

비즈니스 요약

4세대 인텔® 제온® 스케일러블 프로세서
클라우드 컴퓨팅



클라우드 성장의 밝은 전망



인텔®을 사용하면 성숙하고 민첩한 하이브리드, 멀티클라우드 또는 인텔리전트 에지 전략을 이행할 수 있습니다.

클라우드 마이그레이션이 전사에 걸쳐 계속 전개됨에 따라 클라우드 서비스 환경의 전망이 점점 더 복잡해지고 있습니다. 업계 베스트 프랙티스 기반의 동기화된 비즈니스 및 기술 전략은 클라우드 우선 운영을 통해 얻을 수 있는 민첩성, 경제성 및 기타 이점을 최적화할 수 있습니다. 조직은 효율을 개선하고 서비스 수준의 예측 가능성을 높이면서 더 빠른 혁신과 더 나은 근거로 뒷받침되는 비즈니스 의사 결정을 지원해야 합니다.

조직이 규모를 확장하고 비용을 절감하고 새로운 서비스를 제공하려고 함에 따라 기술이 비즈니스 가치를 제공하는 것이 그 어느 때보다도 중요해졌습니다. 기업은 복잡성을 가중시키는 새로운 애플리케이션에 맞춰 시스템을 바꾸는 대신 확장 가능한 플랫폼을 통해 현재와 미래의 매우 다양한 배포를 지원하는 데 필요한 성능을 얻을 수 있습니다.



인텔® 기술로 클라우드 가속화

4세대 인텔® 제온® 스케일러블 프로세서는 AI, 분석, 네트워킹, 스토리지 및 HPC를 포함한 가장 빠르게 성장하는 클라우드 워크로드 전반에 걸쳐 성능 및 전원 효율의 우위를 제공하기 위해 시장에 출시된 CPU 중 가장 많은 가속기를 내장하고 있습니다. 완전히 새로운 인텔® Advanced Matrix Extensions(인텔® AMX)를 탑재한 4세대 인텔® 제온® 스케일러블 프로세서는 AI 학습 및 추론 성능이 탁월합니다. 그 외의 완벽하게 통합된 가속기는 더 빠른 네트워킹을 위해 데이터 이동 및 압축 속도를 높이고, 보다 신속한 분석을 위해 쿼리 처리량을 높이고, 클라우드 플랫폼의 여러 코어에 걸쳐 동적으로 부하를 분산시키기 위해 스케줄링 및 대기열 관리를 오프로드합니다. 새로운 내장된 가속기 기능을 하이퍼스케일 환경 안에서 사용할 수 있도록 하기 위해, 인텔은 가장 일반적인 모든 클라우드 API, 라이브러리 및 OS 레벨 소프트웨어로 에코시스템을 지원합니다.

내장된 가속기 및 소프트웨어 최적화 기능을 통해, 이전 세대의 인텔® 제온® 스케일러블 프로세서는 실제 클라우드 워크로드에서 최고의 와트당 성능을 제공하는 것으로 나타났습니다.³ 그 결과 CPU를 더 효율적으로 사용하고 클라우드의 전력 소비를 줄이고 서비스 ROI를 높이면서 기업의 지속가능성 목표 달성도 지원할 수 있었습니다.

