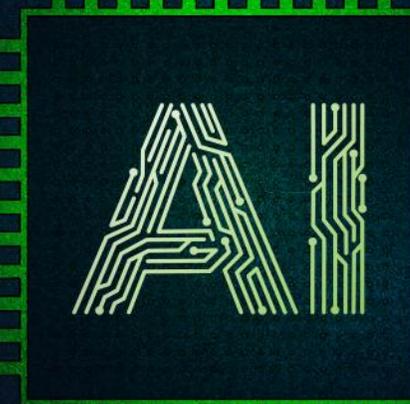


[디지털데일리 금융IT 이노베이션 컨퍼런스 2025]

금융 IT 혁신을 가속화하는 최신 AI인프라 트렌드 및 사례

정문중 매니저
HS호성인포메이션시스템
2025년 12월 11일

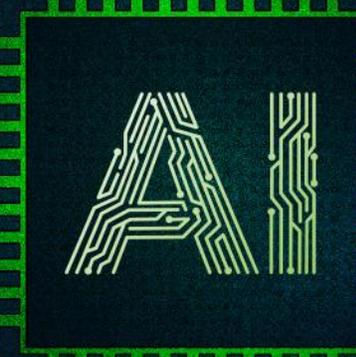


Agenda

1. AI 트렌드: 피지컬AI
2. 금융 산업별 피지컬 AI 전략
3. HS효성 AI플랫폼
4. 슈퍼마이크로
5. 금융 레퍼런스

1. AI트랜드: 피지컬 AI

1. 간부회동
2. 빠르게 성장하는 AI
3. 소버린AI, AI 3대 강국
4. 왜 한국인가: 피지컬 AI를 현실로 만드는 힘
5. GPU 26만장
6. 피지컬 AI로의 이동



깐부회동

I. 시트랜드: 피지컬시

- NVIDIA, HBM·모빌리티 동맹 등 AI 협력 논의



[사진: 뉴스1]

빠르게 성장하는 AI, 빅테크 기업의 과감한 투자

I. 시트랜드: 피지컬AI

빅테크 초대형 컴퓨터 투자 확대



오픈AI 오라클과 5년간 3,000억 달러
오픈AI AWS와 7년간 380억 달러
xAI 20만장 보유에서 최대 백만장 확대

데이터 주권·소버린 AI 필요성



온프레미스 AI 도입 급증세
데이터 주권 확보와 보안 강화
비용 예측 가능성 및 운영 효율 증가

PHYSICAL AI 전환 가속



스마트 팩토리 및 물류 자동화 고도화
엣지, IoT 장비 내장형 AI 탑재
사용자 친화적 AI 로봇 및 장비 보급 확산

소버린AI, AI 3대 강국

I. AI트랜드: 피지컬AI

- 2030년까지 고성능 GPU 5만장 확보를 통한 소버린 AI 구축 목표



2030년까지
5만장 확보



소버린AI, AI 3대 강국

I. AI트랜드: 피지컬AI

- 2030년까지 고성능 GPU 26만장 확보를 통한 소버린 AI 구축과 AI 3대 강국으로 도약



2030년까지
26만장 확보

AI 3대 강국

소버린 AI

- 2030년까지 고성능 GPU 26만장 확보를 통한 소버린 AI 구축과 AI 3대 강국으로 도약

GPU

Physical AI

피지컬AI는 제조 공장, 산업 환경, 로봇, 자동차 등
실제 물리 세계와 상호작용하는 AI를 의미

왜 한국인가: 피지컬 AI를 현실로 만드는 힘

I. AI트랜드: 피지컬AI

- 피지컬AI: 제조 공장, 산업 환경, 로봇, 자동차 등 실제 물리 세계와 상호작용하는 AI



소프트웨어 역량 우수

- ✓ 모델·클라우드·개발자 생태계
강해 두뇌(소프트웨어) 공급



소프트웨어 & 제조 역량 우수

- ✓ H/W+S/W의 신속한 양산 환경
(제조 라인+운영형 소프트웨어)
- ✓ 로봇·스마트팩토리·모빌리티·가전
디바이스형 AI등 피지컬 AI 제품화·
대중화에 최적화



