

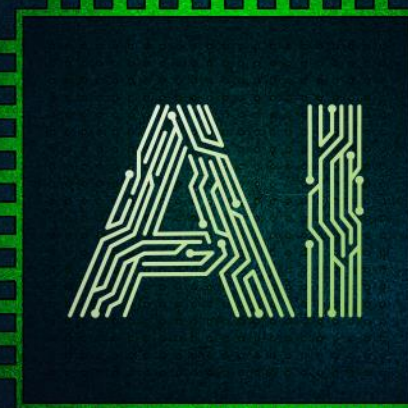
HS  효성인포메이션시스템 | XSLAB 엑세스랩

# CIO Summit 2026

## VIP 조찬 세미나

2026년 2월 26일

그랜드 인터컨티넨탈 서울 파르나스 로즈룸(5층)

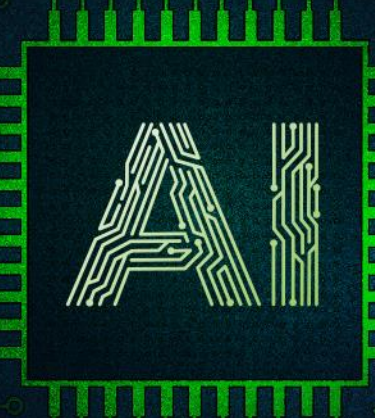


[세션1]

# AI 컴퓨팅의 전환점: AI 운영 환경에서 Arm 서버

---

HS효성인포메이션시스템 이진석 팀장



# Agenda

I. AI 트렌드

II. Arm 이란?

III. AI 시대에 Arm이 적합한 이유

# I. AI 트렌드

---

1. '26 AI Trend
2. HS효성인포메이션시스템 AX 포트폴리오

# '26 AI Trend

- AI를 실무에 적용하고 ROI를 생각하는 시대 -

멀티모달 AI

텍스트, 이미지, 음성 등  
다양한 데이터를  
동시에 처리

Agentic AI

개인화된 AI Agent를 통한  
단순 작업 자동화에서 다중  
단계 업무 까지 수행

Physical AI

물리적 법칙 + 데이터 기반  
학습을 통해 실제 현상을  
보다 정확히 대응

# HS효성인포메이션시스템 AX 포트폴리오

## AX를 위한 전사 시너지 강화



## II. Arm 이란?

---

1. Arm 아키텍처
2. Arm 서버 기대 효과
3. 사례 및 활용영역

# Arm 아키텍처

- X86 대비 적은 수의 단순한 명령 집합을 빠르게 수행하도록 설계된 프로세서
- 단순한 구조로 인해 저전력, 저발열, 내열성 등의 장점을 보유

# ARM

# VS

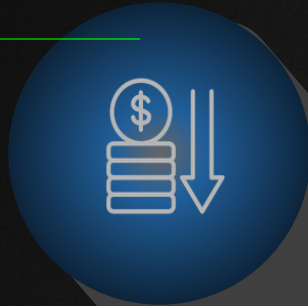


Reduced Instruction Set Computing (RISC)	아키텍처	Complex Instruction Set Computing (CISC)
단순한 작업을 빠르게 수행하여 효율성 증대	개발 목표	다양한 작업을 수행 가능하여 범용성 중시
멀티코어 병렬 구조 단순, 트랜지스터 적음	하드웨어 구조	고성능 단일코어 중심 복잡, 트랜지스터 많음
높음	전력 효율	낮음
적음	발열	많음

# Arm 서버 기대 효과

## 도입 원가 절감

- X86(Intel & AMD) 대비 CPU 칩 자체 비용 저렴
- 서버 팬, 전원 등 x86 대비 저렴한 부품 구성 가능

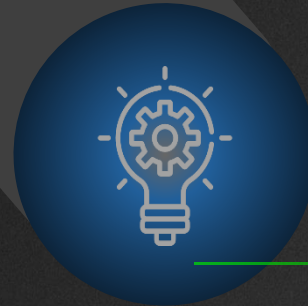


## 저전력/저발열 지원

- RISC의 단순한 구조로 인해 x86 대비 적은 전력으로 원활한 운영 가능
- 저발열로 인해 냉방비 및 공간 활용 측면에서 이점이 있음

## 실시간 병렬 처리 지원

- 매니코어 기반의 실시간 병렬 처리 지원, 데이터 병목현상을 줄여 빠른 응답속도 기대 가능
- 32/64/80/128/192 Core 등 고객 요구에 맞춘 코어 대응 가능



## 환경적 제약 극복

- X86 대비 전원/온도/습도 제약이 적어 열악한 환경 속에서도 운영 가능
- 전장 장비(배, 트럭 등)/사무실/야외 환경 등 상온에서 운영 가능

# 사례 및 적용영역

## Cloud Native



Graviton 5

- 전작 대비 성능 25% 향상, 탄소 배출량 저감 핵심



Google Cloud

Axion

- x86 대비 성능 60%, 전성비 65% 우위



Microsoft Azure

Cobalt 200

- Cobalt 100 대비 성능 50% 향상, 초고밀도 설계

## Edge Computing



맥북

- M2 칩에는 AI 연산을 전담하는 16코어 뉴럴 엔진 탑재
- 전성비 우수 (저전력/저발열)



Mercedes-Benz

자율주행

- 차량 내부에 ARM Neoverse 기반 CPU가 탑재
- 자율주행 AI 모델 구동

## Physical AI



Boston Dynamics

아틀라스

- Nvidia Jetson Thor 이용 로봇 개발
- 현장 로봇과 개발/테스트 단의 생태계 환경 ARM 통일

# III. AI 시대에 Arm이 적합한 이유

---

1. AI 워크로드 변화
2. AI 컴퓨팅 기준 변화
3. Physical AI 확산
4. 적용영역별 GreenCore 라인업

# AI 워크로드 변화

- AI 워크로드는 목적과 특성에 따라 두 가지 영역으로 구분



## 모델 학습 Training Phase

### 핵심 목표

모델 품질 향상 & 대규모 데이터 소화

### 주요 관점

- ✓ GPU 중심 자원 배치
- ✓ 전체 처리량(Throughput) 중심의 지표 관리



## 추론 & 운영 Inference, Service, Control

### 핵심 목표

안정적 서비스 운영 & 비용 효율화

### 주요 관점

- ✓ 실시간 처리에 가까운 Low Latency
- ✓ 24시간 항상 운영되어야 하는 가용성
- ✓ 전력 비용 절감

# AI 컴퓨팅 기준 변화

- AI는 이제 PoC 단계를 넘어 업무 프로세스와 현장 시스템에서 상시 개입
- 인프라를 평가하는 기준이 단순 성능에서 운영 안정성으로 변화 중



과거  
**AI 모델 중심** Model Centric

## 크기 & 속도

얼마나 큰 모델을 빠르게 학습하는가?

## GPU 집약 성능

연산 능력 절대치 중시

"더 큰 데이터와 더 많은 GPU를 부어  
성능을 1%라도 더 올리자."



현재  
**시스템 중심** System Centric

## 가용성 & 지속성

서비스/현장에서 상시 사용 가능한가?

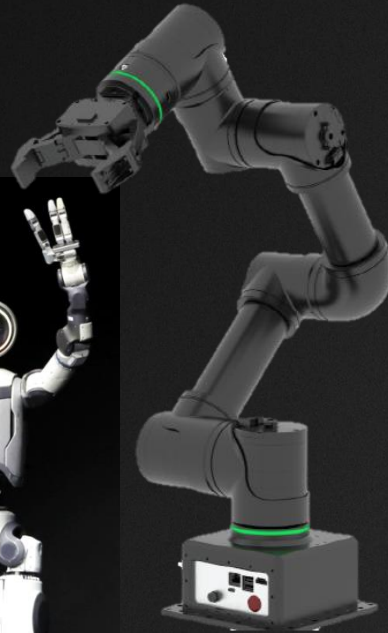
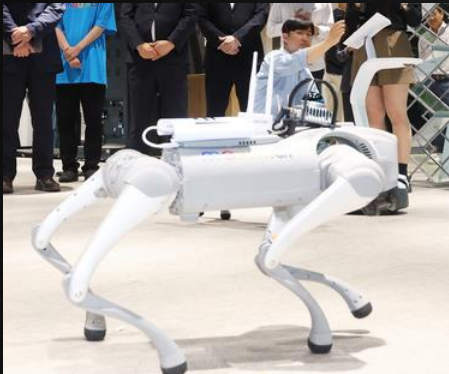
## 운영 효율성 관점

전력 비용, 공간 활용 등 운영 효율성 중시

"답만을 원하는 것이 아니고 실행력을 원한다."

# Physical AI 확산

- 자율주행차, 스마트 공장, 의료 현장 등 Physical AI의 범위 확산 중
- Arm 아키텍처는 낮은 지연 시간, 전력/발열 제약 환경에서의 안정적 운영 등 Physical AI의 요구 조건에 충족



## CHALLENGE 현장의 요구 조건

- ✓ 낮은 지연 시간
- ✓ 전력/발열 제약 환경에서의 운용
- ✓ 장시간 무중단 운용



**GreenCore**

# 적용영역별 GreenCore 라인업

모델	GreenCore SQ mini	GreenCore LX	GreenCore Altra 32/64	GreenCore Altra 80/128
외형				
사이즈	1U	2U	1U/2U/4U	
CPU/Mem	24(1.0)core ~64Gb	16(2.2)core ~128Gb	32(1.7)/64(2.2)Core/ ~2TB	80(3.0)/128(3.0)Core ~4TB
Disk	HDD/SSD 지원	HDD/SSD 지원	HDD/SSD 지원	
포지션	엣지 서버 시장 (엔트리 환경)	데이터센터 서버 (미드레인지)	클라우드 서비스 / 데이터센터 고성능 서버 시장(하이엔드)	
주요 업무 (타겟)	저가용 서버/ 엣지 환경	데이터센터 저장 중심 서버	고성능 서버 클라우드 서버용 (가상화, 오픈스택, K8s 등)	



고성능 워크로드 대응을 위한  
**GPU / NPU 추가 구성 가능**

온라인 및 실제 장비  
**테스트 환경 제공**

H/W부터 S/W까지  
**고객 맞춤 구성 및 지원**

# Summary

“성공적인 AI 도입은 가장 빠른 컴퓨팅이 아닌,  
비즈니스에 적합한 효율적인 컴퓨팅을 구축하는 것입니다.”

