

# 4세대 인텔<sup>®</sup> 제온<sup>®</sup> 스케일러블 프로세서

인텔코리아 나승주 상무

데이터센터와 네트워크, 엣지의  
컴퓨팅 수요는 그 어느 때보다  
커졌습니다.

# 고객의 실 사용 워크로드에 집중

## 성능 가속을 위한 설계

인공지능



네트워킹/5G



스토리지



HPC



데이터 분석



보안



# 인텔만의 차별화된 전략

성능 가속을 위한 설계

고객 워크로드 우선

CPU Cores + 내장 가속기

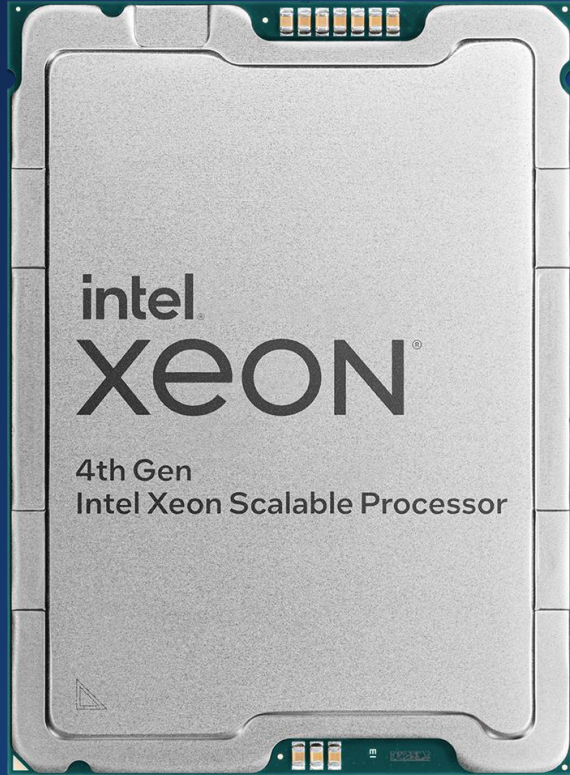
오픈 소프트웨어 생태계 + oneAPI & AI 툴

고성능

향상된  
효율성

최적의 TCO

# 4세대 인텔® 제온® 스케일러블 프로세서



1 ~ 8 소켓 확장성

최대 60 코어  
프로세서당

현존 CPU중 가장 많은 내장 가속기 탑재

DDR5로 메모리 대역폭 증대

PCIe 5로 I/O 대역폭 증대  
80 레인

UPI 2.0로 소켓간 대역폭 증대

