

HPE Apollo Server Systems

고성능 컴퓨팅 및 인공지능/빅데이터를 위한 고집적 최적화 솔루션



목차

고성능 컴퓨팅 및 인공지능/빅데이터를 위한 고집적 최적화 솔루션.....	3
HPE Apollo 시스템 제품군.....	4
HPE Apollo 2000 시스템 제품라인.....	5
HPE Apollo 70 시리즈.....	6
HPE Apollo 2000 시리즈.....	7
HPE Apollo 4000 시리즈.....	9
HPE Apollo 4200 Gen10 서버.....	10
HPE Apollo 4510 Gen10 서버.....	11
HPE Apollo 6000 시스템.....	12
HPE Apollo 6500 Gen10 시스템.....	15
HPE Apollo 6500 Gen10 Plus 시스템.....	16
HPE SGI 8600 시스템.....	17
HPE 기술 서비스.....	20
NVIDIA GPU 시리즈.....	21
NVIDIA A100 Tensor 코어 GPU.....	25

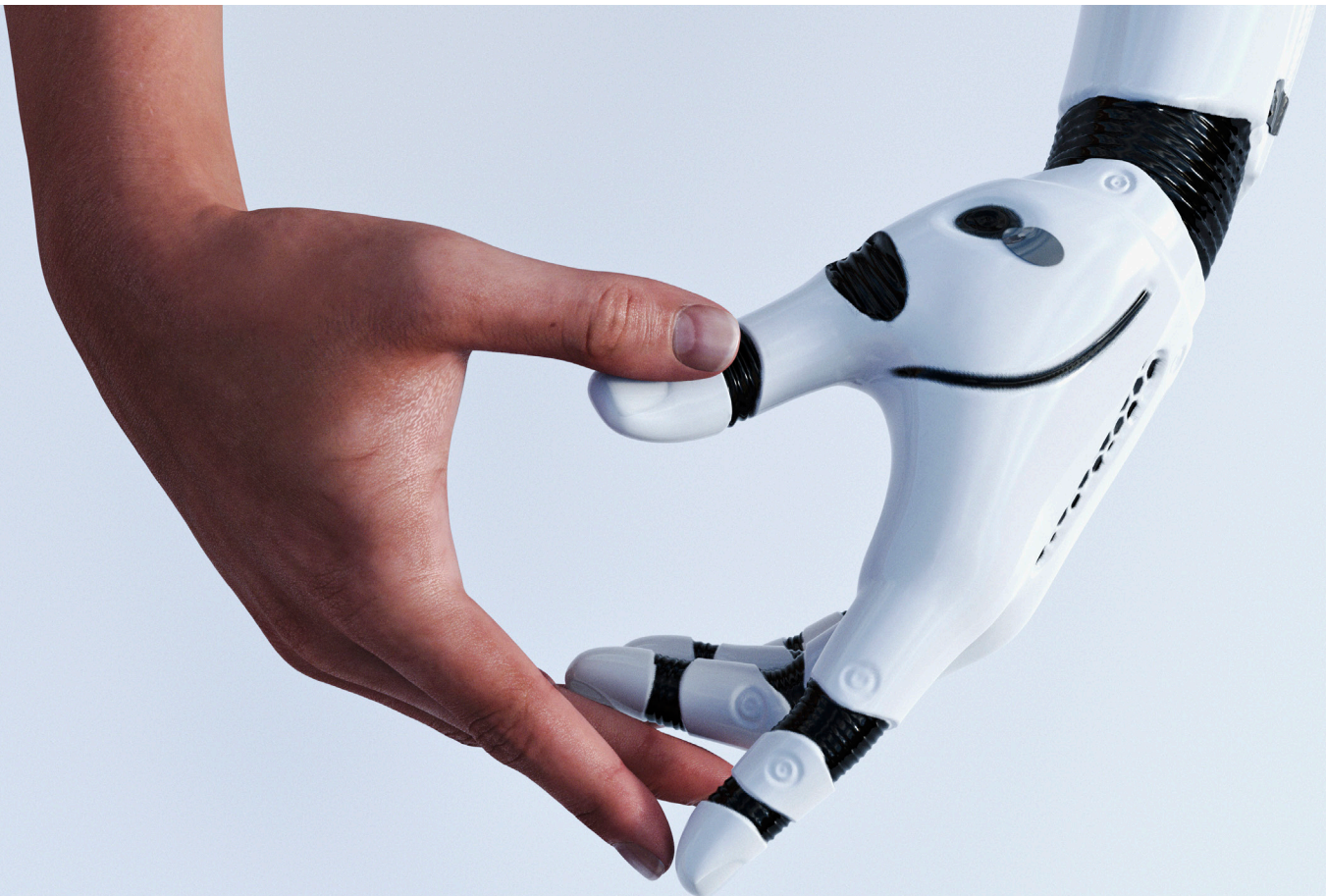
고성능 컴퓨팅 및 인공지능/빅데이터를 위한 고집적 최적화 솔루션

HPE Apollo 시스템 제품군은 빅데이터, 분석, 개체 스토리지 및 고성능 컴퓨팅(HPC) 워크로드에 랙 스케일의 효율적인 컴퓨팅, 스토리지, 네트워킹, 전력 및 냉각 솔루션을 제공하도록 설계되었습니다. 이러한 랙 스케일 효율성을 바탕으로 다음과 같이 탁월한 비즈니스 이점을 제공하고 있습니다.

- 특정 워크로드에 최적화된 시스템에서 확장성, 성능 및 효율성을 적정 수준으로 조정
- 수개월이 걸리던 구현 시간을 단 며칠로 단축함으로써 가치 창출을 가속화
- 스케일업 및 스케일아웃 솔루션에 유연한 아키텍처를 제공
- 자본 지출(CAPEX) 및 운영 비용(OPEX)을 대폭 절감
- 파트너, 통합업체 및 ISV(독립 서비스 공급업체)가 참여하는 확장된 에코시스템을 활용
- 모든 것을 갖춘 서비스 및 지원 기능을 통해 안심하고 사용

HPE Apollo 시스템 포트폴리오에서 스케일아웃 컴퓨팅을 담당하는 솔루션으로는 하이퍼스케일과 범용 스케일아웃 컴퓨팅용 Apollo 2000 시스템, 빅데이터 분석 및 개체 스토리지용 Apollo 4000 시스템 제품군, HPC 및 슈퍼컴퓨터용 Apollo 6000 및 SGI 8600 수냉 시스템이 있습니다.

이렇게 포괄적인 제품군 덕분에 규모에 관계 없이 모든 기업들이 고밀도 서버 스토리지, 관리 및 랙 스케일 효율성을 경험할 수 있게 되었습니다. 또한, 단계적 접근 방식을 통해 데이터 기반 기업을 위한 빅데이터, 개체 스토리지 및 HPC를 활용하고 싶어하는 고객들이 간편하면서도 합리적으로 솔루션을 구현할 수 있도록 시작점을 제공하고 있습니다.





HPE Apollo 시스템 제품군

HPC 및 슈퍼컴퓨터를 위한 솔루션

Apollo 2000 시스템: 기업이 스케일아웃 아키텍처로 나아갈 수 있는 “다리” 역할

공간과 리소스의 제약이 있는 가운데 더 높은 컴퓨팅 성능을 배포하여 클라우드 비즈니스의 이점을 누리고, 웹 기반 애플리케이션을 배포하며, HPC 성능을 높여 연구 속도를 높이고 새 제품 및 서비스의 시장 출시를 가속화하십시오. Apollo 2000 시스템은 기존의 데이터 센터가 스케일아웃 아키텍처로 나아갈 수 있도록 하는 “다리” 역할을 함으로써 기업이 기존 투자를 활용하면서도 공간과 비용을 줄일 수 있는 밀도 최적화 인프라를 자연스럽게 도입할 수 있도록 합니다.

Apollo 6000 시스템: 밀도, 성능 및 전력 효율성을 높이고 소유 비용을 줄인 랙 스케일 솔루션

HPC 및 효율성에 대한 요구가 끊임없이 증가하고 있는 상황에서 HPE는 새로운 접근 방식을 통해 한 발 앞서 나가고 있습니다. 서버에만 매달리지 않고 랙 수준의 솔루션을 설계하여 적절한 비용으로 적절한 컴퓨팅을 제공함으로써 사용자가 인프라와 예산을 최대한 활용할 수 있는 방법으로 눈을 돌린 것입니다. HPE Apollo 6000 시스템은 워크로드 요건을 정확하게 충족하도록 시스템을 맞춤화할 수 있는 유연성을 바탕으로 공간은 줄이고 성능은 업계 최고 수준으로 높였습니다.

Apollo 6500 시스템: AI 및 이기종 컴퓨팅을 위해 설계된 고성능 시스템

최근 고성능 및 다양한 형태의 AI 시스템에 대한 요구가 늘어남에 따라, 이러한 요구를 충족하기 위하여 개발된 Apollo 6500 Gen10의 다양한 Accelerator 지원 및 유연한 확장성을 제공하는 시스템입니다. 컴퓨팅 성능이 향상됨에 따라 고려되어야 하는 전력 효율성 및 냉각 효율성 역시 업계 최고 수준으로 높여 안정적인 시스템 운영이 가능합니다.

SGI 8600 시스템: 슈퍼컴퓨팅 기술의 발전

HPE SGI 8600 시스템은 과학 기술 혁신의 동력이 되어줄 온수를 이용한 수냉식 기술을 통해 “친환경적”인 방식으로 업계의 가장 까다로운 문제들을 해결해주는 HPC 솔루션입니다. HPE SGI 8600 시스템은 탄소배출량을 줄일 수 있도록 실질적인 도움을 주고 블레이드를 통해 서비스 편의성을 제공함으로써 새로운 차원의 고밀도를 실현했습니다.

빅데이터 구현을 위한 솔루션

Apollo 4000 시스템 포트폴리오: 빅데이터를 위해 특별 설계된 서버 솔루션

이 새로운 제품군은 Hadoop과 기타 빅데이터 분석 솔루션 및 개체 스토리지 시스템 등 빅데이터 IT 혁신을 주도하고 있는 기술들을 지원할 수 있도록 특별 설계되었습니다. 이들 시스템은 데이터 센터의 공간, 에너지 및 시간의 제약을 극복함으로써 시장 점유율을 확보하고 비즈니스 성장을 가속화할 수 있도록 경쟁력을 제공합니다.