01

## 워크로드의 변화

Training vs Inference

학습과 추론(서비스) 영역의  
요구사항이 명확히 분리되고,  
각 목적에 맞는 최적의 자원 배치가  
필요



02

## 인프라 기준의 전환

From Model to System

AI 인프라의 평가 기준이 단순 모델  
성능에서 지속적으로 운영 가능한  
시스템의 안정성 및 신뢰성으로 전환



03

## Physical AI의 부상

Rise of Physical AI

Physical AI를 통한 AI의 현실  
세계와의 결합이 가속화됨에 따라,  
지연시간·전력 효율·안정성에  
강점이 있는 Arm 아키텍처의  
역할이 확대되는 중

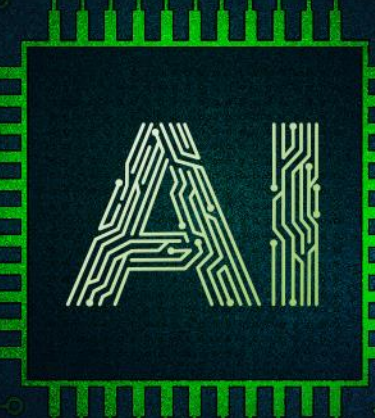


[세션2]

# Beyond DX to Green AX

---

엑세스랩(주) 유명환 대표이사



# Agenda

- I. [현재] TCO & ESG 달성을 위한 Green Computing
- II. [미래] 클라우드 네이티브와 AI 시대를 위한 인프라
- III. [시작] 리스크 없는 단계적 전환

# [결론] Beyond DX to Green AX

## 맥 미니, 갑작스런 품귀 왜?...오픈클로가 만든 새로운 시 열풍

▲ 시리포터 | © 승인 2026.02.20 09:07

AI 요약 **오픈소스 시 에이전트 오픈클로** 인기로 고사양 맥 미니 수요가 급증하며 품귀 현상이 빚어지고 있다. 일부 모델은 배송이 최대 3주까지 지연되며 이베이에선 높은 가격에 되팔리고 있다. 전문가들은 올해 신형 맥 미니 출시 가능성을 언급하며 소비자들에게 신중한 구매를 당부했다.



M4 맥 미니 제품 사양도 [사진: 애플]

[디지털투데이 시리포터] 애플의 소형 데스크톱 '맥 미니'가 최근 품귀 현상을 빚고 있다.

19일(현지시간) 비즈니스인사이드에 따르면, 오픈소스 인공지능(AI) 에이전트 '오픈클로'(OpenClaw)가 인기를 끌면서 고사양 맥 미니 수요가 급증했다. 이 영향으로 애플 스토어에서는 고용량 메모리를 탑재한 모델의 배송이 최대 3주까지 지연되고 있다.

지난 2월 9일 틱톡(TikTok)에 올라온 영상에서 미국 대형 전자제품 유통업체 베스트바이(Best Buy) 직원은 고객에게 "맥 미니가 왜 이렇게 잘 팔리는지 모르겠다. 시 때문인가?"라고 말하기도 했다. 갑작스러운 수요 증가에 일부 매장 직원들도 혼란을 겪는 모습이다.

수요 급증의 배경에는 오픈소스 자율형 AI 에이전트 오픈클로가 있다. 최근 오픈시에 합류한 피터 스타인버거(Peter Steinberger)가 개발한 이 도구는 일정 관리와 AI 에이전트 실행 기능을 제공하며, 구동에 높은 메모리 용량을 요구한다. 이에 따라 대용량 메모리를 지원하는 맥 미니가 개발자들 사이에서 필수 장비로 떠오르고 있다.

이제 **AI** 는

✓ 누가 더 많은 데이터를

✓ 누가 더 싸게 구축할 수 있는지

## 오픈클로 열풍, 맥 미니 이어 초소형 PC '라즈베리 파이' 주가 폭등 이끌어

▲ 임대준 기자 | © 승인 2026.02.19 19:23

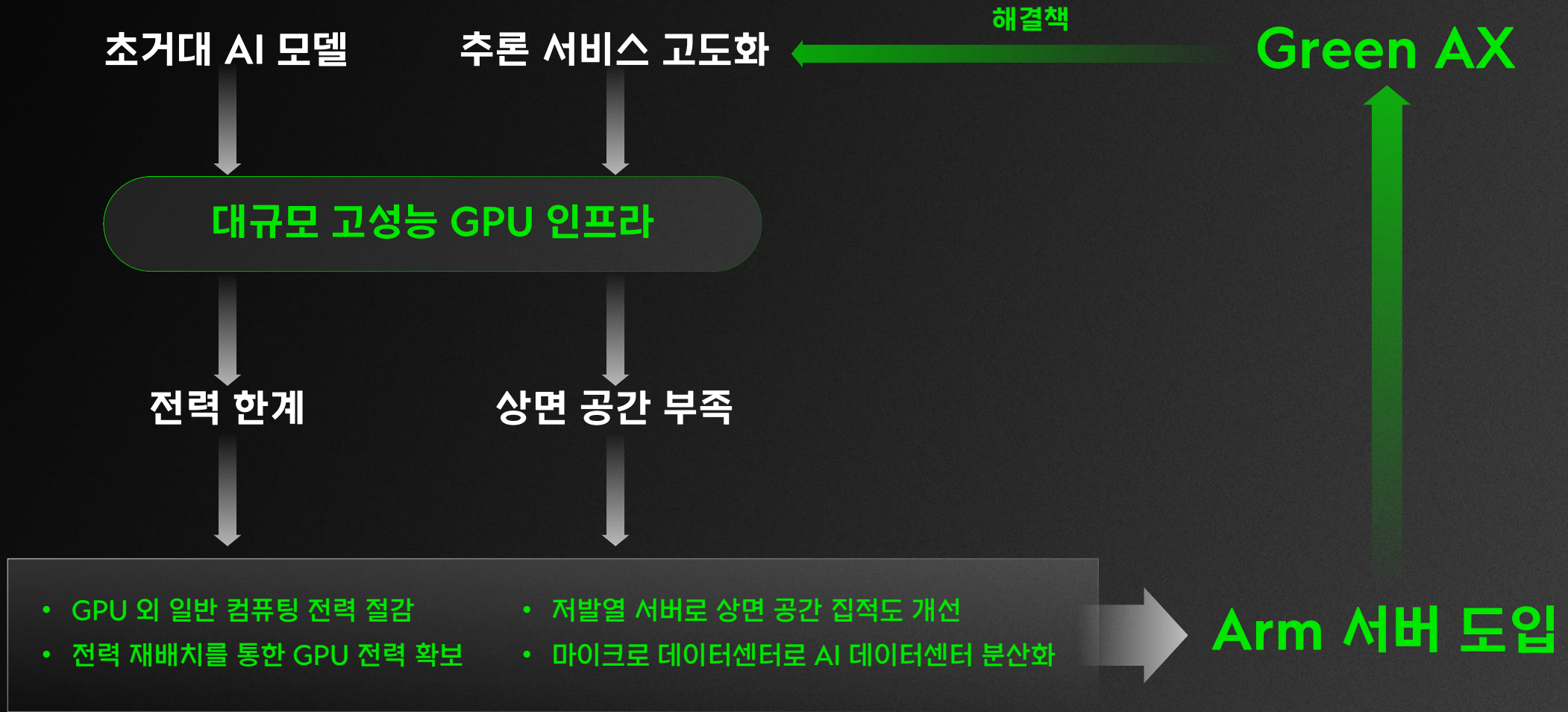


라즈베리 파이 5 (사진=셔터스톡)

맥 미니의 품귀 현상을 이끌었던 '오픈클로'가 이번에는 저가형 PC 업체의 주가를 폭등시켰다.

18일(현지시간) 파이낸셜 타임스에 따르면, 영국의 라즈베리 파이는 9개월 만에 처음으로 기업가치 10억파운드(약 1조9550억원)를 돌파했다.

# [결론] Beyond DX to Green AX



# [현재] TCO & ESG 달성을 위한 Green Computing

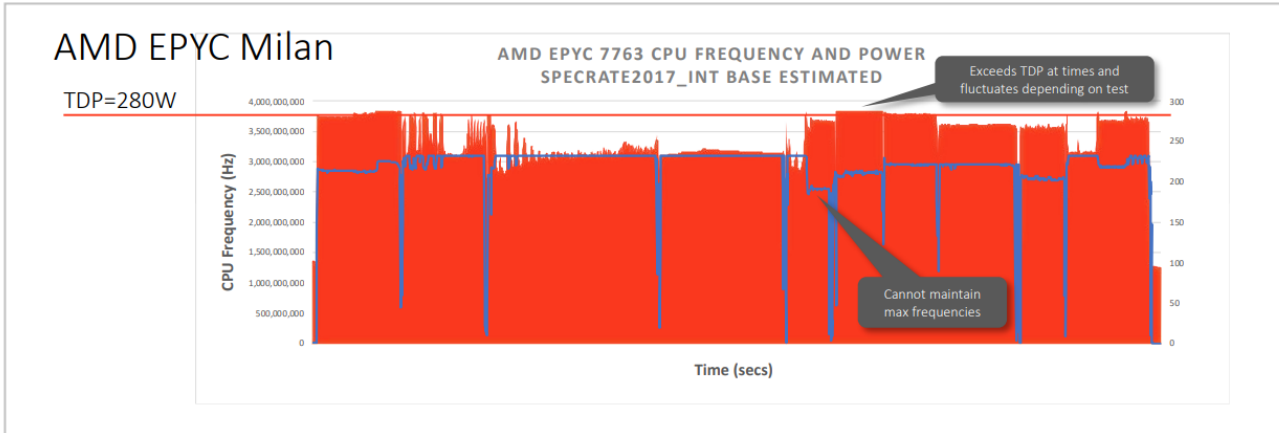
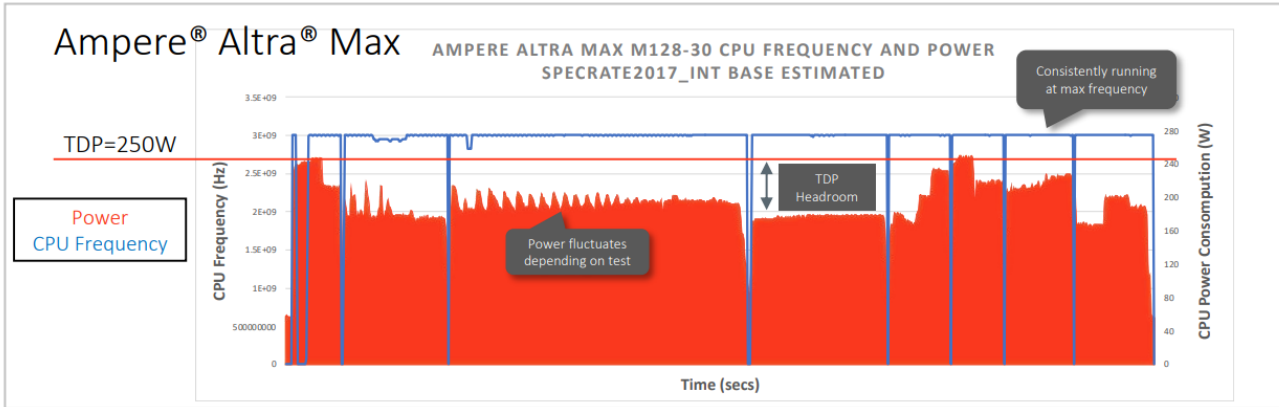
## 기존 x86 서버 vs Arm 서버 인프라 TCO 및 가치 분석