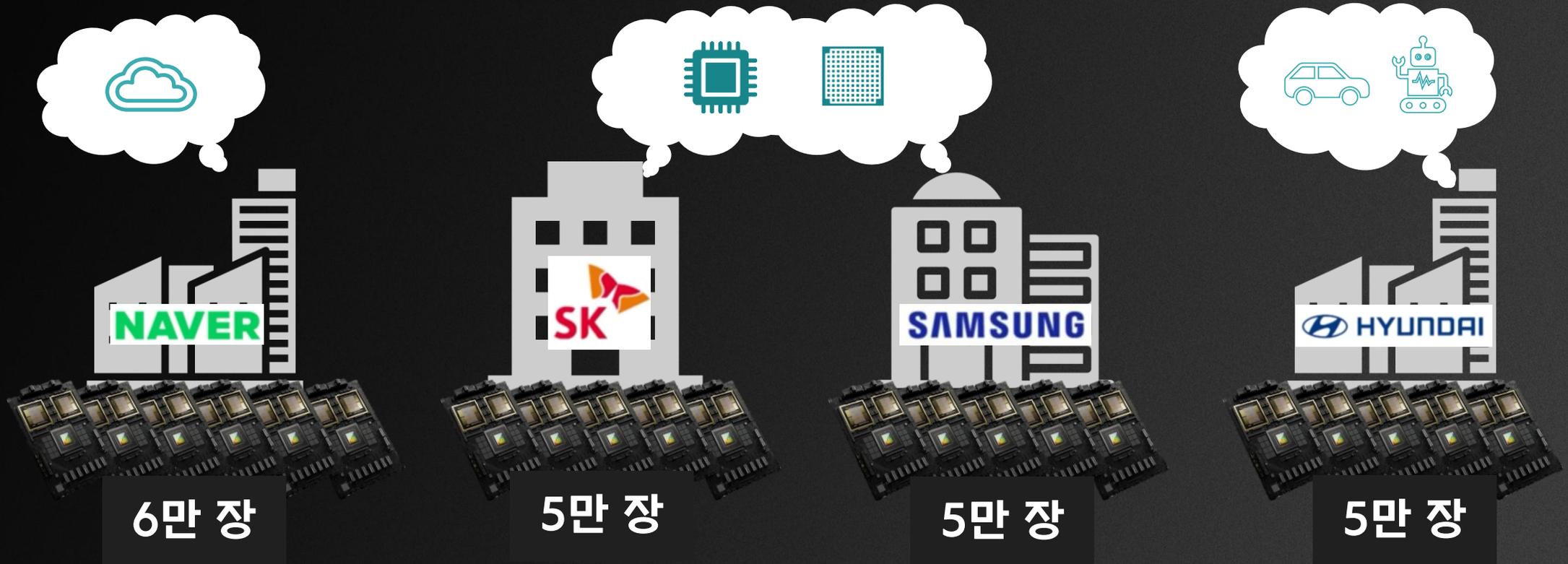
제조 역량 우수

- ✓ 기계·상업장비 표준이 강해
몸체(정밀 하드웨어)에 강점

GPU 26만장

I. 시트랜드: 피지컬시

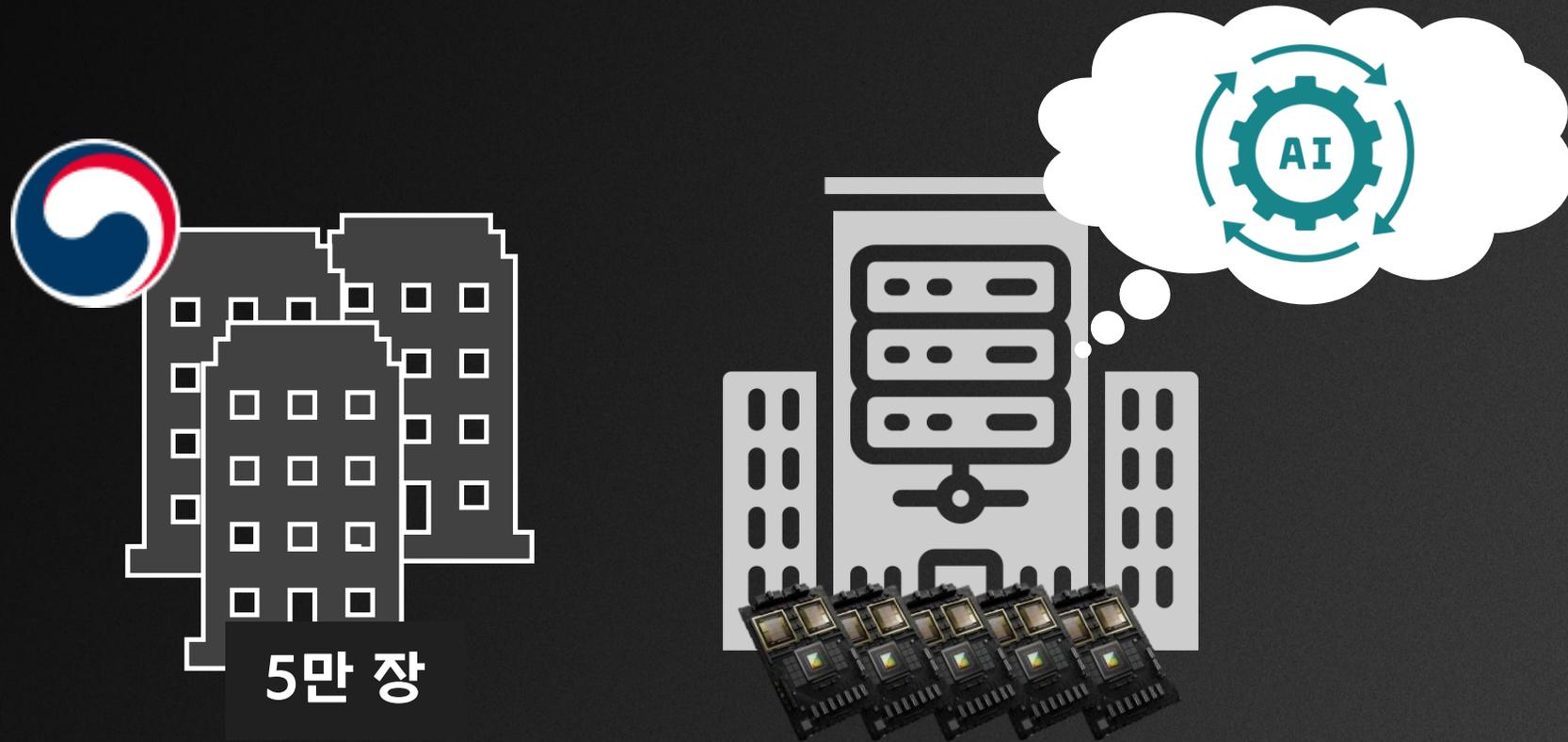
- 클라우드, 반도체 제조, 자율 주행 데이터 및 로봇



GPU 26만장

I. AI트랜드: 피지컬AI

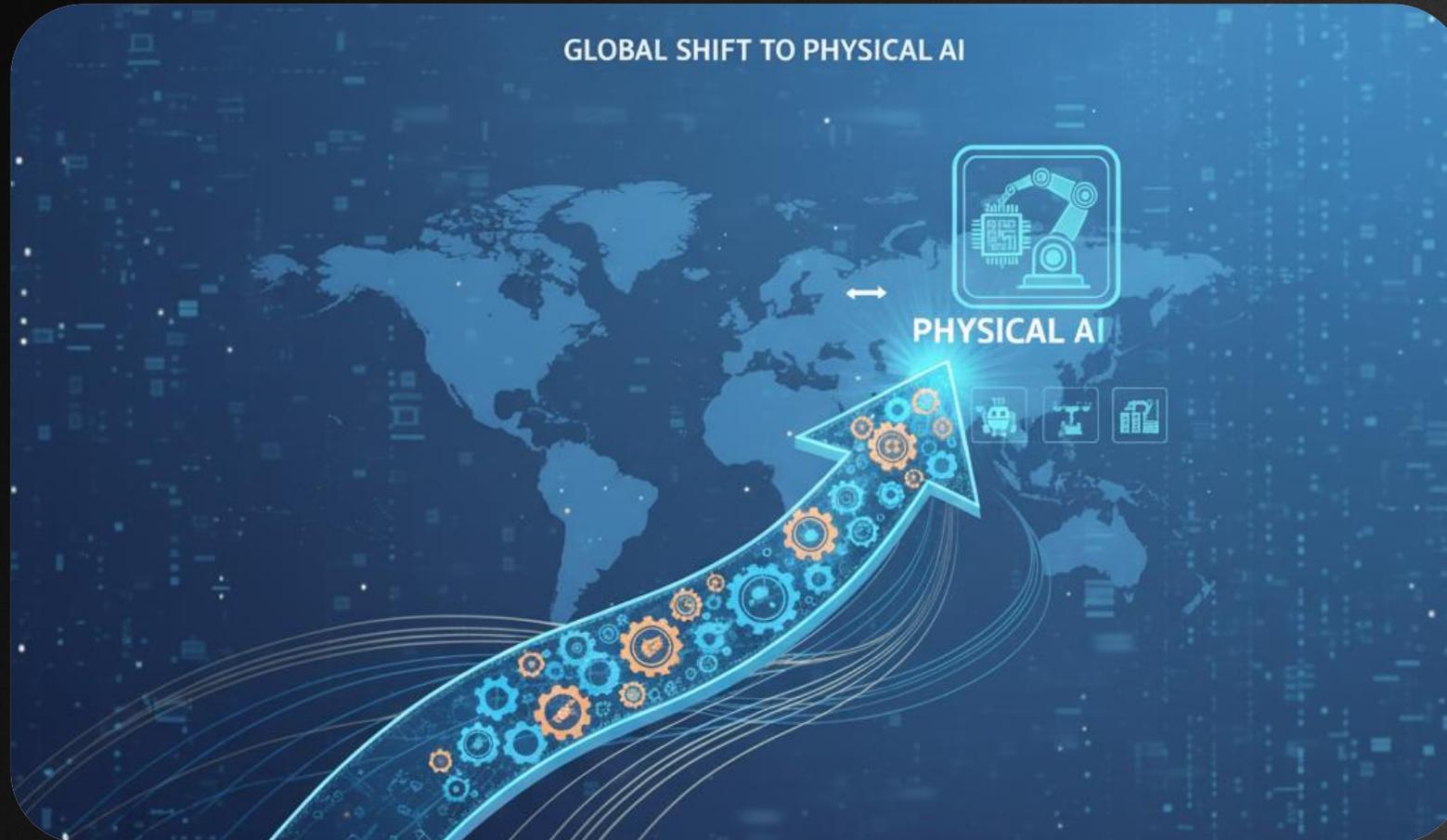
- 국가 AI 데이터센터 구축



피지컬 시로의 이동

I. 시트랜드: 피지컬 AI

- 스마트 지점의 로봇, 드론을 이용한 손해 사정, AI 기반의 물리적 보안 및 데이터센터 운영
- **피지컬 AI**는 금융 산업의 운영 효율성과 고객 경험을 근본적으로 재정의



스마트지점



드론 촬영



AI 보안 및 운영



고객 경험 재정의

피지컬 시로의 이동

I. 시트랜드: 피지컬시

- AI 시대의 혁신은 고성능 GPU 서버에서 시작되며, 슈퍼마이크로는 금융권 AI 인프라의 최적 파트너입니다.
- 최고의 성능·효율·공급 역량으로 AI 인프라를 안정적이고 강력하게 구축합니다.

