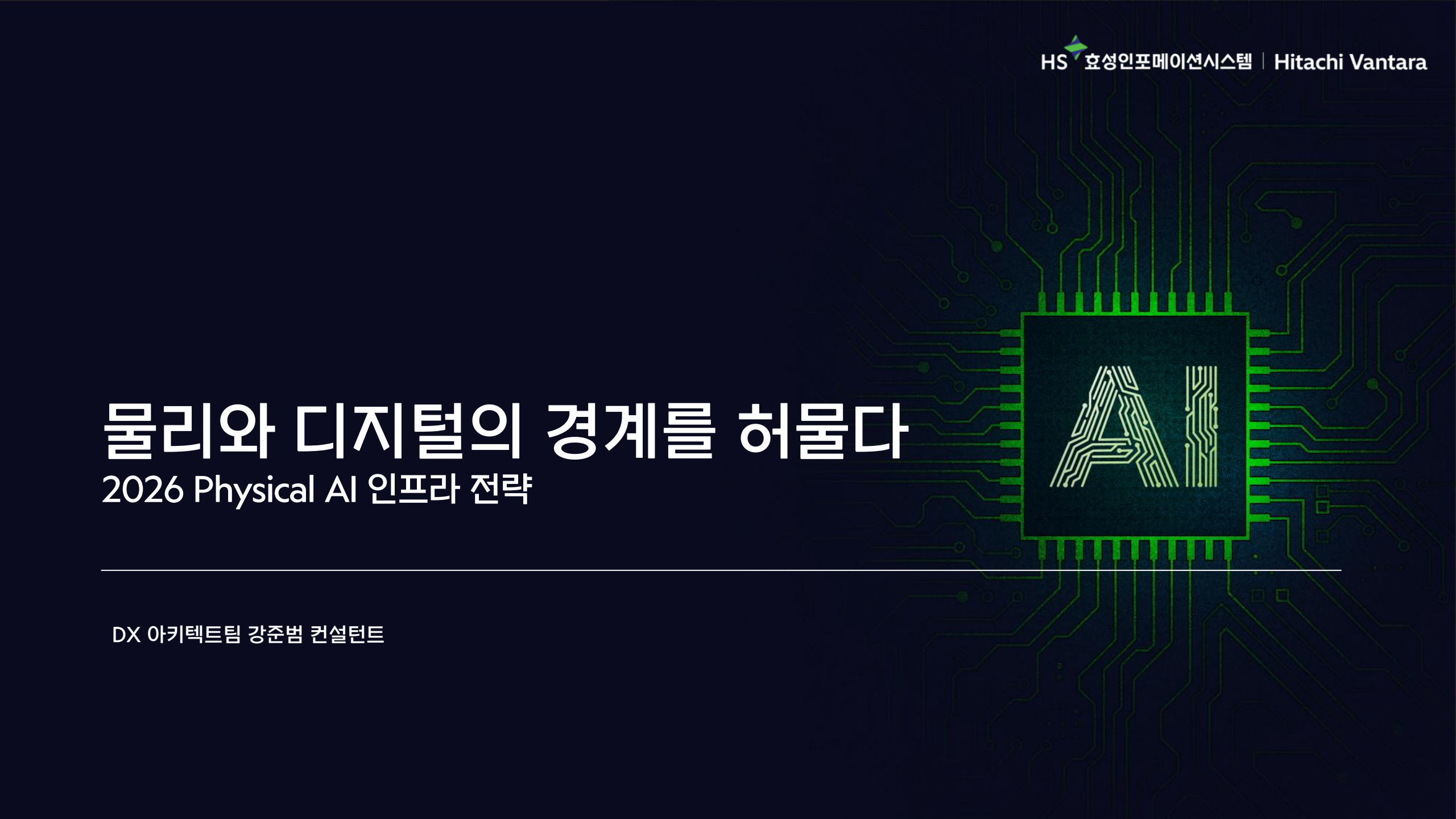
HS 효성인포메이션시스템 | Hitachi Vantara



물리와 디지털의 경계를 허물다

2026 Physical AI 인프라 전략

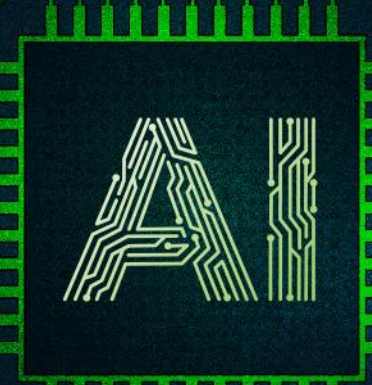
DX 아키텍트팀 강준범 컨설턴트

Agenda

1. AI 트렌드
2. AI를 위한 HS효성 AI플랫폼
3. AI 구축 방안 및 사례

1. AI 트렌드

1. 지능의 완성은 현실과의 상호작용
2. Physical AI의 핵심 가치
3. Physical AI의 워크플로우
4. Physical AI를 위한 인프라 3대 기둥



1. 지능의 완성은 현실과의 상호작용

LLM / 멀티모달 AI
지능의 탄생

텍스트, 이미지, 음성 등
다양한 데이터를
동시에 인식 및 처리

Agent AI
도구의 활용

스스로 **도구**를 **사용**하고 **다중**
워크플로우를 **계획 및 수행**