Compute Express Link (CXL) 1.1

하드웨어로 강화된 보안



# 인텔의 품질에 대한 장점

## 품질

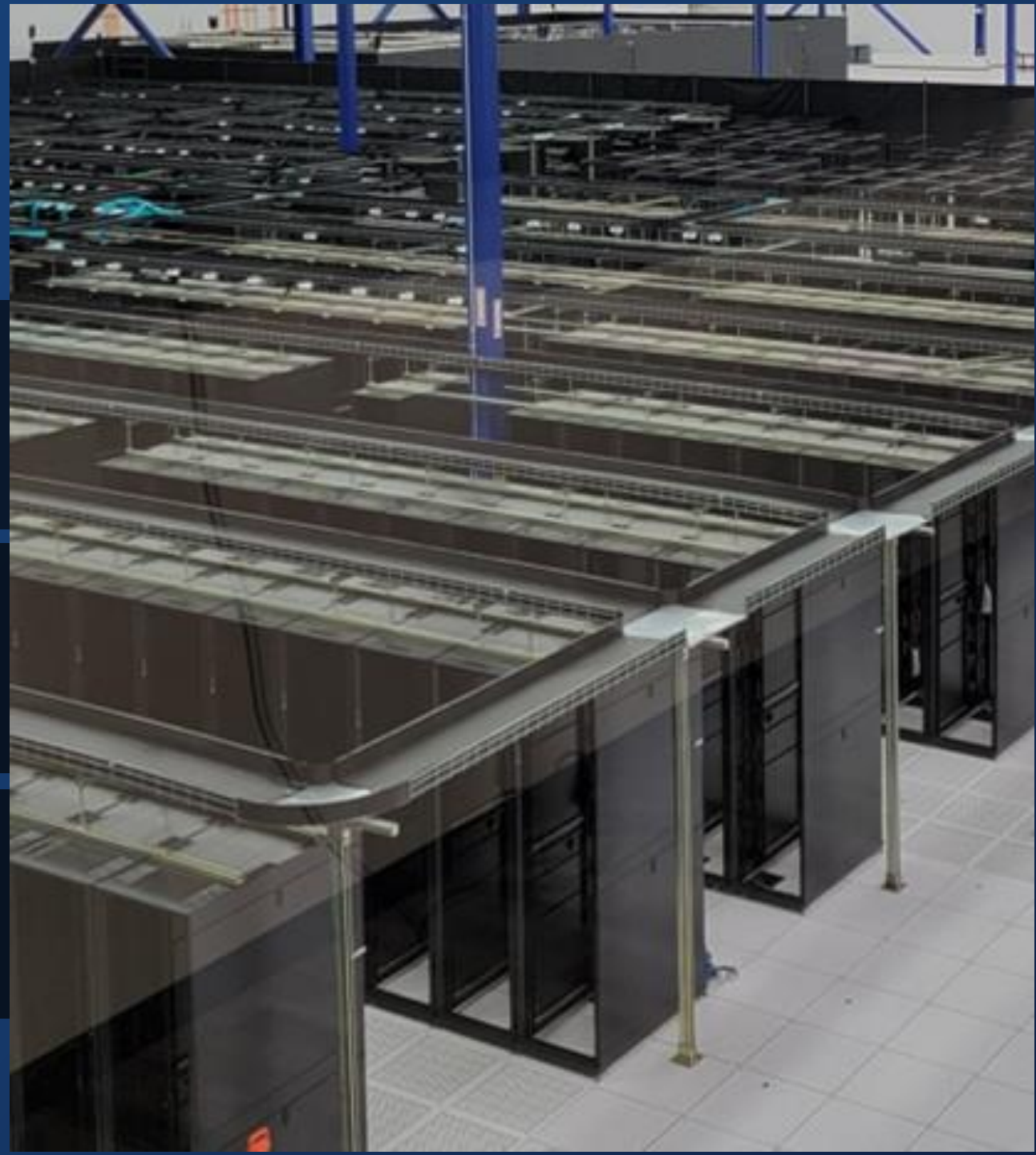
지속적인 개선 및 시장 리더십

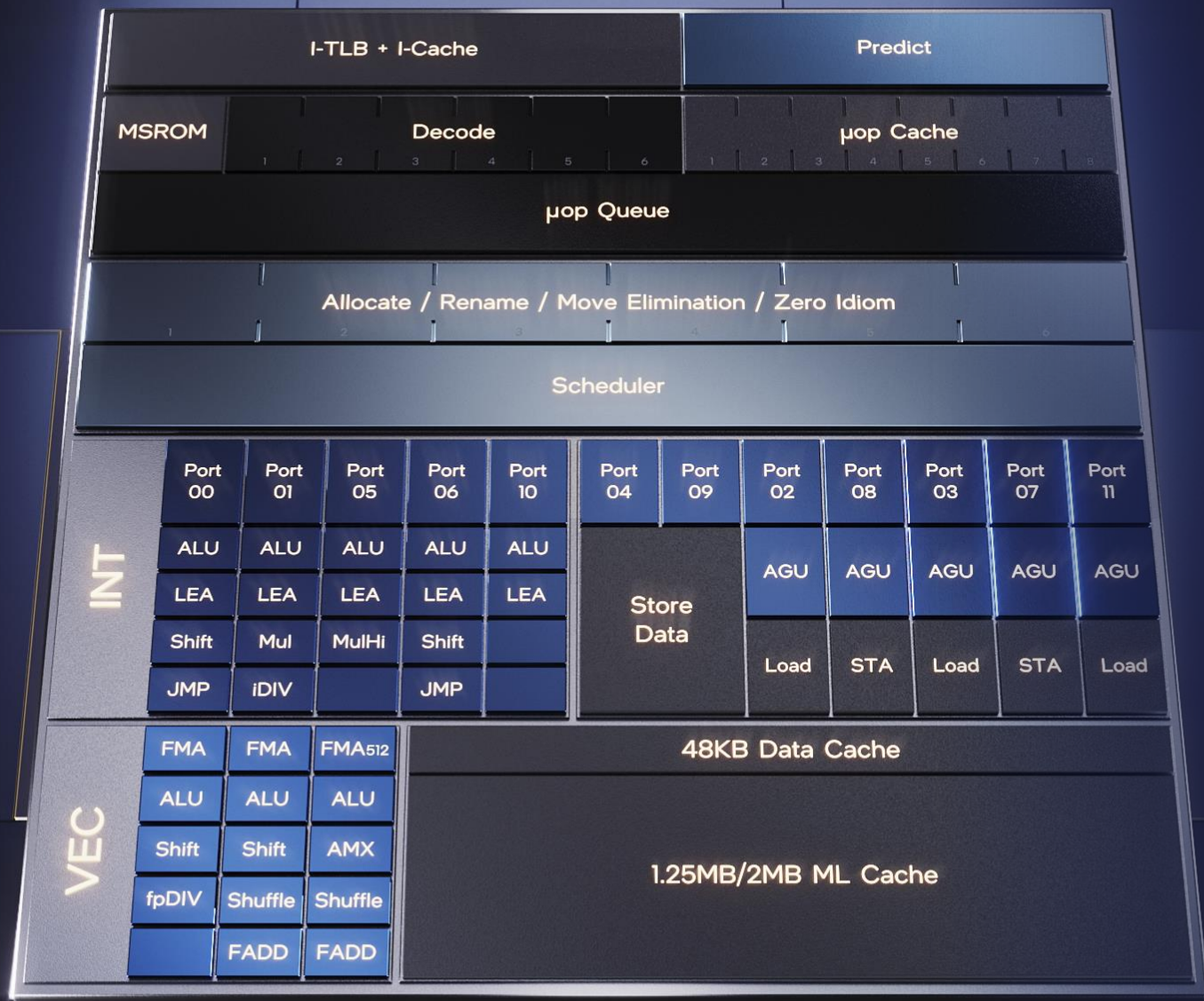
## 사전 검증

대규모 하드웨어 + 소프트웨어

## 관리

전 제품 주기동안 관리





New

# Performance x86 Core

A Step Function in CPU Architecture  
Performance For the Next Decade of  
Compute

A significant IPC boost at high power efficiency

Wider      Deeper      Smarter

Better supports large data set and large code footprint applications