핵심 지표	기존 인프라 (인텔 서버)	차세대 Green 인프라 (Arm 서버)	핵심 체크 포인트 (TCO/효과)
소모 전력 (TDP)	CPU 당 300 ~ 400W	CPU 당 150 ~ 250W	<ul style="list-style-type: none"> <li>전기 요금(OPEX) 약 30 ~ 40% 절감</li> <li>냉각 설비 가동률 동반 하락</li> </ul>
상면 공간 (Rack Density)	높은 발열로 랙당 장착 대수 제한	단일 CPU 에 128 ~ 192코어 고집적 저발열	<ul style="list-style-type: none"> <li>데이터센터 임대료(공간) 절약</li> <li>동일 공간 대비 처리량 극대화</li> </ul>
전력 대비 성능 (Perf/Watt)	기준 (1x)	1.5x ~ 2x 이상	<ul style="list-style-type: none"> <li>대규모 데이터 전처리</li> <li>클라우드 네이티브 워크로드 → 적은 비용으로 처리 가능</li> </ul>
ESG / 지속가능성 (Green)	탄소 배출 규제 대응에 한계	탄소 발자국(Carbon Footprint) 획기적 감소	<ul style="list-style-type: none"> <li>기업 ESG 평가 지표 상승</li> <li>친환경 데이터센터 인증 유리</li> </ul>

# [현재] TCO & ESG 달성을 위한 Green Computing



## Ampere® Altra® Max Energy Efficiency



	Performance (SpecRate2017_INT Base)	Usage Power (W)	Performance /Watt
AMD EPYC Milan	331	280W	1.0x
Ampere® Altra® Max	360	178W	1.71x

Arm 서버 도입 = 탄소 발자국 감소

- 성능은 그대로, 전력은 반으로
- 탄소 배출 저감 목표 달성을 위한 가장 최적의 솔루션

Ampere® Altra® Max maintains **predictable core frequencies** while consuming lower power (below TDP)

Power headroom means workload-driven power capping can lead to huge density improvements!

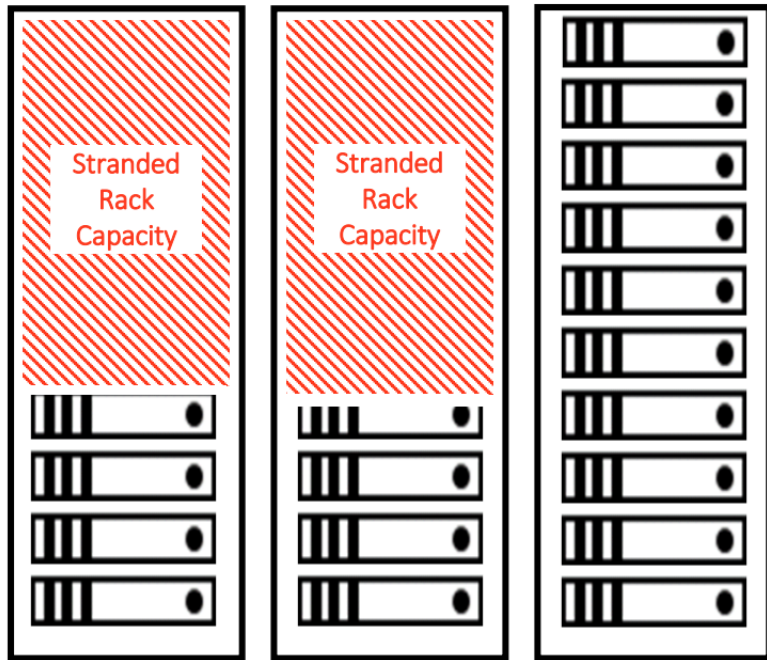
Compelling performance/Watt at competitive levels of performance



# [현재] TCO & ESG 달성을 위한 Green Computing

## Rack Efficiency Using Ampere Cloud Native Processors

Based on 42U rack @12.8 kW



Intel Ice Lake  
8380

AMD Milan  
7763

Ampere Altra Max  
M128-26

### Performance per Rack<sup>1</sup>

Workload	Intel	AMD	Ampere
SIR2017 Est.	1X	1.4X	2X
Redis	1X	1.5X	2.6X
NGINX	1X	1.7X	3.5X
x.264 <sup>2</sup>	1X	1.7X	2.25X
Cassandra	1X	1.1X	1.8X

Cores	1200	1792	4864
Servers	15	14	38

### 데이터센터 공간 효율성 (density)

- 동일 랙 기준 2배 이상 서버 운용
- 동일 서버 기준 2배 이상 CPU 코어 운용

Use 2-3X Fewer Racks vs Legacy x86 for Equivalent Performance



Notes:

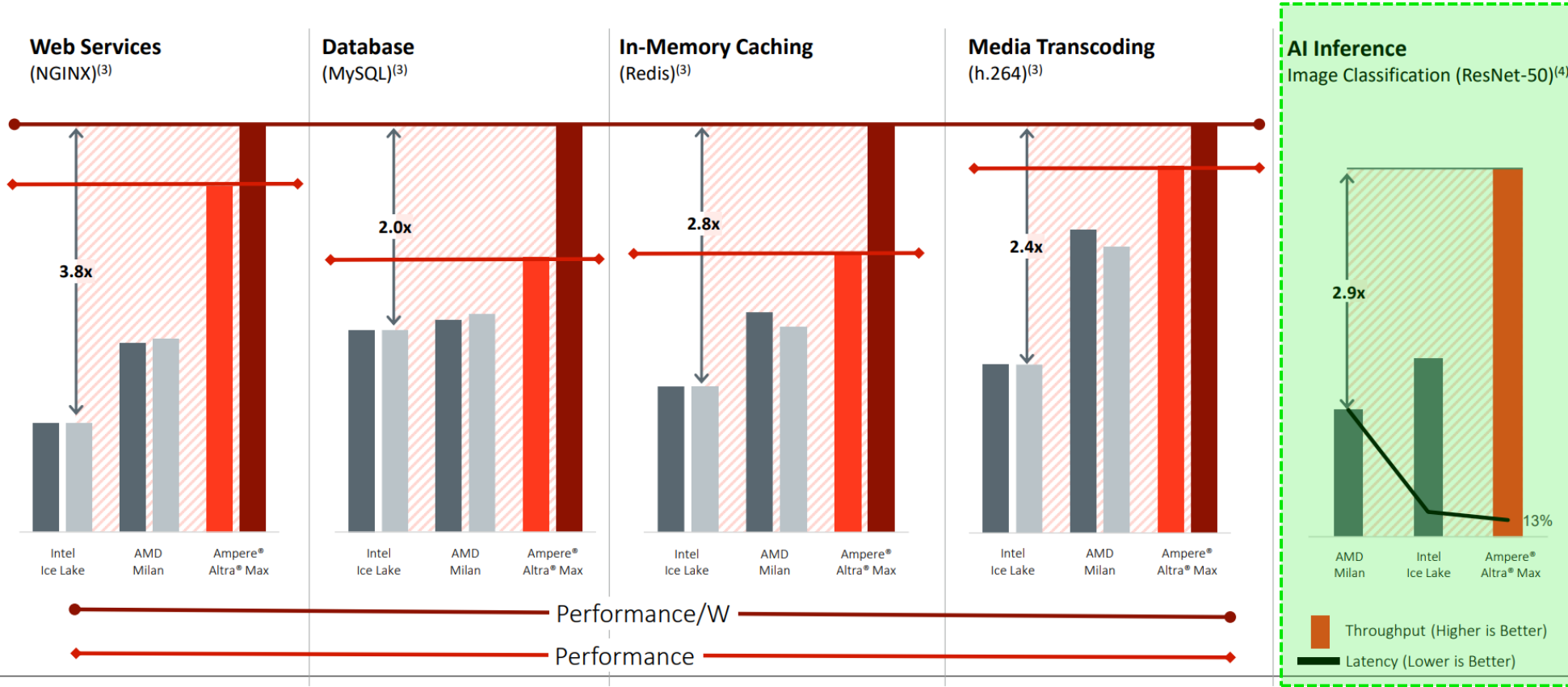
1. Ampere internal models and analysis to identify total compute performance and system usage power consumption numbers, in standard 42U 12.8kW rack, see end notes
2. Data point uses data taken on M128-30 whereas all other data points use the M128-26.



# [현재] TCO & ESG 달성을 위한 Green Computing

## Ampere: Leadership Performance for Cloud Workloads

Highest Performance and Power Efficiency Across Key Cloud Workloads<sup>(1)(2)</sup>