Physical AI
현실 세계의 개입

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

디지털트윈 / 자율주행 등

2. Physical AI의 핵심 가치

LLM

- ‘생각’하는 AI

- 실시간 대응의 어려움
- 텍스트 기반의 학습 데이터
- 환각증상(Hallucination)
- 고차원적인 계획이나 추론을 담당

Physical AI

- ‘행동’하는 AI

- 밀리세컨드(ms) 단위의 즉각적인 반응 필요
- 실시간 데이터 처리(센서데이터)
- 환경적인 요인 변화에 대한 다양한 데이터의 부재
- 무의식적이고 즉각적인 실시간 제어

- tesla Optimus(@Tesla_Optimus) / X) 이미지

2. Physical AI의 핵심 가치

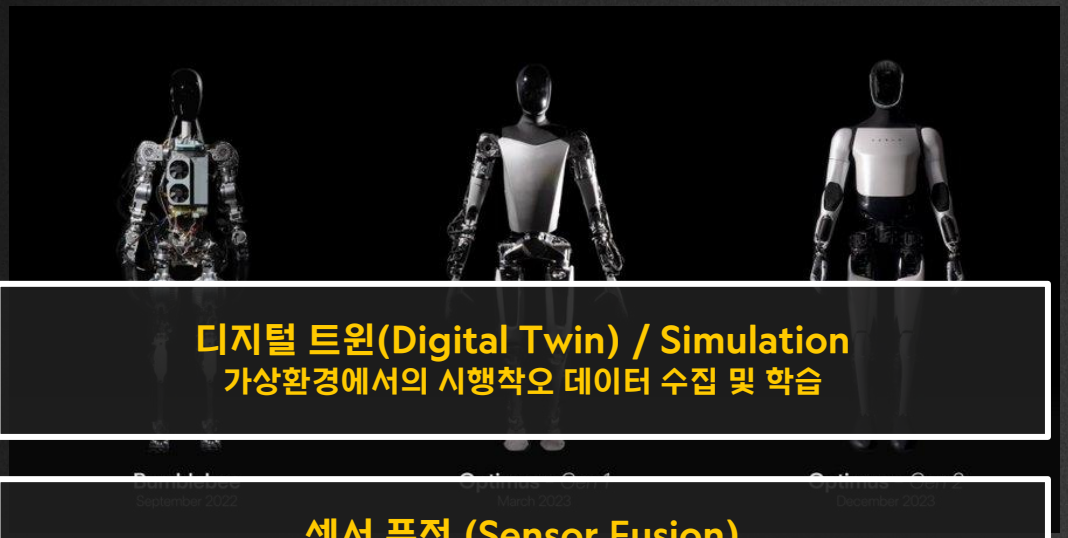
프롬프트(Prompt) 엔지니어링

RAG(Retrieval-Augmented Generation)