Enhanced power management improves frequency and power

Machine Learning Technology: Intel® AMX – Tile Multiplication

All in a tailored scalable architecture to serve the  
full range of Laptops to Desktops to Data Centers

# 모든 코어의 효율성 극대화

## 새롭게 포함된 가속 엔진 IP

인텔® 가속 엔진은 보다 범용적인 컴퓨팅 작업을 위해 코어를 확보하여 전반적인 워크로드 성능과 전력 효율성을 높임

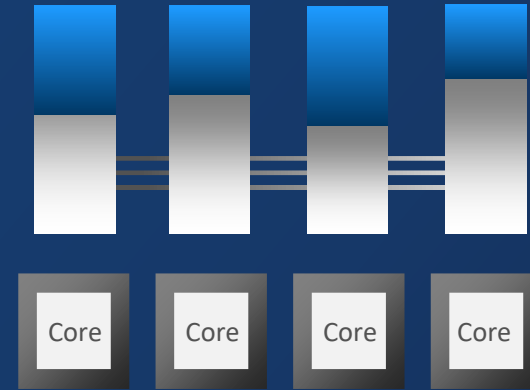
### 내장된 가속 엔진 IP

- 인텔® 퀵어시스트 기술 (인텔® QAT)
- 인텔® 다이내믹 로드 밸런서 (인텔® DLB)
- 인텔® 데이터 스트리밍 가속기 (인텔® DSA)
- 인텔® 인-메모리 분석 가속기 (인텔® IAA)