성능 입증

최대 더 높은 처리량(RPS)
2.5X

최대 더 낮은 P99 대기 시간
74%

최대 더 낮은 CPU 사용률
12%

4세대 인텔® 제온® Platinum 8480+ 프로세서와 인텔® QAT 장치 2개 사용 시, 가속기가 없는 솔루션 대비⁴





주요 클라우드 워크로드

기밀 컴퓨팅

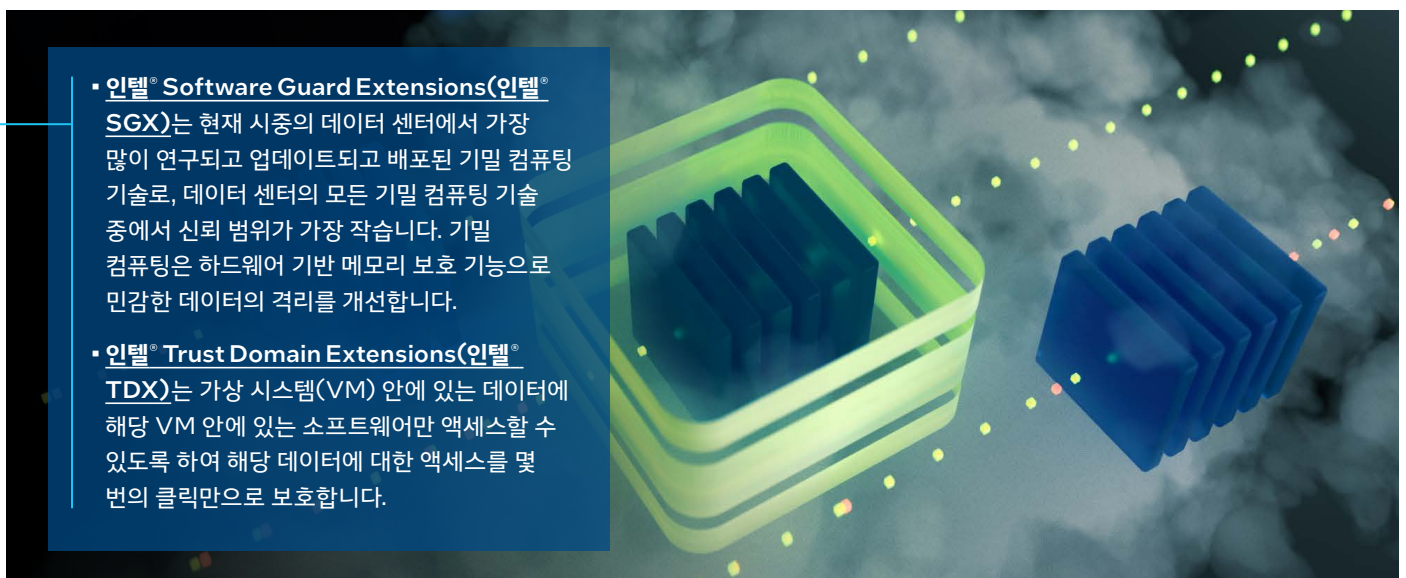
조직은 제로 트러스트 보안 전략의 일부로 철저한 보안, ID 및 컴플라이언스 관리에 초점을 맞추고 있습니다. 기밀 컴퓨팅은 하드웨어 기반 메모리 보호 기능으로 민감한 데이터의 격리를 개선합니다. 기밀 컴퓨팅 시장은 90%~95%의 CAGR로 성장하여 2026년에는 540억 달러에 도달할 것으로 예상됩니다.⁵

필요한 사항:

제 3자 간 데이터 공유는 개인 정보 보호, 규제 또는 지적 재산에 관한 고려 사항 때문에 어렵거나 불가능해지는 경우가 많습니다. 이런 장애물을 극복하면 비용, 확장성 및 민첩성에 관한 클라우드의 이점을 이용하여 거대한 분산 데이터 세트로 연구를 지원하거나 AI 모델을 학습시키는 등의 잠재적인 이점을 얻을 수 있습니다.

해답:

4세대 인텔® 제온® 스케일러블 프로세서를 사용하는 기밀 컴퓨팅은 하드웨어로 보호되는 메모리 엔클레이브 안에서 신뢰할 수 있는 코드만 데이터에 액세스할 수 있도록 하여 데이터 격리를 개선합니다. 이런 개인 정보 보호 기능을 사용하면 하이퍼스케일 플랫폼 및 기타 분산 네트워크에서 조절되는 워크로드를 중심으로 혁신이 가능합니다.



