

제품 개요



HPC 가속화된
인텔® 제온® 스케일러블 프로세서

인텔 HPC® 엔진으로 가장 까다로운 컴퓨팅 작업을 가속화 하십시오

4세대 인텔 제온 프로세서로

최대
1.56배 높아진

4세대 인텔 제온 스케일러블 프로세서의
LAMMPS 성능
(이전 세대 대비)



고객 사례: 인텔 제온 스케일러블
프로세서의 현실에서의 가속

버펄로 대학교는

산업 컴퓨팅 클러스터를 업그레이드하고
HPC 및 AI를 커뮤니티와 공유하고
있습니다.

[자세한 내용 보기 >](#)

CERN은 인텔 제온 스케일러블
프로세서와 인텔® 딥 러닝 부스트를
이용하여 입자 가속기 시뮬레이션의
속도를 높입니다.

[자세한 내용 보기 >](#)

고성능 컴퓨팅(HPC)은 과학적 발견, 엔지니어링 시뮬레이션 및 복잡한 시스템 모델링에 필수적입니다. 가속기는 CPU 코어 수를 늘리는 대신 성능을 높이는 효율적이고 효과적인 대안이 될 수 있습니다. 4세대 인텔® 제온® 스케일러블 프로세서에는 가속화 기능으로 작업을 오프로드 하여 HPC 워크로드 성능과 전력 효율성을 높일 수 있는 특별 제작된 가속기가 장착되어 있습니다.

새로운 시대를 여는 고성능 컴퓨팅

오랜 시간동안 우리는 HPC사용자가 비용을 제어하면서 향상된 컴퓨팅 속도와 액세스에 대한 요구를 어떻게 해결하는지 그 변화를 지켜봐 왔습니다. 업계는 비즈니스 통찰력을 신속하게 파악하고 중요한 비즈니스 의사 결정을 내릴 수 있도록 지원하는 도구로 HPC를 점점 더 많이 사용하고 있습니다.

인텔® HPC 엔진과 함께 새롭게 향상된 4세대 인텔 제온 스케일러블 프로세서의 기능은 시뮬레이션, 모델링 등과 같이 가장 빠르게 증가하는 워크로드 유형에서 성능을 향상시킬 수 있습니다. 4세대 인텔 제온 스케일러블 프로세서의 내장된 가속기는 대상 워크로드에 향상된 성능, 효율성 및 비용 절감을 제공합니다 : 인텔® 어드밴스드 매트릭스 익스텐션 (인텔® AMX), 인텔® 데이터 스트리밍 가속기(인텔® DSA), 인텔® 쿼어리스트 테크놀로지(인텔® QAT) 등.

인텔® 어드밴스드 매트릭스 익스텐션 (인텔® AMX)

AI 가속기

머신러닝(ML) 기술은 워크로드를 더욱더 효율적이고 효과적이며 통찰력 있게 만들어 줍니다. 업계 트렌드는 고객이 HPC 및 AI 기반의 솔루션을 통해 사업에 이익을 얻을 수 있게 합니다. 인텔은 AI 성능을 향상하도록 설계된 새로운 가속기를 도입하였으며, 인텔의 AI 전문성을 HPC와 AI 솔루션을 활용하는 고객에게 공유하고 있습니다.

4세대 인텔 제온 스케일러블 프로세서에 통합된 새로운 내장 가속기 엔진 중 하나인 인텔 AMX는 딥 러닝 인터페이스와 AI학습 성능을 위한 인텔의 새로운 발전을 보여줍니다. 인텔은 이전의 인텔 제온 스케일러블 프로세서의 내장형 AI 가속화 기능을 확장하여 인텔 AMX가 딥 러닝 워크로드의 핵심인 대용량 매트릭스 수학 계산을 한번의 작업으로 완료할 수 있도록 하였습니다. 또한 인텔 AMX는 2차원 레지스터 파일을 이용하여 대용량 데이터를 저장합니다. AI 워크로드를 가속하기 위해 구축된 인텔 AMX는 HPC와 AI가 융합되는 워크로드에 성능을 제공하는 데 매우 중요합니다.