HPE Apollo 2000 시스템 제품라인

고객이 스케일 아웃 아키텍처로 나아갈 수 있게 도와드리는 역할

클라우드, 웹 기반 애플리케이션 및 HPC의 이점을 누리기 위해서는 더 높은 컴퓨팅 성능을 배포해야 하지만, 공간 및 리소스의 제약이 여기에 걸림돌이 되고 있습니다. Apollo 2000 시스템은 기존의 랙 서버 데이터 센터가 스케일아웃 아키텍처로 나아갈 수 있는 “다리” 역할을 합니다. 따라서 기업은 기존 투자를 활용하면서도 공간과 비용을 줄일 수 있는 밀도 최적화 인프라를 자연스럽게 도입할 수 있습니다.

다양한 워크로드를 위한 유연한 구성

HPE Apollo 2000 시스템은 작은 데이터 센터에서 높은 성능과 워크로드 용량을 지원하는 고밀도 솔루션으로써, 공간의 제약이 있는 데이터 센터나 원격 사이트에 적합합니다. 실제로, 단일 2U 새시에 4개의 독립적인 핫 플러그 방식의 Apollo 2000 서버가 탑재되기 때문에 비슷한 가격대에서 표준 1U 서버의 2배에 달하는 밀도를 제공합니다(4U가 아닌 2U의 랙 공간에 4개의 서버 탑재).

유연한 구성 옵션을 갖춘 Apollo 2000 시스템은 HPC 클러스터를 포함한 다양한 워크로드에 적합합니다. 동일한 새시에 여러 서버를 조합할 수 있는 능력과 고유의 드라이브 매핑 유연성 덕분에 다양한 애플리케이션에 최적화된 구성이 가능합니다. 새시 또는 새시 그룹은 규모에 맞는 구현과 향후 확장을 위해 비용 효율적인 모듈형의 2U 빌딩 블록 역할을 하도록 맞춤 구성이 가능합니다.

데이터 센터에 원활하고 손쉽게 통합

Apollo 2000 시스템은 서비스 중단이나 환경 변경 없이도 기존의 엔터프라이즈 데이터 센터에 원활하게 배포되도록 설계되었습니다. 기존 랙 서버와 동일한 하드웨어 및 소프트웨어 도구를 이용하고 동일한 서비스 절차 및 관행에 따라 개별 서버 차원에서 시스템을 관리할 수 있기 때문에 변경으로 인해 발생하는 비용을 대폭 줄일 수 있습니다.

HPE Apollo 2000 시스템은 기업이 비용을 절감할 수 있는 효율적인 스케일아웃 아키텍처로 나아가는 데 필요한 가치와 적합한 특성을 제공합니다.

주요 기능 및 이점

기존 데이터 센터를 위한 밀도 최적화

- 2U 새시에 최대 4개의 서버 탑재(1U 서버보다 2배 높은 밀도)
- 기존 데이터 센터를 위한 기존 랙 및 케이블링 지원
- 비용 효율적 구성 제공

다양한 워크로드를 위한 유연한 구성 옵션

- 워크로드 최적화를 위한 서버 조합
- 가속기, top bin CPU 및 고속 HPC 클러스터링을 통한 HPC 성능 지원
- 워크로드 최적화를 위해 스토리지 유연성과 다양한 I/O 옵션 제공

HPE ProLiant Gen10을 통해 손쉽게 확장

- HPE ProLiant Gen10과 동일한 엔터프라이즈급 관리 및 운영 도구 지원
- HPE iLO 관리를 통해 관리 시간과 비용 절약
- HPE Advanced Power Manager를 통해 랙당 용량을 더욱 효율적으로 사용 가능
- HPE Insight Cluster Management Utility를 통해 모든 규모의 컴퓨팅 클러스터를 모니터링, 관리 및 최적화

모든 HPC 옵션을 완벽하게 지원하는 Apollo 2000 시스템

제품 설계, 테스트 시뮬레이션, 금융 리스크 모델링, Monte Carlo 시뮬레이션, 과학 연구 모델링 등 다양한 애플리케이션에서 HPC 클러스터를 최적화합니다. HPC 옵션에는 top bin CPU, 고속 메모리, 통합 가속기(GPU 또는 보조 프로세서), 고속 클러스터 패브릭 및 I/O 상호 연결 등 HPC 워크로드에서 최적의 성능 및 최고의 가성비를 손쉽게 달성할 수 있도록 해주는 기능들이 포함되어 있습니다.

HPE Apollo 70 시리즈

새롭게 선보이는 ARM Processor 고집적 시스템



전면



서버 내부



후면

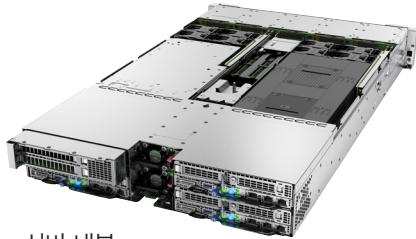
Apollo AR64z 1U

Apollo AR64z 2U

지원 프로세서	Cavium's ThunderX2™ Arm® processor
프로세서 수	2
프로세서당 코어	28/32
최대 프로세서 성능/캐시	2.0/2.2 Ghz
최대 내장 메모리	16 DIMM slots, 2666 MT/s, 512 GB Max(16 x 32 GB)
스토리지 드라이브 베이	최대 2 LFF/SATA/SAS SSD/HDD
최대 내장 스토리지 용량	16TB
PCI 슬롯	1x PCIe 16배속 Gen3
Onboard 네트워크	Dual Port SFP+: 10 GbE Mellanox® CX4LX Single Port RJ-45: 1GbE NIC - Management only 1x USB 3.0, 1xSUV, 1x UART Single Port Mellanox CX-5 100 Gb/s VPI adaptor(InfiniBand or Ethernet)
시스템 관리	IPMI 2.0/UEFI
보증기간 (연 기준: parts/labor/onsite)	3/3/3

HPE Apollo 2000 시리즈

엔터프라이즈급 스케일아웃 아키텍처를 위한 유연성 최적화



서버 내부

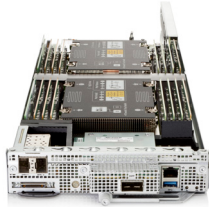


전면

	HPE Apollo r2200	HPE Apollo r2600	HPE Apollo r2800
종류	Gen10 12 LFF 디스크 또는 SSD 샷시	Gen10 24 SFF 디스크 또는 SSD 샷시	드라이브 매핑 기능이 있는 Gen10 24 SFF 디스크 또는 SSD 샷시
스토리지 구성	서버 노드에 동일하게 할당 된 12개의 LFF 핫 플러그 SAS/SATA HDD/SSD	24개의 SFF 핫 플러그 SAS/SATA HDD/SSD 또는 16 SFF SAS/SATA + 8 NVMe HDD, 서버 노드에 동일하게 할당	16 SFF NVMe BP로 사용자 정의 드라이브 할당을 워크로드 부하에 맞출 수 있습니다.
크기	2U: 너비 17.64 in. x 깊이 34.62 in.	2U: 너비 17.64 in. x 깊이 33.02 in.	2U: 너비 17.64 in. x 깊이 33.02 in.
전원 공급 장치	(2) N 또는 N + N 이중화가 있는 800W/1400W 277VAC/1600W 플래티넘 전원 공급 장치	(2) N 또는 N + N 이중화가 있는 800W/1400W 277VAC/1600W 플래티넘 전원 공급 장치	(2) N 또는 N + N 이중화가 있는 800W/1400W 277VAC/1600W 플래티넘 전원 공급 장치

HPE ProLiant XL170r/XL190r Gen10 서버

인클로저에는 서버를 대, 을 대까지 설치할 수 있습니다.



HPE ProLiant XL170r Gen10 Server

HPE Apollo r2000 시리즈 새시는 새시 당 1~4개의 서버가 있는 42U 랙에 최대 80개의 서버가 장착된 독립적으로 관리 가능한 4개의 HPE ProLiant XL170r Gen10 서버를 수용합니다.



HPE ProLiant XL190r Gen10 Server

HPE Apollo r2000 시리즈 새시는 새시 당 하나 또는 두개의 서버가 있는 42U 랙에 최대 40개의 서버와 80개의 통합 가속기를 갖춘 독립적으로 서비스 가능한 2개의 HPE ProLiant XL190r Gen10 서버를 수용합니다.

최대 수	1U half-width, 2P, Apollo 2000 사시당 최대 4개	2U half-width, 2P, Apollo 2000 사시당 1~2개
프로세서	8/10/12/14/16/18/20/22/24 코어, 최대 3.6GHz CPU 속도 및 최대 150 W의 전력 등급을 지원하는 Intel Xeon Scalable Processors	8/10/12/14/16/18/20/22/24 코어, 최대 3.6GHz CPU 속도 및 최대 150 W의 전력 등급을 지원하는 Intel Xeon Scalable Processors
칩셋	Intel® C622	Intel C622
메모리	최대 1.5 TB의 16 DIMM 슬롯 2933 MT/s DDR4 SmartMemory (DIMM 당 12개의 DIMM x 128 GB)	최대 1.5 TB의 16 DIMM 슬롯 2933 MT/s DDR4 SmartMemory (DIMM 당 12개의 DIMM x 128 GB)
네트워크 모듈	네트워킹 및 클러스터링 옵션을 선택할 수 있는 두 개의 I/O 슬롯: <ul style="list-style-type: none"> • 1/10/40 Gb 이더넷 • 100 Gb/s EDR InfiniBand 또는 Intel OPA • 파이버 채널 	네트워킹 및 클러스터링 옵션을 선택할 수 있는 두 개의 I/O 슬롯: <ul style="list-style-type: none"> • 1/10/40 Gb 이더넷 • 100 Gb/s EDR InfiniBand 또는 Intel OPA • 파이버 채널
PCIe 3.0 슬롯	PCIe 슬롯 1개와 FlexibleLOM 또는 2개의 PCIe 슬롯	추가 구성 기능을 제공하기 위해 다중 구성에서 최대 4개의 PCIe 슬롯: <ul style="list-style-type: none"> • 서버 당 최대 2개의 통합 가속기(NVIDIA® GPU) 또는 추가 I/O 옵션 • 다양한 PCIe 및 FlexibleLOM 조합을 포함한 네트워킹 및 클러스터링 옵션 선택 가능
스토리지	노드 당 최대 24개의 드라이브 듀얼 SATA 호스트 기반 M.2 2280 NGFF SSD - 내부 핫 플러그 HDD 지원 내부 USB 포트 r2800 새시의 하드 드라이브 매핑 가능	노드 당 최대 24개의 드라이브 듀얼 SATA 호스트 기반 M.2 2280 NGFF SSD - 내부 핫 플러그 HDD 지원 내부 USB 포트 r2800 새시의 하드 드라이브 매핑 가능
스토리지 컨트롤러	통합 스마트 어레이 S100i 스토리지 컨트롤러 HPE SmartCache 및 RAID 10 Advanced Data Mirroring과 같은 고급 배열 기능이 있는 PCIe 호스트 버스 어댑터 및 Smart Array 컨트롤러(선택 사양)	통합 스마트 어레이 S100i 스토리지 컨트롤러 HPE SmartCache 및 RAID 10 Advanced Data Mirroring과 같은 고급 배열 기능이 있는 PCIe 호스트 버스 어댑터 및 Smart Array 컨트롤러(선택 사양)
지원되는 가속기	N/A	서버당 최대 2개까지 지원 NVIDIA, AMD, and Intel GPGPUs
관리	HPE iLO 5 관리로 관리 시간 및 비용 절약 랙 수준의 관리를 제공하는 HPE Apollo 플랫폼 관리자, HPC 클러스터의 효율적인 배치 및 관리를 가능하게 해주는 CMU(Insight Cluster Management Utility)	HPE iLO 5 관리로 관리 시간 및 비용 절약 랙 수준의 관리를 제공하는 HPE Apollo 플랫폼 관리자, HPC 클러스터의 효율적인 배치 및 관리를 가능하게 해주는 CMU(Insight Cluster Management Utility)
공통 워크로드	고성능 컴퓨팅, 중소 규모 HPC 클러스터, 기존 데이터 센터의 대형 HPC 클러스터 대기업, 중소 규모 기업 및 원격 위치, 프라이빗 클라우드 배치를 위한 스케일 아웃 아키텍처 전통적인 데이터 센터를 운영하는 서비스 제공 업체	고성능 컴퓨팅, 중소 규모 HPC 클러스터, 기존 데이터 센터의 대형 HPC 클러스터 대기업, 중소 규모 기업 및 원격 위치, 프라이빗 클라우드 배치를 위한 스케일 아웃 아키텍처 전통적인 데이터 센터를 운영하는 서비스 제공 업체