워크로드 별  
와트 당 성능

- 조건 : 멀티 코어 운용
- AI 서비스(추론) 운용에 최적화 된 솔루션



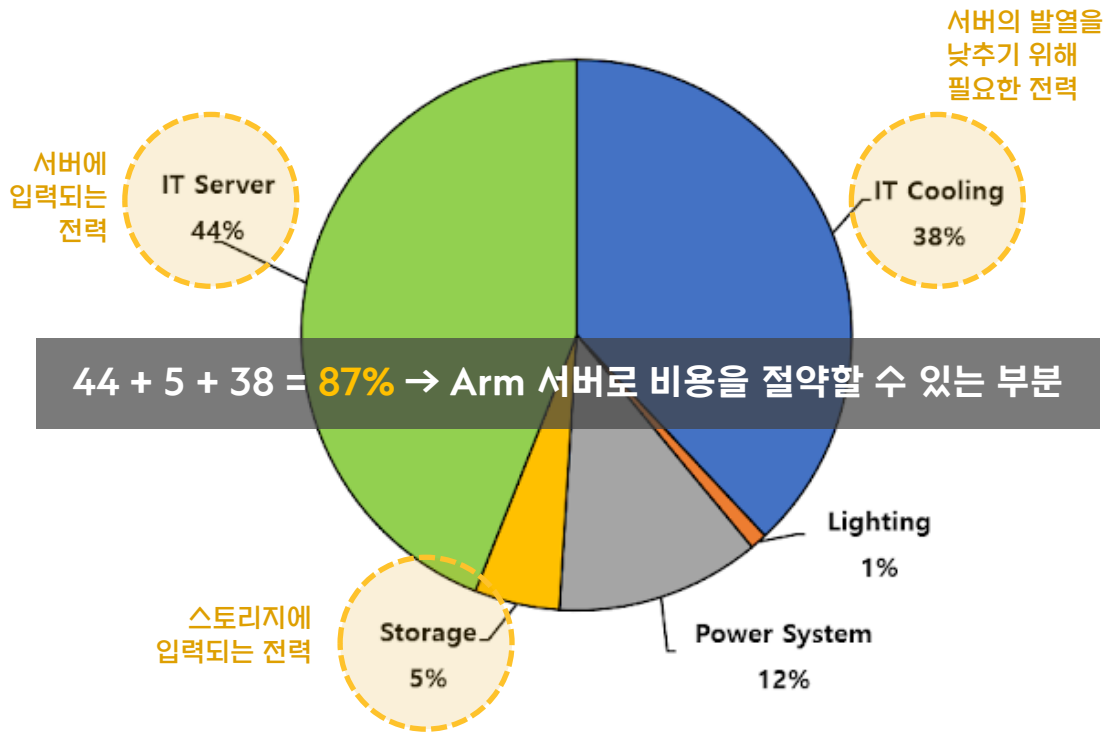
Notes:  
 1. Based on Company benchmarking  
 2. Intel Ice Lake represents Intel 8380 SKU; AMD Milan represents AMD 7763 SKU.  
 3. Percentages represent AMD Milan and Ampere® Altra® Max indexed against Intel Ice Lake  
 4. Percentages represent Intel Ice Lake and Ampere® Altra® Max indexed against AMD Milan



# [현재] TCO & ESG 달성을 위한 Green Computing

- 샘 올트먼 오픈AI는 “AI는 **예상을 뛰어넘는 전력을 소모할 것**” 이라고 경고
- 빌 게이츠 마이크로소프트(MS) 창업자도 “AI 데이터센터의 수익성을 결정하는 핵심 요소는 **전력이 될 것**” 이라며 전력 관리의 중요성을 강조한 바 있다.  
(출처 : ITBizNews, 2024.07.15)

데이터센터의 전형적인 에너지 소비 구성도

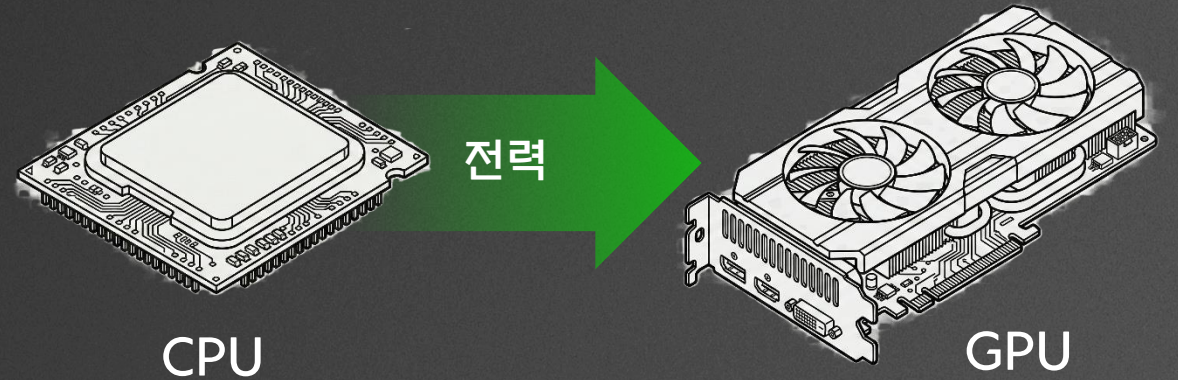


데이터센터의 경쟁력



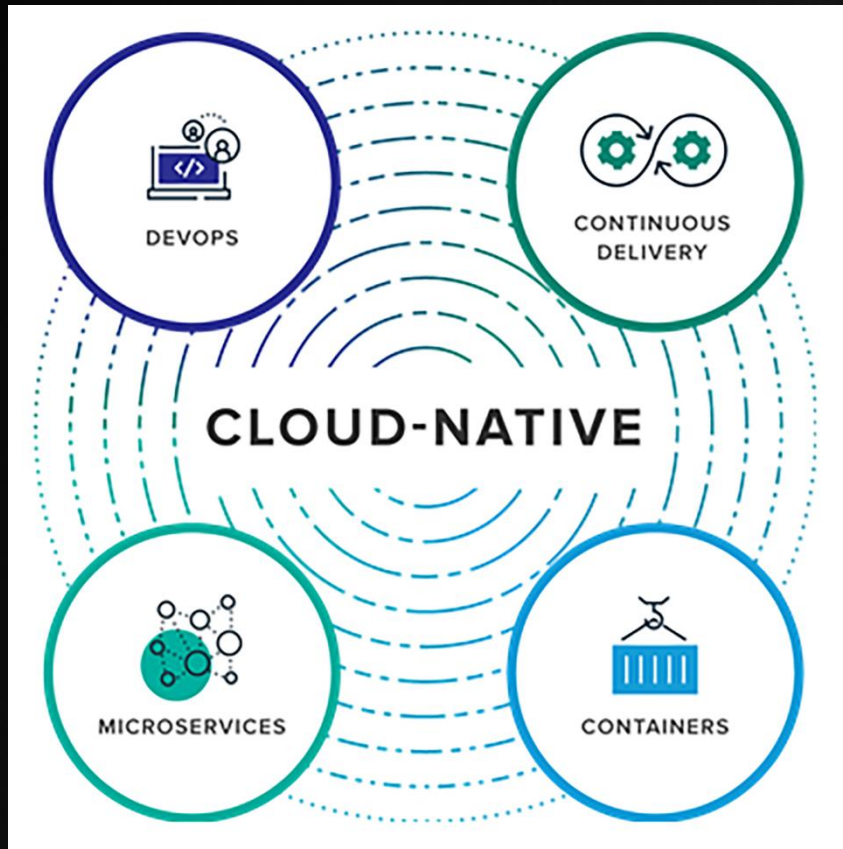
## 전력 예산의 재배치

Power Budget Reallocation



# [미래] 클라우드 네이티브와 AI 시대를 위한 인프라

## 현대적인 어플리케이션 구조에 가장 적합한 컴퓨팅 아키텍처



### MSA 와 컨테이너에 최적화 된 아키텍처

- Scale-Up 지향적인 x86 : 여러 개의 컨테이너가 한 코어에 몰릴 경우 Noisy Neighbor 현상  
→ 하이퍼스레딩으로 인한 성능 저하
- Scale-Out 지향적인 Arm : 각 컨테이너에 독립된 물리 코어를 1:1 로 매핑  
→ 항상 예측 가능한 성능 (Predictable Performance) 보장

### 압도적인 가성비 (Price / Performance)

- 동일 vCPU, 메모리 기준 Arm 인스턴스는 기존 x86 인스턴스 대비  
→ 비용은 약 20% 저렴, 성능(처리량)은 15 ~ 20% 높게 나옴

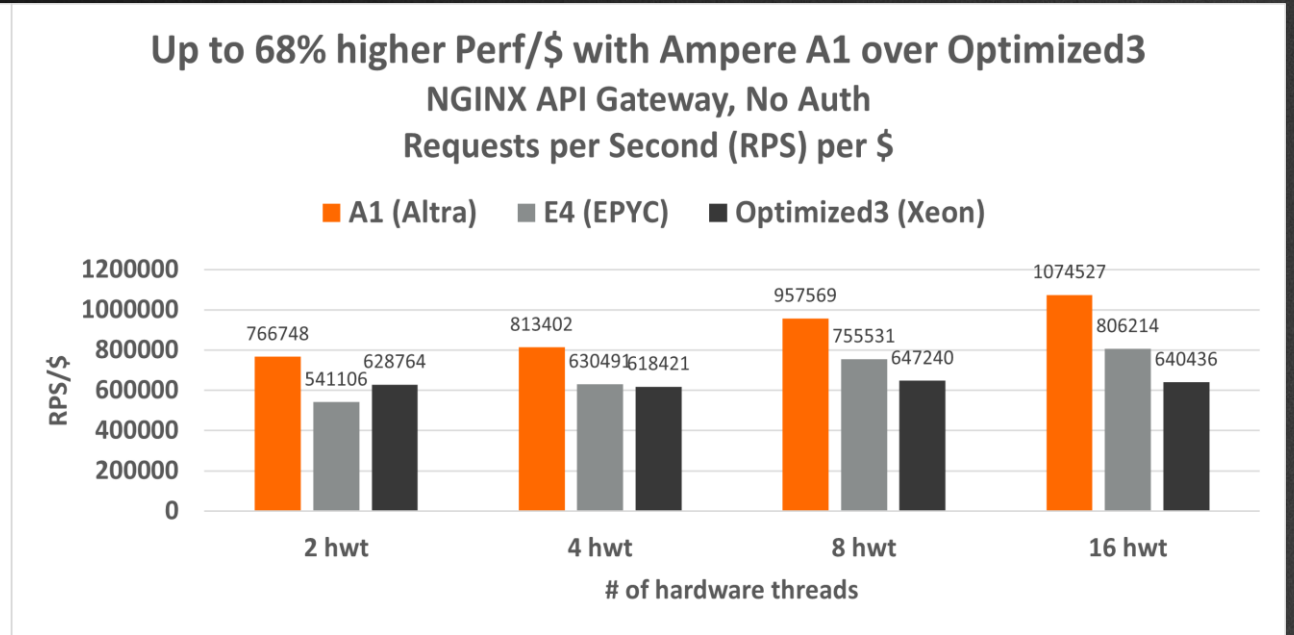
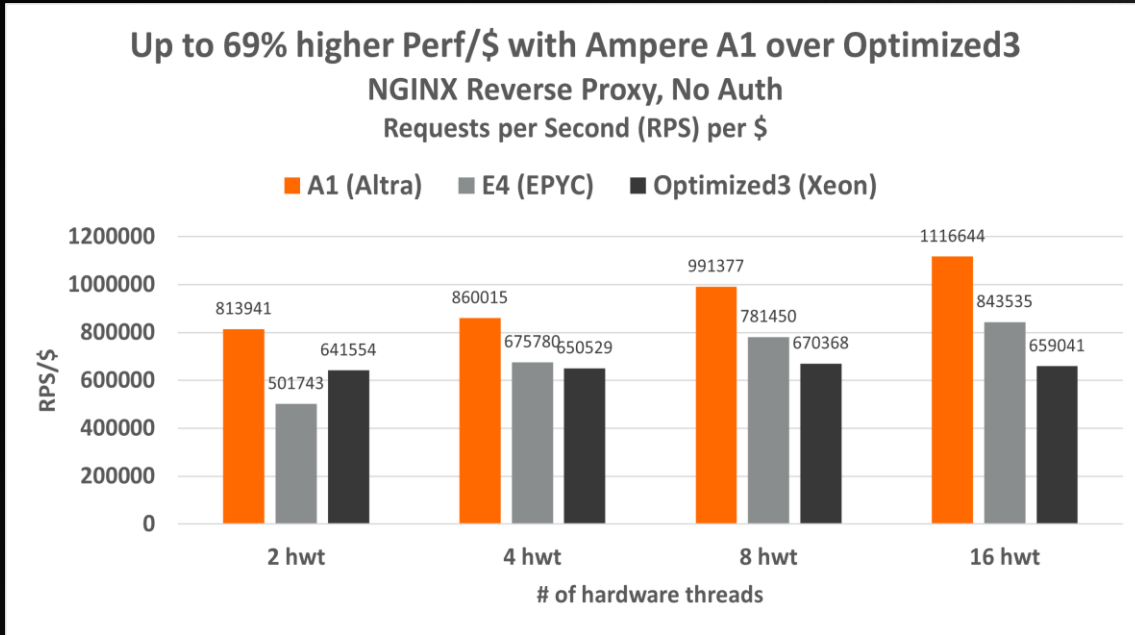
### 고집적도 & 전력 효율 (TCO & ESG)

