



서버 가속기

애플리케이션 성능 향상



10명 중 7명

향후 5-10년 동안 AI를 가장 중요한 기술 3개 중 하나로 선택한 기술 경영진의 비중¹

최대 100개

단일 GPU 내의 CPU 수²

600개 이상 애플리케이션

가속화되고 전체 딥 러닝 프레임워크로 구성³

신속한 통찰력 확보 및 혁신 가속화

디지털 엔터프라이즈의 성공은 대량의 빠른 데이터를 활용하는 데 달려 있습니다. 그러나 데이터 세트가 증가함에 따라 기존 데이터 센터는 특히 실시간 데이터 소스를 수집하고 쿼리하는 데 있어 성능 및 확장의 한계에 도달하기 시작합니다.

일부는 오랫동안 시각화, 모델링 및 시뮬레이션 속도를 높이기 위해 가속기를 활용했지만, 오늘날에는 그 어느 때보다 더 많은 주류 애플리케이션이 가속기를 활용하여 통찰력과 혁신을 향상할 수 있게 되었습니다. GPU(Graphic Processing Unit) 및 FPGA(Field Programmable Gate Array)와 같은 가속기는 병렬 프로세싱을 통해 대량의 데이터를 더 빠르게 처리함으로써 CPU를 보완하고 가속화합니다. 또한 가속화된 데이터 센터는 경제성이 더 뛰어나 더 적은 수의 서버로 획기적인 성능을 제공하므로 결과적으로 통찰력을 높이고 비용을 절감할 수 있습니다.

여러 산업 분야의 조직은 경쟁 우위를 확보하기 위해 서버 가속기를 채택하며, 데이터에서 확보한 통찰력으로 제품과 서비스 오퍼링을 개선하고 더 나은 애플리케이션 성능으로 생산성을 향상하며, 빠르고 강력한 분석으로 운영을 최적화하고 그 어느 때보다 시장 출시 기간을 단축합니다.

Dell EMC는 Dell EMC PowerEdge 서버에서 다양한 서버 가속기를 제공하므로 애플리케이션 성능을 향상할 수 있습니다.

가속기 활용 사례

- **머신 러닝 및 딥 러닝** - 가속기는 교육 및 추론 워크로드를 가속화하는 데 필요한 병렬 프로세싱 성능을 지원함으로써 AI를 이론에서 주류로 이끌었습니다.
- **예측 분석** - 가속기로 활성화되는 AI는 분석을 강화하여 동적 상관관계를 형성하고 엄청난 속도, 정확성 및 확장성으로 예측적인 결과를 제공할 수 있습니다.
- **가속화된 데이터베이스** - 가속기는 취합, 정렬 및 그룹화 작업을 가속화하여 기존 데이터베이스에 과부하를 일으키는 복잡한 분석 작업을 해결하는 데 도움이 됩니다.
- **데이터 스트리밍** - IoT(Internet of Things)로 인해 많은 양의 데이터를 스트리밍하는 것이 가능해졌습니다. 가속기는 실시간 분석을 위해 데이터 스트리밍의 동시 수집, 탐색 및 시각화를 가능하게 합니다.

¹ "2019 Edelman AI 설문조사," 2019년 3월.

² "NVIDIA Tesla V100 Tensor Core GPU," 2019년 5월.

³ "NVIDIA Tesla P100," 2019년 5월.

GPU

GPU(Graphics Processing Unit)는 컴퓨팅 성능을 가속화하도록 설계된 보조 프로세서입니다. GPU는 일반적으로 수학적 기능을 효율적으로 실행하도록 설계된 수천 개의 코어로 구성됩니다. 워크로드의 일부는 CPU에서 GPU로 오프로드되며, 나머지 코드는 CPU에서 실행되어 전체 애플리케이션 성능이 향상됩니다.

Dell EMC는 서버 PCIe 슬롯에 적합한 PCIe 카드와 서버 마더보드에 탑재된 SXM2 모듈로서 다양한 GPU를 제공합니다.

SXM2 모듈이 탑재된 PowerEdge C4140 서버는 GPU 간 NVLINK 고속 상호 연결을 활용할 수 있습니다.