컨텍스트(Context) 엔지니어링

MCP (Model Context Protocol)

하네스(Harness) 엔지니어링



디지털 트윈(Digital Twin) / Simulation
가상환경에서의 시행착오 데이터 수집 및 학습

센서 퓨전 (Sensor Fusion)
다양한 센서 데이터를 실시간으로 통합하여 환경을 이해

- '행동'하는 AI

실시간 추론(Real-time Inference)
밀리세컨드(ms)단위의 환각 없는 즉각적인 의사결정

물리적 임바딩 (physical Embodiment)
AI모델의 실제 물리력 행사 단계

- tesla Optimus(@Tesla_Optimus) / X) 이미지

3. Physical AI의 워크플로우

Edge 데이터 수집 → 시뮬레이션(합성데이터) → 학습 → 배포 → 추론
즉각적인 대응과 모델 학습을 위한 고성능 저장소, 고성능 GPU 인프라의 필요

01 Edge Data Collection

실제 데이터 수집

- **Sensor Ingestion**
  카메라·LiDAR 등
- **Teleoperation Demo**
  성공/실패 사례 기록
- **Edge Processing**
  비식별화·노이즈 제거 후 전송

02 Real2Sim & Digital Twin

가상 환경 재구성

- **Scene Reconstruction**
  3D 환경 복제
- **Asset Creation**
  Isaac Sim / Cosmos 변환
- **SDG(Synthetic Data Generation)**
  가상 데이터 수천 배 증폭

03 Large-scale Training

HPC 클러스터 학습

- **Foundation Model**
  World Model + VLA 학습
- **Reinforcement Learning**
  수백만 번 시행착오 정책 습득
- **Validation**
  시뮬레이션 내 안전성 검증

04 Model Optimization

최적화 및 배포

- **Quantization**
  FP32→FP8/FP4
  Blackwell 최적화
- **TensorRT Compile**
  하드웨어 가속 엔진 생성
- **GitOps/OTA**
  전 세계 엣지 무선 배포

