

5세대 인텔® 제온® 프로세서



더 빠른 메모리와 더 큰 라스트 레벨 캐시를 갖춘 5세대 인텔® 제온® 프로세서는 이전 세대보다 최대 42% 향상된 AI 성능을 제공하며, 3세대 인텔® 제온® 프로세서 보다 최대 14배 향상된 성능을 제공합니다.

신뢰할 수 있는 성능. 탁월한 효율성.

디지털 세상은 데이터로 가득 차 있습니다. 에지 컴퓨팅에서 수집한 원격 측정 데이터부터 온라인 거래, 클라우드 데이터 저장소에 이르기까지 우리 세상은 정보로 가득합니다. 하지만 그 양과 상관없이 원시 데이터만으로는 미래를 내다볼 수 없습니다. 진정한 혁신을 이루려면 이 풍부한 데이터를 의미 있는 정보로 변환해야 합니다.

인텔은 이러한 혁신을 가능하게 하고, 기업이 소프트웨어와 하드웨어를 통해 세상에서 가장 복잡한 문제를 해결할 수 있도록 도와줌으로써 실행 가능한 통찰력을 얻는 시간을 단축합니다. 고객들은 시장에 나와 있는 가장한 폭넓은 제품 포트폴리오 중 하나를 이용하여 다양하고 까다로운 워크로드를 실행하는데 필요한 성능과 전략 효율성에 액세스합니다. 인텔은 광범위한 파트너들의 에코 시스템들과 함께 기업이 모든 과제에 정면으로 대처할 수 있게 지원합니다.

이제 인텔의 고객들은 5세대 인텔® 제온® 프로세서의 힘을 활용하여 데이터를 지식으로 전환하고 혁신을 추진할 수 있습니다. 이 새로운 프로세서는 모든 워크로드에서 와트당 뛰어난 성능과 더불어 AI, 데이터베이스, 네트워킹, 스토리지 및 고성능 컴퓨팅(HPC)을 위한 뛰어난 성능 및 낮은 총소유비용(TCO)을 제공합니다. 이전 세대와 동일한 전원 엔벨로프에서 더 많은 컴퓨팅, 더 큰 공유 라스트 레벨 캐시 및 더 빠른 메모리를 제공합니다. 또한 이전 세대의 4세대 인텔® 제온® 프로세서와 소프트웨어 및 플랫폼 호환되므로 AI 및 기타 워크로드를 위한 새로운 시스템을 구축할 때 테스트 및 검증을 최소화할 수 있습니다.

AI(인공지능)을 위해 설계된 프로세서

AI는 우리의 일하는 방식을 바꾸고 있습니다. 기업들은 언어, 텍스트 이미지 및 코드 생성을 위한 최첨단 인공지능 모델을 사용하여 비용과 시간을 절약하고 새로운 기회를 창출합니다. 모든 코어에 인공지능 가속이 적용된 5세대 인텔® 제온® 프로세서는 인텔® 어드밴스드 매트릭스 확장과 같은 인공지능 가속기가 내장되어 있습니다. (인텔® AMX), 더 빠른 메모리, 더 큰 라스트 레벨 캐시는 딥 러닝(DL) 추론 및 훈련을 위해 향상된 성능으로 까다로운 AI 워크로드를 처리할 준비가 되어 있습니다.

PyTorch에서 3세대 인텔 제온 프로세서에 비해 5세대 인텔® 제온® 프로세서로 추론 및 훈련 시 최대 14배 더 높은 성능을 얻을 수 있습니다. 또한 인텔® AMX가 탑재된 5세대 인텔® 제온® 프로세서는 4세대 AMD EPYC 프로세서에 비해 2.5배 더 높은 와트당 성능을 실현합니다.

최대 200억 파라미터를 보유한 빠른 거대 언어 모델(LLMs)과 함께, 5세대 인텔® 제온® 프로세서는 100ms 이하의 토큰 지연시간으로 실시간 사용자 경험을 제공하기 위한 서비스 수준 협약(SLAs)을 충족시킵니다.