Power your AI Factories with Ultra Performance
- Supermicro Systems -



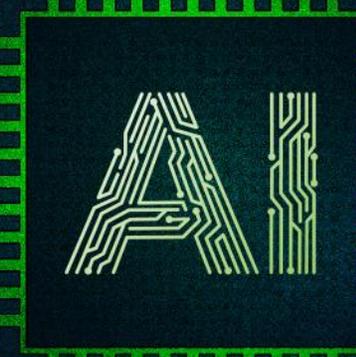
1. 업계 최고 수준의 성능

2. 에너지 효율성

3. FTM(First to Market) 전략

II. 금융 산업별 AI전략

1. 공통 전략
2. 산업별 피지컬 AI 전략
-은행, 보험, 증권, 카드, 캐피탈
3. AI 데이터센터



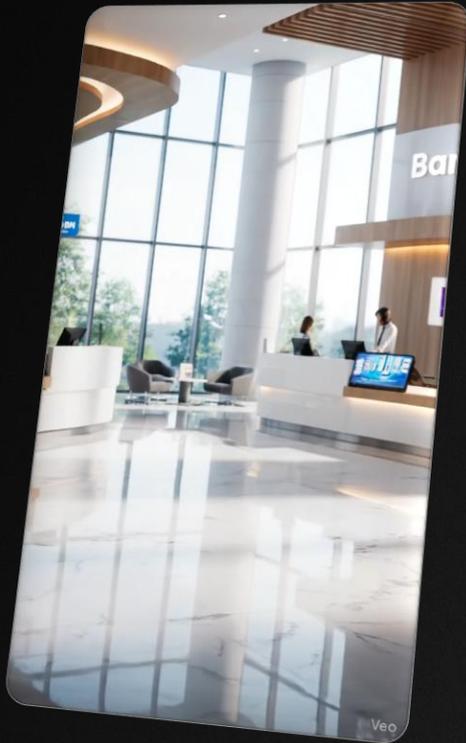
강력한 컴퓨팅 인프라 확보

AI 모델의 개발, 학습, 추론,
피지컬 AI가 수집하는
방대한 데이터 처리

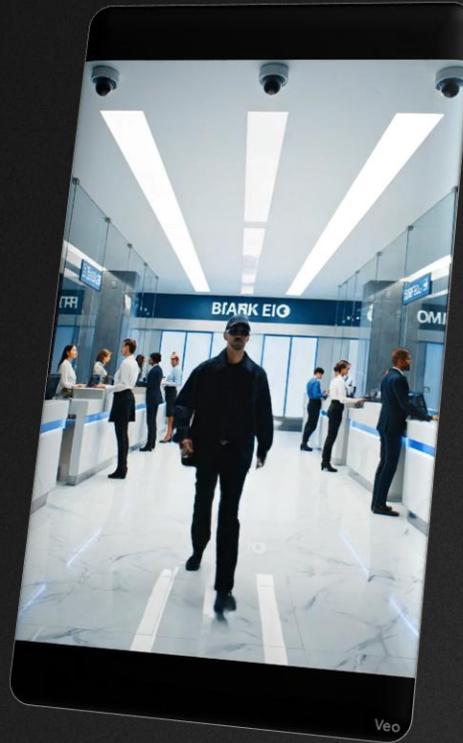
고성능, 고효율, 고확장성 인프라에 대한
선제적 투자가 필수



고객 경험의 물리적 자동화 및 지능화



스마트 브랜치

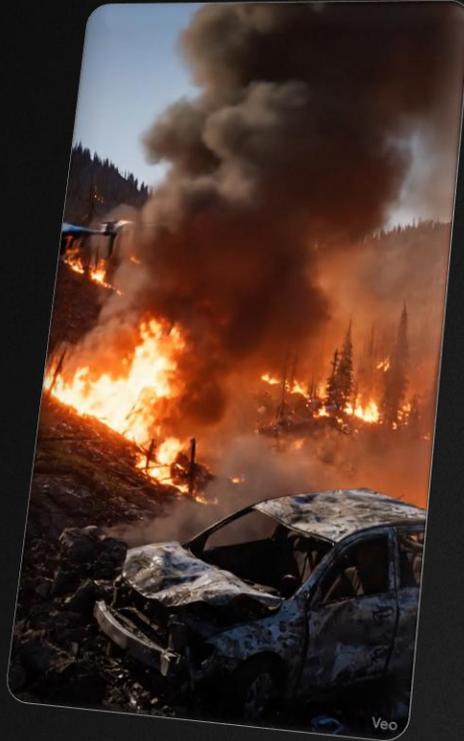


지능형 보안



데이터센터 자동화

물리적 손해 사정의 신속화 및 자동화



AI 손해 사정



텔레메틱스 고도화

초고속 연산 및 물리적 인프라의 극한 효율화



고성능 AI기반 HFT



데이터센터 운영 고도화

물리적 접점에서의 인증 및 사기 방지



물리적 AI FDS



생체 인증 결제

물리적 자산(차량/설비)의 가치 극대화 및 지능형 관리



AIoT 예지 정비



시비전 가치 산정



지능형 여신 심사

AI 데이터센터

고성능 컴퓨팅 (HPC)

