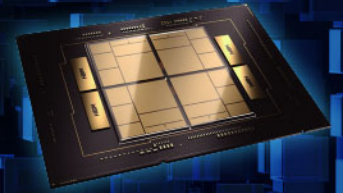


제품 소개

컴퓨팅 가속화
시스템 및 그래픽



인텔® 제온® CPU MAX 시리즈



지난 10년간 AI 통합과 마찬가지로 피크 컴퓨팅(Peak Compute)이 크게 증가했지만 코어에 데이터를 공급하는 데 비효율적인 작업으로 인해 워크로드 성능이 보조를 맞추지 못했습니다. 인텔® 제온® CPU Max 시리즈는 인텔® 제온® 플랫폼을 강화하고 메모리에 한정된 HPC 및 AI 워크로드 제약을 해제하고 가속화하도록 설계된 고대역폭 메모리를 갖춘 유일한 x86 기반 프로세서입니다.



날씨 예보부터 인간 게놈 지도 작성 및 세계에서 가장 치명적인 질병 치료에 도움이 되는 것, 더 에너지 효율적인 재료 설계에 이르기까지 고성능 컴퓨팅(HPC)은 우리 삶의 모든 부분에 영향을 미칩니다. HPC와 AI의 발전은 경쟁력을 높이고 과학 컴퓨팅 수요를 새로운 수준으로 끌어 올리지만, 모든 상황에 맞는 솔루션은 없습니다. 기존 HPC 소프트웨어에는 놀라운 다양성이 있으며, 수직 및 특성별로 일반적인 워크로드를 살펴보면 일부는 메모리 제약이 있습니다. 또다른 이슈는 CPU 바인딩(컴퓨팅 바인딩)입니다. 일부는 제어 흐름이 많은 작은 커널을 지고 있습니다. 다른 것들은 큰 데이터 병렬 커널을 가지고 있습니다. 대부분은 매우 큰 데이터 세트를 포함합니다.

인텔® 제온® CPU Max 시리즈는 인텔® 제온® 스케일러블 프로세서에 고대역폭 메모리(HBM)를 추가로 탑재하여 모델링, 인공지능, 딥러닝, 고성능 컴퓨팅(HPC) 및 데이터 분석과 같은 데이터 집약적인 워크로드에서 성능을 높이고 검색 속도를 높이도록 설계되었습니다.

향상된 대역폭으로 성능 극대화

인텔® 제온® CPU Max 시리즈는 새로운 마이크로아키텍처를 내장하고 있으며, 코어 수 증가, 고급 I/O 및 메모리 하위 시스템, 삶을 변화시킬 수준의 검색을 신속하게 제공하는 내장 가속기 등 다양한 플랫폼 향상 기능을 지원합니다. 인텔® Max 시리즈 CPU는 다음의 주요 특징을 가지고 있습니다:

- 4개의 타일로 구성되고 인텔의 EMIB(다이 인터커넥트 브리지) 기술을 사용하여 연결된 **최대 56개의 성능 코어** 및 350W TDP 지원
- **64 GB**의 고대역폭 인 패키지 메모리와 PCI Express 5.0 및 CXL 1.1 I/O가 포함. 제온 제온 MAX CPU는 대부분의 HPC 워크로드에 충분히 대응 할 수 있는 코어당 메모리 (HBM) 용량을 제공
- 다른 CPU에 비해 HBM을 사용하여 자연어 처리 (NLP)용 Numenta AI 기술에서 **최대 20배**의 성능 향상.²

5배 더 나은 성능

경쟁사 제품 및 동사 이전 세대 제품 대비
메모리 대역폭 비교 기준¹

듀얼 인텔 제온 Max CPU 대 듀얼 AMD EPYC 7773X
듀얼 3세대 인텔 제온 8380

과학적 혁신의 가속화

더 빠른 연구 결과 도출과 효과적인 연구를 가능하게 합니다. 인텔® 제온® Max 시리즈 및 4세대 인텔® 제온® 스케일러블 프로세서를 사용하면, 귀하는 가장 어려운 워크로드에 요구되는 성능과 전력 효율 및 현존하는 어느 CPU보다도 많은 내장 가속기를 얻게 됩니다. HPC 및 AI 워크로드를 위한 다음의 핵심 내장 가속기를 활용하여 더 효율적으로 CPU를 사용하고, 전력 소모를 더 낮추고, 더 높은 ROI를 확보하십시오.