마이크로서비스

마이크로서비스 채택은 계속하여 개발 조직 안에서 변환에 기여하는 중요한 요인이 되고 있으며, 특히 하이퍼스케일 플랫폼과 인텔리전스 에지 환경의 경우 더욱 그렇습니다. 최근에 인텔에서 실시한 연구에서는 모든 새로운 클라우드 네이티브 애플리케이션과 SaaS(서비스형 소프트웨어) 솔루션 중 83% 이상이 마이크로서비스를 사용하는 것으로 나타났습니다.⁶

필요한 사항:

비즈니스는 앱 보안을 강화하고 대규모로 더 빠르고 효과적으로 일하여 멀티 클라우드 환경에서 원활하게 운영해야 합니다. 클라우드 하드웨어 및 소프트웨어는 처리량을 위주로 최적화하여 서비스 품질, 인프라 오버헤드 및 관측 가능성 같은 주요 마이크로서비스 문제를 지원해야 합니다.

성능 입증

최대
1.88X 더 높은 처리량

4세대 인텔® 제온® 스케일러블 프로세서 사용 시, 이전 세대 대비, DeathStarBench 기준 —
소셜 네트워크 마이크로서비스 워크로드(홍 타임라인 읽기), 100ms SLA⁷

해답:

4세대 인텔® 제온® 스케일러블 프로세서에 내장된 가속기는 클라우드 플랫폼을 통해, 그리고 그 안에서 데이터가 이동하는 속도를 높여 서비스 품질, 인프라 오버헤드 및 데이터/프로세스 관측 가능성에 대한 운영 목표를 지킵니다. 그리고 클라우드 비즈니스 서비스와 수신되는 수백만에 이르는 운영 요청의 부하를 분산시키는 데 도움이 되는 하드웨어 지원 대기열과 고속 분산 네트워크 통신의 빠른 시작 및 종료를 보장하는 데에도 도움이 됩니다.

- **인텔® Data Streaming Accelerator(Intel® DSA)**는 스트리밍 데이터 이동 및 변환 작업을 개선하여 스토리지, 네트워킹 및 데이터 집약적인 워크로드를 위한 고성능을 제공합니다. 데이터 센터 규모의 배포에서 오버헤드를 초래하는 가장 일반적인 데이터 이동 작업의 부하를 덜도록 설계된 인텔® DSA는 CPU, 메모리 및 캐시와 모든 연결된 메모리, 스토리지 및 네트워크 장치 간 데이터 이동 속도를 높이는 데 도움이 됩니다.
- **인텔® Dynamic Load Balancer(인텔® DLB)**는 멀티코어 인텔® 제온® 스케일러블 프로세서의 네트워크 데이터 처리와 관련된 시스템 성능을 개선합니다. 인텔® DLB를 사용하면 네트워크 처리를 여러 CPU 코어/스레드에 효율적으로 배분할 수 있고, 처리할 네트워크 데이터가 여러 CPU 코어에 시스템 부하 변동에 따라 동적으로 분산됩니다. 또한 인텔® DLB는 CPU 코어에서 동시에 처리되는 네트워킹 데이터 패킷의 순서를 복원합니다.

성능 입증

최대 더 낮은 대기
96% 시간

같은 처리량(RPS) 기준, 4세대 인텔® 제온® 스케일러블 프로세서와 인텔® DLB 사용 시, 6 코어(12 스레드)에서 작동하는 Istio 수신 게이트웨이용 소프트웨어 대비⁸

성능 입증

최대 **2.33X** 더 높은 싱글코어 처리량 [MTU1500, 1500B] 및 최대 **2.63X** 더 높은 싱글코어 처리량 [MTU 9000, 2400B]