FPGA

FPGA(Field Programmable Gate Array)는 기존 소프트웨어 솔루션보다 최대 1,000배 빠른 속도로 특정 유형의 알고리즘을 실행할 수 있는 특정한 기술적 특성을 가진 가속기입니다. FPGA는 특정 업무나 애플리케이션의 요구 사항에 정확히 일치하도록 구성할 수 있습니다.

Dell EMC는 서버 PCIe 슬롯에 적합한 PCIe 카드로 다양한 FPGA를 제공합니다.

병렬 처리

병렬 프로세싱은 여러 마이크로프로세서에서 프로그램 작업을 동시에 나누어 실행하여 작업 시간을 단축하는 방법입니다. 병렬 프로세싱은 두 개 이상의 프로세서가 있는 컴퓨터 또는 컴퓨터 네트워크를 통해 수행할 수 있습니다. 병렬 프로세싱은 병렬 컴퓨팅이라고도 합니다.

코드 최적화

서버 가속기를 최대한 활용하려면 소프트웨어 코드를 최적화하십시오.

컴퓨팅 커널 실행을 위해 CUDA(Compute Unified Device Architecture)는 GPU 가상 명령 집합 및 병렬 컴퓨팅 요소에 직접적인 액세스 권한을 부여합니다.

FPGA 구성은 일반적으로 HDL(Hardware Description Language)로 지정됩니다. HDL을 통해 ASIC(Application-Specific Integrated Circuit)을 모방하는 특정 작업이나 애플리케이션의 요구 사항에 맞게 기본 사항을 구성할 수 있습니다.

- **시각화** - 가속기는 CAD(Computer Aided Design)와 같은 3D 시각화 애플리케이션의 성능을 향상하여 사용자가 소프트웨어를 사용함에 따라 실시간으로 모델을 그릴 수 있도록 합니다.
- **모델링 및 시뮬레이션** - 가속기는 조기 평가, 설계 수정 사항에 대한 신속한 테스트 및 설계에 대한 반복 작업을 위해 실제보다 빠르게 모델링 시뮬레이션을 제공할 수 있습니다.
- **금융 모델링** - 가속화된 HPC 및 AI(Artificial Intelligence) 솔루션은 분석 톨을 혁신하며, 금융 업계에서 대규모 데이터 요소를 활용하여 위험과 수익을 더 잘 이해할 수 있도록 지원합니다.
- **지질 측정 데이터 처리** - 석유 및 가스 회사는 결과 산출 시간을 단축하고 비용을 절감하기 위해 가속기를 통해 대규모 지질 측정 데이터 저장소에서 정보를 추출할 수 있는 새롭고 개선된 방법을 찾고 있습니다.
- **신호 처리** - 가속기를 통해 공급업체는 컴퓨터, 라디오, 비디오 및 휴대전화에서 들어오는 대량의 신호 데이터 스트림을 실시간으로 모델링하고 분석할 수 있습니다.

Dell EMC PowerEdge 서버를 위한 GPU 및 FPGA

일부 Dell EMC PowerEdge 타워 및 랙 서버에서 사용 가능한 성능 가속기를 통해 애플리케이션을 강화할 수 있습니다. 다음 표는 [PowerEdge 서버](#)에 적합한 가속기(GPU 및 FPGA)의 수량과 모델을 보여줍니다. Xilinx® Alveo™ U200 및 Arria® 10 GX 포함 인텔® PAC(Programmable Acceleration Card)는 FPGA이며 나머지는 NVIDIA® GPU입니다.

2S 타워	1U, 2S 랙		2U, 2S 랙		4U, 4S 랙	
T640	R640	C4140	R740	R7425	R840	R940xa
가속기 4개	가속기 1개	가속기 4개	가속기 3개	가속기 3개	가속기 2개	가속기 4개
			P4			
	T4		T4	T4		
P40		P40	P40	P40	P40	
M10			M10	M10	M10	
V100 16GB 또는 32GB		V100 16GB 또는 32GB PCIe 또는 SXM2	V100 16GB 또는 32GB	V100 16GB 또는 32GB	V100 16GB 또는 32GB	V100 16GB 또는 32GB
			U200	U200	U200	U200
	10GX		10GX		10GX	10GX

U = 랙 유닛, S = 소켓 또는 CPU
 표에 나열된 가속기의 수는 더블 와이드 전체 높이 카드의 경우입니다. T4는 로우 프로파일, 절반 높이 카드입니다.