언제 어디서나 AI를 활용할 수 있는 소프트웨어 도구 및 에코 시스템

가장 큰 생태계와 공급 가용성, 친숙한 도구 세트, 클라우드, 데이터 센터 및 에지 전반에 걸쳐 확장할 수 있는 기능을 제공하는 인텔® 제온® 프로세서로 AI를 빠르게 배포합니다. 인텔® 제온® 프로세서에 대한 최적화는 TensorFlow 및 PyTorch를 포함한 인기 있는 AI 프레임워크의 주류 배포판에 이미 통합되어 있습니다. 수십 개의 미리 훈련되고 최적화된 인텔 AI 모델을 사용하고 쉽게 사용자 지정할 수 있습니다. 개발자들은 이러한 리소스를 사용하여 다양한 하드웨어 환경에서 코드를 쉽게 마이그레이션할 수 있습니다.

5세대 인텔® 제온® 프로세서는 다음과 같은 다양한 AI 사용 사례에 탁월합니다:

- **생성형 AI(Generative AI):** LLM 및 텍스트에서 이미지를 생성하는 것 처럼 생성형 AI 모델을 CPU 기반에서 실행합니다.
- **추천 시스템:** 사용자 경험을 늦추지 않는 빠르고 개인화된 제품이나 콘텐츠 추천을 제공합니다.
- **자연어 처리(NLP):** NLP 추론의 성능 도약으로 보다 반응성이 뛰어난 스마트 어시스턴트, 챗봇 및 예측 텍스트를 가능하게 합니다.
- **기계 학습(ML):** 고전적인 ML을 HPC 및 AI 애플리케이션에 적용할 때 더 빠르게 결과를 얻을 수 있습니다.
- **이미지 분류:** 제조 라인에서 결함을 감지하고 소매 애플리케이션에서 제품을 식별하는 등의 방법으로 생산성과 효율성을 향상 시킵니다.

기타 활용 사례는 다음과 같습니다:

일반계산

5세대 인텔® 제온® 프로세서를 사용하여 만든 저지연 시스템으로 성능을 높이고 투자 수익(ROI)을 높입니다. 가속기는 워크로드당 필요한 코어 수를 줄여서 CPU 코어로부터 태스크 오프로딩을 할 수 있습니다. 이를 통해 각 서버에서 용량을 확장하거나 더 많은 애플리케이션을 실행할 수 있습니다.

평균 1.84배의 성능 향상
5세대 제온® 프로세서 탑재 시스템과 3세대 제온® 프로세서 탑재 시스템 사용을 비교한 경우

AI

인텔® AI 엔진이 내장된 5세대 인텔® 제온® 프로세서를 사용하여 on-premises 또는 클라우드 인스턴스에서 다른 CPU와 비교할 수 없는 AI 성능을 달성하세요. DL 훈련 및 추론을 크게 향상시키는 내장형 가속기 인텔® AMX를 사용하세요. 프로세서 최적화는 TensorFlow 및 PyTorch를 포함하여 많이 사용되는 AI 프레임워크에 통합되어 있습니다.

최대 14배 더 높은 성능 제공
5세대 제온® 프로세서 탑재 시스템과 3세대 인텔® 제온® 프로세서 탑재 시스템에서 추론 및 훈련을 위한 PyTorch를 실행하여 비교한 경우

HPC

제조 시뮬레이션에서 지구 시스템 모델링에 이르기까지 HPC 애플리케이션의 성능을 향상 시킵니다. 더 큰 라스트 레벨 캐시와 더 빠른 메모리를 갖춘 5세대 인텔® 제온® 프로세서는 높은 수준의 정밀도를 제공하는 동시에 워크로드 속도를 높입니다. 초광대역 512비트 벡터 연산 기능을 갖춘 내장 가속기인 인텔® Advanced Vector Extension 512(인텔® AVX-512)는 특히 HPC 세그먼트에서 까다로운 연산 작업에 적합합니다.