4세대 인텔® 제온® 스케일러블 프로세서와 인텔® DSA 메모리 복사, 소프트웨어 메모리 복사 대비⁹

AI와 고급 데이터 분석

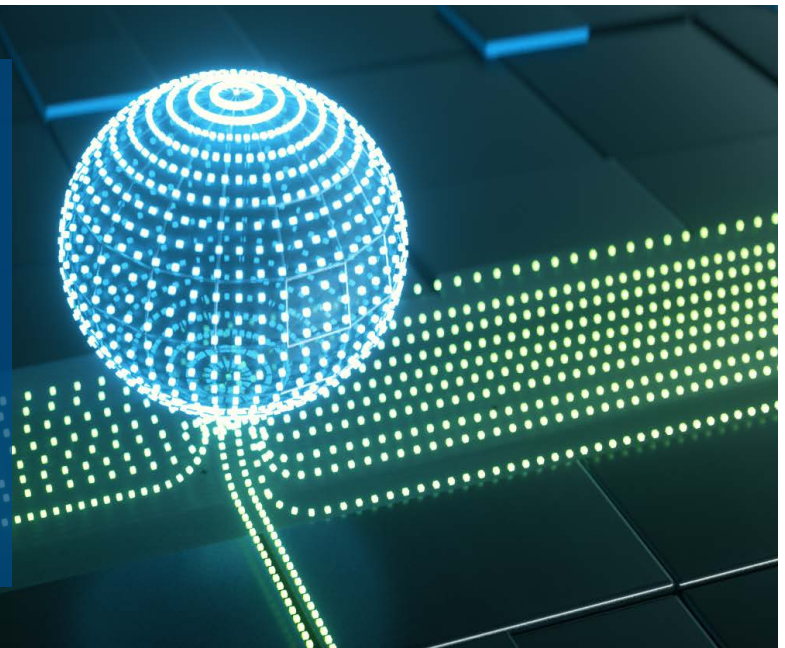
AI와 분석의 통합을 수용하면 비즈니스가 운영 인사이트와 데이터에 기반한 전략적 의사 결정에 사용되는 데이터를 볼 수 있는 가시성을 얻는데 도움이 됩니다. 비즈니스 임원의 84%는 성장 목표를 달성하려면 AI를 사용해야 한다고 믿지만, 76%는 AI를 비즈니스와 서비스 전체에 걸쳐 확장할 방법이 고민이라고 인정합니다.¹⁰

필요한 사항:

모든 유형과 규모의 비즈니스가 AI와 분석을 점점 많이 사용함에 따라, 성능 및 기타 운영 결과를 최적화하고 양의 ROI를 실현해야 합니다.

해답:

4세대 인텔® 제온® 프로세서를 사용하는 비즈니스는 데이터에서 더 많은 인사이트와 가치를 얻어 AI를 데이터 기반 의사 결정에 사용할 수 있습니다. 금융 서비스 산업에서는 투자 및 위험 관리에 대한 더 나은 결정을 내릴 수 있습니다. 의료 서비스 조직은 환자 치료를 개선하고, 더 효율적인 청구와 사전 승인을 통해 의료비를 줄이고, 수술 후 합병증을 더 정확하게 예측할 수 있습니다. 온라인 업체와 소매업체는 고객의 특정 한 니즈를 이해하고 연결하여 수요 예측 개선, 사전 마케팅 및 개인별 맞춤 마케팅, 실시간 물류 및 재고 관리를 통해 수익을 늘릴 수 있습니다.



- **인텔® Advanced Matrix Extensions(인텔® AMX)**는 4세대 인텔® 제온® 스케일러블 프로세서의 AI 기능을 가속하여 추가 하드웨어 없이 학습 및 추론의 속도를 높입니다. 이 가속기는 자연어 처리, 추천 시스템 및 이미지 인식 같은 워크로드에 적합합니다.
- **인텔® Advanced Vector Extensions 512 (인텔® AVX-512)**는 집약적 워크로드의 성능을 가속합니다. 이 최신 x86 벡터 명령 집합은 가장 까다로운 컴퓨팅 작업에 맞춰 성능을 가속화할 수 있도록 최대 2개의 단일 곱셈-누산기(FMA) 유닛과 기타 최적화를 포함하고 있습니다.

성능 입증

5.7X~10X 더 높은 PYTORCH 실시간 추론 성능¹¹ **3.5X~10X** 더 높은 PYTORCH 학습 성능¹²

내장된 인텔® AMX(BF16) 사용 시, 이전 세대(FP32) 대비

더 많은 작업 수행

여러 컴퓨팅 노드에 걸친 하이퍼스케일 클라우드 서비스 및 애플리케이션의 성능 향상.