### 이미 완성된 소프트웨어 생태계

- 운영체제, 개발 언어, CI/CD 등 클라우드 네이티브 생태계 전체가 Arm 을 완벽하게 지원

# [미래] 클라우드 네이티브와 AI 시대를 위한 인프라

## OCI 기반 Altra (Arm 서버) 워크로드 성능: No Auth 상태의 NGINX 가격 대비 성능 (가성비)



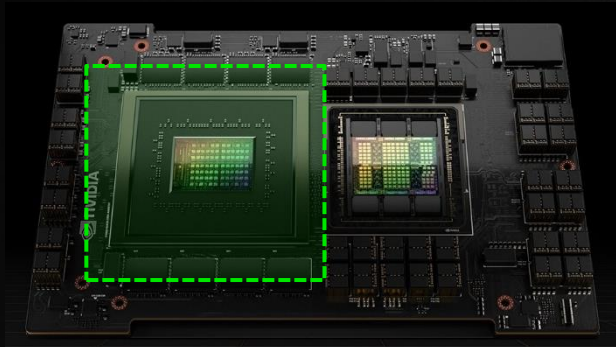
### 성능 평가 조건

- A1.Flex : Ampere® Altra® CPUs based on Arm Neoverse N1 cores, 3.0GHz
- E4.Flex: AMD EPYC third generation processors, 2.55GHz base, 3.5GHz single-core turbo
- Optimized3.Flex: latest third Gen Intel Xeon (Ice Lake) scalable processors, 3.0GHz base, 3.6GHz single-core turbo

### CPU 종류별 클라우드 서비스 가격 요금 비교

OCI virtual machine	Per OCPU per hour(\$)	Per GB memory(\$)
Intel.VM.Optimized3	0.054	0.0015
AMD.VM.E4	0.025	0.0015
Ampere.VM.A1	0.01	0.0015

# [미래] 클라우드 네이티브와 AI 시대를 위한 인프라



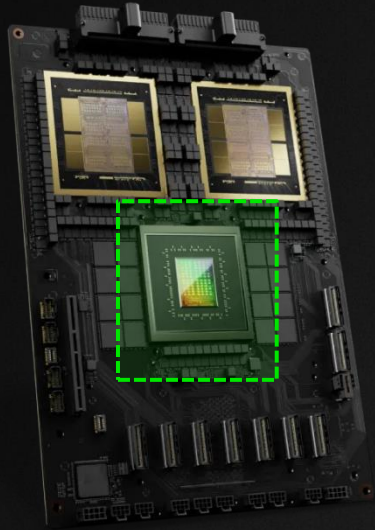
## AI 인프라에 Arm 서버를 적용해야 하는 이유

막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속

AI 추론 서비스의 컨테이너 인프라 확장성

피지컬 AI 분야의 학습 환경과 배포 환경의 일관성


랙 당 GPU 밀도 개선



# [Arm] 막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속

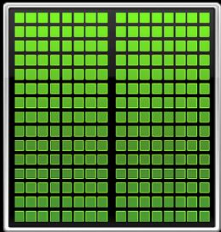
## Nvidia GPU architecture

**CPU**  
Optimized for Serial Tasks



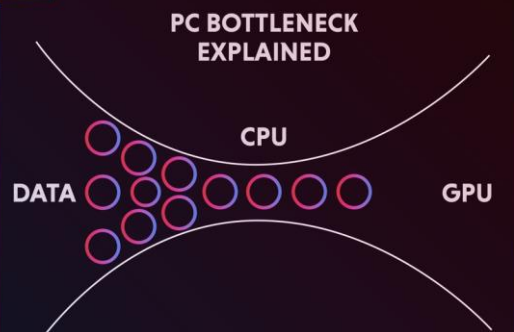
CPU

**GPU**  
Optimized for Many Parallel Tasks



GPU

### PC BOTTLENECK EXPLAINED

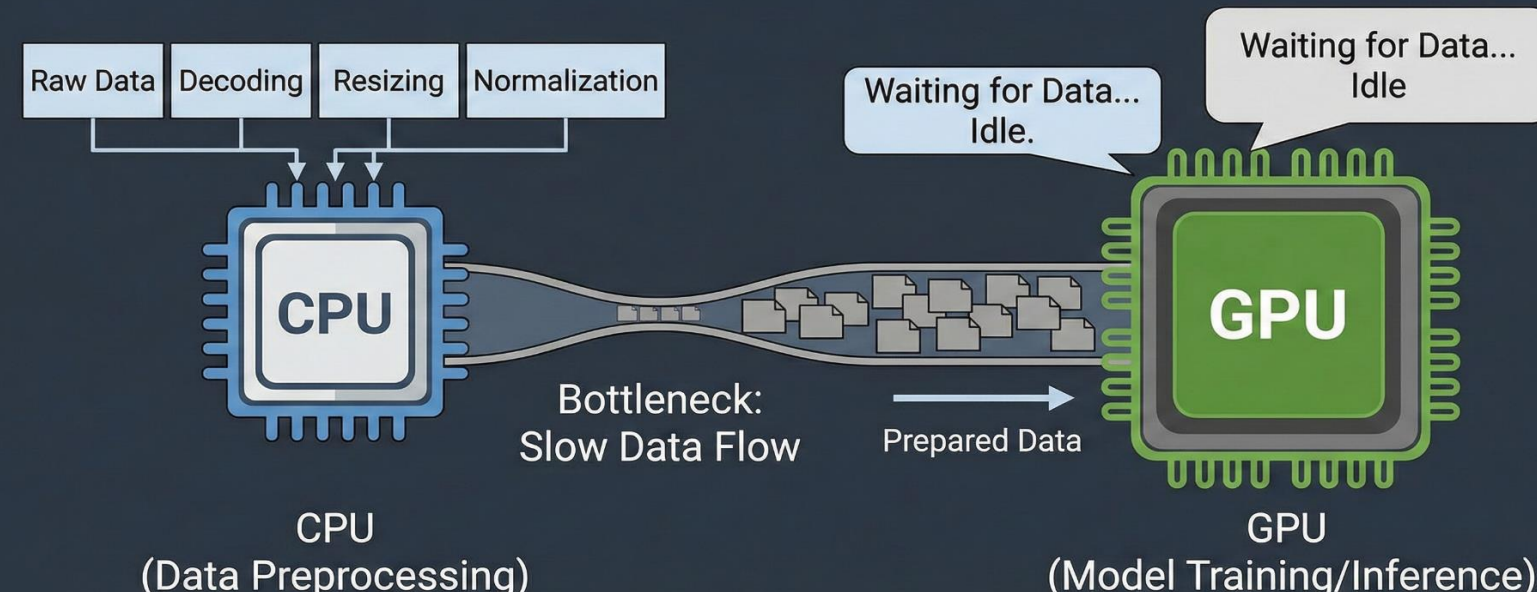


Bottleneck occurs when there is a limit on how much data can be processed by a computer. If there is differences in processing speed, bottleneck will happen. CPU bottleneck happens when the processor isn't fast enough to process/transfer the data received. The same concept applies to GPUs - only this happens if you pair a fast processor with a low level GPU.

#4

## 데이터 기아 (Data Starvation) 현상

### CPU Bottleneck in AI Server



Raw Data → Decoding → Resizing → Normalization

Waiting for Data... Idle.