평균 최대 2.1배 향상
5세대 제온® 프로세서 탑재 HPC 시스템 성능과 3세대 인텔® 제온® 프로세서 탑재 HPC 시스템 성능과 비교한 경우

웹 및 마이크로서비스

고객들에게 온라인 고객의 참여를 유도하는 반응형 웹 경험을 제공합니다. 인텔® QuickAssist 기술(인텔® QAT) 및 인텔® Dynamic Load Balancer(인텔® DLB)와 같은 인텔® 가속기 엔진을 활용하여 CPU 효율성을 높입니다. 이러한 기본 제공 가속기는 마이크로서비스 네트워킹 및 스토리지 애플리케이션 성능을 향상시키는 데 도움이 됩니다.

처리량 최대 1.3배 향상
5세대 제온® 프로세서와 4세대 인텔® 제온® 프로세서를 탑재한 시스템에서 100ms SLA10 소셜 네트워크 마이크로서비스 워크로드 (DeathStarBench의 Read Home Timeline)을 비교한 경우

데이터베이스 및 분석

5세대 인텔® 제온® 프로세서와 인텔® 분석 엔진으로 구동되는 빠른 데이터베이스와 분석을 통해 통찰력과 생산성을 가속화합니다. 인텔® In-Memory Analysis Accelerator (인텔® IAA)를 활성화하여 쿼리 처리량을 늘릴 수 있습니다. 데이터 복사 및 변환 작업을 수행하여 CPU 주기를 늘리는 인텔® Data Streaming Accelerator(인텔® DSA)의 기능을 활용할 수 있습니다. 더 빠른 메모리와 더 큰 최종 레벨 캐시와 함께 이 두 가속기는 인메모리 데이터 베이스, 빅 데이터 분석 및 데이터 웨어하우스의 성능을 향상시킵니다.

최대 3.7배 향상
내장된 인텔® IAA를 이용한 5세대 제온® 프로세서 탑재 시스템과 Zstd를 이용한 3세대 인텔® 제온® 프로세서 탑재 시스템에서 RockDB 성능을 비교한 경우

네트워킹

데이터 이동, 암호화 및 압축 속도를 높여 네트워크에 용량을 추가합니다. 인텔® QAT 및 인텔® DLB와 같은 인텔® 네트워크 엔진이 탑재된 5세대 인텔® 제온® 프로세서를 사용하면 효율적인 네트워크 데이터 배치와 가속화된 네트워크 암호화를 통해 초당 더 많은 웹 연결을 안전하게 보호할 수 있습니다.

최대 1.7배 더 높음
인텔® QAT를 내장한 5세대 제온® 프로세서 탑재 시스템과 3세대 인텔® 제온® 프로세서 탑재 시스템에서 NGINX TLS 핸드셰이크 성능을 비교한 경우

비용을 절감하면서 성능 및 효율성 향상

에너지 효율성은 특히 지속 가능성 이니셔티브를 발전시키기 위해 노력하는 모든 종류의 조직에서 점점 더 큰 관심을 받고 있습니다. 따라서 기술 인프라에서 전력 소비를 줄이는 것이 최우선 과제입니다.

5세대 인텔® 제온® 프로세서는 이전 세대에 비해 34% 향상된 성능/전력을 제공합니다. 이점을 얻을 수 있는 워크로드에 대해 플랫폼 BIOS에서 최적화된 전원 모드를 활성화하면 전력 효율성과 절감 효과를 더욱 확대할 수 있습니다.

인텔® 가속기 엔진을 탑재한 5세대 인텔® 제온® 프로세서는 대상 워크로드에서 와트당 최대 10배 더 높은 성능을 제공합니다. 시중의 그 어떤 CPU보다 더 많은 내장형 가속기를 갖춘 5세대 인텔® 제온® 프로세서는 AI, 데이터베이스, 네트워킹 및 HPC 워크로드에 탁월한 성능과 TCO를 제공합니다.

5세대 인텔® 제온® 프로세서를 서버 교체 전략에 통합하여 전력 소비를 줄이는 동시에 데이터 센터를 효율적으로 현대화 할 수 있습니다. 최신 세대 프로세서로 전환하면 성능 향상과 비용 절감을 실현하여 조직의 현재 및 미래 요구 사항을 지원할 수 있습니다. 5세대 인텔® 제온® 프로세서는 더 많은 코어와 코어당 더 많은 성능을 제공함으로써 필요한 서버 수를 줄여 성능 요구 사항을 충족하는 동시에 전력 소비와 운영 비용을 줄입니다.