[Google](#), [Meta](#), [Amazon](#), [Microsoft](#)를 비롯한 주요 고객은 데이터 센터 전반에서 시간 동기화 정확도를 획기적으로 향상시키는 것의 가치를 인식하고 있으며, 많은 고객이 정밀 네트워크 시간 동기화에 의존하는 분산형 애플리케이션을 배포했습니다(IEEE Std. 1588-2019 및 그와 유사함). 4세대 인텔® 제온® 스케일러블 프로세서에는 PCIe 네트워크 장치부터 CPU에서 실행되는 애플리케이션 소프트웨어에 이르기까지 서버 내의 정밀한 네트워크 시간을 확장하는 PTM(정밀 시간 측정)이 포함되어 있어 서버 간에 달성할 수 있는 작업량의 개선을 보완합니다.

PTM에 대한 자세한 내용은 다음을 참조하십시오 [“Adopting Time-Sensitive Networking \(TSN\) for Automation Systems.”](#)

기존 인프라와 쉽게 통합

인텔®을 사용하면 비즈니스가 알고 사용하는 가장 큰 파트너 에코시스템으로 배포 시간을 단축할 수 있습니다. 전세계의 하드웨어 및 소프트웨어 공급업체와 솔루션 통합업체는 인텔® 제온® 스케일러블 프로세서를 기반으로 자사 제품을 제작하여 선택권과 상호운용성을 극대화하고 수천에 이르는 실제 구현을 통해 신뢰를 높입니다.

수십 년에 걸친 에코시스템 지원은 운영 민첩성 위주의 제작, 확장 및 변환을 위해 필요한 모든 것을 사용하여 인텔의 신뢰 컴퓨팅 기초를 하이퍼스케일 데이터 센터 및 새로운 에지의 개척지 전체로 확장하는 데 도움이 됩니다. 이 접근방식의 개방성은 하드웨어, 소프트웨어, 클라우드 및 서비스 제공업체들에게 많이 선택 받는 원인입니다. [인텔® 파트너 얼라이언스](#)를 통해 AI, 클라우드, 고성능 컴퓨팅 및 기타 솔루션 분야를 위한 전용 리소스에 액세스하여 계획 및 제작하고 고객에게 더 많은 가치를 제공할 수 있습니다.

특정 워크로드 요구 사항을 지원하기 위해 다양한 클라우드 서비스, 형태 및 규모를 가장 유연하게 선택할 수 있는 인텔® 아키텍처는 전세계 모든 주요 클라우드 제공업체로 확장됩니다.

지원 상황

인텔의 **50,000** 가지가 넘는 고유 인스턴스 유형, 크기 및 지역으로 가장 많은 선택이 가능합니다. (경쟁사 대비 **6배 더 많은 수**)¹³

리더들이 디지털 변환 과정에서 가장 중요하게 생각하는 비즈니스 우선 과제

조직 리더(기술 및 비즈니스 모두)들의 디지털 변환 투자는 2022년과 2024년 사이에 총 6.3조 달러에 달해 2024년에는 전체 IT 지출에서 차지하는 비중이 최대 55%에 이를 것으로 예상됩니다.¹⁴ 본 비즈니스 요약은 이 변환의 미래에 비즈니스의 성공을 실현하기 위해 리더들이 초점을 맞출 최우선 비즈니스 과제와 4세대 인텔® 제온® 프로세서를 포함한 인텔® 하드웨어, 소프트웨어 및 서비스가 이런 우선 과제를 달성하는 데 도움이 되는 방법을 조명하는 시리즈의 일부입니다.



- **클라우드(본 요약):** 하이브리드, 멀티클라우드 및 인텔리전트 에지에 걸쳐 전략 활성화
- **보안:** 철저한 보안을 실현하고 제로 트러스트 보안 전략 보강
- **AI:** 데이터 분석과 AI를 채택하여 중요한 성과 주도
- **직원 경험 재정의:** 경계가 없는 대화형 직원 경험 포용
- **ESG:** 환경 | 사회 | 거버넌스(ESG)의 공정한 결과 및 책임 증진