Waiting for Data... Idle

**CPU**  
(Data Preprocessing)

**GPU**  
(Model Training/Inference)

Bottleneck: Slow Data Flow

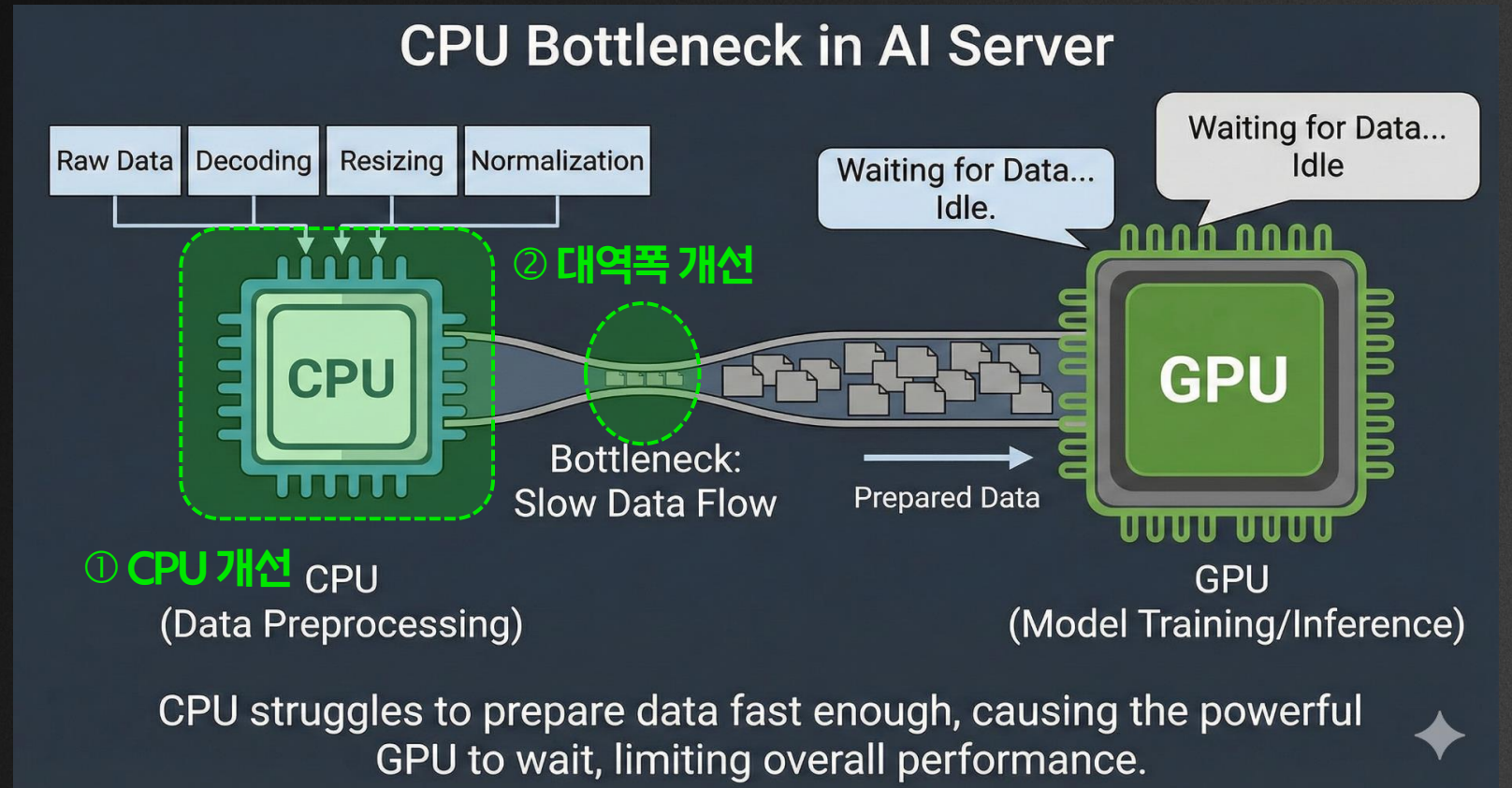
Prepared Data

CPU struggles to prepare data fast enough, causing the powerful GPU to wait, limiting overall performance.

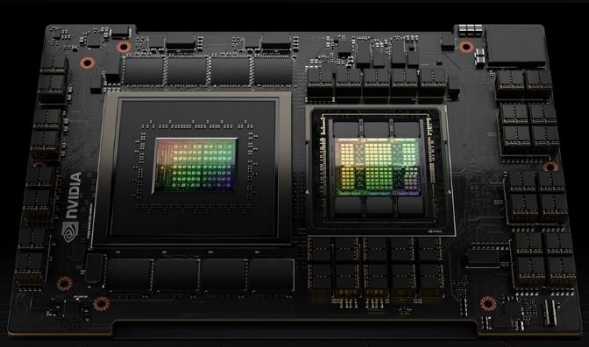
# [Arm] 막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속

- GPU 로 향하는 데이터 파이프라인과 I/O 레인을 효율적으로 관리하기 위해 **코어 수와 PCIe 레인이 압도적으로 많은** 서버용 프로세서가 필요
- 전력 대비 효율이 높고 **병렬 데이터 I/O 처리에 강점이 있는** 고성능 Arm 서버 아키텍처가 적합
- 대형 시뮬레이션과 AI 데이터 전처리에 Arm 서버가 활용되고 있음

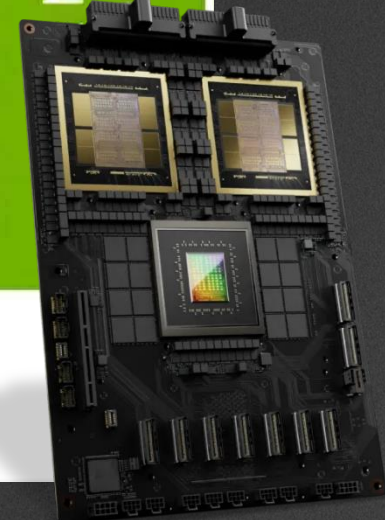
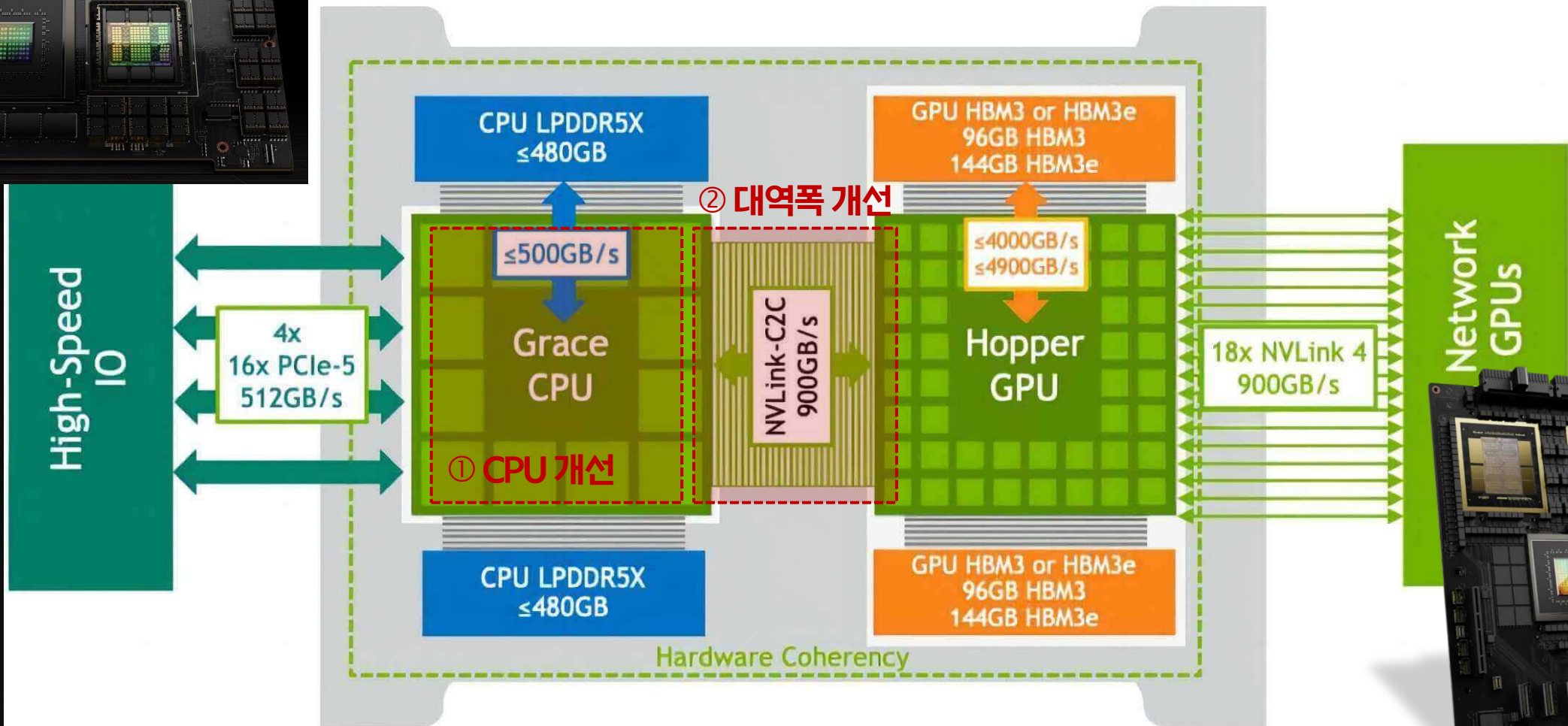
## 데이터 기아 (Data Starvation) 현상 → 해결 방안



# [Arm] 막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속



## NVIDIA GH200 Grace Hopper Superchip

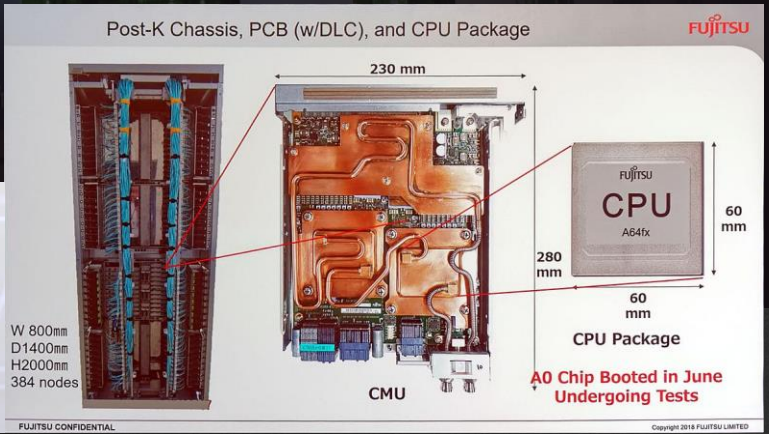
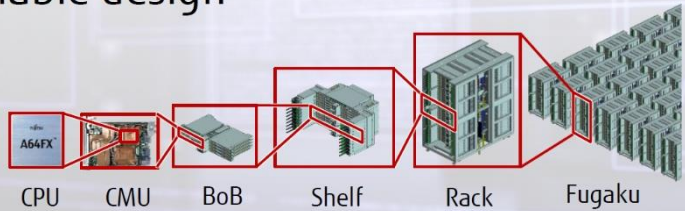


