타플롭스

74,000

코어

500

초당 기가바이트

"이 사람들이 만족할 만한 충분한 컴퓨팅 성능을 제공하는 것은 불가능합니다. 뭐든 주는 대로 소비할 테니까요. 케임브리지대학교의 슈퍼컴퓨터는 연구원들에게 AI 작업에 필요한 빠르고 경제적인 슈퍼컴퓨팅 성능을 제공합니다."

- Dr. Paul Calleja, Director of Research Computing Services,
케임브리지대학교

[사례 연구 보기.](#)

4 AI의 잠재력을 확장한 피사대학교
Dell Technologies, VMware 및 NVIDIA의 솔루션 덕분에 피사대학교는 기존 워크로드와 AI 워크로드를 동일한 시스템에서 실행하여 IT 리소스의 활용도를 극대화하는 동시에 요구 사항을 충족할 수 있는 유연성을 갖추게 되었습니다.

제로

AI 시스템의
사일로 수

1

가상 데스크탑 및
앱을 배포하기 위한
플랫폼의 수

여러 개

동일한 인프라스트럭처에서 지원되는 워크로드의 수

"가상화된 GPU의 가장 큰 이점은 유연성으로, 엔터프라이즈 인프라스트럭처를 AI 워크로드에 맞게 설계하고 조정할 수 있습니다."

- Maurizio Davini, Chief Technology Officer, 피사대학교

[사례 연구 보기.](#)

Dell Technologies를 선택해야 하는 이유

전 세계 Customer Solution Center에서 협업

전 세계 [Customer Solution Center](#)에서 Dell Technologies 엔지니어링 팀과 협업하거나, [HPC & AI Center of Excellence](#)의 리소스를 활용하거나, [HPC & AI Innovation Lab](#)에서 실제 시스템을 테스트하고 튜닝할 수 있습니다.

Dell APEX를 통해 AI as-a-Service 사용

Dell APEX를 사용하면 aaS(as-a-Service)로 제공되는 간편하고 일관된 클라우드 경험을 통해 어디서나 지능적인 성과를 빠르게 실현하는 데 필요한 AI 최적화 솔루션을 구현할 수 있습니다. [Dell APEX](#)는 온프레미스, 오프프레미스 및 엣지에서 AI를 위한 클라우드 운영 모델을 제공할 수 있으므로 고객이 어떤 규모로든 데이터를 활용하여 측정 가능한 가치를 창출할 수 있습니다.

서비스로 성공 가속화

[Dell Technologies Services](#)에는 초기 설정과 인력의 숙련도 향상부터 지속적인 지원에 이르기까지 AI 환경의 신속한 도입 및 최적화를 추진하는 데 도움이 되는 컨설팅, 구축, 지원, 교육 서비스 등이 있습니다. [Managed Services](#)와 [Residency Services](#)는 IT 관리의 비용, 복잡성 및 위험을 줄여 주므로 고객은 디지털 혁신 및 전환에 리소스를 집중할 수 있습니다.

35K+

AI 성공을 위한 로드맵 작성을 도와주는 서비스 및 지원 직원 수¹⁵

\$0

Dell Technologies AI 전문가와의 협업 비용¹⁶

10

전 세계 Dell HPC and AI Center of Excellence 수¹⁷



지능적인 성과 가속화

Dell Technologies는 유형과 규모에 관계없이 모든 기업과 조직이 새로운 기회를 발굴하고 데이터의 무한한 잠재력을 최대한 활용하도록 지원합니다. 450개 이상의 AI 프로젝트를 추진하는 35개 이상의 데이터 사이언스 팀과 데이터에서 통찰력을 추출하는 데 전념하는 1,800명 이상의 팀원을 보유한 Dell Technologies는 검증된 AI 전문 지식을 기반으로 IT 효율성을 개선하고 위험을 완화하여 향상된 고객 통찰력과 경험을 제공합니다. 또한 온프레미스, 오프프레미스, 엣지 환경의 하이브리드 클라우드 전반에 걸쳐 일관된 방식으로 이를 지원합니다.

AI 시대에 성공을 거둘 수 있도록 Dell Technologies가 도와드리겠습니다.

자세한 정보

Dell.com/PowerEdge

Dell Technologies와 NVIDIA

AI 워크로드 지원 및 가속화

Dell Technologies와 NVIDIA는 협력을 통해 엔지니어링 검증을 거친 하드웨어 및 소프트웨어를 제공하여 AI, ML 및 DL 워크로드를 가속화합니다. Dell Technologies는 또한 최첨단 NVIDIA GPU, SmartNIC 및 DPU, NVIDIA AI Enterprise 소프트웨어를 통합하는 서버 및 솔루션에 막대한 투자를 하고 있습니다. NVIDIA 그리고 Dell Technologies와 함께하면 이전에는 불가능하다고 생각했던 곳에서도 AI를 활용할 수 있습니다.