GPU는 CUDA 코어 수, 메모리 용량, 전력 및 냉각 요구 사항에 따라 다릅니다. 예를 들어, NVIDIA T4 GPU는 CUDA 코어가 2560개, 메모리가 16GB이며 최대 70와트를 사용합니다. NVIDIA Tesla® V100은 CUDA 코어가 5120개, 메모리가 최대 32GB이며 최대 300와트를 사용합니다. 인텔 및 Xilinx FPGA 사양에는 10와트~225와트를 사용하는 35MB~64MB의 내장형 및 오프 칩 메모리 용량이 포함됩니다.



47배 향상

딥 러닝 추론의 CPU보다 높은 처리량⁴



1.5 TFLOP

(초당 테라바이트 단위 부동 소수점 작업) 및 최대 40GFLOP/와트의 전력 효율성⁵



최대 90배

머신 러닝 추론, 비디오 트랜스코딩 및 데이터베이스 검색과 분석을 위한 CPU보다 높은 성능⁶

NVIDIA Tesla GPU

NVIDIA Tesla GPU는 시각화 및 기타 워크로드를 이전보다 빠르게 실행하는 데 필요한 성능을 제공합니다. 또한 Tesla는 VDI(Virtual Desktop Infrastructure)에 높은 성능 및 사용자 직접도를 제공합니다.

- [T4 텐서 코어 GPU](#)
- [Tesla V100 GPU 가속기](#)
- [NVIDIA NVLink™ Fabric](#) 상호 연결
- [NVIDIA GPU CLOUD™](#) 컨테이너
- [소프트웨어 애플리케이션 카탈로그](#)
- [NVIDIA 개발자](#) 리소스

NVIDIA GPU는 R640, R740, R740xd, R7425, R840, R940xa, C4140를 포함한 [Dell EMC PowerEdge 서버](#)와 [HPC 및 AI를 위한 Dell EMC Ready Solutions](#)에서 사용할 수 있습니다.

인텔 FPGA

인텔 FPGA는 데이터 분석, 이미지 추론, 암호화 및 압축과 같은 워크로드와 정확히 일치하는 데이터 경로를 통해 동적으로 재프로그래밍할 수 있습니다.

- [인텔 Arria 10 GX FPGA가 포함된 인텔 PAC](#)
- [소프트웨어 가속화 스택](#)
- [인텔 FPGA 가속화 허브](#)

R640, R740, R740xd, R840, R940xa를 포함한 [Dell EMC PowerEdge 서버](#)와 [HPC 및 AI 솔루션](#)에서 사용할 수 있습니다. [구성 가이드](#)를 다운로드하십시오.

Xilinx FPGA

Xilinx 16nm UltraScale™ 아키텍처를 기반으로 구축된 Xilinx Alveo 가속기 카드는 변화하는 가속화 요구 사항 및 알고리즘 표준에 적용 가능하며, 하드웨어를 변경하지 않고도 어느 워크로드든지 가속화할 수 있으며 총 소유 비용을 절감할 수 있습니다.

- [Alveo U200](#)
- [소프트웨어 생태계](#)
- [개발자 툴](#)

R740, R740xd, R7425, R840, R940xa를 포함한 [Dell EMC PowerEdge 서버](#)와 [HPC 및 AI 솔루션](#)에서 사용할 수 있습니다.

가속화된 Dell EMC 솔루션

GPU와 FPGA가 포함된 Dell EMC 및 파트너 솔루션으로 시간을 절약하십시오.

Dell EMC Ready Solution

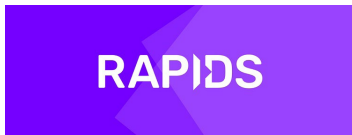
[Dell EMC Ready Solutions for HPC](#)는 고급 컴퓨팅을 보다 빠르고 간편하게 도입하도록 지원합니다. Dell EMC는 서버, 네트워킹, 스토리지, 솔루션 및 서비스가 함께 최적화되어 다양한 산업의 활용 사례를 처리할 수 있는 유연하고 확장 가능한 여러 HPC(High Performance Computing) 솔루션을 제공합니다.

[Dell EMC Ready Solutions for AI](#)는 AI 이니셔티브를 가속화하는 데 필요한 모든 것을 제공합니다. AI를 단순화한 통합 시스템은 머신 러닝과 딥 러닝에 적합하므로 고객과 비즈니스에 대한 심층적인 정보를 더 빠르게 파악할 수 있습니다.