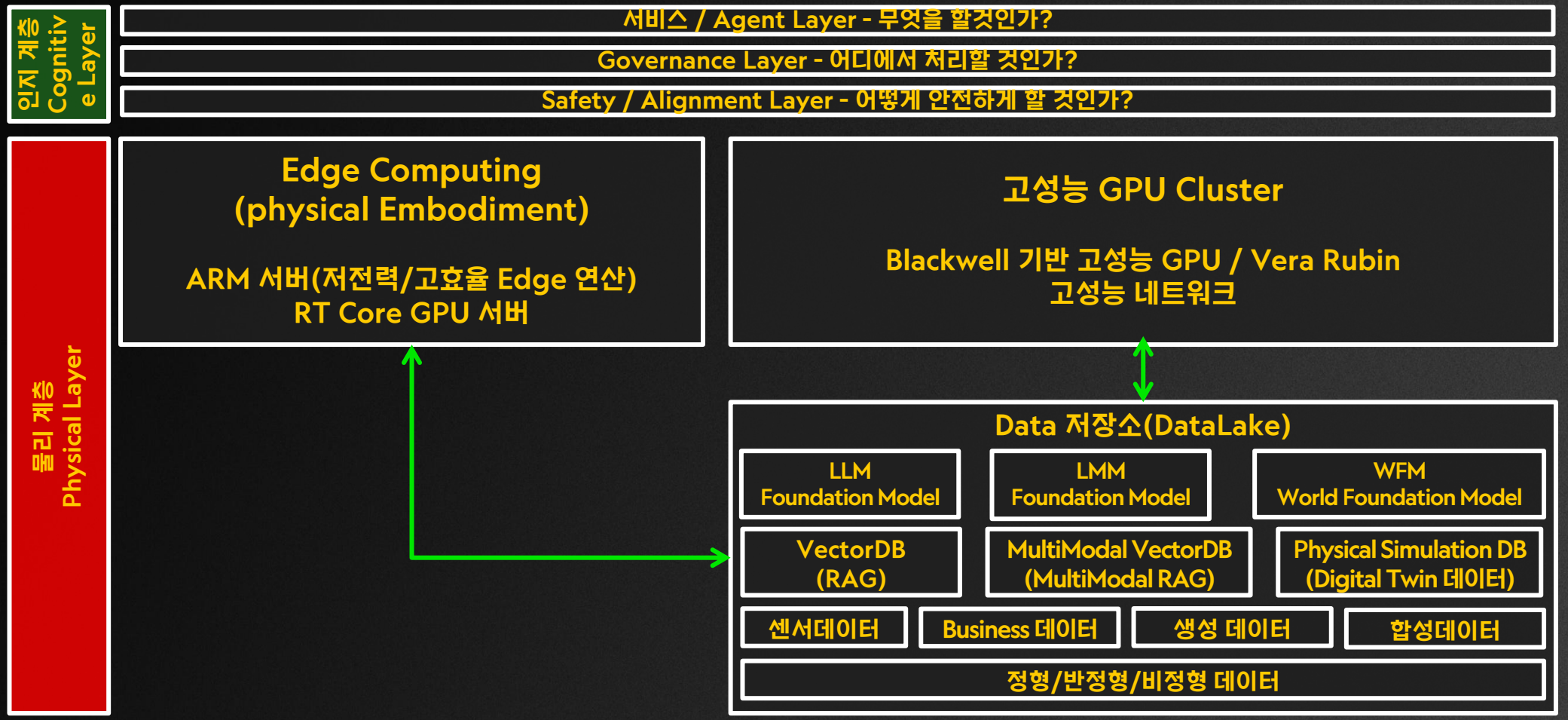
05 Sim2Real & Inference

엣지 추론 및 피드백

- **Real-time Inference**
  실시간 로봇 제어 명령
- **Error Monitoring**
  실패 데이터 → 재학습 피드백
- **Continuous Loop**
  1단계로 데이터 환류