- **인텔 고급 매트릭스 익스텐션(인텔 AMX)** - 인텔® AMX를 통해 딥 러닝 추론 및 교육을 크게 가속화 할 수 있습니다. 이는 AI 성능을 향상시키고 INT32 누적 작동으로 INT8용 AVX-512에 비해 8배의 피크 처리량을 제공합니다.³
- **인텔 데이터 스트리밍 가속기(인텔 DSA)** - 스트리밍 데이터 이동을 개선하여 데이터 집약적인 워크로드를 위한 고성능을 제공합니다. 인텔® DSA를 사용하면 NVMe over TCP를 사용할 때 대기 시간을 최대 45% 단축하고 초당 스토리지 I/O(IOPS)를 최대 79% 향상할 수 있습니다.⁴
- **인텔 Advanced Vector Extensions 512(인텔 AVX-512)** - 벡터화를 통해 성능을 가속화하여 과학 시뮬레이션, AI/딥 러닝, 3D 모델링 및 분석 및 기타 집중적인 워크로드를 위한 대규모 데이터 세트의 신속한 계산을 지원합니다. 인텔® AVX-512는 가장 까다로운 계산 작업의 성능을 가속화하는 최신 x86 벡터 명령 집합입니다.
- **I/O 및 메모리 하위 시스템이 다음과 같이 향상되었습니다:**
 - **DDR5** - 더 높은 메모리 대역폭으로 데이터 병목 현상을 극복하여 컴퓨팅 성능을 향상시킵니다. DDR5는 DDR4에 비해 최대 1.5배의 대역폭 향상을 제공합니다.⁴
 - **PCI Express Gen 5 (PCIe 5.0)** - CPU와 장치 간에 최고의 처리량을 제공할 수 있는 기회를 통해 빠른 I/O 속도를 실현합니다. 4세대 인텔® 제온® 스케일러블 및 인텔® 제온® Max 시리즈 프로세서는 PCIe 4.0의 I/O 대역폭의 두배인 최대 80레인을 지원하는 PCIe 5.0을 지원합니다.⁴
 - **Compute Express Link (CXL) 1.1** - 높은 패브릭 대역폭 및 연결된 가속기 효율성을 지원합니다.
- **인텔® 제온® 플랫폼의 손쉬운 통합** - 대부분의 구축에 있어 코드 변경 없이 동일한 소켓 구성 상태에서 Max 시리즈 CPU를 4세대 인텔® 제온® 스케일러블 플랫폼에 추가할 수 있습니다.

모든 HPC 와 AI 워크로드를 위한 유연성

인텔 Max 시리즈 CPU는 워크로드 특성에 따라 다양한 메모리 모드 또는 구성에서 실행할 수 있는 유연성을 제공합니다:

- **HBM 전용 모드** - 64GB 용량에 맞는 워크로드와 코어당 1-2GB 메모리로 확장할 수 있는 기능을 지원하는 HBM 전용 모드는 코드 변경 및 DDR 없이 시스템 부팅을 지원합니다.
- **HBM 플랫폼 모드** - 대용량 메모리가 필요한 애플리케이션에 유연성을 제공하는 HBM 플랫폼 모드는 HBM 및 DRAM과 함께 플랫폼 메모리 영역을 제공하며 코어당 2GB 이상의 메모리가 필요한 워크로드에 적용할 수 있습니다. 코드 변경이 필요할 수 있습니다.
- **HBM 캐시 모드** - 64GB 이상의 용량 또는 코어당 2GB 이상의 메모리가 필요한 워크로드에 대해 성능을 향상시키도록 설계되었습니다. 코드를 변경할 필요가 없으며 HBM은 DDR을 캐시합니다. 다른 CPU에 비해 HBM을 사용하여 자연어 처리 (NLP)용 Numenta AI 기술에서 최대 20배의 성능 향상됩니다.

인텔® 제온® Max CPU 시리즈	
코어 수	32-56
HBM2E 메모리	64 GB
피크 HBM 전송 속도	3200 MT/s
피크 DDR5 전송 속도	4800 MT/s (1DPC) 4400 MT/s (2DPC)
가속기	AMX, 4 DSA Devices
AI/ML 명령어	INT8 and BFLOAT16

다양한 아키텍처 전반에 걸쳐 HPC 및 AI 워크로드 가속

모든 인텔 Max 시리즈 제품군은 oneAPI로 통합되어 생산성과 성능을 향상시키는 공통의 개방형 표준 기반 프로그래밍 모델을 제공합니다. 개발자는 인텔 oneAPI 베이스 툴킷 및 인텔 oneAPI HPC + 도메인별 툴킷을 사용하여 여러 유형의 아키텍처에서 일반 컴퓨팅, HPC 및 AI 애플리케이션을 보다 쉽게 구축, 분석, 최적화 및 확장할 수 있습니다.