HPE Apollo 4000 시리즈 - 빅데이터용 설계

HPE Apollo 4200 시스템은 기존의 랙 서버 데이터 센터를 운영 중인 대기업과 중소기업 고객들이 밀도 최적화된 빅데이터 솔루션으로 나아갈 수 있는 “다리”를 제공합니다.

빅데이터가 폭발적으로 증가하면서 다양한 기업들이 데이터를 수집해서 이를 경쟁 우위로 전환하기 위한 방안을 모색하고 있습니다. 여기서 핵심은 Hadoop 기반의 빅데이터 분석 솔루션 및 개체 스토리지 솔루션 시스템과 같이 오늘날 빅데이터의 양과 속도, 다양성을 처리하기 위해 등장한 새로운 기술들을 어떻게 활용하는 것인가입니다. 이러한 새로운 기술들은 효과적이지만, 이에 따른 워크로드가 페타바이트(PB) 규모로 커지면 기존의 범용 인프라는 한계에 봉착하게 되고 데이터 센터에 용량의 제한과 에너지 비용의 상승, 인프라 복잡성 심화, 비효율성 같은 문제가 발생할 수 있습니다. 빅데이터의 가치를 극대화하려면 빅데이터 워크로드용으로 특별 설계된 시스템이 필요합니다. HPE Apollo 4000 시스템은 Hadoop과 기타 빅데이터 분석 및 개체 스토리지 시스템을 위해 특별 설계된 새로운 시스템 제품군입니다.

Apollo 4000 시스템을 사용하면 페타바이트 수준을 능가할 정도로 증가하고 있는 데이터를 비용 효율적으로 분석하고 이 정보를 통해 통찰력을 얻어 전략적인 결정을 보다 신속하게 내릴 수 있습니다. 또한, 이러한 이점을 실현하면서 동시에 데이터 센터가 안고 있는 공간, 에너지 및 시간의 제약을 극복할 수 있습니다.

HPE Apollo 4200 Gen10 서버

- 기업이 빅데이터 솔루션으로 나아갈 수 있는 “브릿지” 역할

HPE Apollo 4200 Gen9 서버는 대기업 및 SME(중소기업)의 기존 랙 서버 데이터 센터를 위해 설계된 전혀 새로운 방식의 밀도 최적화 서버 스토리지 솔루션입니다. 2U 크기의 이 다목적 빅데이터 서버는 기존의 데이터 센터에 원활하게 통합되기 때문에 기존과 동일한 랙 치수, 케이블링, 서비스 기능, 관리 절차 및 도구를 이용할 수 있습니다. 이 모든 것을 고려해 볼 때, 이 제품은 기업 성장에 따라 적절한 가격으로 확장할 수 있는 능력을 갖추고 빅데이터를 위해 특별 설계된 서버 인프라로 전환하기에 적합한 시스템입니다.

주요 기능 및 이점

업계 최고의 스토리지 용량

- LFF 서버는 서버당 최대 336TB의 직접 연결 스토리지와 최대 28개의 핫스왑 LFF SAS 또는 SATA HDD/SSD를 지원
- SFF 서버는 서버당 최대 90TB의 직접 연결 스토리지를 제공하고 최대 50개의 핫스왑 SFF SAS 또는 SATA HDD/SSD 지원

기존의 랙 서버 데이터 센터에 맞게 설계

- 표준 크기의 랙, 전면/후면에서 핫 플러그 방식으로 디스크 교체, 뒷면 케이블링, 표준 랙 서버 시스템 관리
- 동일한 랙, 케이블링 및 서비스 기능을 바탕으로 기존의 데이터 센터에서 플러그 앤 플레이 방식으로 통합할 수 있기 때문에 서비스 중단이나 재교육 없이도 손쉽게 구현 및 지원이 가능하고 동일한 시스템 관리 기능을 이용 가능

성과 비용 효율성의 적절한 균형 고려

- 비용 효율적으로 2U에 대기업 및 SME 개체 스토리지 솔루션을 구축해서 점진적으로 확장
- 프로세서 서버 구성 옵션 제공:
 - Intel® Xeon® 8200/6200/5200 시리즈 프로세서
 - 최대 2,933MHz에서 최대 1TB DDR4 메모리가 장착되는 16개의 메모리 DIMM 슬롯(소형 개체에서 빠른 성능이 필요하거나 실시간에 가까운 분석 소프트웨어에서 인메모리 데이터 처리가 필요한 개체 스토리지에 적합)
 - 스토리지 성능 옵션(SFF HDD 모델은 분석 워크로드를 위해 데이터를 신속하게 전송할 수 있도록 12G 출력 및 15k RPM의 SAS 및 SSD 드라이브를 지원)
 - 최대 5개의 PCIe Gen3 슬롯을 통해 초고속 I/O가 필요한 애플리케이션의 네트워킹 및 클러스터 성능 요건 충족

개체 스토리지용

HPE Apollo 4200 LFF 시스템은 소규모의 개체 스토리지를 구현하거나 기존의 엔터프라이즈 랙 서버 데이터 센터에 “플러그 앤 플레이” 방식으로 통합하는 데 적합합니다.

Hadoop 및 빅데이터 분석용

HPE Apollo 4200 SFF 시스템은 Hadoop 기반의 병렬식 데이터 마이닝을 토대로 하는 분석 솔루션을 비롯해 NoSQL 기반의 빅데이터 분석 솔루션에 적합합니다.

HPE Apollo 4200 Gen10 서버

Moonshot 서버 카트리지



HPE Apollo 4200 Gen10 LFF 서버



HPE Apollo 4200 Gen10 SFF 서버

프로세서 제품군	Intel® Xeon® 6200 시리즈	Intel® Xeon® 6200 시리즈
폼 팩터	2U 랙 서버	2U 랙 서버
스토리지 유형	최대 24 LFF 핫 플러그 SAS/SATA/SSD + 4 LFF 또는 2 SFF 옵션(후면의 드라이브 케이징)	최대 48 SFF 핫 플러그 SAS/SATA/SSD + 2 SFF 옵션(후면의 드라이브 케이징)
스토리지 용량	최대 336TB(24 + 4 LFF 8TB HDD)	최대 384TB(48 + 2 SFF 7.68TB HDD)
스토리지 컨트롤러	Flexible Smart Array P840ar 및 Dynamic Smart Array B140i HPE Flexible Smart Array 또는 Smart HBA 컨트롤러 옵션	Flexible Smart Array P840ar 및 Dynamic Smart Array B140i HPE Flexible Smart Array 또는 Smart HBA 컨트롤러 옵션
프로세서 수	서버당 1~2개	서버당 1~2개
가용 프로세서 코어	8/10/12/14/16/18/20/22/24	8/10/12/14/16/18/20/22/24
프로세서 주파수(최대)	1.6GHz~3.6GHz	1.6GHz~3.6GHz
메모리	HPE SmartMemory DIMM 슬롯 16개 최대 1.0TB DDR4 메모리	HPE SmartMemory DIMM 슬롯 16개 최대 1.0TB DDR4 메모리
네트워킹	2 x 1Gb 이더넷 + FlexibleLOM 및 PCIe 옵션	2 x 1Gb 이더넷 + FlexibleLOM 및 PCIe 옵션
확장 슬롯	최대 5개의 PCIe 슬롯 + FlexibleLOM 지원	최대 5개의 PCIe 슬롯 + FlexibleLOM 지원
규모에 최적화된 관리를 위해 권장되는 관리 도구	HPE iLO 4 HPE Advanced Power Manager HPE Insight Cluster Management Utility	HPE iLO 4 HPE Advanced Power Manager HPE Insight Cluster Management Utility
시스템 팬 기능	최대 10개 팬(이중화용)	최대 10개 팬(이중화용)
전원 공급 장치 유형	최대 2개의 전원 공급 장치, 800W 및 1,400W Flex 슬롯, 핫 플러그 방식의 이중화된 전원 공급 장치	최대 2개의 전원 공급 장치, 800W 및 1,400W Flex 슬롯, 핫 플러그 방식의 이중화된 전원 공급 장치
QuickSpecs URL	h20195.www2.hpe.com/v2/gethtml.aspx?docname=a00056091enw	