Nvidia GPU /
AMD GPU, NPU 등

고성능 데이터 저장소

고성능 AI 분석을 위한
효율적인
고성능 스토리지

Resource 효율화 솔루션

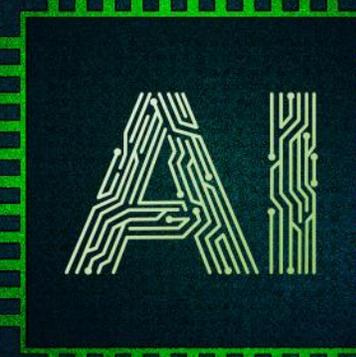
GPU 자원 효율을
높이기 위한
AIOps Stack

ESG를 위한 전력 효율화 솔루션

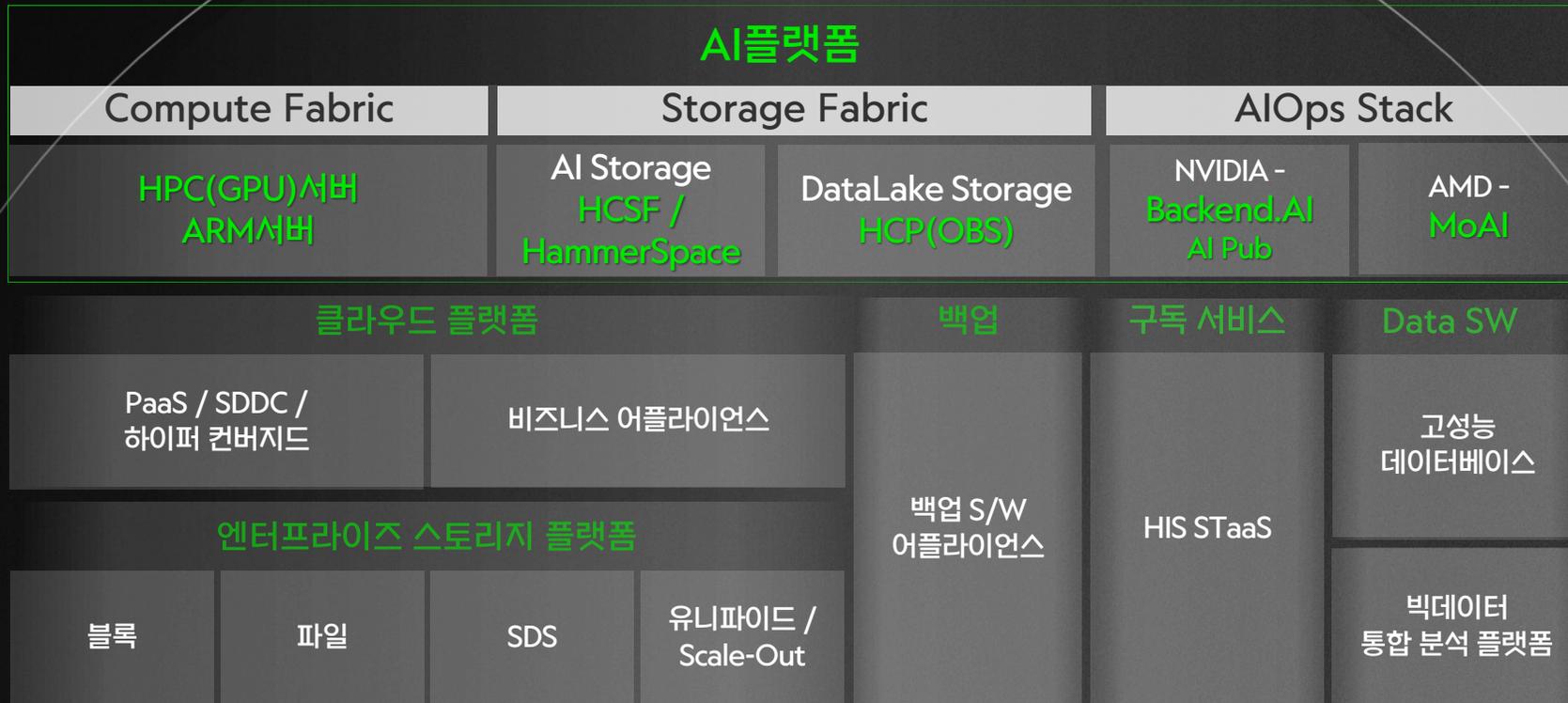
Arm 기반 친환경 서버
Liquid Cooling 냉각 방식

III. HS효성 AI플랫폼

1. HS효성 비즈니스 포트폴리오



AI인프라를 위한 시너지 강화



파트너 에코시스템 협업

AI인프라를 위한 시너지 강화



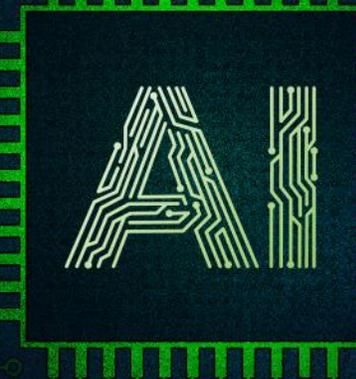
AI인프라를 위한 시너지 강화



파트너 에코시스템 협업

IV. 슈퍼마이크로

1. Supermicro Overview
2. Supermicro Liquid-Cooled Solution
3. Supermicro Blackwell GPU Systems



Supermicro Overview

- 가장 광범위한 애플리케이션에 최적화된 제품을 제공하는 업계 최고의 혁신 서버 전문 기업 Supermicro(슈퍼마이크로)



Revenue	\$33B+ (FY2026 Guidance) \$22B (FY2025), \$14.9B (FY2024)-Ranked #2 Server Market
Worldwide HQ	Silicon Valley (HQ)
Human Resource	7,000+ Headcount Worldwide, ~50% R&D staff
Corporate Growth	#1 in Generative AI and LLM Platforms 200%+ YoY Growth in Accel. Computing
Product Volume	6M+ Sq ft. Facilities Worldwide Silicon Valley (HQ), Taiwan, Netherlands, Malaysia and others \$50B/yr Production Capacity (FY25) Offering the Best AI, Cloud, Storage and 5G Technologies

Introduced world's first
NVIDIA-Certified Server Portfolio with
New NVIDIA Optimized GPU Systems



First to Market - Innovation

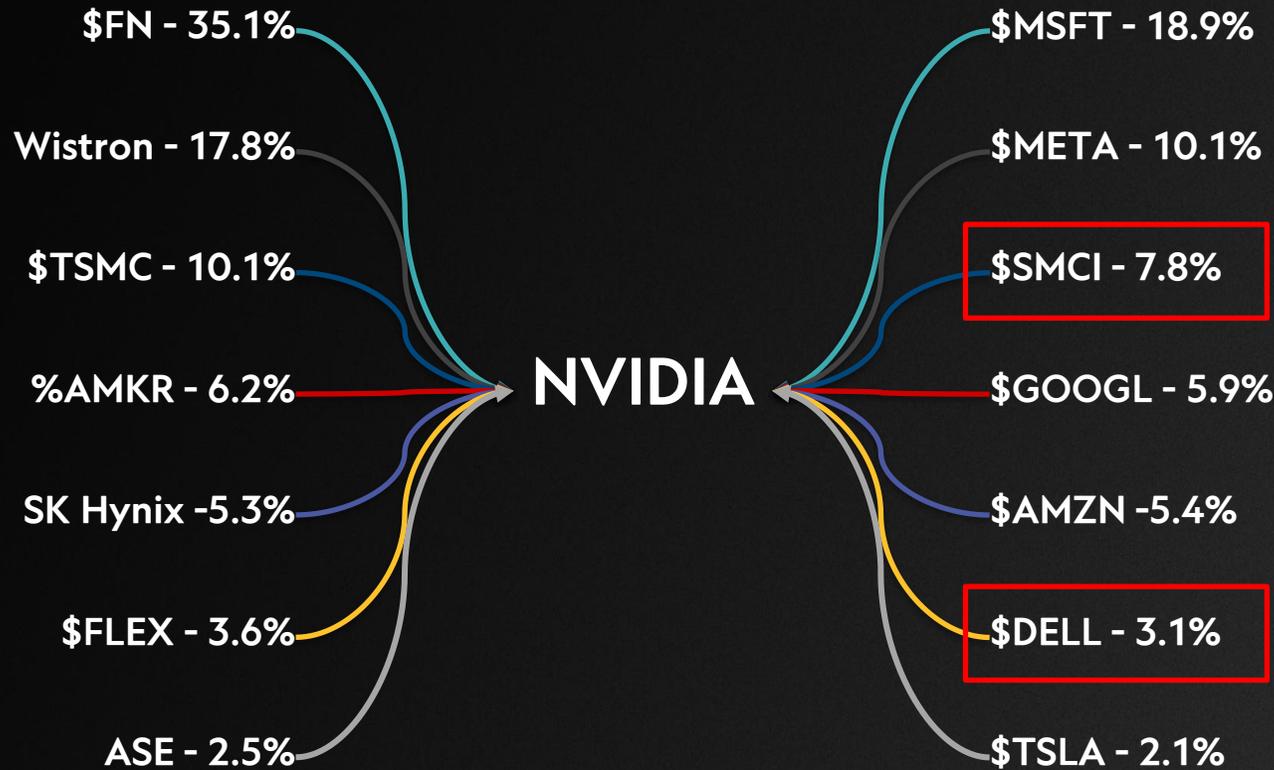
Total Solutions

Direct Liquid Cooling
(Keep IT Green)

US-Based Engineering and
Manufacturing

Supermicro Overview

- NVIDIA의 상위 3대 고객사 중 하나이며, 하드웨어 제조사 가운데 가장 높은 GPU 서버 시장 점유율을 보유
- AI 시장에서의 리더십을 기반으로 2024년 4분기 서버 시장 점유율 1위를 달성



Now ranked as the No. 1 Server vendor in the world*

Supermicro achieved this leadership position based on its revenue share

As per IDC's latest research, we have closed the gap

Source: IDC Worldwide Quarterly Server Tracker, 2024 Q4

*Ranking refers to IDC declared revenue share and takes in to account revenue of ODM Direct consumers.