Copyright © 2023 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell 및 기타 상표는 Dell Inc. 또는 해당 자회사의 상표입니다. NVIDIA®, CUDA®, NVLink™, BlueField®, ConnectX® 및 NVIDIA-Certified Systems™는 미국 및 기타 국가에서 NVIDIA Corporation의 상표 및/또는 등록 상표입니다. 인텔® 및 제온®은 미국 및/또는 기타 국가에서 인텔사 또는 그 자회사의 상표입니다. AMD® 및 EPYC™는 Advanced Micro Devices, Inc.의 상표입니다. VMware®는 미국 및 기타 국가에서 VMware, Inc.의 등록 상표 또는 상표입니다. Taboola®는 Taboola, Inc.의 등록 상표입니다. Duo Technologies®는 Duo Technologies, Inc.의 상표 및 브랜드입니다. 기타 상표는 해당 소유주의 자산일 수 있습니다. Published in the USA 2023년 2월 eBook dell-nvidia-ai-EB-101

Dell Technologies는 본 문서의 정보가 해당 발행일 현재 정확한 것으로 간주합니다. 이 정보는 예고 없이 변경될 수 있습니다.



¹ ESG 인포그래픽, Dell Technologies 서버와 NVIDIA를 통한 컴퓨팅 현대화로 AI 기반 미래 대비, 2022년.

² "AI 평가 그룹"과 비교. 출처: Dell Technologies와 NVIDIA의 후원으로 작성된 IDC Analyst Brief, *Scaling Skills for AI: Lessons from Early Adopters*, 2022년 8월.

³ "AI 평가 그룹"과 비교. 출처: Dell Technologies와 NVIDIA의 후원으로 작성된 IDC 백서, *운영 환경에서 AI를 활용하는 기업이 뒤쳐진 기업에 주는 가르침*, 2022년 8월.

⁴ Dell Precision Data Science Workstation 사용. *DSW Ready Day One Guide* 참조.

⁵ Dell Validated Designs for AI 사용. Forrester, *The Total Economic Impact™ Of Dell Validated Designs For AI*, 2022년 8월.

⁶ Dell Precision Data Science Workstation 사용. Dell Technologies 사례 연구, *AI deep learning extends data science horizons*, 2021년 2월.

⁷ 성능 테스트에서 Dell Technologies 및 VMware를 사용한 구성은 동일한 서버에서 베어 메탈 성능의 최대 97.5%를 달성했습니다. 출처: *Principled Technologies 보고서, Achieve near bare metal inference throughput for image classification workloads with the Dell PowerEdge R7525 server using virtual GPUs*, 2022년 7월.

⁸ NVIDIA H100 구성을 사용하는 Dell PowerEdge R750xa는 A100 구성에 비해 와트당 성능이 66% 향상되었습니다. 출처: Dell Technologies 기술 노트, *PowerEdge R750xa and NVIDIA H100 PCIe GPU: 66% Increase in HPC Performance per Watt*, 2022년.

⁹ NVIDIA H100 구성을 사용하는 PowerEdge R750xa는 NVIDIA A100 구성에 비해 HPL 벤치마크 성능이 67% 향상되었습니다. 출처: Dell Technologies 기술 노트, *PowerEdge R750xa and NVIDIA H100 PCIe GPU: 66% Increase in HPC Performance per Watt*, 2022년.

¹⁰ H100은 4세대 Tensor 코어와 FP8 정밀도의 Transformer Engine을 탑재하여 MoE(Mixture-of-Experts) 모델을 이전 세대보다 최대 9배 더 빠르게 훈련시킵니다. 출처: *NVIDIA, NVIDIA H100 Tensor 코어 GPU*, 2023년 1월 참조.

¹¹ 이전 세대와 비교. 출처: *NVIDIA, NVIDIA H100 Tensor 코어 GPU*, 2023년 1월 참조.

¹² 구조적 희소성이 지원되는 NVIDIA H100 SXM GPU의 경우, 희소성이 지원되지 않는 경우 사양이 반 정도 낮습니다. 출처: *NVIDIA, NVIDIA H100 Tensor 코어 GPU*, 2023년 1월 참조.

¹³ NVIDIA 웹사이트, *이 시대의 가장 중요한 작업 가속화*, 2022년 6월 참조.

¹⁴ NVIDIA 웹사이트, *NVIDIA NVLink*, 2022년 6월 참조.

¹⁵ Dell Technologies, *Key Facts*, 2022년.

¹⁶ *Dell Technologies Customer Solution Center* 및 *HPC & AI Innovation Lab*. 자세한 내용은 영업 담당자에게 문의하십시오.

¹⁷ 자세한 내용은 [dell.com](https://Dell.com)을 참조하십시오.