이러한 리소스에는 벡터화, 멀티스레딩, 멀티노드 병렬화 및 메모리 최적화에 대한 최첨단 기술이 포함되어 있어 HPC에 적합한 고성능 멀티 아키텍처 소프트웨어를 쉽게 구축할 수 있습니다. 최신 HPC 소프트웨어 개발 도구를 보려면 [4세대 인텔® 제온® 및 인텔® 제온® CPU Max 시리즈 프로세서용 소프트웨어 및 HPC 소프트웨어 및 도구 리소스 페이지](#)를 방문하십시오.

인텔® 제온® CPU Max 시리즈 프로세서

SKU	Core	Base (GHz)	All-Core Turbo (Ghz)	Max Turbo (GHz)	캐시 (MB)	TDP (Watts)	최대 지원 프로세서 수	DDR5 메모리 속도	Intel® SGX Enclave Capacity (Per Processor)
9480	56	1.9	2.6	3.5	112.5	350	2S	4800	512 GB
9470	52	2	2.7	3.5	105	350	2S	4800	512 GB
9468	48	2.1	2.6	3.5	105	350	2S	4800	512 GB
9460	40	2.2	2.7	3.5	97.5	350	2S	4800	128 GB
9462	32	2.7	3.1	3.5	75	350	2S	4800	128 GB

최신 정보를 보려면 [Intel.com/MaxSeriesCPU](https://www.intel.com/MaxSeriesCPU) 을 방문하십시오



¹ [intel.com/performanceindex](https://www.intel.com/performanceindex) 을 방문하십시오 (이벤트: 워크로드 및 구성을 위한 슈퍼컴퓨팅 22). 결과는 다를 수 있습니다.

² Numenta BERT-Large

- AMD Milan: 2022년 11월 28일 기준으로 Numenta에 의해 테스트됨. AWS m6a.48xlarge, 768GB DDR4-3200, Ubuntu 20.04 커널 5.15, OpenVINO 2022.3, BERT-Large, 시퀀스 길이 512, 배치 크기 1
- Intel® Xeon® 8480+: Numenta에 의해 11/28/208.1 노드, 2x Intel Xeon 8480+, 512GB DDR5-4800, Ubuntu 22.04 커널 5.17, OpenVINO 2022.3, Numenta 최적화 BERT-Large, 시퀀스 길이 512, 배치 크기 1로 평가됨
- Intel® Xeon® Max 9468: Numenta에 의해 11/30/165.1 노드, Intel Xeon Max 9468, 128GB HBM2e 3200 MT/s, Ubuntu 22.04 커널 5.15, OpenVINO 2022.3, Numenta에 최적화된 BERT-Large, 시퀀스 길이 512, 배치 크기 1 기준으로 테스트됨

³ 소스: AMX Performance-core 설명 및 자세한 내용은 Intel을 참조하십시오. "성과지표: 건축의 날 2021." [edc.intel.com/content/www/tw/zh/products/performance/benchmarks/architecture-day-2021/](https://www.intel.com/content/www/tw/zh/products/performance/benchmarks/architecture-day-2021/)

⁴ 자세한 내용은 [intel.com/processorclaims](https://www.intel.com/processorclaims) 에서 확인하십시오. 성능은 용도, 구성 및 기타 요인에 따라 다릅니다. 결과는 다를 수 있습니다.

성능 결과는 구성에 표시된 날짜의 테스트를 기반으로 하며 공개적으로 사용 가능한 모든 업데이트를 반영하지 않을 수 있습니다. 어떤 제품이나 구성 요소도 절대적으로 안전할 수 없습니다. 인텔은 타사 데이터를 제어하거나 감사하지 않습니다. 정확도를 평가하려면 다른 출처를 참조해야 합니다. 인텔 기술은 활성화된 하드웨어, 소프트웨어 또는 서비스 활성화를 요구할 수 있습니다.

© 인텔 코퍼레이션. 인텔, 인텔 로고 및 기타 인텔 상표는 인텔 Corporation 또는 자회사의 상표입니다. 기타 다른 이름 및 브랜드는 다른 회사의 재산으로 청구될 수 있습니다.