HPE Apollo 4510 Gen10 서버

확장형 오브젝트 스토리지를 위한 최적의 솔루션



서버 내부



전면



후면

샤시	<ul style="list-style-type: none"> - 4 RU, HPE 표준 시리즈 1075 mm 랙에 설치할 수 있음 - Gen10 플렉스 슬롯 전원 공급 장치(AC 및 DC 버전), 최대 1600 W - 각각 30개의 핫 플러그 가능한 LFF 베이이 있는 듀얼 HDD 드로어즈, 측면 로드가 가능한 전면, 총 60개의 LFF 디스크 베이
프로세서	<ul style="list-style-type: none"> - 최대 2개의 Intel® Xeon® Processor Scalable Family, 최대 28 코어의 CPU, 최대 2x 150 W 프로세서 (전체 150 W 프로세서 지원 목록은 QuickSpecs 참조) - 최대 512 GB(8 x 64 GB)가 각 프로세서와 연결되어 서버 노드 당 최대 512 GB
메모리	<ul style="list-style-type: none"> - DDR4-2933 MHz 저전압 메모리, 16 DIMM 슬롯, 채널당 2 슬롯, 프로세서 당 6 채널 - 2-1-1 DIMM 구성, 프로세서 당 최대 2개의 NVDIMM
OS drive controller/drives	<ul style="list-style-type: none"> - 옵션: 서버 노드의 Gen10 AROC - 옵션: 대량 보관 서랍 용 Gen10 PCIe 스탠드 업 카드 2개
스토리지	<ul style="list-style-type: none"> - 최대 60 LFF 데이터 드라이브 12G SAS 지원 - 최대 12TB HDD 드라이브 지원 - SSD 드라이브, NVMe 드라이브 및 NVMe 가속기 카드에 대한 여러 가지 옵션. 현재 옵션은 HPE 솔리드 스테이트 디스크 드라이브(SSD 및 Accelerator) QuickSpecs
PCIe 슬롯	2개의 프로세서가있는 최대 1 x 16로우 프로파일 및 2 x 16 전체 높이 PCIe 3.0 슬롯과 1x FlexibleLOM 슬롯 (x8)을 지원
네트워킹	<ul style="list-style-type: none"> 듀얼 포트 1GbE - FlexibleLOM 및 PCIe 스탠드 업 카드(선택 사양)
시스템 관리	<ul style="list-style-type: none"> - HPE iLO 5 관리 엔진 및 전면 iLO 서비스 포트 - HPE 어드밴스드 전원 관리(옵션)
데이터 센터 지원(옵션)	HPE OneView
클러스터 관리(옵션)	HPE Insight Cluster Management Utility(CMU)
보증기간	3/3/3

“HPE는 EDA 워크로드에서 성능을 최대 35% 높일 수 있을 것으로 기대하고 있습니다. 이러한 서버를 5,000대 이상 배포한 결과, 더 높은 랙 밀도와 전력 효율성을 얻은 동시에 Intel® Silicon 설계 엔지니어들에게 향상된 애플리케이션 성능을 제공할 수 있었습니다.”

-Kim Stevenson, Intel CIO

³ Synopsys SNUG 2014 컨퍼런스에서 발표된 Synopsys의 벤치마크를 통해 문서화

⁴ HPE의 내부 분석 결과 대 Dell 계산기 (AC 입력), Dell M620 1 x 2P대 1 x 1P, HPE 전력 소모량: 153W로 측정, Dell: 전력 계산기 측정 결과 281W, 153/281 = 0.54(46% 절약)

⁵ HPE 내부 계산: (1.2/0.54)/0.51 = 4.36 배 향상

⁶ HPE 내부 계산: HPE 대 Dell M620 10U/16 서버, 48U의 HPE 160, 160/16 = Dell 인클로저 10개, 10 x 10 = 100U, 100U/48 = 랙 2.08개 > 랙 3개, 1/3 = 67% 랙 공간 절약, 최소 60%의 랙 공간 절약

⁷ HPE 내부 계산: 서버당 153W AC 입력, Dell: 전력 계산기를 통해 서버당 281W AC 입력(281W-153W) = 128W, 128x160개 서버 = 20.47kW, 20.47kW x 0.14 달러 kWh x 24시간 x 365일/년 = 25,100 달러/년, 랙당, 25,100달러/160개 서버 = 서버당 연간 157달러에 상응하는 전력 절약(Dell 전력 계산기 결과는 0.14달러)

HPE Apollo 6000 Gen10 시스템

밀도, 성능 및 전력 효율성을 높이고 TCO를 절감한 랙 스케일 솔루션

HPC 및 효율성에 대한 요구가 끊임없이 증가하고 있는 상황에서 HPE는 새로운 접근 방식을 통해 한 발 앞서 나가고 있습니다. 서버에만 매달리지 않고 랙 수준의 솔루션을 설계하여 적절한 비용으로 적절한 컴퓨팅을 제공함으로써 사용자가 인프라와 예산을 최대한 활용할 수 있는 방법으로 눈을 돌린 것입니다.

와트 및 달러당 최고 수준의 성능

- 단일 스레드 애플리케이션에 20% 향상된 성능 제공³
- 시스템의 에너지 소비를 46% 절감⁴
- 와트 및 달러당 성능을 4배 향상⁵

랙 스케일 효율성

- 경쟁사의 블레이드보다 공간을 60% 절약⁶
- 서버당 연간 157달러의 운영 비용(OPEX) 절감⁷
- 다음을 통해 단순화된 랙 스케일 관리 효율성 달성:
 - Smart Update
 - 통합 관리 도구
 - 네트워킹 유연성
 - 비용 효율적인 이중화를 통한 전력 효율성 향상
 - Advanced Power Manager

TCO를 절감할 수 있도록 워크로드에 맞게 솔루션을 설계할 수 있는 유연성

- 1,000개의 서버로 3년간 300만 달러에 달하는 비용 절감
- 혁신 존(zone)을 통해 비용 절감 효과를 높이는 동시에 워크로드 요건에 최적화된 NIC, FlexibleLOM 옵션을 선택할 수 있도록 지원
- 워크로드에 따라 인프라를 유연하게 맞춤 설계:
 - 모듈형의 단일 인프라에서 컴퓨팅, 스토리지 및 GPU/가속기 트레이를 선택하여 새시 또는 랙당 확장을 손쉽게 수행
 - 랙 수준에서 컴퓨팅 및 스토리지를 선택할 수 있는 유연성(SL4540/기타 스토리지)
 - Advanced Power Manager를 통해 손쉽게 관리

Apollo 6000 Systems

랙 크기의 공냉식 고성능 컴퓨팅은 밀도 최적화, 향상된 전력 효율성 및 낮은 소유 비용을 제공합니다.



다양한 성능, 효율성 및 가치 창출 시간

매우 복잡한 문제를 신속하게 해결해야 하는 IT 관리자에게, HPE Apollo 6000 Gen10 시스템의 새로운 통합 설계는 수백 대의 서버를 신속하게 배치하고 즉시 클러스터를 사용할 수 있는 기능을 제공합니다. 이 시스템의 첨단 기술은 성능을 제공하면서 에너지를 절약하도록 최적화되어 있습니다. 통합 설계는 배치를 간소화하며 포괄적인 관리 도구는 IT 오버헤드를 줄입니다.

첨단 기술과 성능이 제공하는 혜택

- 이전 세대에 비해 최대 2배 향상된 성능²을 제공하는 Intel® Xeon® Scalable 프로세서 제품군으로 랙당 323 TFLOPS³를 달성할 수 있습니다.
- 향후 기술을 지원하도록 미래형 아키텍처에서 최신 컴퓨팅, 스토리지 및 패브릭 기술을 지금 이용할 수 있습니다.
- 랙 규모의 효율성 향상
- 통합 설계를 통해 랙 레벨의 신뢰성, 가용성, 서비스 용이성 및 관리성 기능의 혜택을 제공합니다.
- 냉기통로(cold aisle) 전면 액세스 노드를 이용하여 컴퓨팅 및 스토리지를 신속하게 배치, 서비스하고 관리할 수 있습니다.

² 32배 정밀도 플롭 x 28 코어 x 2.5 GHz x 144 프로세서/랙 = 323 TFLOPS/랙

³ Intel AVX2(Haswell 및 Broadwell) 프로세서 대비 피크 성능 향상. Intel 내부 테스트, 2016년 9월.



TCO 절감

- 최적화된 통합 아키텍처 설계로 전력 및 공간 요구사항을 낮춥니다.
- 패브릭 기반에 연계된 최적 설계로 고립된 스위치 포트를 제거합니다.

HPE Apollo k6000 새시의 특징

- 12U 공랭식 새시
- 24개의 전면 액세스 노드

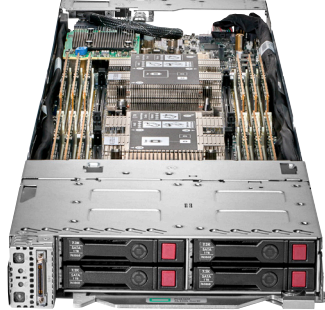
고도로 통합된 새시:

- 공통 냉각 기능을 갖춘 이중구조의 핫플러그 전원공급장치
- 고속 InfiniBand EDR(Enhanced Data Rate) 또는 Intel OPA(Omni-Path Architecture) 스위치 중에서 선택 가능
- 10GbE 스위치 또는 패스스루 모듈

HPE Apollo k6000 Chassis

Extreme compute configuration N + 0 (15.9 kW)	Typical configuration N + 1 (13.2 kW)
Node with 2 x 165 W	Node with 2 x 145 W
12 DDR4 DIMMs	12 DDR4 DIMMs
2 SATA HDDs	2 SATA HDDs
2 x EDR Mezzanine Adapter + switch	EDR Mezzanine Adapter + switch
2 x 10/40GbE switch	10/40GbE switch
APM module	APM module
4 x front and 12 x rear fans	4 x front and 12 x rear fans