Source: IDC Worldwide Quarterly Server Tracker, march 13, 2025.

Notes:

* IDC declares a statistics tie in the worldwide server market when there is a difference of one percent or less in the share of revenues among two or more vendors.

Supermicro Liquid-Cooled Solution

- 랙 스케일 솔루션을 통해 월간 공냉 5,000랙, 수냉 2,000랙 생산 가능
- Liquid-Cooled 시스템을 위한 Total Solution 공급과 사전 테스트 및 검증을 통한 안정적인 공급으로 구축 기간 단축



Liquid-Cooled Systems



CDU/Sidecar



Chilled Door



Cooling Manifold/
Cold Plates



Water/Dry
Cooling Tower



PnP Solution
Integration



Testing &
Validation

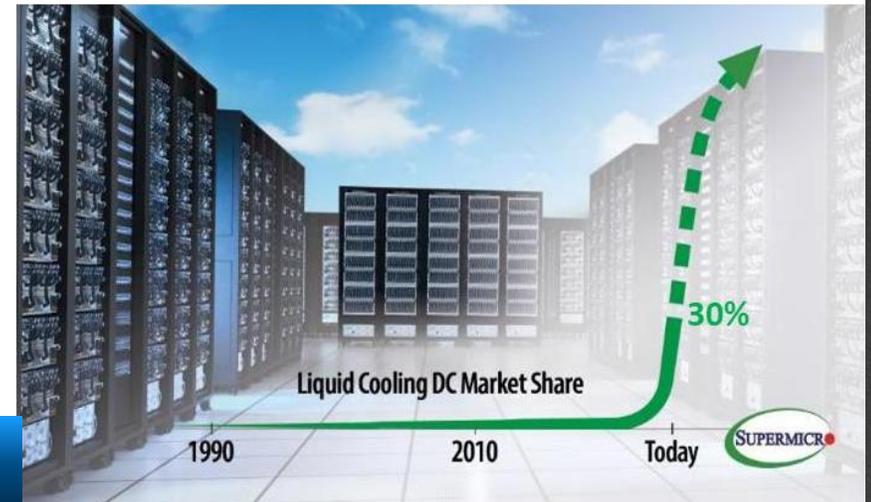


Onsite
Deployment



Mgt Software &
Services

5,000 Racks per month production NOW
2,000 DLC Racks per month
150kW (250, 500kW) Racks ship in volume
20MW Facilities operating today (**30MW** soon)

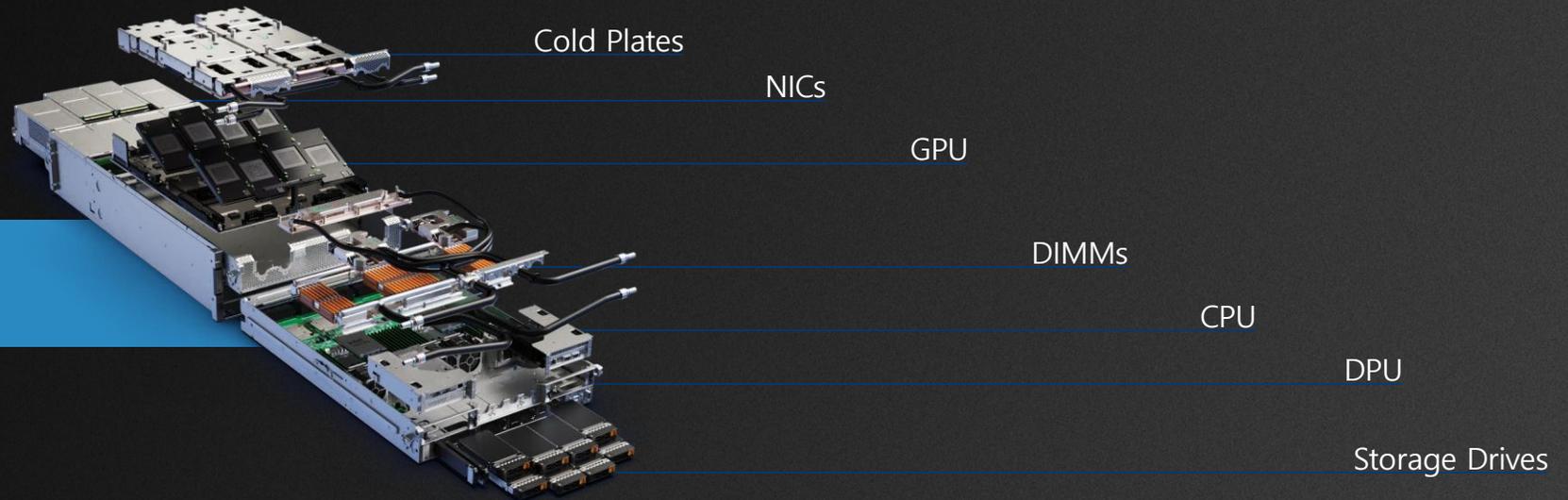
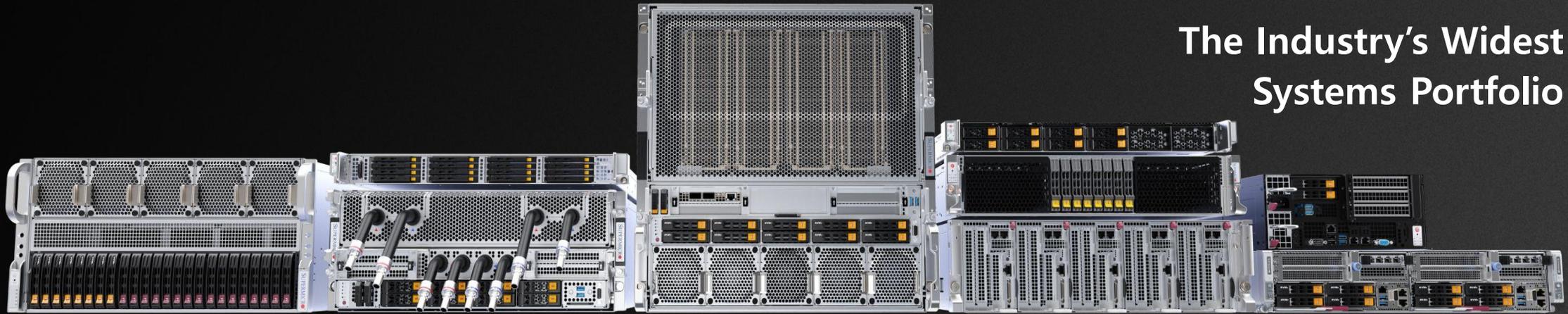