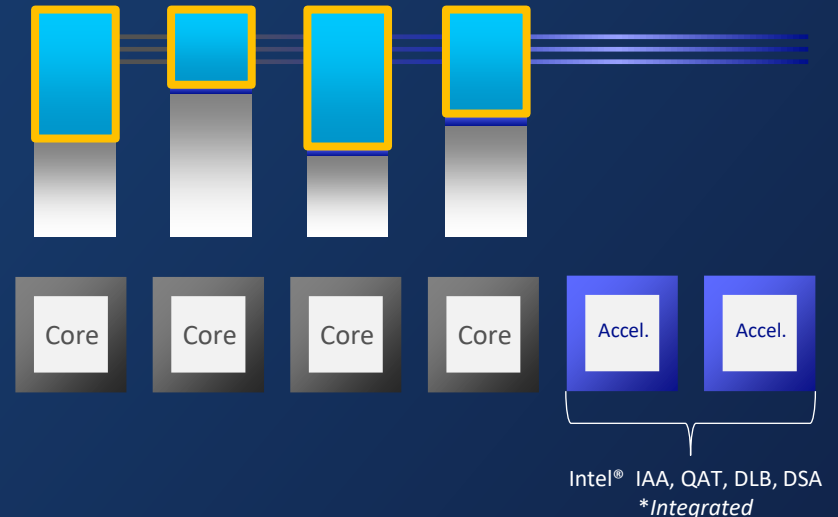
### 새로운 CPU 명령어(ISA)

- 인텔® 어드밴스드 매트릭스 익스텐션 (인텔® AMX)
- vRAN용 인텔® 어드밴스드 벡터 익스텐션 (인텔® AVX-512)

가속기 없는 CPU 사용 예



가속기 사용하는 CPU 사용 예



중요 작업

일반 작업

추가적인 작업 허용량

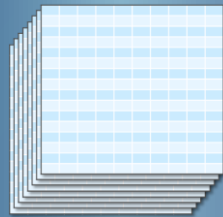
**Hewlett Packard**  
Enterprise



# 인텔® 어드밴스드 매트릭스 익스텐션 (인텔® AMX)

모든 코어에 딥러닝 성능 가속

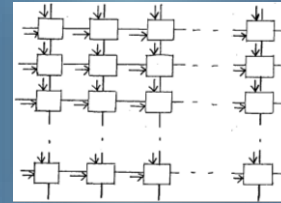
“Tiles”



2D 레지스터 파일



“TMUL”



타일 매트릭스 곱셈

더 큰 데이터 단위로  
저장



한번에 더 많은  
매트릭스를 계산하는  
명령어

# 인텔® 어드밴스드 매트릭스 익스텐션 (인텔® AMX)

## 가속 엔진

이전 세대 대비  
성능 향상 비교

이전 세대 대비  
성능 향상 비교

### 기능

- 광범위한 하드웨어 및 소프트웨어 최적화를 제공하여 AI 가속화를 가능

### 비즈니스에 제공하는 가치

- AI/딥러닝 추론 및 학습 워크로드에 대한 상당한 성능 향상
- 하드웨어 가속화로 공통 애플리케이션을 더 빠르게 함

### 소프트웨어 지원

- AI 프레임워크, 툴킷, 라이브러리 (PyTorch, TensorFlow), 인텔® oneAPI 딥 뉴럴 네트워크 라이브러리 (oneDNN)

### 활용 사례

- 이미지 인식, 추천 시스템, 기계/언어 번역, NLP, 미디어 처리 및 전달

### 음성 인식 추론

내장된 인텔® AMX (BF16)  
와 이전 세대 (FP32) 비교시

최대 **8.6배**

높은 음성 인식 추론 성능

### PyTorch 학습과 추론

내장된 인텔® AMX  
(BF16)와 이전 세대 (FP32)  
비교시

실시간 추론 및  
학습 성능 모두에서

최대 **10배**

높은 PyTorch 성능

See backup for workloads and configurations. Results may vary.