HPE ProLiant XL230k Gen10 서버



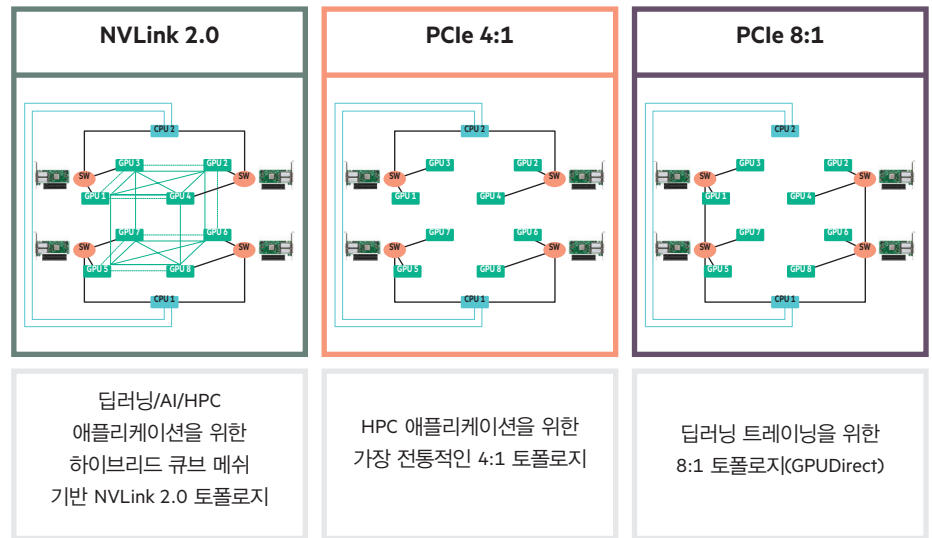
폼 팩터	HPE ProLiant XL230k Gen10 서버는 HPE Apollo k6000 샤시에 탑재되는 싱글 슬롯 서버입니다.
프로세서	Intel Xeon Platinum Processor를 포함한 Intel Xeon Processor Scalable 제품군
Cores	4/6/8/12/14/16/18/20/22/24/26/28
칩셋	Intel C622 series chipset
최대 프로세서 속도	3.6 GHz
드라이브 구성	4 SFF SAS/SATA/NVMe
지원 드라이브	Hot-plug 2.5-inch SAS/SATA
메모리 슬롯	16 x DDR4-2666 DIMMs 12 x RDIMMs/LRDIMMs 4 x 3D XPoint
최대 메모리	1024 GB(16 x 64 GB) per server tray
메모리 유형, ECC	DDR4; RDIMM/LRDIMM 2666 MHz
네트워크 옵션	2 x 10GbE integrated ports, InfiniBand EDR or Intel OPA Mezzanine cards
스토리지 컨트롤러	HPE Dynamic Smart Array S100i SATA Controller HPE Smart Array E208i-p Controller HPE Smart Array P408i-p Controller
확장 슬롯	2 x InfiniBand EDR or Intel OPA Mezzanine Adapters PCIe slot options: <ul style="list-style-type: none"> • 1 x external x 16 low profile • 1 x internal and 1 x external x 8 low profile PCIe • Supports -F Omni-Path directly off CPU(independent of above I/O options)
USB 포트	1 internal and 2 external via serial, USB, video(SUV) port
관리	New HPE Integrated Lights-Out(iLO) 5 remote manageability
OS 지원	Microsoft® Windows Server® Red Hat® Enterprise Linux® 7 and 8 SUSE Linux Enterprise Server(SLES) VMware® ESX® CentOS

HPE Apollo 6500 Gen10 시스템

딥러닝과 HPE Apollo 6500 - 완벽한 조합

한층 다양한 애플리케이션 영역에 AI를 실현하는 딥러닝은 고성능 컴퓨팅에 기반하여 방대한 양의 데이터 내에서 패턴과 관계를 도출합니다. 지금까지, 기존 HPC시스템은 딥러닝이 요구하는 전례 없는 수준의 요건을 부응할 수 없었습니다. 따라서, 딥러닝의 이점은 사실상 이용되지 않은 상태로 남아 있습니다.

HPE Apollo 6500 Gen10 시스템은 다양한 워크로드에 따른 토폴로지를 제공합니다.



HPE ProLiant XL270d Gen10 서버



그래픽 처리 장치	서버당 최대 8개의 GPU, 최대 350 와트 GPU 지원, NVLink 2.0 또는 표준 PCIe를 이용한 고속 전용 GPU간 통신 옵션 제공
지원 가속기	NVLink 및 PCIe 구성의 NVIDIA® Tesla V100, T4, Quadro RTX 6000, NEC Vector Engine IOB, AMD MI25
어댑터	4가지 고속 패브릭 어댑터(이더넷, Intel Omni-Path Architecture, InfiniBand EDR, 향후 InfiniBand HDR)
프로세서	Intel® Xeon® 확장형 프로세서(8200 및 6200), 최대 165 와트, 최대 28 코어 또는 최대 3.6 GHz 주파수 제공
메모리	24개의 2933 MHz DDR4 SmartMemory DIMM, 3 TB(24 x 128 GB) 메모리 제공
스토리지	최대 16개의 SFF 장치: 최대 4 NVMe SSD 전면 액세스 스토리지를 탑재한 핫플러그 SAS/SATA SSD
네트워크 옵션	<ul style="list-style-type: none"> • 4개의 1GbE 단일 포트 RJ-45 • InfiniBand EDR 또는 Intel OPA 또는 고속 이더넷을 지원하는 4개의 16 PCI Express 슬롯, 최대 100 Gbps
I/O 슬롯	<ul style="list-style-type: none"> • 1 x 16 FHHL PCIe Gen3 시스템 보드 모듈 • GPU 모듈의 4 x 16 PCIe 로우 프로파일 Gen3
전원공급장치	최대 4개의 2200W 80 Plus Platinum(2 + 2)
랙	표준 1075 mm 깊이 랙에 대하여 설계

HPE Apollo 6500 Gen10 Plus 시스템

가속화된 컴퓨팅의 새로운 유형



그림 1. HPE Apollo 6500 Gen10 Plus 시스템

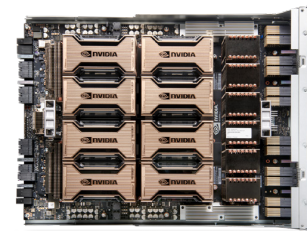


그림 2. NVIDIA A100 Tensor Core GPUs

Hewlett Packard Enterprise는 기업의 AI 여정에 힘을 실어줄 컴퓨팅 혁신을 가속하여 다가오는 엑사스케일 시대를 맞이하고 있습니다.

HPE는 에지에서 클라우드에 이르기까지 모든 데이터 센터 워크 로드를 처리할 수 있는 엔드 투 엔드 솔루션을 배포하는 것입니다. 이를 달성하기 위해 성능과 밀도를 염두에 두고 설계된 HPE Apollo 6500 Gen10 Plus 시스템(그림 1)을 소개합니다.

시장에서 가장 효과적인 액셀러레이터를 기반으로 하는 HPE 시스템은 광범위한 미션 크리티컬 HPC 및 AI 애플리케이션에 CPU-GPU 다차원적 컴퓨팅을 활용합니다.

최신 NVIDIA® A100 Tensor Core GPUs(그림 2) 및 2세대 AMD EPYC™ 시리즈 프로세서의 탁월한 처리 용량을 사용하여 기존 인프라의 부담을 없애고 엑사스케일을 지원합니다. 새로운 A100 GPU는 이러한 강력한 가속 컴퓨팅 솔루션을 개선하기 위해 높은 처리량에서 낮은 대기 시간을 보장합니다. 3세대 Tensor Core를 사용하는 A100은 최대 수천 개의 GPU까지 효율적으로 확장할 수 있습니다.

NVIDIA 다중 인스턴스 GPU(MIG) 기술을 사용하면 7개의 격리된 GPU 인스턴스로 나눌 수 있어 다양한 워크 로드를 가속할 수 있습니다. GPU 활용도를 극대화하면 Volta에서 Ampere 아키텍처까지 성능을 최대 20배가량 획기적으로 개선할 수 있습니다. A100은 대규모 모델 훈련을 위해 최대 6배의 기본 성능을 제공하고, 추론 작업 시 MIG로 성능을 최대 7배까지 개선하여 인사이트 도출 시간을 대폭 단축합니다.

HPE Apollo 6500 Gen10 Plus 시스템

제품 사양	HPE ProLiant XL645d server	HPE ProLiant XL675d server
밀도	단일 XL675d 또는 2개의 XL645d를 지원하는 6U 시스템	
GPU	NVIDIA HGX A100 4-GPU는 NVIDIA A100 Tensor Core GPU와 새로운 NVIDIA NVLink 및 NVSwitch를 결합합니다 Infinity Fabric이 적용된 AMD MI100 사용 가능한 기타 주요 액셀러레이터 최대 4개의 Double Wide PCIe 또는 8개의 Single Wide PCIe GPU	NVIDIA HGX A100 8-GPU는 NVIDIA A100 Tensor Core GPU와 새로운 NVIDIA NVLink 및 NVSwitch를 결합합니다 Infinity Fabric이 적용된 AMD MI100 사용 가능한 기타 주요 액셀러레이터 최대 10개의 Double Wide PCIe 또는 16개의 Single Wide PCIe GPU
상호연결장치	이더넷, InfiniBand 또는 HPE Slingshot 등 최대 3개의 고속 패브릭 상호연결 지원 이더넷 옵션을 위한 추가 포트	이더넷, InfiniBand 또는 HPE Slingshot 등 최대 6개의 고속 패브릭 상호연결 지원
프로세서	노드 당 단일 AMD EPYC 7002 시리즈 프로세서, 수요되는 워크 로드에서 최고 주파수 및 코어 수에 대해 최대 280W 제공	노드 당 두개의 AMD EPYC 7002 시리즈 프로세서, 수요되는 워크 로드에서 최고 주파수 및 코어 수에 대해 최대 280W 제공
메모리	8 3200MT/s DDR4 SmartMemory	32 3200MT/s DDR4 SmartMemory
저장소	총 8개의 SFF 드라이브 (옵션) M.2	총 16개의 SFF 드라이브 (옵션) M.2
시스템 관리 및 시스템 보안	HPE iLO5(Integrated Lights Out), HPCM(HP Performance Cluster Manager), HPE 컨테이너 플랫폼, HPE OneView iLO5 Silicon Root of Trust, iLO Advanced (옵션)	
OS 지원	HPE Cray OS, Microsoft Windows Server, Red Hat®, Ubuntu, VMware®	
전원	새시 당 최대 6개의 3000W 플래티넘 핫 플러그 전원 공급 장치로 모든 구성을 위한 완전 중복 전원, 전력 캐핑은 서버 및 새시 차원에서, 그리고 랙 및 행 차원에서는 APM으로 제공.	
냉각	15개의 80mm 듀얼 로터 핫 플러그 가능 새시 팬 신규: HPE와 완전 통합, 설치 및 지원되는 직접 액체 냉각 시스템을 갖춘 HPE Apollo 6500 Gen10 Plus 시스템	
저장 컨트롤러	내장형 SATA 또는 다양한 HPE Smart Array 컨트롤러	
보증(부품, 공임, 현장 지원)	3/3/3	

HPE SGI 8600 시스템

업계 최고 수준의 기술, 성능 및 시스템 아키텍처 제공



HPE SGI 8600 E-셀

HPE SGI 8600 시스템 아키텍처

HPE SGI 8600은 처음부터 최고의 성능, 규모, 밀도를 제공하도록 설계된 수냉식, 트레이 기반, 고밀도, 클러스터링 컴퓨팅 시스템입니다. HPE SGI 8600 시스템의 기본 구성 요소는 E-셀입니다. E-셀은 냉각 랙으로 구분된 2개의 42U 높이 E-랙으로 구성되어 있습니다. E-셀은 폐쇄 루프 냉각 기술을 사용하며 데이터 센터로 가열된 공기를 배출하지 않는 밀폐형 유닛입니다.

직접 연결된 수냉식 콜드 싱크는 프로세서, GPU, 스위치 등 고출력 기기의 열을 보조 CDU(Cooling Distribution Unit)를 통해 효율적으로 제거합니다. E-셀은 냉각을 위해 설비에서 공급하는 물을 사용하며 물 온도가 7-32°C(45-90°F) 이내인 경우 데이터 센터의 열이 상승하지 않습니다. 물 온도가 실내 환경에 영향을 받지 않도록 무려 90°F에 달해 기존 공냉식 설계에 비해 냉각 비용을 크게 절감할 수 있습니다.