# [Arm] 막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속

한국 국가 슈퍼컴퓨터 6호기  
2026년 GH200 기반으로 구축될 예정



Scalable design



## TOP500 주요 순위

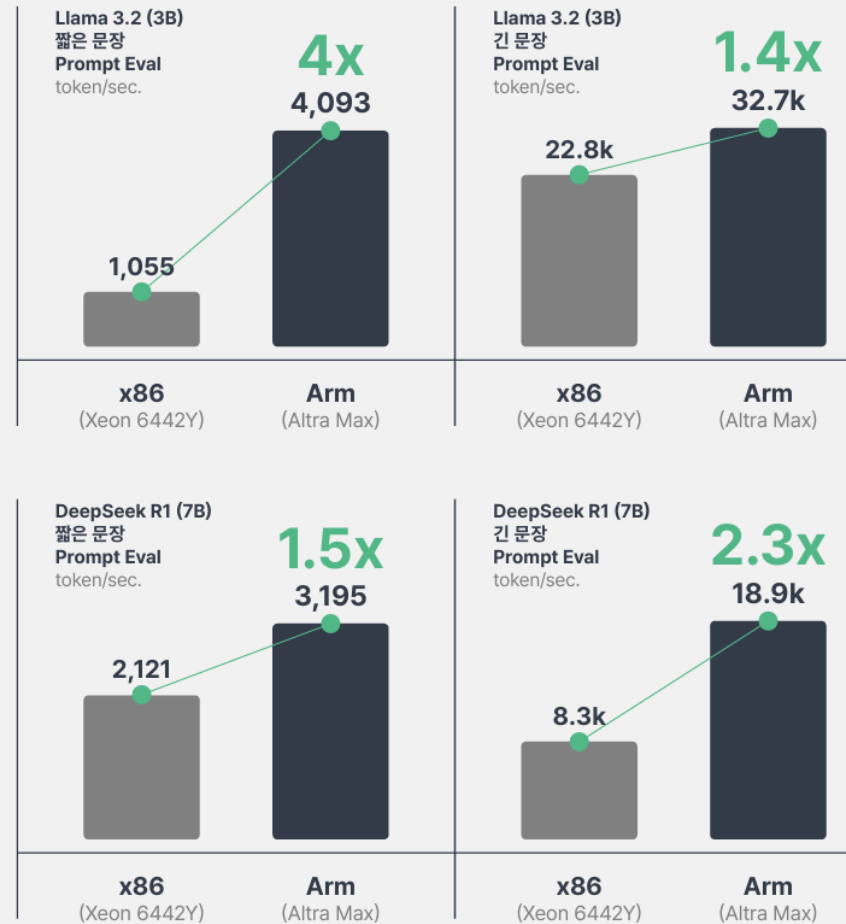
(2025년 6월 기준, 단위: PFLOPS)

순위	명칭	보유국가	보유기관	제조사	이론성능	실측성능
1(-)	El Capitan	미국	LLNL	HPE	2,746.38	1,742
2(-)	Frontier	미국	ORNL	HPE	2,055.72	1,353
3(-)	Aurora	미국	ANL	Intel	1,980.01	1,012
4(신규)	JUPITER Booster	독일	FZJ	EVIDEN	930	793.4
5(▼1)	Eagle	미국	Microsoft Azure	Microsoft Azure	846.84	561.2
6(▼1)	HPC6	이탈리아	Eni S.p.A.	HPE	606.97	477.9
7(▼1)	Supercomputer Fugaku	일본	RIKEN	Fujitsu	537.21	442.01
8(▼1)	Alps	스위스	CSCS	HPE	574.84	434.9
9(▼1)	LUMI	핀란드	CSC	HPE	531.51	379.7
10(▼1)	Leonardo	이탈리아	CINECA	EVIDEN	306.31	241.2
11(신규)	Isambard-AI phase 2	영국	Univ. of Bristol	HPE	278.58	216.5
12(▼2)	Tuolumne	미국	LLNL	HPE	288.88	208.1
13(신규)	ISEG2	네덜란드	Nebius	Nebius AI	338.49	202.4
14(▼3)	MareNostrum 5 ACC	스페인	BSC	EVIDEN	249.44	175.3
15(신규)	ABCI 3.0	일본	AIST	HPE	181.49	145.1
16(▼4)	Eos NVIDIA DGX SuperPOD	미국	NVIDIA	NVIDIA	188.65	121.4
17(신규)	Discovery 6	미국	ExxonMobil	HPE	222.86	118.6
18(신규)	SSC-24	대한민국	삼성전자	HPE	151.1	106.2
19(▼6)	Venado	미국	LANL	HPE	130.44	98.51
20(▼6)	Sierra	미국	LLNL	IBM/NVIDIA Mellanox	125.71	94.64

# [Arm] 막강한 병렬 처리 능력을 통한 '데이터 전처리' 가속

PERSONA AI		
항목	ARM 기반 서버 (효성 'GreenCore')	Intel 기반 서버 (일반 x86)
전력 효율성	● 우수 (저전력 설계)	○ 보통
코어 밀도	● 높음 (많은 코어 수)	○ 보통
성능 일관성	● 안정적	○ 가변적
확장성	● 유연함	○ 보통
호환성	○ 보통	● 높음

## 같은 GPU를 장착하고도 최대 4배 빠른 Arm 서버



### 테스트 서버 사양

#### Arm 서버

- Ampere® Altra® Max 128스레드, 250W
- NVIDIA® A40, 48GB

#### x86 서버

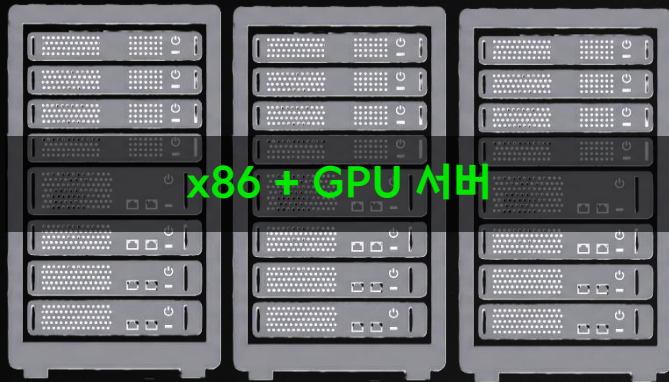
- intel® Xeon® 6442Y 96스레드, 450W
- NVIDIA® A40, 48GB

TDP는 낮지만 속도는 더 빠른 Arm 서버 프로세서

엑세스랩 내부 시험 결과에 따름

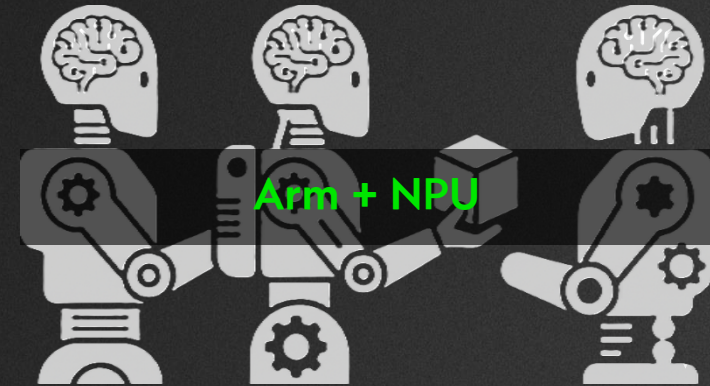
# [Arm] 피지컬 AI 분야의 학습 환경과 배포 환경의 일관성

피지컬 AI 학습 환경



피지컬 AI 모델 배포

피지컬 AI 환경



- 피지컬 AI 가 구동될 타겟 디바이스는 99% Arm 아키텍처
- 피지컬 AI 디바이스에 사용되는 Arm 칩은 보통 INT8 로 양자화 하여 연산
- 고성능 x86 서버와 GPU 는 FP32 또는 FP16 에 최적화
- FP32 / FP16 에서 INT8 로 변환하는 과정에서 정보 손실 발생 → 피지컬 AI 모델의 정확도가 급격히 떨어질 수 있는 이슈
- 피지컬 AI 디바이스에 사용되는 NPU 등의 호환성 문제로 피지컬 AI 디바이스의 CPU 로 강제 전환 (fallback) 되어 추론 성능 저하

# [Arm] 피지컬 AI 분야의 학습 환경과 배포 환경의 일관성

01.

## 대규모 병렬 디지털 트윈 시뮬레이션 최적화

- 피지컬 AI 는 현실 세계에 배치되기 전 가상환경에서 수많은 강화학습과 테스트를 위해 디지털 트윈 환경 필요
- 수많은 Arm 코어가 독립적인 시뮬레이션 환경을 스레드 간섭 없이 안정적으로 구동되어 대규모 병렬 시뮬레이션 처리량을 극대화

02.

## 대규모 센서 데이터 I/O 처리 및 GPU 병목 완화

- 피지컬 AI 는 고해상도 비전 카메라, LiDAR, 레이더 등 수많은 센서로부터 초당 테라바이트 급의 방대한 데이터를 수집
- Arm 서버 프로세서는 x86 대비 상대적으로 많은 수의 PCIe 레인을 제공
- 막대한 I/O 대역폭 및 매니 코어의 병렬 처리를 통해 외부에서 들어오는 대규모 센서 데이터를 병목 없이 전처리 과정을 거쳐 GPU 로 지연 없이 공급 가능

03.

## 예측 가능하고 일관된 워크로드 실행

- 피지컬 AI 는 지연시간(delay)에 매우 민감
- 로봇 제어나 자율주행의 판단은 예측된 시간 안에 반드시 실행되어야 함
- Arm 코어의 클럭 속도는 부하에 따라, 그리고 복잡한 스레드 스케줄링에 따라 크게 요동치지 않아 예측이 가능함
- 실시간 피지컬 AI 워크로드를 처리하는데 중요한 안정성을 제공

**Arm 서버는 수많은 독립된 코어를 통해 압도적인 시뮬레이션 처리 능력, I/O 를 통해 피지컬 AI 인프라의 백본에 매우 적합!**

# [Arm] 랙 당 GPU 밀도 개선

x86 350 ~ 500W

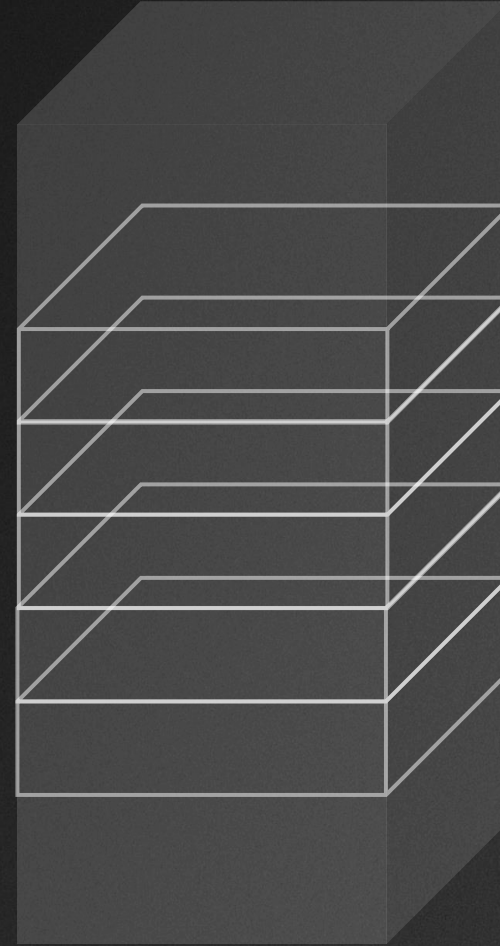
x86 350 ~ 500W

Arm 180 ~ 400W

700 ~ 1000W

GPU 서버 한대당  
520 ~ 600W 차이

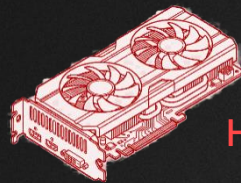
180 ~ 400W



랙 당 GPU 서버 5대씩 설치  
= 랙 당 H200 5개 추가 장착



데이터센터 당 H200 ???