# 인텔® 퀵 어시스트 기술(인텔® QAT)

## 가속 엔진

가속기를 사용하지 않았을 때와  
성능 향상 비교

이전 세대 대비  
성능 향상 비교

### 기능

- 암호화 데이터 압축/압축 해제 가속

### 비즈니스에 제공하는 가치

- 압축/압축 해제 오프로드 가속화로 CPU 효율성 향상
- 더 적은 오버헤드로 장치 간 더 많은 암호화된 연결 및 웹 보안 연결

### 소프트웨어 지원

- 암호화 가속화용 인텔® QAT 엔진

### 활용 사례

- 분산 스토리지 시스템, 파일 시스템, RocksDB, 데이터 레이크, 아파치 스파크, 하둡, NGINX, IPSec

### 네트워크 시큐어 게이트웨이

NGINX에서 아웃오브 박스  
소프트웨어 대비  
내장된 QAT를 사용할 때

최대 **84%**

적은 코어로  
동일한 초당 연결 가능

### 엔터프라이즈 스토리지 및 데이터 분석

이전 세대 대비  
내장 QAT를 사용하면,  
코어수를

최대 **95%**  
작게 사용하고,

레벨 1 압축 처리량을

**2배**  
향상

# 인텔® 데이터 스트리밍 가속기 (인텔® DSA)

## 가속 엔진

가속기를 사용하지 않았을 때와  
성능 향상 비교

이전 세대 대비  
성능 향상 비교

### 기능

- 스트리밍 데이터의 이동 및 변환 작업을 최적화

### 비즈니스에 제공하는 가치

- CPU를 오프로드하여 데이터 스토리지 애플리케이션의 효율성을 향상하고 NVMe/TCP의 데이터 보호를 가속화

### 소프트웨어 지원

- 인텔® 데이터 이동 라이브러리 (Intel® Data Mover Library)

### 활용 사례

- 가상화, non-transparent 브리지 간의 빠른 복제, ERP, 인메모리 데이터베이스

### 데이터 무결성 (처리량)

대용량 패킷 순차  
읽기에서

내장된 인텔®DSA와  
ISA-L 소프트웨어를  
비교할 경우

최대 **1.7배**  
높은 IOPs

### 데이터 무결성 (처리량과 지연시간)

대용량 패킷 순차  
읽기에서

내장된 인텔®DSA와  
이전 세대를 비교할  
경우

최대 **1.6배**  
더 높은 IOPs 와

**37 %**

지연시간 단축



# 인텔® 다이نام믹 로드 밸런서 (인텔® DLB)

## 가속 엔진

가속기를 사용하지 않았을 때와  
성능 향상 비교

이전 세대 대비  
성능 향상 비교

### 기능

- 정적인 NIC 분배로 인해 로드 불균형이 발생할 경우, 데이터 로드를 코어 간에 동적으로 재분배

### 비즈니스에 제공하는 가치

- 멀티코어 인텔® 제온® 스케일러블 프로세서에서 네트워크 데이터 처리와 관련된 시스템 성능 향상
- 분산 처리, 동적 로드 밸런싱 및 동적 네트워크 프로세싱의 재정렬을 위한 성능 향상

### 소프트웨어 지원

- 인텔® 데이터 이동 라이브러리 (Intel® Data Mover Library)