Powering the most powerful AI training/Inference cluster in the world -xAI, Coreweave and more

DCBBS System-Level Building Blocks

IV. 슈퍼마이크로

The Industry's Widest
Systems Portfolio



Component-Level
Customization

Supermicro Liquid-Cooled Solution_DLC-2

- Supermicro DLC-2 : 고성능과 에너지 효율을 동시에 실현하는 차세대 직접 수냉식 솔루션
- 서버, CDU 등 부품 일체를 Supermicro 직접 생산

구분	DLC (1.0)	DLC 2.0 (신규)
PUE	1.05 (목표)	
냉각 커버리지	CPU, GPU	CPU, GPU, Memory, PSU, VRM, PCIe switch등 (약98%)
유입 온도	30~35°C	최대 45°C
CDU 용량	랙당 100kW	랙당 250kW
소음	약 73dB	약 50dB
서버 밀도	표준	약 60% 상면 개선
냉각 시스템 관리	부분적 통합	SuperCloud Composer 완전 통합 관리
전력 절감	10~20% 전력 절감	최대 40% 전력 절감
냉각용수 절감	-	최대 40% 액체 절감



*CDU: Coolant Distribution Unit, 수냉식 시스템에서 냉각수를 관리하고 분배하는 핵심 장치

NVIDIA DGX B200

AI Center of Excellence의 기반입니다.

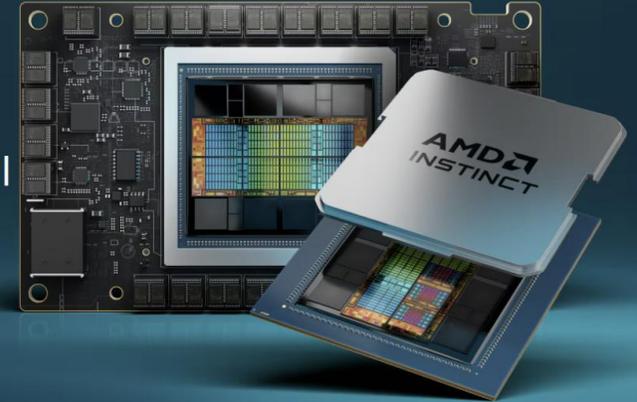
CUDA



AMD Instinct™ MI300 시리즈 가속기

최고의 생성형 AI 가속기 및 데이터 센터 APU

ROCm



인텔® Gaudi® AI 가속기

인텔® Gaudi AI 가속기 및 인텔® Gaudi® 소프트웨어는 데이터 센터 생성형 AI에 새로운 차원의 생산성 이점과 선택권을 제공하도록 설계되었습니다.

oneAPI



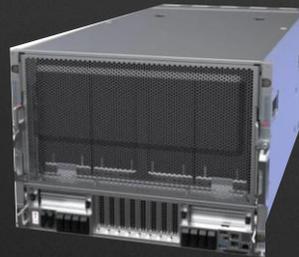
현대 데이터 센터를 위한 딥 러닝 최적화 서버

SYS-A22GA-NBRT



DP Intel 10U System with NVIDIA HGX B200 8-GPU

SYS-822GS-NBRT



DP Intel 8U System with NVIDIA HGX B200 8-GPU supporting 32 DIMMs

AS-A126GS-TNBR



DP AMD 10U System with NVIDIA HGX B200 8-GPU

AS-8126GS-TNMR



DP AMD 8U System with AMD Instinct™ MI350X 8-GPU

Max GPU	Up to 8 onboard GPUs	Up to 8 onboard GPUs	Up to 8 onboard GPUs	Up to 8 onboard GPUs
Supported GPUs	NVIDIA SXM: HGX B200 8-GPU (180GB)	NVIDIA SXM: HGX B200 8-GPU (180GB)	NVIDIA SXM: HGX B200 8-GPU (180GB)	AMD: Instinct™ MI350X Accelerators, Instinct™ MI325X Accelerators
GPU-GPU	NVIDIA® NVLink® with NVSwitch™	NVIDIA® NVLink® with NVSwitch™	NVIDIA® NVLink® with NVSwitch™	AMD Infinity Fabric™ Link
CPU	2	2	2	2
CPU Type	Intel® Xeon® 6900 series processors with P-cores	Intel® Xeon® 6700 series processors with P-cores	Dual AMD EPYC™ 9005/9004 Series Processors	Dual AMD EPYC™ 9005/9004 Series Processors
DIM Slots	24	32	24	24
Drive Size	2.5"	E1.S	2.5"	2.5
Drives	10	8	10	8
Networking	2x 10G	2x 10G	2x 10G, AIOM	2x 10G

Supermicro B200 GPU Systems

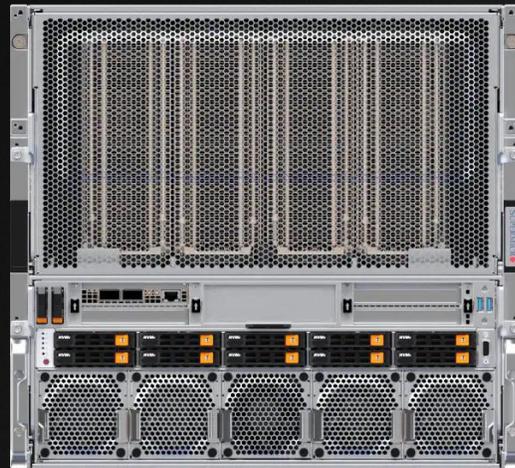
Air-Cooled System, Evolved

랙에 최대 32개
GPU(공랭식)