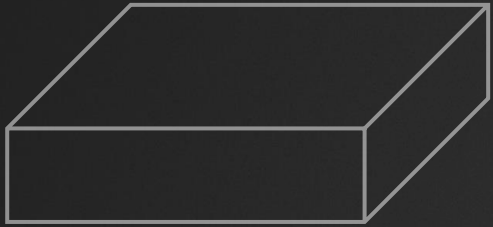
H200, RTX PRO 6000 = 600W

# [시작] 리스크 없는 단계적 전환

## 하이브리드 (Hybrid Architecture) 전략으로 단계적으로 전환

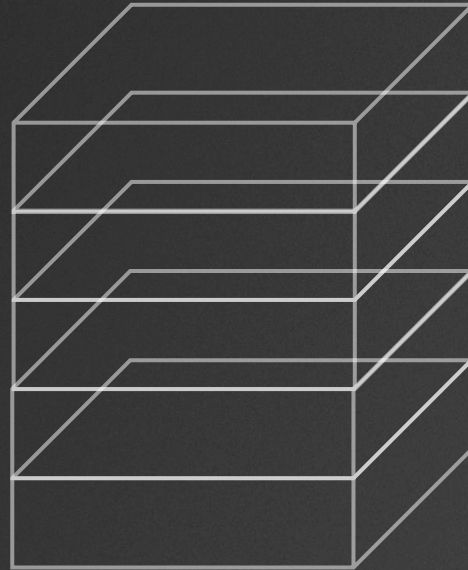
### 1단계

- 호환성 이슈가 없는 인프라부터 교체
- 비용 절감 효과 검증



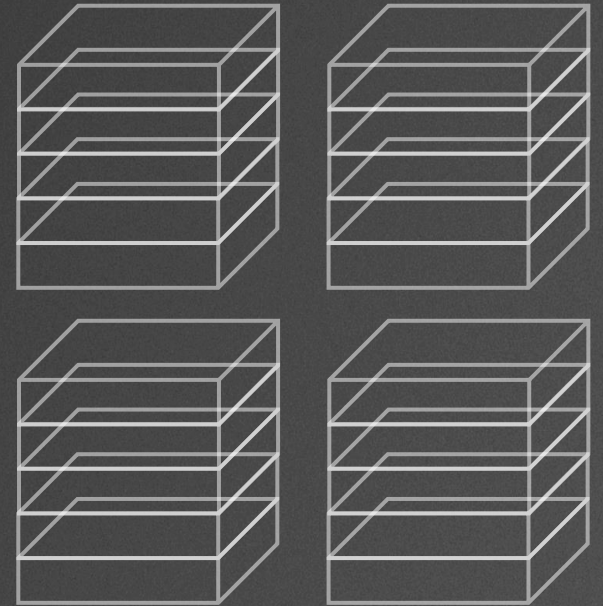
### 2단계

- 컨테이너 기반 (MSA) 서버로 확장 전환



### 3단계

- 실 서비스 및 핵심 코어 워크로드 전환



# [시작] 리스크 없는 단계적 전환

Arm 서버 적용 분야



# [시작] 리스크 없는 단계적 전환

이스트 시큐리티  
알약 리눅스 (보안)  
→ Arm 서버 공식 지원

EST SECURITY 기업    보안제품    보안솔루션    제품구매    보안정보    보안센터    고객지원    개인고객 | 회사소개 | 회사공고

리눅스/유닉스 서버를 위한 최적의 백신 솔루션  
**알약 리눅스/유닉스**

Endpoint Security  
**알약 리눅스/유닉스**

알약 리눅스/유닉스는 윈도우용 환경에서 이미 검증된 알약의 기술력을 바탕으로 기업 내 서버 환경에 침투한 악성코드를 검사, 치료하는 Linux 및 Unix 서버 전용 백신입니다.

#엔드포인트 강력보호 #악성코드 차단/치료 #리눅스/유닉스 환경 최적화

구매문의

Q. 엔드포인트 서버를 위협하는 악성코드가 증가하는 가운데, 무겁고 호환이 어려운 보안솔루션들로 도입을 주저하셨나요?

A. 리눅스/유닉스 환경에서도 알약! 최적화된 보안기능을 만나보세요.

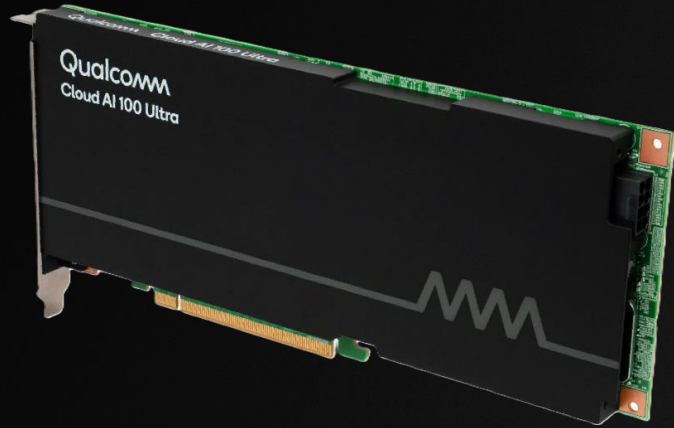
알약의 리눅스/유닉스는 검증된 알약의 기술력을 바탕으로 리눅스/유닉스 환경에서도 악성코드 탐지, 치료 기능을 효율적으로 제공합니다.

## 사용환경

## 알약 리눅스/유닉스 Specs

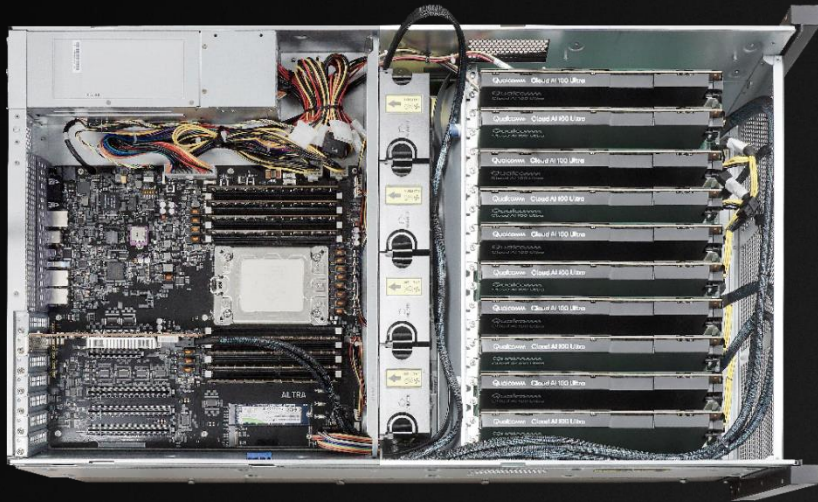
CPU	RAM
x86 (32bit), x64 (64bit) Dual Core 이상 Arm64 (AArch64) Dual Core 이상	1GB 이상

# [시작] 리스크 없는 단계적 전환



## 퀄컴 AI 반도체 → Qualcomm Cloud AI 100 Ultra

- TDP : 150W → 870 TOPS, 288 TFLOPS
- VRAM : 128GB → NPU 카드 1장으로 120B 구동 가능



```

Date & time: year 2026 month 02 day 01 - 17:27:58 --- LRT Image Version: LRT.AIC.13.2.1.20.4.0_LRT.AIC.REL
Device Selection: Scroll-Up-Down[
  o All-Devices
  o DeviceID-0
  o DeviceID-1
  o DeviceID-2
  o DeviceID-3
  o DeviceID-4
  o DeviceID-5
  o DeviceID-6
  o DeviceID-7
DeviceID-0 --- Status: Ready --- FW Version: 1.20.4.0
NSP Used / Total 0 / 16  NW Loaded / Active 0 / 0  DRAM Used / Total 253361 KB / 313917  DRAM Bandwidth 153232.00 KBps  NSP Frequency 595.20 Mhz
DDR Frequency 2133.00 Mhz  Temperature 44.00 C  Brd. Power / TDP 45.00 W / 150.00 W  Soc Power 10.00 W
DeviceID-1 --- Status: Ready --- FW Version: 1.20.4.0
NSP Used / Total 0 / 16  NW Loaded / Active 0 / 0  DRAM Used / Total 253361 KB / 313917  DRAM Bandwidth 154208.00 KBps  NSP Frequency 595.20 Mhz
DDR Frequency 2133.00 Mhz  Temperature 46.00 C  Brd. Power / TDP 46.00 W / 150.00 W  Soc Power 10.00 W
DeviceID-2 --- Status: Ready --- FW Version: 1.20.4.0
NSP Used / Total 0 / 16  NW Loaded / Active 0 / 0  DRAM Used / Total 253361 KB / 313917  DRAM Bandwidth 153232.00 KBps  NSP Frequency 595.20 Mhz
DDR Frequency 2133.00 Mhz  Temperature 47.00 C  Brd. Power / TDP 46.00 W / 150.00 W  Soc Power 9.00 W
DeviceID-3 --- Status: Ready --- FW Version: 1.20.4.0
NSP Used / Total 0 / 16  NW Loaded / Active 0 / 0  DRAM Used / Total 253361 KB / 313917  DRAM Bandwidth 153232.00 KBps  NSP Frequency 595.20 Mhz
  
```

# Green AX를 위한 협력

HS  호성인포메이션시스템

 GreenCore

X< XSLAB 액세스랩

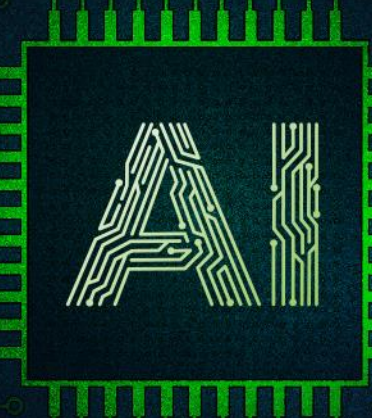
AI·데이터 인프라 솔루션 전문기업

40년 이상 IT 인프라 전 영역에  
전문 기술력 제공

15년 이상 국내 Arm 서버를 연구 개발한  
Arm 서버 전문 기업

다수의 소프트웨어 개발 인력,  
전문 기술 지원 엔지니어 보유

# Q&A



감사합니다.