⁴ "NVIDIA Tesla V100 Tensor Core GPU." 2019년 5월.

⁵ "인텔 Arria 10 FPGA 기능." 2019년 5월.

⁶ "Alveo U200 Data Center Accelerator Card." 2019년 5월.



Dell EMC 파트너에서 제공하는 솔루션

[Amulet Hotkey® 가상 데스크탑 솔루션](#)은 엔터프라이즈급 서버와 가상 GPU 가속기를 결합하여 데이터 센터에 최적화된 고집적 솔루션을 제공하고 Windows® 10으로의 전환을 간소화합니다. 또한 가상 GPU는 Windows 10, Microsoft® Office 365®, YouTube® 등과 같은 일상적인 프로그램의 가속 그래픽 가상화에 대한 늘어나는 수요를 충족하고 뛰어난 가상 데스크탑 환경을 구축하도록 도울 수 있습니다. [Amulet Hotkey의 고객 성공 사례](#)에 대해 알아보십시오.

[BittWare Stratix® 10 기반 FPGA 가속화 컴퓨팅 노드](#)는 1U당 최대 4개의 인텔 Stratix 10 FPGA를 지원하는 FPGA 가속기에 최적화된 고집적 랙 마운트 서버를 사용하여 가장 까다로운 데이터 센터 워크로드를 실행할 수 있도록 합니다. Dell EMC 또는 BittWare에서 직접 시스템을 구매할 수 있습니다. [비디오 보기](#).

[Kinetica®](#)는 GPU 가속 데이터베이스, 시각적 검색 및 머신 러닝 기능 그리고 가속 병렬 컴퓨팅 기능을 포함하는 인사이트 엔진입니다. Dell PowerEdge 서버와 NVIDIA GPU에서 실행되는 Kinetica는 양이 방대하며 복잡하고 예측 불가능한 데이터로 인해 발생하는 문제를 해결하도록 도움을 줍니다. 문서 보기: [CEO에게 GPU 설명하기: 제품화의 이점](#).

[Tracewell Systems®](#)는 해상이나 육지 어디서든 환경적 요소로 고유한 컴퓨팅 문제가 발생하는 장소, 고정 위치 및 모바일 설치 또는 특수 하드웨어나 소프트웨어와의 통합이 필요한 상황에서 기업, 정부 기관 및 OEM에 강력하고 바로 사용 가능한 컴퓨팅 기술을 제공합니다. [Data Sheet](#), [비디오](#) 및 [리소스 가져오기](#).

Dell EMC Acceleration 소프트웨어 파트너

[NVIDIA GRID™](#) 가상 PC 및 GRID® 가상 애플리케이션은 NVIDIA® Tesla® GPU 기반의 검증된 성능으로 가상 데스크탑 및 애플리케이션을 개선합니다.

[Kinetica](#) 소프트웨어는 병렬 컴퓨팅용 GPU를 통해 기존의 OLAP(Online Analytics Processing) 워크로드 속도를 크게 향상합니다.

[BitFusion](#) 소프트웨어는 GPU, FPGA 및 ASIC를 분리하고 데이터 센터의 어느 곳이나 동적으로 연결합니다.

[SQream Technologies®](#) GPU 가속 데이터 웨어하우스는 모든 규모와 워크로드에 맞게 테라바이트에서 페타바이트 단위까지 확장할 수 있습니다.

[FASTDATA.io](#) PlasmaENGINE® GPU 기반의 소프트웨어는 여러 개의 GPU를 통해 다중 노드에서 이동 중인 무한 데이터를 실시간으로 처리할 수 있습니다.

[RAPIDS](#)는 NVIDIA GPU에서 포괄적인 데이터 과학 교육 파이프라인을 실행하기 위해 NVIDIA CUDA-X를 기반으로 구축된 데이터 과학 라이브러리 모음입니다.

Dell Technologies 파트너가 되기

[Dell Technologies Partner Program](#)에 참여하면 날마다 상호 협력을 통해 전 세계 조직의 디지털, IT, 업무 환경 및 보안을 혁신하기 위해 노력하는 파트너 협력 체계에 합류하게 됩니다. Dell Technologies Partner Program은 네트워크 말단 환경부터 중앙의 데이터 센터와 클라우드에 이르기까지 업계 유수의 성능을 자랑하는 포트폴리오를 뒷받침하고 있으며, 간편한 절차, 예측 가능한 결과, 높은 수익성을 제공하도록 설계되었습니다.