더 빠른 HPC를 위한 기반, 인텔® 어드밴스드 벡터 익스텐션 512 (인텔® AVX-512)

모든 x86 CPU는 공통된 명령어 집합 아키텍처(ISA)를 공유합니다. 인텔은 2011년에 인텔® 어드밴스드 벡터 익스텐션(인텔® AVX)을 시작으로 세대를 거쳐 기본적인 x86 명령어를 새로운 워크로드로 확장하고 인텔의 역량을 확장하였습니다. 오리지널 인텔 AVX 명령어와 더불어 인텔 AVX-512¹ 및 인텔 AVX2²는 일반적인 컴퓨팅, AI 프로세싱 및 연산 집약적인 HPC 워크로드를 가속화 합니다. 예를 들어, 3세대 인텔 제온 스케일러블 프로세서의 경우 코어 별로 더 높은 성능을 제공함으로써 이전 세대보다 더 나은 컴퓨터 이용 공학(CAE) 애플리케이션 성능을 제공하였습니다.²

연산의 단계가 줄어들면 처리 속도는 더 빨라집니다

인텔 AVX-512의 "확장(extensions)"은 일반적인 컴퓨팅 작업을 보다 적은 단계로 줄이고, 통합 및 결합합니다. 단순한 예로, CPU에 3 x 3 x 3 x 3 x 3을 계산하도록 명령할 경우에는 5번의 클럭 주기가 필요합니다. 또는 같은 계산을 CPU가 한 번의 클럭 주기에서 실행할 수 있도록 3⁵에 대한 명령어를 수행할 수도 있습니다. 인텔 AVX-512는 이러한 로직을 이용하고 이를 FMA(Fused MultiPLY Add)와 같은 수백 개의 과제별 작업에 적용합니다. 인텔 제온 스케일러블 프로세서에는 코어당 2개의 FMA 장치가 있어 한번의 연산에 곱셈과 덧셈을 결합하여 계산 속도를 가속 합니다.

HPC를 위한 신경망 가속화, 인텔® 딥 러닝 부스트(인텔® DL 부스트)

딥 러닝 추론 및 학습과 함께 머신 러닝은 처리 속도를 가속화하고 정확도를 높이며 완전히 새로운 모델링과 분석 방법을 통해 HPC의 역량을 확장하고 있습니다. 인텔 DL 부스트에는 컴퓨팅 리소스와 캐시 사용률을 극대화하고 클럭 주기별 작업 수를 줄이기 위해 설계된 AVX-512 명령어 세 개가 결합된 벡터 신경망 명령어(VNNI)가 포함되어 있습니다. VNNI는 잠재적인 대역폭 병목 현상을 극복하여 딥 러닝 추론 워크로드를 가속할 수도 있습니다. 인텔 DL 부스트는 INT8(VNNI) 및 BF16 데이터 타입을 지원합니다. 세 개의 명령어 집합을 하나로 결합함으로써 인텔 DL 부스트는 워크로드 전반에 걸쳐 AI의 평균 성능을 더 높일 수 있습니다.

인텔 DL 부스트 익스텐션은 HPC 시스템에서 AI 융합을 지원하기 위해 메모리 요구 사항을 확장하지 않고도 AI 성능을 가속화 할 수 있습니다.

인텔® 퀵어시스트 테크놀로지 (인텔® QAT)

4세대 인텔 제온 스케일러블 프로세서를 위한 새로운 내장 가속기인 인텔 QAT를 통해 컴퓨팅 집약적 워크로드를 오프로드하여 공간을 확보하고 비용을 절감하십시오. 인텔 QAT는 가속화 된 암호화, 키 보호 및 데이터 압축을 제공하여 시스템 리소스 소비를 줄입니다. 이를 통해 고객에게 빅데이터 및 데이터베이스 애플리케이션에서 더 많은 Gbps와 Ops/Sec 성능을 제공하는 등 혜택을 제공합니다.