기밀 컴퓨팅(Confidential Computing)

신뢰할 수 있는 실행 환경(TEE)을 갖춘 기밀 컴퓨팅은 데이터와 AI 모델을 보호하는 데 도움이 됩니다. 5세대 인텔® 제온® 프로세서를 사용하면 현재 시장에 나와 있는 데이터 센터에서 가장 많이 조사되고 업데이트된 기밀 컴퓨팅 옵션을 선택할 수 있습니다. 인텔® Software Guard Extension (인텔®SGX)은 애플리케이션 격리 기능을 제공하며 정지 상태, 이동 중 및 사용 중 데이터 보호 기능을 강화하도록 설계되었습니다. 인텔® Trust Domain Extension(인텔®TDX)은 가상 머신(VM) 수준의 격리 기능과 기밀성을 제공합니다. 고객은 인텔®Trust Authority attestation Services 사용하여 여러 데이터 센터, 클라우드 공급자 및 에지 위치에서 인텔 기밀 컴퓨팅 환경의 무결성을 독립적으로 확인할 수 있습니다.

기술개요

4세대 인텔® 제온® 프로세서와 공유 아키텍처 플랫폼을 기반으로 구축된 5세대 인텔® 제온® 프로세서는 와트당 성능과 성능, TCO 향상 및 실리콘 기반 보안 기능을 향상시킵니다. 동일한 열 설계 전력(TDP)에서 이전 세대에 비해 전체 성능이 21% 향상되어 향상된 ROI(투자 수익률)를 달성할 수 있습니다. 5세대 인텔® 제온® 프로세서는 이전 세대에 비해 더 빠른 메모리와 더 큰 마지막 레벨 캐시를 통해 메모리 바인딩 및 지연 시간에 민감한 워크로드의 성능을 향상시킵니다.

이전 세대보다 더 많은 코어와 CPU 캐시를 갖춘 광범위한 SKU 중에서 선택하면 1소켓 또는 2소켓 서버 설계에서 전체적인 성능을 향상시킬 수 있습니다. 5세대 인텔® 제온® 프로세서는 워크로드 요구사항의 변화에 따라 확장 및 조정할 수 있는 메모리 및 입출력(I/O) 서브시스템의 향상된 기능을 제공합니다. 또한 이러한 프로세서를 통해 다음과 같은 이점을 얻을 수 있습니다:

- 4세대 인텔® 제온® 프로세서보다 16% 향상된 최대 5,600메가 전송(MT/s)(1DPC)을 지원하여 DDR5 메모리를 더욱 빠르게 사용할 수 있습니다.
- 일부 SKU의 경우 이전 세대보다 최대 3배 증가한 최대 320MB의 마지막 레벨 캐시를 모든 코어에서 공유하여 공유 마지막 레벨 캐시를 늘립니다.
- 인텔® Ultra Path Interconnect(Intel® UPI) 2.0을 통해 socket 간 대역폭을 확장하고 초당 최대 20기가 전송(GT/s)을 제공하여 이전세대보다 25% 향상되었습니다.
- 인텔® TDX를 사용하여 기밀 컴퓨팅을 VM 수준의 워크로드 격리로 확장하는 동시에 Intel SGX를 사용하여 애플리케이션 수준의 워크로드 격리를 지속적으로 구현할 수 있습니다.
- CXL(Compute Express Link) Type 1 또는 2 장치를 사용하여 지연 시간이 짧고 CPU와의 일관된 메모리 통신을 촉진합니다. Type 1 장치는 스마트 네트워크 인터페이스 카드(NIC) 및 가속기이고, Type 2 장치는 캐시가 있는 가속기입니다.
- CXL Type 3 메모리 장치로 메모리 용량을 확장하여 대상 하이퍼스케일러 지원을 통해 시스템 내 메모리 대역폭을 늘릴 수 있습니다.
- NVMe(NVMe) 솔리드 스테이트 드라이브(SSD)를 위해 특별히 설계된 기업용 RAID 솔루션인 CPU(Intel® VROC)에 인텔® Virtual RAID를 탑재한 기존의 하드웨어 RAID 호스트 버스 어댑터(HBA) 카드가 필요하지 않습니다.
- 인텔® Speed Select Technology(인텔® SST)를 통해 특정 워크로드 요구를 충족하도록 CPU를 구성합니다.
- 고급 RAS(신뢰성, 가용성 및 서비스 가능성) 기능을 통해 시스템 가동 시간을 늘리고 계획되지 않은 다운타임의 지속 시간을 단축하며 데이터 무결성을 유지함으로써 모든 플랫폼 아키텍처 요소에서 가용성을 향상시키고 데이터 신뢰성을 유지합니다.
- 인텔® Ethernet 800 Series Network Adapters를 사용하여 우선순위가 높은 애플리케이션, 패킷 처리 및 지연 시간에 민감한 워크로드를 가속화할 수 있습니다.
- 원활한 펌웨어 업데이트, 인텔® Platform Monitoring Technology 및 인텔® Resource Director Technology(인텔® RDT)와 같은 기능을 통해 운영 효율성 문제를 해결합니다.
- 가속기 또는 하드웨어 강화 기능을 업그레이드할 수 있는 대부분의 5세대 인텔® 제온® 프로세서에서 사용할 수 있는 서비스 인텔® On Demand를 사용하여 필요할 때 보다 유연하게 가속화 또는 보안을 추가할 수 있습니다.