리소스

PowerEdge 서버와 함께 모든 워크로드를 처리할 수 있도록 데이터 센터를 혁신하십시오. PowerEdge 타워 서버는 조직의 비즈니스 성장 속도에 맞추어 확장할 수 있도록 설계되었습니다. PowerEdge 랙 서버는 확장성이 뛰어난 아키텍처와 최적의 균형을 이루도록 구성된 컴퓨팅 및 메모리를 결합하여 다양한 애플리케이션에 걸쳐 성능을 극대화합니다.

성능 결과 보기

hpcatdell.com 및 [서버 지원 솔루션 리소스](#) 페이지에서 워크로드, 레퍼런스 아키텍처 및 엔지니어링 관련 블로그 게시물별로 벤치마킹 데이터를 확보하십시오. 다음 백서를 살펴보세요.

- [딥 러닝 성능 비교 - PowerEdge C4140 스케일 업과 스케일 아웃 비교](#)
- [PowerEdge R940xa의 GPU 데이터베이스 가속화](#)
- [PowerEdge R7425의 딥 러닝 추론](#)
- [Dell EMC 인프라스트럭처의 인텔 PAC\(Programmable Acceleration Card\)를 통한 딥 러닝 추론](#)

교육 서비스 액세스

education.emc.com 페이지에서 필요한 기술, 교육 및 인증을 받으십시오. Dell EMC의 Deep Learning Institute에서 [딥 러닝으로 문제를 해결하는 방법을 알아보십시오](#).

HPC and AI Innovation Lab 방문

세계적 수준의 솔루션 엔지니어링, 기술 테스트 및 협업: dell EMC.com/ko-kr/innovationlab

CoE(Center of Excellence)와 연결

혁신과 전문성의 글로벌 허브: dell EMC.com/ko-kr/coe

문의

자세한 내용은 dell EMC.com/ko-kr/accelerators, dell EMC.com/ko-kr/servers 페이지를 참조하거나 해당 지역 영업 대표 또는 공인 리셀러에게 문의하시기 바랍니다.

리소스 센터

[서버 솔루션 리소스](#)를 다운로드

기술 문서

성능 결과, 레퍼런스 아키텍처 및 HPC 엔지니어링 관련 블로그를 참조
hpcatdell.com

Virtual Rack

Virtual Rack에서 서버 및 솔루션 확인
esgvr.dell.com

Dell EMC HPC 커뮤니티 참여

적극적인 아이디어 교환을 촉진하는 세계적인 기술 포럼
dellhpc.org

© 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC 및 기타 상표는 Dell Inc. 또는 해당 자회사의 상표입니다. 기타 모든 상표는 해당 소유주의 상표일 수 있습니다. 참조 번호: 05/19 DELL-EMC-BRO-Accelerators-USLET-101

인텔®은 미국 및/또는 기타 국가에서 인텔사의 상표입니다. NVIDIA®, Tesla® 및 GRID™는 NVIDIA Corporation의 등록 상표입니다. Microsoft®, Windows® 및 Office 365®는 미국 및/또는 기타 국가에서 Microsoft Corporation의 등록 상표 또는 상표입니다. YouTube®는 Google Inc의 등록 상표입니다. Arria® 및 Stratix®는 미국 특허청과 Trademark Office 및/또는 기타 국가에서 Altera Corporation의 상표로 등록되어 있습니다. Xilinx®, UltraScale™, 및 Alveo™는 Xilinx, Inc의 상표이며, SQream Technologies®는 SQream Technologies Ltd.의 등록 상표입니다. Kinetica®는 미국 및/또는 기타 국가에서 Kinetica와 자회사의 상표입니다. FASTDATA.io®는 미국 및/또는 기타 국가에서 FASTDATA.io의 등록 상표입니다. Amulet Hotkey®는 영국, 미국 및/또는 기타 국가에서 Amulet Hotkey Ltd.의 상표입니다. Tracowell Systems은 미국 및/또는 기타 국가에서 Tracowell Systems의 등록 상표입니다.

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