각 E-랙에는 최대 9개의 컴퓨팅 트레이에 전력, 냉각, 시스템 제어, 네트워크 패브릭을 제공하는 최대 4개의 컴퓨팅 트레이 인클로저가 포함되어 있습니다. 따라서 E-셀 하나에 최대 72개의 컴퓨팅 트레이를 포함할 수 있습니다. 컴퓨팅 트레이와 패브릭 스위치 간 통합 미드프레임은 인클로저 내 패브릭과 전력을 케이블 없이 연결할 수 있는 인프라를 제공합니다.

기술 사양: HPE SGI 8600 시스템

HPE SGI 8600 E-랙

치수	89.6"(H)x24.0"(W)x48.0"(D)
컴퓨팅 트레이 인클로저	랙당 컴퓨팅 트레이, 스위치, 전력이 통합된 최대 4개의 컴퓨팅 트레이 인클로저. 컴퓨팅 트레이 인클로저당 최대 8 + 1의 이중 3000 W 전원 공급 장치
냉각	실내 환경에 영향을 받지 않는 최대 32°C의 데이터 센터 공급수 및 많은 전력을 사용하는 구성 요소를 위한 직접 수냉 방식 및 폐쇄 루프 공기 흐름

컴퓨팅 트레이 옵션

HPE XA730i Gen10 Server

폼 팩터	HPE XA730i는 HPE SGI 8600 컴퓨팅 인클로저 어셈블리를 위한 싱글 슬롯 트레이입니다.
프로세서	Intel Xeon Processor Scalable Family(전체 SKU 스택 지원)
컴퓨팅 노드/트레이	2소켓 CPU 노드 4개
메모리/노드	노드당 최대 1536 GB, 노드당 12개 DIMM 슬롯(CPU 소켓당 6개)
메모리 기술	8, 16, 32, 64, 128 GB DDR4 2666 MT/s ECC 등록 DIMM
확장 공간 옵션(노드당)	최대 (2) 2.5" SATA HDD/SSD (1) HDD/SSD 및 (1) x16 소규격 PCIe 슬롯
패브릭 옵션(메자닌 카드를 통해)(노드당)	싱글 또는 듀얼 포트 Mellanox Connect-IB(FDR) 싱글 또는 듀얼 포트 Mellanox ConnectX-5(EDR) 싱글 또는 듀얼 포트 Intel Omni-Path

HPE XA780i Gen10 Server

폼 팩터	HPE XA780i는 HPE SGI 8600 컴퓨팅 인클로저 어셈블리를 위한 싱글 슬롯 트레이입니다.
프로세서	Intel Xeon Processor Scalable Family
컴퓨팅 노드/트레이	2소켓 CPU 노드 1개
메모리/노드	노드당 최대 1536 GB, 노드당 14개 DIMM 슬롯(CPU 소켓당 8개)
메모리 기술	8, 16, 32, 64, 128 GB DDR4 2666 MT/s ECC 등록 DIMM
GPU 옵션	NVLink 탑재 SXM2 GPU에 대해 최대 4개의 NVIDIA Tesla
확장 공간 옵션(노드당)	노드당 최대 4개의 SSD
패브릭 옵션(메자닌 카드를 통해)(노드당)	노드당 최대 4개의 고속 연결(Mellanox Connect-IB [FDR], Mellanox ConnectX-5 [EDR] 또는 Intel Omni-Path)

기술 사양(계속): HPE SGI 8600 시스템

통합 스위치 옵션

표준 InfiniBand	단일 36포트 FDR 또는 EDR InfiniBand 스위치 및 외부 18개 포트
프리미엄 InfiniBand	듀얼 36포트 FDR 또는 EDR InfiniBand 스위치 및 외부 36개 포트
Intel Omni-Path	싱글 48포트 Intel Omni-Path 스위치 및 외부 30개 포트

시스템 소프트웨어

운영 체제	SUSE Linux Enterprise Server 12 Red Hat Enterprise Linux 7 CentOS 7
HPE 관리	HPE SGI Management Suite
HPC 워크로드 관리자	Altair® PBS Professional Adaptive Computing Moab® HPC Suite Slurm Workload Manager
프로그래밍 언어, 디버거, 라이브러리	Intel® Parallel Studio XE, Intel® VTune™ Amplifier XE PGI® Professional Edition Rogue Wave Software® TotalView® Allinea Forge HPE Performance Software - 메시지 전달 인터페이스 Mellanox HPC-X: Open MPI, OpenSHMEM, Berkeley UPC, MXM, FCA NVIDIA CUDA 툴킷, OpenCL, OpenACC

HPE 기술 서비스 컨설팅 및 지원에 대한 자세한 내용은 hpe.com/services에서 확인하십시오.

HPE Factory Express는 구매하신 스토리지 및 서버 제품에 대해 맞춤형 설정과 배포 서비스를 제공합니다. 공장에서 하드웨어의 사양을 원하는 대로 설정함으로써 배포 속도를 높일 수 있습니다.

참조 사이트: www.hpe.com/kr/ko/product-catalog/services/services/pip.3777594.html

새로운 IT 제품 구매, 기존 자산 관리, 불필요한 장비 제거 등 맞춤형 IT 수명 주기 관리를 계획해 보십시오.

HPE가 제공하는 ExpertOne 교육 및 인증을 통해 필요한 기술을 배울 수 있습니다. HPE ProLiant 교육을 통해 빠르게 새로운 기술을 도입하고 운영 효율성을 향상시키며 HPE 제품에 대한 투자 가치를 최상으로 끌어올릴 수 있습니다. HPE 교육은 유연한 전송 옵션과 글로벌 교육 기능을 지원하여 언제 어디서든 이용하실 수 있습니다

자세히 알아보기

www.hpe.com/kr/ko/solutions/hpc-high-performance-computing.htm

HPE 기술 서비스

HPE 기술 서비스는 HPE Apollo를 최대한 활용할 수 있도록 지원하는 단일 소스 솔루션을 통해 신뢰감을 주고 위험성을 줄이며 고객에게 더욱 뛰어난 민첩성과 안정성을 실현하도록 합니다. 고객은 유연한 옵션을 통해 필요한 요건에 맞는 서비스 수준을 자유롭게 선택할 수 있습니다. HPE 기술 서비스의 컨설팅 및 지원을 활용하면 기존 환경에서도 문제없이 HPE Apollo 솔루션을 성공적으로 배포 및 운영하여 그 이점을 누릴 수 있습니다.

HPE와 연계하면 문제 예방 및 신속한 문제 해결이 가능합니다. HPE의 지원 기술은 수백만 개의 장치와 수천 명에 이르는 전문가들의 지식을 이용하여 언제 어디서든 최신 정보를 얻고 모든 것을 제어할 수 있는 토대가 될 것입니다.

유연한 서비스 옵션을 자유롭게 선택

- **설치 및 시작 서비스** - 신속하고 원활하게 가동할 수 있도록 도와줍니다.
- **HPE Datacenter Care** - 제어, 성능 및 간편성을 최대한 높일 수 있도록 향상된 콜 관리, 사전 예방적 서비스, 하드웨어 및 소프트웨어 지원을 적절하게 혼합하여 전체 IT 환경을 지원하는 가장 유연한 서비스입니다.
 - **HPE Flexible Capacity** - HPE Datacenter Care의 옵션 중 하나로 퍼블릭 및 온프레미스 IT의 이점과 함께 퍼블릭 클라우드 기능을 제공합니다. 이러한 "확장형 과금(pay-as-you-grow)" 솔루션 덕분에 제품이 조달되기를 기다릴 필요 없이 성장에 따라 즉각적인 확장이 가능합니다.
- **HPE Proactive Care 서비스**는 2가지 버전으로 제공됩니다.
 - **HPE Proactive Care** - HPE와 연계된 제품을 활용하여 개인 맞춤형 문제 예방 기능을 사용할 수 있는 것은 물론이고 다운타임을 최대 77%까지 줄이고¹¹ 100%에 가까운 진단 정확성을 달성하며¹² 단일 창구를 통해 IT 환경을 통합적으로 관리할 수 있습니다. 연중무휴 모니터링과 사전 장애 경고, 자동화된 콜 로깅 및 자동 부품 발송 같은 기능을 활용할 수 있습니다. 문제가 발생하는 경우, 전문가에게 신속하게 연락하여 완벽한 콜 관리를 통해 IT를 안정화할 수 있습니다.
 - **HPE Proactive Care Advanced** - 비즈니스에 중요한 IT를 실행하고 있는 서버용으로 개발된 서비스입니다. 해당 지역을 담당하는 고객 지원 매니저(ASM)를 배정함으로써 HPE Proactive Care 서비스를 한 단계 확장했습니다. ASM은 모범 사례에 대한 조언을 제공하고 전 세계 기술 전문가를 동원하는 것은 물론, 중요 이벤트 관리를 통해 복잡한 문제를 신속하게 해결함으로써 시스템이 최고의 성능을 유지하도록 돕고 있습니다.
- **HPE Foundation Care** - 지원 적용 범위와 응답 시간을 자유롭게 선택하여 합리적인 비용으로 하드웨어 및 소프트웨어 지원을 받을 수 있는 서비스입니다. 여기에는 업계 최고의 x86 운영 체제 소프트웨어를 통한 협력적 콜 관리가 포함되어 있습니다.
- **HPE Education Services** - 비용 및 리소스 관리 문제를 해결하고 최신 기술로 업데이트할 수 있도록 해주는 서비스입니다.
- **HPE Lifecycle Event Services** - 전략 수립부터 설계, 배포 및 운영에 이르기까지 전 과정에서 전문가의 지원을 받을 수 있는 서비스입니다.

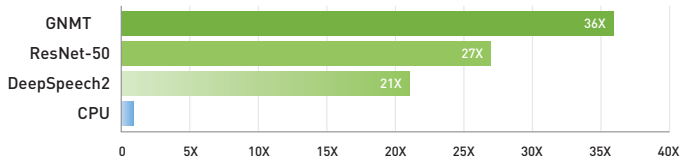


NVIDIA T4 TENSOR CORE GPU

더욱 강력한 스케일아웃 AI 트레이닝 및 추론

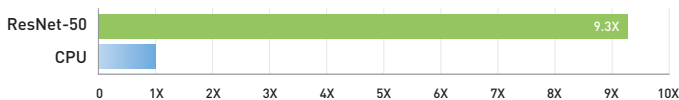
세계 최고 성능을 자랑하는 스케일 아웃 가속 그래픽 카드인 NVIDIA® T4 GPU로 서버의 속도를 높이십시오. NVIDIA Turing™ 텐서 코어가 탑재된 로우 프로파일 70W 설계로 획기적인 다중 정밀도 성능을 구현하여 다양한 최신 애플리케이션을 가속화할 수 있습니다. 에너지 효율이 뛰어난 70W의 소형 PCIe 폼팩터를 갖춘 이 고성능 GPU는 스케일아웃 서버에 최적화되어 있으며 최첨단 AI 구현을 위해 설계되었습니다.