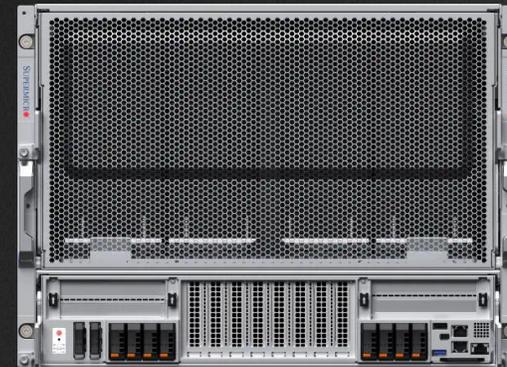
1.8TB/s GPU-GPU
상호 연결
GPU당 180GB
HBM3e

듀얼 Intel Xeon 6
또는 AMD EPYC
9005/9004 프로세서

차세대 공랭식 방열판
및 서비스 용이성



대규모 언어 모델 훈련 및 대용량 추론을
위한 후면 I/O 공랭식 시스템



향상된 시스템 메모리 구성 유연성 및
냉간 통로 서비스 용이성을 갖춘 전면
I/O 공랭식 시스템

Supermicro B300 GPU Systems

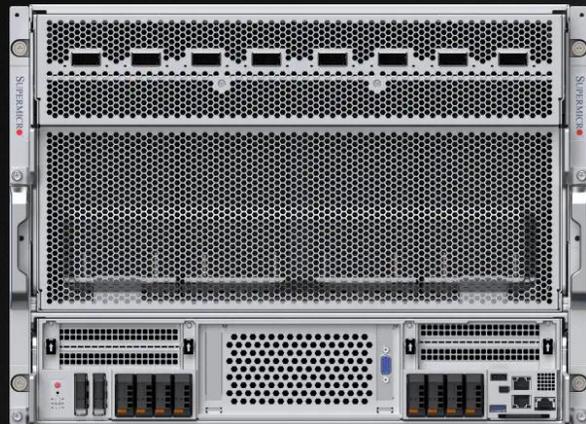
Ultra Performance for AI Reasoning

랙에 최대 64개
(수냉식) GPU

1.8TB/s GPU-GPU
상호 연결
GPU당 288GB
HBM3e

듀얼 Intel Xeon 6
또는 AMD EPYC
9005/9004 프로세서

800Gb/s 컴퓨팅 패브
릭을 위한 통합
NVIDIA ConnectX-8
SuperNIC



NVIDIA ConneX-8 네트워킹이 통합된
NVIDIA Blackwell Ultra B300 8-GPU용
I/O 공랭식 시스템



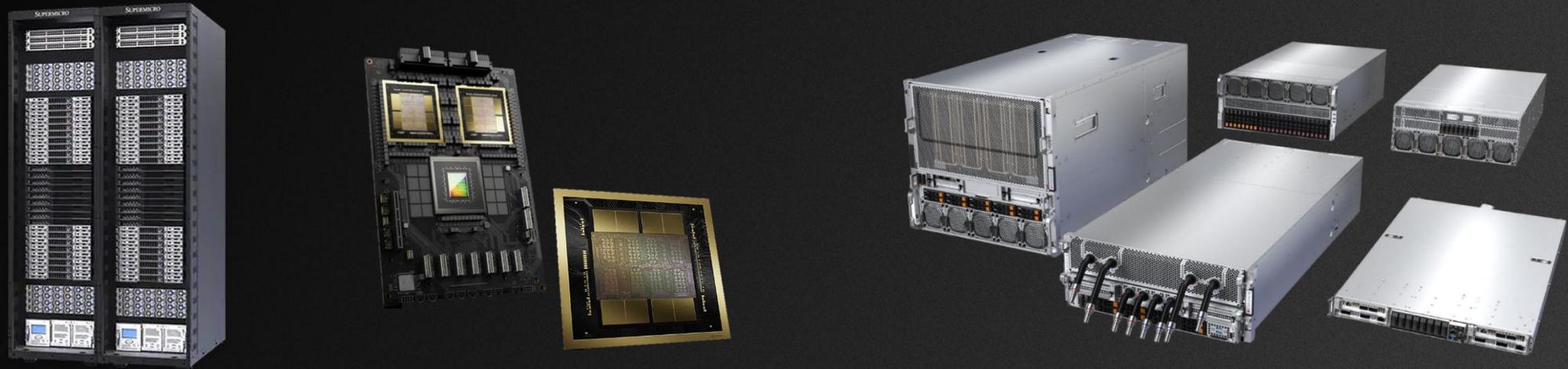
NVIDIA ConnectX-8 800Gb/s 네트워킹이
통합된 전면 I/O 수냉식 시스템



NVIDIA Blackwell Architecture Solutions

IV. 슈퍼마이크로

End-to-End AI Datacenter Building Block Solutions



방대한 경험

Supermicro AI데이터 센터
빌딩 블록 솔루션,
세계 최대 규모
액체 냉각 AI 데이터 센터 구축

유연한 제공

GPU 최적화 및
공기 또는 액체 냉각,
다중시스템 및 랙 폼 팩터,
CPU, 스토리지, 네트워킹 옵션 등 필요에
맞는 최적화 가능

액체 냉각의 선구자

AI혁명을 지속하기 위한
검증되고 확장 가능하며
플러그 앤 플레이 방식의
액체 냉각 솔루션

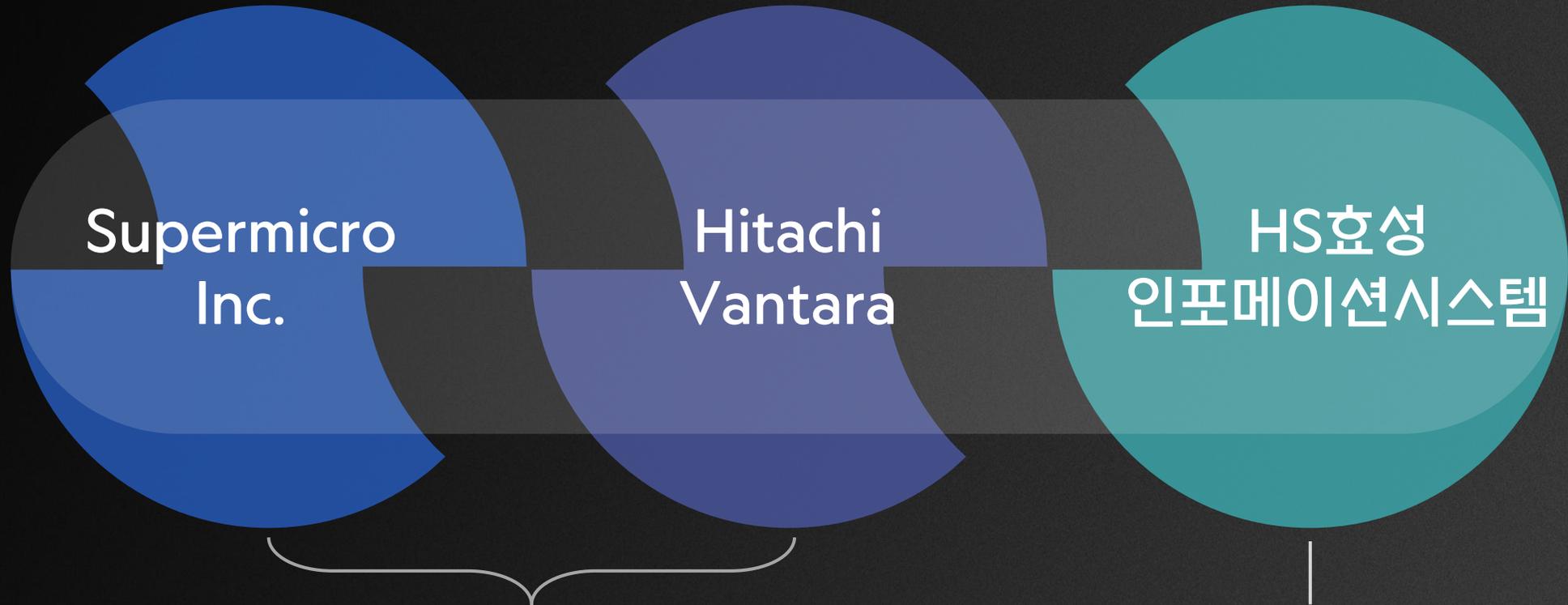
NVIDIA Blackwell 아키텍처를 위한 특별한 설계

빠른 온라인 시간

글로벌 역량을 통한
세계적 수준의 배포 전문성,
원스톱 서비스를 통해
AI를 빠르게 프로덕션에 적용

Supermicro + HS효성인포메이션시스템

- HS효성인포메이션시스템은 슈퍼마이크로 공식 총판사로서 GPU 및 x86 서버를 공급 합니다



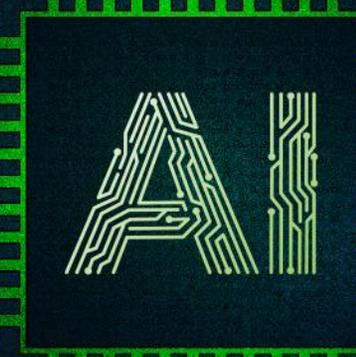
Supermicro의 GPU·AI 역량과 Hitachi Vantara의 VSP One 통합 스토리지 플랫폼, Hitachi iQ AI 포트폴리오를 결합하는 전략적 파트너십