인텔 QAT는 암호화 및 압축과 관련된 오버헤드를 감소시켜 결과적으로 클러스터 성능 향상에 중요한 역할을 합니다.

인텔 QAT를 사용하면 암호화 및 데이터 압축 성능을 개선하는 동시에 데이터 공간을 줄여 각 코어는 더 많은 클라이언트에 서비스를 제공할 수 있습니다.

인텔® 데이터 스트리밍 가속기 (인텔® DSA)

데이터가 메모리, 스토리지, 네트워킹 서브시스템으로 이동하면 CPU 에 부담을 주게 됩니다.

인텔 제온 프로세서에 통합된 가속기인 인텔 DSA는 스트리밍 데이터 이동과 변환 작업을 향상하여 스토리지 및 네트워킹 그리고 데이터 집약적 워크로드에 높은 성능을 제공합니다. 인텔 DSA는 CPU, 메모리 및 캐시는 물론 연결된 모든 메모리와 스토리지, 네트워킹 장치 간에 데이터를 빠르게 이동할 수 있도록 지원합니다.

인텔 DSA는 성능을 향상할 수 있고, Open vSwitch(OVS)의 데이터 이동시간 지연을 줄여 보다 효과적인 네트워크 자동화를 가능하게 지원하는 것은 물론 중요한 작업을 위해 CPU 코어를 확보할 수도 있습니다. 인텔 DSA는 데이터 이동에 대해 높은 대역폭과 짧은 지연 시간을 제공할 뿐만 아니라, OVS 등 수많은 사용 사례를 위해 향상된 전력 효율성을 제공합니다.

고객은 OVS를 인텔® 인프라 프로세싱 유닛(인텔® IPU)에 오프로드 하여 성능을 더욱 향상하고 CPU 효율을 최적화할 수 있습니다.

인텔 제온 스케일러블 프로세서에는 HPC 가속기가 내장되어 있습니다

HPC 가속을 위한 핵심 기반은 모든 인텔 제온 스케일러블 프로세서에 내장되어 있으며 거의 모든 소프트웨어에서 사용할 수 있습니다. HPC 고객은 최소한의 노력으로 가속 기술이 주는 이점을 누릴 수 있습니다.

인텔® oneAPI HPC 툴킷은 벡터화, 멀티스레딩, 다중 노드 병렬화 및 메모리 최적화에 최신 기술을 이용하는 HPC 애플리케이션을 구축하기 위한 인텔® oneAPI 베이스 툴킷의 추가 기능입니다. 이 툴킷에는 개방형 메시지 전달 인터페이스(Open MPI) 라이브러리를 기반으로 하는 클러스터 분석과 튜닝 툴이 포함되어 있습니다.

차세대 HPC를 위한 가속화된 성능

HPC의 사용이 더욱 쉬워지고 비용이 감소함에 따라 슈퍼컴퓨팅 리소스의 상대적 가치는 기하급수적으로 증가하게 됩니다. 한때 국립 연구소와 글로벌 제조업체만 제한적으로 사용하던 컴퓨팅 파워는 클라우드 인스턴스와 하이브리드 HPC 클러스터를 통해 사용이 가능하게 되었습니다. 인텔® HPC 엔진은 전반적으로 HPC 성능을 향상시켜 더 많은 조직이 새로운 개발, 혁신 및 시장 출시를 가속화 하는 데 필요한 컴퓨팅 리소스에 액세스 할 수 있도록 지원합니다.

인텔 제온 프로세서에 내장된 인텔 HPC 엔진을 이용하여 가장 까다로운 컴퓨팅 작업을 정복하십시오.