추론 성능

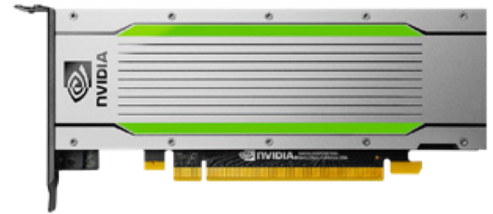


NVIDIA T4 GPU 17개와 XEON GOLD 6140 CPU 탑재 서버의 비교

트레이닝 성능



듀얼 NVIDIA T4 GPU와 듀얼 소켓 XEON GOLD 6140 CPU 탑재 서버의 비교



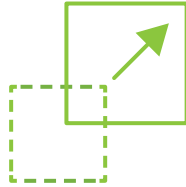
GPU 아키텍처	NVIDIA TURING
NVIDIA Turing 텐서 코어 수	320
NVIDIA CUDA 코어 수	2,560
단정밀도	8.1TFLOPS
혼합 정밀도(FP16/FP32)	65TFLOPS
INT8	130TOPS
INT4	260TOPS
GPU 메모리	16GB GDDR6 300GB/s
ECC	지원
인터커넥트 대역폭	32GB/sec
시스템 인터페이스	x16 PCIe Gen3
폼팩터	로우 프로파일 PCIe
열 솔루션	수동적
컴퓨팅 API	CUDA, NVIDIA TensorRT™, ONNX

데이터 센터 가속화를 실현하는 스케일아웃 성능

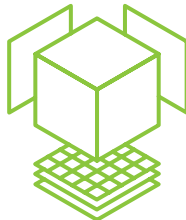


70W의 소형 폼팩터 설계로 스케일아웃 서버에 최적화된 T4는 CPU 대비 에너지 효율이 50배 높으며 그 결과 운영 비용을 크게 절감할 수 있습니다.

NVIDIA의 추론 플랫폼은 지난 2년간 10배 넘게 효율성을 높이면서 분산형 AI 트레이닝 및 추론을 위한 에너지 효율 솔루션으로서 최고의 자리를 지키고 있습니다.



NVIDIA T4 데이터 센터 GPU는 분산 컴퓨팅 환경에 이상적인 범용 가속 그래픽 카드입니다. 획기적인 다중 정밀도 성능으로 딥러닝 및 기계 학습의 트레이닝과 추론, 비디오 트랜스코딩, 가상 데스크톱 속도를 대폭 향상합니다. 모든 AI 프레임워크와 네트워크 유형을 지원하는 T4는 성능과 효율성을 크게 높여 대규모 배포의 효율성을 극대화합니다.



Turing 텐서 코어 기술은 AI용 다중 정밀도 컴퓨팅과 더불어 FP32부터 FP16, INT8 및 INT4 정밀도에 이르기까지 획기적인 성능을 제공합니다.

트레이닝에서는 CPU 대비 최대 9.3배, 추론에서는 최대 36배나 높은 성능을 발휘합니다.



NVIDIA TESLA V100 GPU 가속기

HPC 및 AI 전용 데이터센터 GPU

NVIDIA® Tesla® V100은 지금까지 개발된 데이터센터 GPU 중 세계에서 가장 앞선 데이터센터 GPU로서 AI, HPC 및 그래픽을 가속합니다. 최신 GPU 아키텍처인 NVIDIA Volta로 지원되는 Tesla V100은 단일 GPU로 최대 100개의 CPU에 성능을 제공하여 불가능의 영역에 머물러있던 도전 과제를 해결할 수 있도록 데이터 과학자, 연구자 및 엔지니어를 지원합니다.

사양

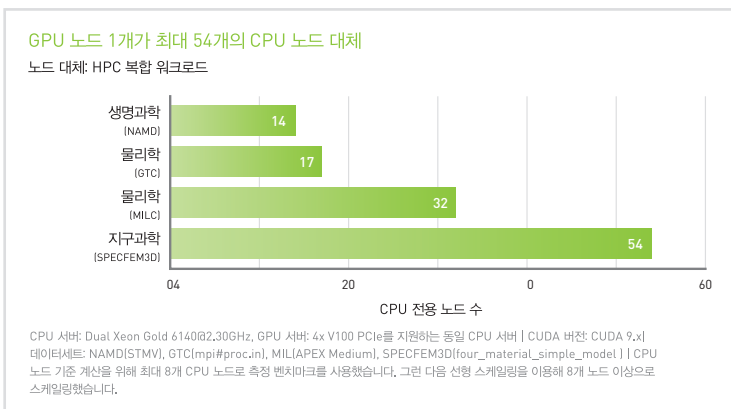
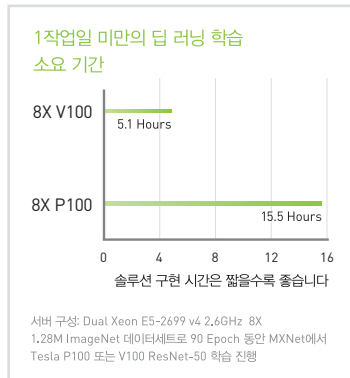
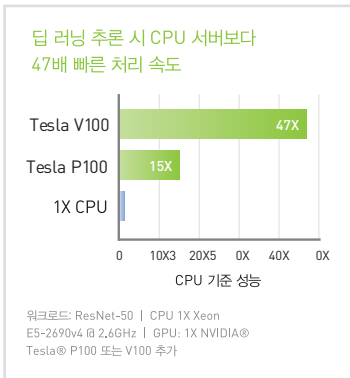


Tesla V100 PCIe

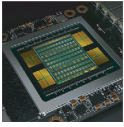


Tesla V100 SXM2

GPU 아키텍처	NVIDIA Volta	
NVIDIA Tensor 코어	640개	
NVIDIA CUDA® 코어	5,120개	
다중정밀도 성능	7 TFLOPS	7.8 TFLOPS
단정밀도 성능	14 TFLOPS	15.7 TFLOPS
Tensor 성능	112 TFLOPS	125 TFLOPS
GPU 메모리	32GB /16GB HBM2	
메모리 대역폭	900GB/초	
ECC	지원	
상호연결 대역폭	32GB/초	300GB/초
시스템 인터페이스	PCIe Gen3	NVIDIA NVLink
폼 팩터	PCIe 전체 높이/길이	SXM2
최대 전력 소비	250 W3	00 W
열 솔루션	패시브	
Compute APIs	CUDA, DirectCompute, OpenCL™, OpenACC	

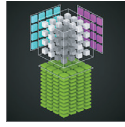


획기적인 혁신



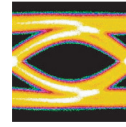
Volta 아키텍처

통합 아키텍처 내에서 CUDA 코어와 Tensor 코어를 페어링함으로써 Tesla V100 GPU가 탑재된 단일 서버는 수백 대의 범용 CPU 서버를 대체하는 성능으로 종래의 HPC와 딥 러닝을 수행할 수 있습니다.



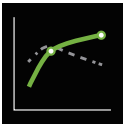
Tensor 코어

640개의 Tensor 코어가 탑재된 Tesla V100은 125 teraFLOPS의 딥 러닝 성능을 제공합니다. 이는 NVIDIA Pascal™ GPU와 비교해 DL 학습의 경우 12X Tensor FLOPS, DL 추론의 경우 6X Tensor FLOPS에 상응합니다.



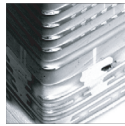
차세대 NVLink

Tesla V100에 탑재된 NVIDIA NVLink는 이전 세대 대비 2배 빠른 처리 속도를 지원합니다. 최대 8대의 Tesla V100 가속기를 최대 300GB/s의 속도로 상호 연결할 수 있어 단일 서버에서 구현 가능한 최고의 애플리케이션 성능이 발휘됩니다.



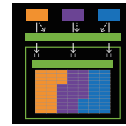
최대 효율 모드

새로운 최대 효율 모드를 이용해 데이터센터는 기존 전력 예산 범위 내에서 랙당 최대 40% 향상된 컴퓨팅 성능을 달성할 수 있습니다. 본 모드에서 Tesla V100은 최고 처리 효율로 가동되며 절반의 전력 소비로 최대 80%의 성능을 제공합니다.



HBM2

향상된 900GB/s의 기본 대역폭과 95%의 DRAM 사용 효율을 결합한 Tesla V100은 STREAM 측정기준으로 Pascal GPU보다 1.5배 향상된 메모리 대역폭을 제공합니다. 현재 Tesla V100은 표준 16GB 제품 메모리의 2배인 32GB 구성을 지원합니다.



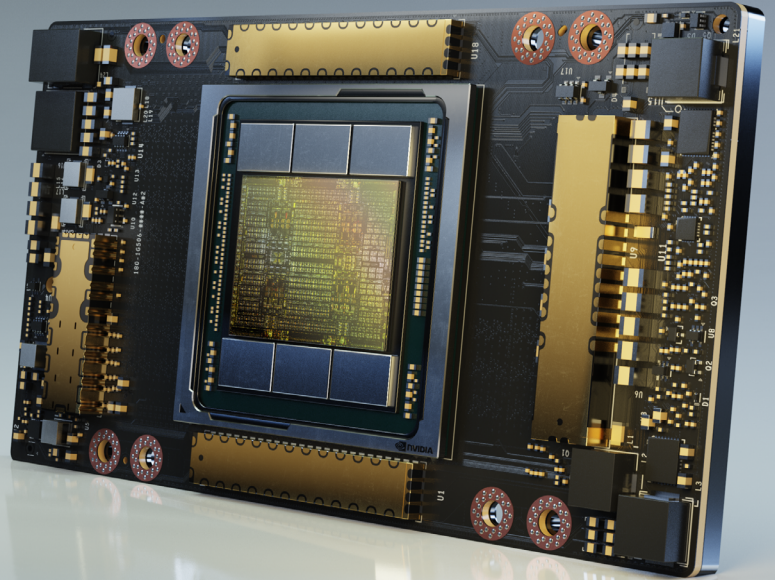
프로그래밍 기능

Tesla V100은 처음부터 철저하게 프로그래밍 기능 단순화를 위해 설계되었습니다. 새로운 독립 스레드 스케줄링 기능은 소규모 작업들 간 리소스 공유를 통해 미세 동기화를 지원하며 GPU 사용률을 향상시킵니다.