### 활용 사례

- IPSec 보안 게이트웨이, VPP 라우터, UPF, vSwitch, 스트리밍 데이터 처리, Elephant 흐름 처리

### 마이크로서비스

동일한 처리량을  
Istio ingress gateway  
소프트웨어와 내장된  
인텔®DLB 로 비교할 경우

최대 **96%**

지연 시간 단축

### 마이크로서비스

이전 세대 대비 내장된  
인텔® DLB를 통해 동일한  
코어에서 지연시간은

최대 **89%**

단축되고,

CPU 활용률은

**57%** 감소

# 인텔® 인-메모리 분석 가속기 (인텔® IAA)

## 가속 엔진

가속기를 사용하지 않았을 때와  
성능 향상 비교

이전 세대 대비  
성능 향상 비교

### 기능

- 내장된 가속기는 분석, CRC 계산, 압축 및 압축해제를 가속

### 비즈니스에 제공하는 가치

- 인메모리 DB 및 분석 워크로드의 쿼리 처리량 증가
- 분석 워크로드를 위한 메모리 및 대역폭 사용량을 줄여 CPU 공간 확보

### 소프트웨어 지원

- 인텔® 쿼리 프로세싱 라이브러리, 인텔® 데이터 이동 라이브러리 (Intel® Data Mover Library)

### 활용 사례

- 상용 인메모리 DB, 오픈소스 인메모리 DB(RocksDB, Redis, Cassandra, MySQL, MongoDB), 빅데이터 분석을 위한 컬럼 형식(columnar formats)

## 임베디드 데이터베이스

내장된 인텔® IAA 와  
Zstd 소프트웨어 사용  
비교시

RocksDB 에서

최대 **2.1배**  
성능 향상

## 임베디드 데이터베이스

이전 세대 대비  
내장된 인텔® IAA를  
사용할 때

최대 **3배**  
높은 RocksDB 성능과

**66%**  
지연시간 단축

See backup for workloads and configurations. Results may vary.

# 고객 워크로드 가속화

가장 많은 내장 가속기로 최고의 성능 제공

4세대 인텔® 제온® 스케일러블 프로세서

인텔® 제온® CPU  
맥스 시리즈

범용 컴퓨팅

53%

평균 성능  
향상\*

인공지능

최대 10배

높은 추론과  
학습 성능\*

네트워크 5G  
vRAN

최대 2배

동일한 전력  
엔벨롭에서 vRAN  
워크로드 용량\*

네트워킹  
& 스토리지

최대 2배

95% 적은  
코어로 더 높은  
데이터 압축\*

데이터 분석

최대 3배

더 높은 성능\*

HPC

최대 3.7배

메모리 집중  
워크로드\*\*

와트 당 성능 **2.9x** 내장 가속기로 평균 향상치\*

\*4세대 인텔 제온 스케일러블 프로세서 vs. 3세대 인텔 제온 스케일러블 프로세서

\*\* 인텔 제온 CPU 맥스 시리즈 vs.  
인텔 제온 8380 프로세서.