5세대 인텔® 제온® 프로세서에 탑재된 메모리 및 I/O 기술:

DDR5 지원

높은 메모리 대역폭으로 데이터 병목 현상을 극복하여 컴퓨팅 성능을 향상시킵니다. DDR5는 DDR4 보다 최대 1.7배 향상된 대역폭을 제공하여 성능, 용량, 전력 효율 및 비용을 개선할 수 있습니다. 5세대 인텔® 제온® 프로세서는 DDR5를 통해 최대 5,600MT/s(1DPC) 또는 4,400MT/s(2DPC)의 메모리 속도를 향상시켜 메모리 바인딩 및 지연 시간에 민감한 워크로드의 성능을 향상시킵니다.

PCIe 5.0 지원

PCIe 4.0에 비해 I/O 대역폭을 두 배로 늘려 CPU와 연결된 장치 간에 가능한 최고의 처리량을 제공합니다. 5세대 인텔® 제온® 프로세서는 80개의 PCIe 5.0 레인을 갖추고 있어 빠른 네트워킹, 고대역폭 가속기 및 고성능 스토리지 장치에 이상적입니다. PCIe 5.0은 PCIe 4.0보다 I/O 대역폭을 두 배로 늘렸고, 하위 호환성을 유지하며 CXL을 위한 기본 솔루션을 제공합니다. 증가하는 PCIe 에코시스템은 고객이 하드웨어를 사용자 지정하고 성능을 확장할 수 있는 유연성을 제공합니다.

CXL 지원

CXL 1.1을 사용하면 데이터 센터의 컴퓨팅 지연 시간을 줄이고 차세대 워크로드에 대한 TCO를 절감할 수 있습니다. CXL은 표준 PCIe 물리 계층 전체에서 실행되는 대체 프로토콜로 동일한 링크에서 표준 PCIe 디바이스와 CXL 디바이스를 모두 지원할 수 있습니다. CXL은 CPU와 가속기 사이에 일관성 있는 통합 메모리 공간을 생성하는 데 중요한 기능을 제공하며, 향후 수년간 데이터 센터 서버 아키텍처가 구축될 방식에 혁신을 일으킬 것입니다.

5세대 인텔® 제온® 프로세서 개요

인텔® 제온® 플래티넘 8500 프로세서는 on-premise 및 멀티 클라우드 모두에서 보안이 지원되고 민첩한 데이터 센터의 기반입니다. AI, 고급 데이터 분석, 고밀도 인프라 및 멀티 클라우드 워크로드용으로 설계되었습니다. 이 프로세서는 최대 2개 프로세서 탑재가 가능하여서 높은 수준의 성능, 향상된 플랫폼 기능, 업계 최고의 워크로드 가속화를 제공합니다. 향상된 하드웨어 기반 보안과 뛰어난 멀티 소켓 처리 성능을 제공합니다. 신뢰할 수 있는 하드웨어로 강화된 데이터 서비스 제공과 최신 I/O 및 연결 기술을 갖춘 이러한 프로세서는 I/O, 메모리, 스토리지 및 네트워크 기능을 향상시켜 점점 더 데이터 중심이 되는 세상에서 실행 가능한 통찰력을 활용합니다. 이 프로세서는 다음과 같은 특징을 가지고 있습니다.