Hitachi Vantara 전제품 국내 공급 Supermicro 14년 경력 전문가 기술지원 (전국 8개 사무소 및 조직)

V. 금융 레퍼런스

1. 도입 사례

- 은행, 보험, 증권, 카드
- A은행 구축 사례



도입 사례

IV. 금융 레퍼런스



도입 사례

IV. 금융 레퍼런스

은행

LLM 모델 운영

은행

안면인식(바이오 정보추출)

은행

STT/TA 고도화

은행

생성형 AI 시스템 구축

은행

오픈소스 생성형AI 모델

은행

개인화 마케팅 업무

은행

문서 분석 및 한글 임베딩

은행

생성형 AI 도입

은행

AI 개발 시스템

은행

검색용 LLM 구축

보험

LLM 파인튜닝

보험

AI-OCR 및 사내 개발용

보험

녹취 요약 STT 학습용

보험

AI-OCR 증설 & 추론

증권

검색용 LLM 구축

카드

FDS(이상거래탐지시스템)

카드

LLM 추론시스템

도입 사례

IV. 금융 레퍼런스

은행

LLM 모델 운영

은행

안면인식(바이오 정보추출)

은행

STT/TA 고도화

은행

생성형 AI 시스템 구축

은행

오픈소스 생성형AI 모델

은행

개인화 마케팅 업무

은행

문서 분석 및 한글 임베딩

은행

생성형 AI 도입

은행

AI 개발 시스템

은행

검색용 LLM 구축

보험

LLM 파인튜닝

보험

AI-OCR 및 사내 개발용

보험

녹취 요약 STT 학습용

보험

AI-OCR 증설 & 추론

증권

검색용 LLM 구축

카드

FDS(이상거래탐지시스템)

카드

LLM 추론시스템

도입 사례

IV. 금융 레퍼런스

은행

LLM 모델 운영

은행

안면인식(바이오 정보추출)

은행

STT/TA 고도화

은행

생성형 AI 시스템 구축

은행

오픈소스 생성형AI 모델

은행

개인화 마케팅 업무

은행

문서 분석 및 한글 임베딩

은행

생성형 AI 도입

은행

AI 개발 시스템

은행

검색용 LLM 구축

보험

LLM 파인튜닝

보험

AI-OCR 및 사내 개발용

보험

녹취 요약 STT 학습용

보험

AI-OCR 증설 & 추론

증권

검색용 LLM 구축

카드

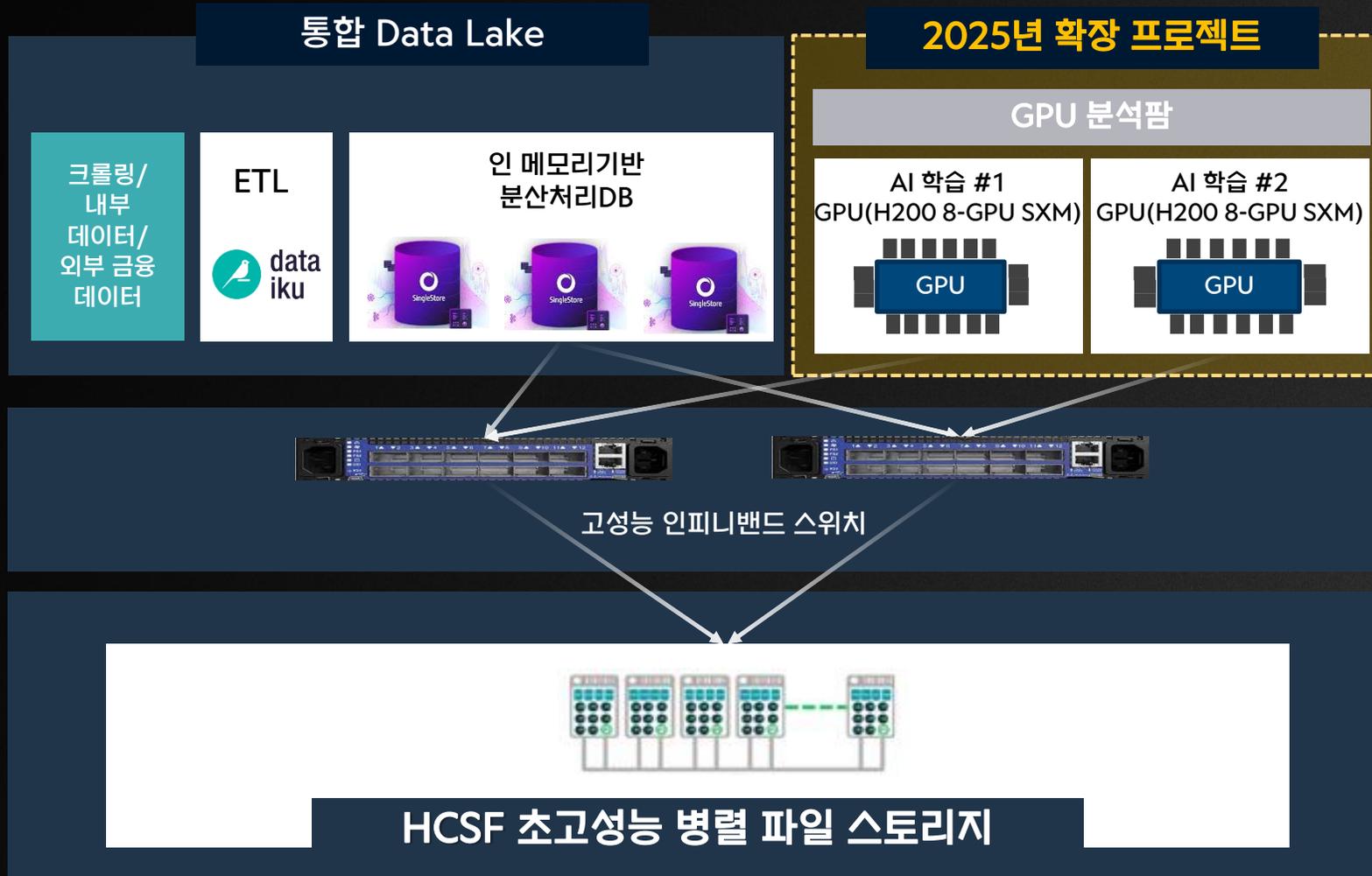
FDS(이상거래탐지시스템)

카드

LLM 추론시스템

A은행_AI 플랫폼 구축 사례

- H200 8-GPU SXM 기반 AI 분석 환경 및 HCSF(고성능 저장소) 구축 사례



사업 목적

- 외부기관 데이터/국내금융기관데이터 통합(최종 80개 기관 데이터수집 목표)으로 효과적인 데이터분석 및 분석체계 확립

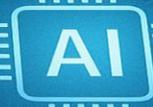
구축 내용

- 정형/반정형 데이터 분석을 위한 데이터 레이크 구축
- 사업인 GPU기반 딥러닝 분석환경으로 확장 가능한 아키텍처 수립

도입 효과

- 통합 플랫폼 (CPU → + GPU) 을 이용해 관리 포인트 최소화 및 확장 유연성 제공
→ 단 2명의 소수 운영자로 관리가 가능해짐
- 고성능 단일 데이터 레이크 저장소 구축 운영
→ 여러 대의 업무 별 스토리지를 하나로 통합
- 고성능 쿼리 분석 엔진 도입

안정적인 AI 인프라
HS효성인포메이션시스템과 함께



감사합니다.