자세히 알아보기

[인텔 고성능 컴퓨팅 >](#)

[인텔 AVX-512 >](#)

[인텔 딥 러닝 부스트 >](#)

[AI 및 HPC 컨버전스 >](#)

[인텔 제온 스케일러블 프로세서의 AI 및 딥 러닝 >](#)

지금 4세대 인텔 제온 스케일러블 프로세서로 클라우드 또는 자체 인프라에서 HPC 워크로드 가속화를 시작해 보십시오.

intel.com/hpc 방문하기

인텔 AMX가 성능에 미치는 '깊은' 영향

인텔 AMX를 이용하는 4세대 인텔 제온 스케일러블 프로세서와 3세대 인텔 제온 스케일러블 프로세서 비교

최대

7.3배 높아진

실시간 자연 언어 처리
추론 성능³

최대

4.1배 높아진

TensorFlow INT8
추론 성능⁴



¹intel.com/processorclaims에서 [H1] 참조: 4세대 인텔® 제온® 스케일러블 프로세서. 결과는 상이할 수 있습니다.

²인텔® 제온® 프로세서 HPC 성능: 제작

³intel.com/processorclaims에서 [A19] 참조: 4세대 인텔 제온 스케일러블 프로세서. 결과는 상이할 수 있습니다.

⁴intel.com/processorclaims에서 [A19] 참조: 4세대 인텔 제온 스케일러블 프로세서. 결과는 상이할 수 있습니다.

주의사항 및 면책 조항

성능은 사용, 구성 및 기타 요인에 따라 달라집니다. intel.com/PerformanceIndex에서 자세히 알아보기.

성능 결과는 환경 설정에 표시된 날짜 기준의 테스트를 기반으로 하며 공개적으로 제공되는 모든 업데이트를 반영하지 못할 수 있습니다. 구성 세부 정보는 백업을 참조하십시오. 어떠한 제품 또는 구성 요소도 절대적으로 안전할 수는 없습니다.

워크로드와 환경 설정에 대한 자세한 정보는 www.intel.com/processorclaims에서 4세대 제온 스케일러블 프로세스를 참조하십시오. 결과는 상이할 수 있습니다.

인텔® 어드밴스드 벡터 익스텐션(인텔® AVX)는 특정 프로세서 작업에 보다 많은 처리량을 제공합니다. 다양한 프로세서 파워의 특성으로 인해 AVX 명령을 활용할 경우 a) 일부가 정격 주파수 미만에서 작동하고 b) 인텔® 터보 부스트 테크놀로지 2.0을 이용하는 일부가 최대 터보 주파수 또는 어떠한 터보 주파수도 달성하지 못할 수 있습니다. 성능은 하드웨어, 소프트웨어 및 시스템 구성에 따라 다르며 자세한 정보는 intel.com/content/www/us/en/architecture-and-technology/turbo-boost/intel-turbo-boost-technology.html에서 확인할 수 있습니다.

인텔® 기술을 사용하려면 이용이 가능한 하드웨어, 소프트웨어 또는 서비스 활성화가 필요할 수 있습니다.

비용 및 결과가 다를 수 있습니다.

인텔은 인권을 존중하고 인권 남용에 가담하지 않습니다. 인텔의 글로벌 인권 원칙을 확인하십시오. 인텔 제품 및 소프트웨어는 국제적으로 인정되는 인권의 침해로 유발하거나 기여하지 않는 애플리케이션에서만 사용되어야 합니다.

© Intel Corporation. 인텔, 인텔 로고 및 기타 인텔 마크는 인텔사 또는 인텔 계열사가 등록한 상표입니다. 기타 이름 및 상표명은 해당 소유주가 재산을 주장할 수 있습니다. 0922/MP/CMD/PDF

가속기의 가용성은 SKU에 따라 상이합니다. 제품에 대한 자세한 내용은 인텔 제품 사양 페이지를 확인하십시오.