자세한 내용

www.intel.com/xeon/scalable
www.intel.com/cloud



참고 자료

¹ Flexera. "2022 State of the Cloud Report." <https://resources.flexera.com/web/pdf/Flexera-State-of-the-Cloud-Report-2022.pdf>.
² IDC Market Forecast, 2021년 9월 (IDC #US47397521). "Worldwide Whole Cloud Forecast, 2021-2025: The Path Ahead for Cloud in a Digital-First World." <https://www.idc.com/getdoc.jsp?containerId=US47397521>.
³ 3세대 인텔® 제온® 스케일러블 프로세서와 AMD EPYC 비교. www.intel.com/3gen-xeon-config에서 구성 세부 정보 [126-130]을 참조하십시오.
⁴ intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [W5]를 참조하십시오. 결과는 다를 수 있습니다.
⁵ Everest Group, 2021년 10월. "Confidential Computing – The Next Frontier in Data Security." https://confidentialcomputing.io/wp-content/uploads/sites/85/2021/10/Everest_Group_-_Confidential_Computing_-_The_Next_Frontier_in_Data_Security_-_2021-10-19.pdf.
⁶ 인텔® 마이크로서비스 인사이트 연구, 2021년 6월.
⁷ intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [W1]을 참조하십시오. 결과는 다를 수 있습니다.
⁸ intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [W6]을 참조하십시오. 결과는 다를 수 있습니다.
⁹ intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [W7]을 참조하십시오. 결과는 다를 수 있습니다.
¹⁰ Accenture, 2019년 11월 19일. "AI: Built to Scale." <https://www.accenture.com/us-en/insights/artificial-intelligence/ai-investments>.
¹¹ intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [A17]을 참조하십시오. 결과는 다를 수 있습니다.
¹² intel.com/processorclaims: 4세대 인텔® 제온® 스케일러블 프로세서에서 [A16]을 참조하십시오. 결과는 다를 수 있습니다.
¹³ 출처: Historical Littr Insights Component tracker data + Intel internal preliminary analysis, 09/02/22 기준.
¹⁴ IDC, 2021년 10월. "IDC FutureScape: Worldwide Digital Transformation 2022 Predictions." <https://www.idc.com/getdoc.jsp?containerId=US47115521>.

공지 및 법적 고지 사항

가속기 사용 가능 여부는 SKU에 따라 다릅니다. 제품 세부 정보를 더 보려면 [인텔® 제품 사양 페이지](#)를 방문하십시오.
 성능은 사용, 구성 및 기타 요인에 따라 다릅니다. 자세한 내용은 <https://www.intel.com/PerformanceIndex>를 참조하십시오.
 성능 결과는 구성에 표시된 날짜의 테스트를 기반으로 하며 공개된 모든 업데이트가 반영되어 있지 않을 수도 있습니다. 자세한 구성 내용은 공개된 구성 정보를 참조하십시오.
 어떤 제품 또는 구성 요소도 절대적으로 안전할 수는 없습니다.
 인텔은 타사 데이터를 제어하거나 감사하지 않습니다. 정확성 평가를 위해서는 기타 소스를 참고해야 합니다.
 비용과 결과는 다를 수 있습니다.
 인텔® 기술은 지원되는 하드웨어, 소프트웨어 또는 서비스 활성화가 필요할 수 있습니다.
 여기에 설명된 인텔® 제품에 대한 침해 또는 기타 법적 분석과 관련하여 이 문서를 사용하거나 사용을 조정해서는 안 됩니다. 귀하는 여기에 공개된 주제를 포함하여 작성된 이후의 모든 특허 청구에 대해 인텔에 로열티가 없는 비독점적인 라이선스를 부여하는 데 동의합니다.
 기술된 제품에는 정오표로 알려진 오류나 설계 결함이 있을 수 있으며, 이로 인해 계산된 사양과 다르게 작동할 수도 있습니다. 현재 정리된 'errata'는 요청할 경우 제공받을 수 있습니다.
 © 인텔사. 인텔, 인텔 로고 및 기타 인텔 마크는 인텔사 또는 그 자회사의 상표입니다. 기타 명칭 및 브랜드는 해당 소유업체의 자산일 수 있습니다.
 1122/MMH/MESH/350495-002US