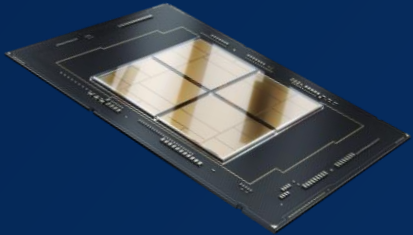
# 인텔® 제온® 프로세서 기반 서버로 교체 및 통합

4세대 인텔® 제온® 프로세서로 총소유비용을 크게 낮추고 4개월 만에 비용 회수가능





# 인텔® 제온® 프로세서 로드맵



4세대 인텔®  
제온® 프로세서  
Intel 7  
2022



에메랄드  
래피즈  
Intel 7  
2023



그라나이트  
래피즈  
Intel 3  
2024



차세대

P-Core

Perf/core optimized for mainstream & premium cloud and data-center applications



시에라  
포레스트  
Intel 3  
2024

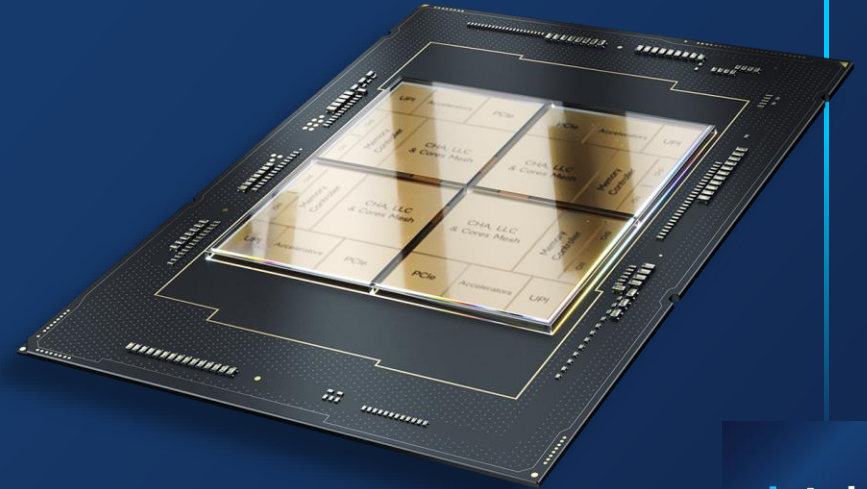
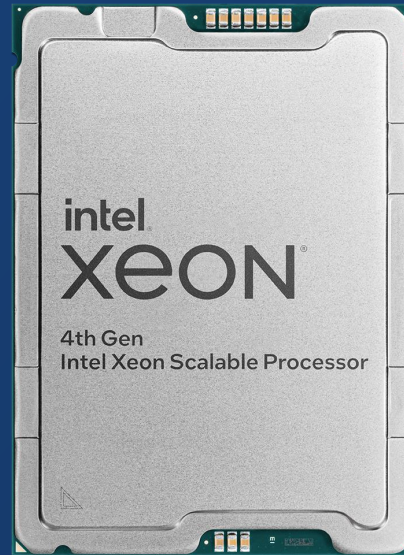
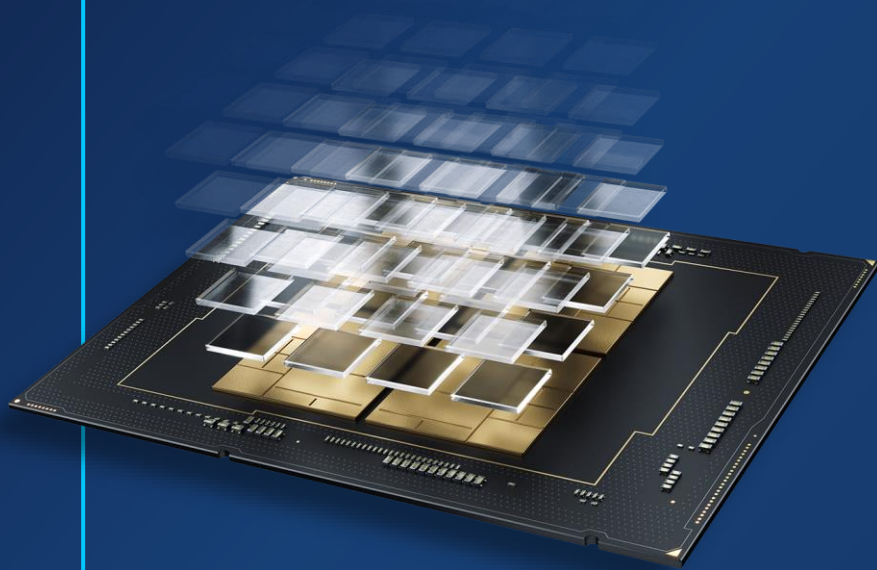


클리어워터  
포레스트  
Intel 18A  
2025

E-Core

Power/perf optimized to support high-density, ultra-efficient compute for the cloud

# 감사합니다



intel.  
XEON®

intel®