- 프로세서당 최대 64개 코어 지원
- 최대 5,600 MT/s(1 DPC)의 프로세서당 8개의 메모리 채널
- DL 추론 및 훈련 성능의 큰 도약을 위한 인텔® AMX를 통한 AI 가속화

최대 2개 프로세서 탑재가 가능한 인텔 제온 골드 6500 및 인텔® 제온® 골드 5500 프로세서는 까다로운 메인스트림 데이터 센터, 멀티 클라우드 컴퓨팅, 네트워크 및 스토리지 워크로드에 최적화되어 있습니다. 더 빠른 메모리 속도와 향상된 메모리 용량을 지원하는 이 프로세서들은 이전 세대에 비해 향상된 성능과 우수한 메모리 기능을 제공합니다. 또한 하드웨어 향상된 보안 및 워크로드 가속 기능을 제공합니다.

인텔® 제온® 실버 4500 프로세서는 필수적인 성능, 향상된 메모리 속도 및 향상된 전력 효율을 제공합니다. 이들은 초급 단계 데이터 센터 컴퓨팅, 네트워크 및 스토리지에 필요한 하드웨어 향상된 성능을 제공합니다. 또한 에지 강화(EE) SKU는 에지 컴퓨팅 사용에 적합한 전력 효율성을 제공합니다.



인텔® 제온® 플래티넘	인텔® 제온® 골드	인텔® 제온® 실버
최대 2개 프로세서 탑재 지원	최대 2개 프로세서 탑재 지원	최대 2개 프로세서 탑재 지원
20 GT/s Intel UPI 포트 3개	20 GT/s Intel UPI 포트 3개	16 GT/s Intel UPI 포트 2개
PCIe 5.0/CXL 1.1 80개 레인	PCIe 5.0/CXL 1.1 80개 레인	PCIe 5.0/CXL 1.1 80개 레인
최대 5,600MT/s 지원 DDR5(채널당 1DIMM) 또는 최대 4,400MT/s 지원 DDR5(채널당 2DIMM)	최대 5,200MT/s 지원 DDR5(채널당 1DIMM) 또는 최대 4,400MT/s 지원 DDR5(채널당 2DIMM)	최대 4,400MT/s 지원 DDR5(채널당 1DIMM) 또는 최대 4,400MT/s 지원 DDR5(채널당 2DIMM)
인텔® AVX-512 (512비트 FMA 2개)	인텔® AVX-512 (512비트 FMA 2개)	인텔 AVX-512 (512비트 FMA 2개)
인텔® HT기술 및 인텔® 터보 부스트 기술	인텔® HT기술 및 인텔® 터보 부스트 기술	인텔 HT기술 및 인텔 터보 부스트 기술
인텔® 딥러닝 부스티 및 인텔® AMX	인텔® 딥러닝 부스티 및 인텔® AMX	인텔 딥러닝 부스티 및 인텔 AMX
인텔® SST	인텔® SST	표준 RAS 기능
고급 RAS 기능	고급 RAS 기능	인텔® SGX 최대 64GB 엔클레이브 크기
인텔® SGX 최대 512GB 엔클레이브 크기	인텔® SGX 최대 128GB 엔클레이브 크기	인텔® TDX
인텔® TDX	인텔® TDX	인텔® QAT, 인텔® DLB, 인텔® DSA 및 인텔® IAA로 워크로드 가속화
인텔® QAT, 인텔® DLB, 인텔® DSA 및 인텔® IAA로 워크로드 가속화	인텔® QAT, 인텔® DLB, 인텔® DSA 및 인텔® IAA로 워크로드 가속화	