Tesla V100은 딥 러닝, HPC 및 그래픽을 위한 Tesla 데이터센터 컴퓨팅 플랫폼 제품입니다. Tesla 플랫폼은 550여 개의 HPC 애플리케이션과 모든 주요 딥 러닝 프레임워크를 가속화합니다. 데스크톱부터 서버 및 클라우드 서비스에 이르기까지 어디서든 사용할 수 있으며, 놀라운 성능 향상과 비용 절감 효과를 동시에 선사합니다.

모든 딥 러닝 프레임워크

550여 개의 GPU 가속 애플리케이션



NVIDIA A100 TENSOR 코어 GPU

모든 스케일에 걸친 전례없는 가속

NVIDIA A100 Tensor 코어 GPU는 모든 스케일의 AI, 데이터 분석, HPC 분야에서 전례없는 가속을 제공하여 세상에서 가장 어려운 컴퓨팅 문제를 해결합니다. A100은 NVIDIA 데이터 센터 플랫폼의 엔진으로서 수천 개의 GPU로 수직 확장할 수 있습니다. 또는 새로운 멀티-인스턴스 GPU(MIG) 기술을 사용해 7개의 독립된 GPU 인스턴스로 나뉘어 다양한 크기의 작업을 가속할 수 있습니다. A100의 3세대 Tensor 코어 기술은 다양한 작업을 위해 여러 종류의 정밀도를 지원하므로 빠르게 인사이트를 얻고 시장에 출시할 수 있게 합니다.

시스템 사양 (최고 성능)

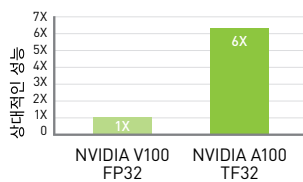
	NVIDIA HGX™ 용 NVIDIA A100 SXM4	NVIDIA A100 PCIe GPU
GPU 아키텍처	NVIDIA Ampere	
배정밀도 성능	FP64: 9.7 TFLOPS FP64 Tensor 코어: 19.5 TFLOPS	
단정밀도 성능	FP32: 19.5 TFLOPS Tensor Float 32 (TF32): 156 TFLOPS 312 TFLOPS*	
반정밀도 성능	312 TFLOPS 624 TFLOPS*	
Bfloat16	312 TFLOPS 624 TFLOPS*	
정수 성능	INT8: 624 TOPS 1,248 TOPS* INT4: 1,248 TOPS 2,496 TOPS*	
GPU 메모리	40 GB HBM2	
메모리 대역폭	1.6 TB/sec	
오류 정정 부호	Yes	
인터커넥트 인터페이스	PCIe Gen4: 64 GB/sec 3세대 NVIDIA® NVLink®: 600 GB/sec**	PCIe Gen4: 64 GB/sec 3세대 NVIDIA® NVLink®: 600 GB/sec**
폼 팩터	NVIDIA HGX™ A100에 있는 4/8 SXM GPU	PCIe
멀티-인스턴스 GPU (MIG)	최대 7 GPU 인스턴스	
최대 전력 소비	400 W	250 W
Delivered Performance for Top Apps	100%	90%
쿨링 솔루션	패시브 (Passive)	
컴퓨팅 API	CUDA®, DirectCompute, OpenCL™, OpenACC®	

* 구조적 회소성 사용

** HGX A100 서버 보드를 사용한 SXM GPU; 최대 2 GPU까지 NVLink Bridge를 사용한 PCIe GPU

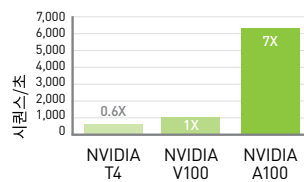
TF32를 사용한 기본 AI 훈련¹ 성능 6배 향상

대규모 BERT 훈련

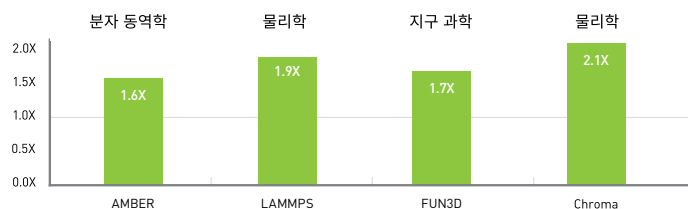


멀티-인스턴스 GPU(MIG)를 사용하여 AI 추론² 성능 7배 향상

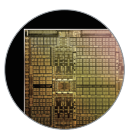
대규모 BERT 추론



HPC 성능³ 2배 이상 향상

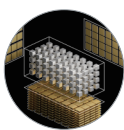


놀라운 혁신



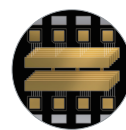
NVIDIA AMPERE 아키텍처

A100은 크고 작은 워크로드를 가속합니다. MIG를 사용해 A100 GPU를 작은 인스턴스로 나누거나 NVLink를 사용해 여러 GPU를 연결해 대규모 워크로드를 가속할 수 있습니다. A100은 작은 업무에서 멀티-노드 워크로드까지 다양한 사이즈의 가속 요구사항을 손쉽게 처리합니다. A100의 다재다능함 덕분에 IT 관리자는 데이터 센터에 있는 모든 GPU의 사용률을 항상 극대화할 수 있습니다.



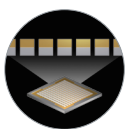
3세대 TENSOR 코어

A100은 312 TFLOPS의 딥러닝 성능을 제공합니다. 이는 NVIDIA Volta™ GPU보다 딥러닝 훈련에서 20배 높은 Tensor FLOPS이고 딥러닝 추론에서 20배 높은 Tensor TOPS입니다.



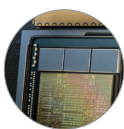
4세대 NVLINK

A100의 NVIDIA NVLink는 이전 세대보다 2배 높은 처리량을 제공합니다. NVIDIA NVSwitch™를 사용하면 최대 초당 600 기가바이트로 A100 GPU를 16개까지 연결하여 단일 서버에서 가능한 가장 높은 성능에 도달할 수 있습니다. HGXA100 서버 보드를 통한 A100 SXM GPU와 최대 2개 GPU까지 NVLink Bridge로 연결한 PCIe GPU에서 사용할 수 있습니다.



멀티-인스턴스 GPU (MIG)

하나의 A100 GPU는 7개의 GPU 인스턴스로 나눌 수 있습니다. 하드웨어 수준에서 완전히 독립적이며 각자의 고대역 메모리, 캐시, 컴퓨팅 코어를 가집니다. MIG 덕분에 개발자들은 모든 애플리케이션을 가속할 수 있고 IT 관리자는 모든 작업에 적절한 크기의 GPU 가속을 제공할 수 있습니다. 동시에 사용률을 최적화하고 모든 사용자와 애플리케이션에 가속 서비스를 제공할 수 있습니다.



HBM2

40 기가바이트 고대역 메모리 (HBM2) 덕분에 A100은 1.6TB/sec의 향상된 대역폭을 제공합니다. 또한 95%의 높은 DRAM 사용 효율을 제공합니다. A100은 이전 세대보다 1.7배 높은 메모리 대역폭을 제공합니다.



구조적 희소성

A100은 수백 만에서 수십억 개에 이르는 파라미터를 가진 대규모 네트워크입니다. 정확한 예측에 모든 파라미터가 필요하지 않기 때문에 정확도를 희생하지 않으면서 일부 파라미터를 0으로 바꾸어 희소(sparse) 모델을 만들 수 있습니다. A100의 Tensor 코어는 희소 모델의 성능을 2배로 높입니다. 이런 희소 기능은 AI 추론에 당연히 도움이 되고 모델 훈련의 성능도 향상시킬 수 있습니다.

NVIDIA A100 Tensor Core GPU는 딥러닝, HPC, 데이터 분석을 위한 NVIDIA 데이터 센터 플랫폼의 대표 제품입니다. 이 플랫폼은 700개 이상의 HPC 애플리케이션과 모든 주요 딥러닝 프레임워크를 가속화합니다. 데스크톱에서 서버, 클라우드 서비스에 이르기까지 어디에서나 사용할 수 있으며 성능을 극적으로 향상시키고 비용을 절감시켜 줍니다.

모든 딥러닝 프레임워크

700개 이상의 GPU 가속 애플리케이션

NVIDIA A100 Tensor 코어 GPU에 대한 더 자세한 내용은 www.nvidia.com/a100 을 참고하세요.

- 파이토치를 사용한 BERT 사전 훈련 처리량, 단계 1(2/3)과 단계 2(1/3) | 단계 1의 시퀀스 길이 = 128, 단계 2의 시퀀스 길이 = 512 | V100: NVIDIA DGX-1™ 서버, NVIDIA V100 텐서 Core GPU 8개, FP32 정밀도 | A100: NVIDIA DGX™ A100 서버, A100 8개, TF32 정밀도.
- BERT 대규모 추론 | NVIDIA T4 Tensor 코어 GPU: NVIDIA TensorRT™ (TRT) 7.1, INT8 정밀도, 배치 크기 256 | V100: TRT 7.1, FP16 정밀도, 배치 크기 256 | A100, 1g.5gb 7 MIG 인스턴스; TRT 시제품, 배치 크기 94, 희소한 INT8 정밀도.
- 사용된 V100은 단일 V100 SXM2. 사용된 A100은 단일 A100 SXM4. PME-Cellulose 기반의 AMBER, Atomic Fluid LJ-2.5를 사용한 LAMMPS, dpw를 사용한 FUN3D, szsc121_24_128을 사용한 Chroma.



Sign up for updates

★ Rate this document

휴렛팩커드 엔터프라이즈

제품정보 및 구입안내: 080-703-0700

©Copyright 2018 Hewlett Packard Enterprise Development LP.

본 문서에 게재된 정보는 사전고지 없이 변경될 수 있습니다. HPE 제품 및 서비스에 대한 보증은 오직 해당 제품 및 서비스에 첨부된 보증서 내용에만 상응합니다. 본 문서의 어떠한 부분도 추가적인 보증으로 유추될 수 없으며, HPE는 본 문서의 기술적 오류나 편집 상 오류, 또는 누락에 대한 책임을 지지 않습니다.

© 2020 NVIDIA Corporation. All rights reserved. NVIDIA, NVIDIA 로고, CUDA, DGX, HGX, HGX A100, NVLink, NVSwitch, OpenACC, TensorRT, Volta는 미국과 다른 나라에 있는 NVIDIA Corporation의 상표 및/또는 등록 상표입니다. OpenCL은 Khronos Group의 허락하에 사용하는 Apple Inc.의 상표입니다. 다른 상표와 저작권은 모두 각 소유자의 재산입니다. JUL20

4AA5-8958KOP,