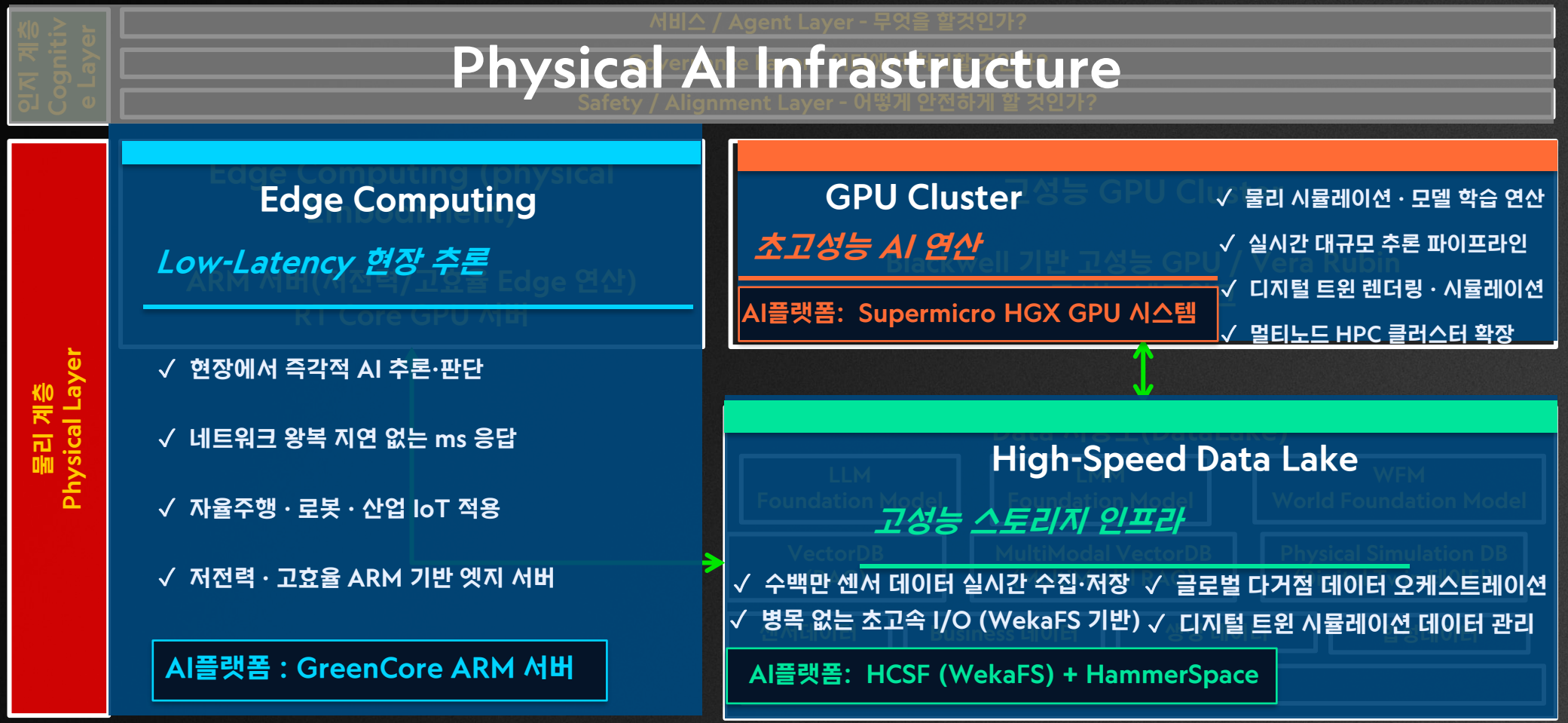
4. Physical AI를 위한 인프라 3대 기둥

Edge에서 센터까지, 데이터로 잇고 인프라로 가속하는 Physical AI 실현



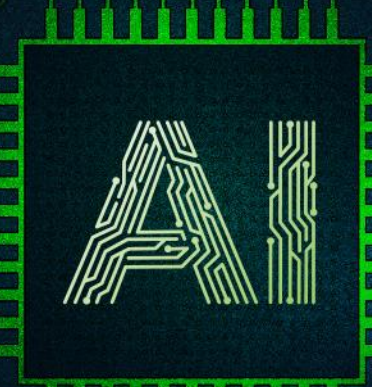
4. Physical AI를 위한 인프라 3대 기둥

Edge에서 센터까지, 데이터로 잇고 인프라로 가속하는 Physical AI 실현



2. AI를 위한 HS효성 AI플랫폼

1. AI 인프라 구성의 복잡성
2. HS효성 AI플랫폼
3. AI플랫폼 Component
4. AI 도입 이슈에 대한 고민 해결



1. AI 인프라 구성의 복잡성

Physical AI / 멀티에이전트 인프라 설계 시 HPC 클러스터부터 Edge Computing · 고성능 스토리지 · GPU 활용도까지, 복합적인 HW 및 솔루션 구성에 대한 검증 필요 → Reference 기반 최적의 구성안 설계 필요!

이슈 1. AI 솔루션 기술 부족

- AI플랫폼은 복잡한 인프라 및 솔루션 조합으로 구성
  (모델링 알고리즘, 클라우드, 컨테이너, GPU/서버 가상화)

이슈 2. 초기 투자 비용 부족

- H/W 인프라에 더해 AI 솔루션에 대한 비용 부담, BigBang 형태의 투자에 대한 부담감
  (서버, 스토리지, 네트워크, AI/ML Ops 솔루션과 구축비용)

이슈 3. 전문 인력 및 역량 부족

- 기업내 내부 AI역량 부족에 대한 우려, 역량 있는 AI 파트너사 중요
  (구축 및 안정적 운영을 위한 기업내 AI역량 확보 이슈)

AI 시작은?

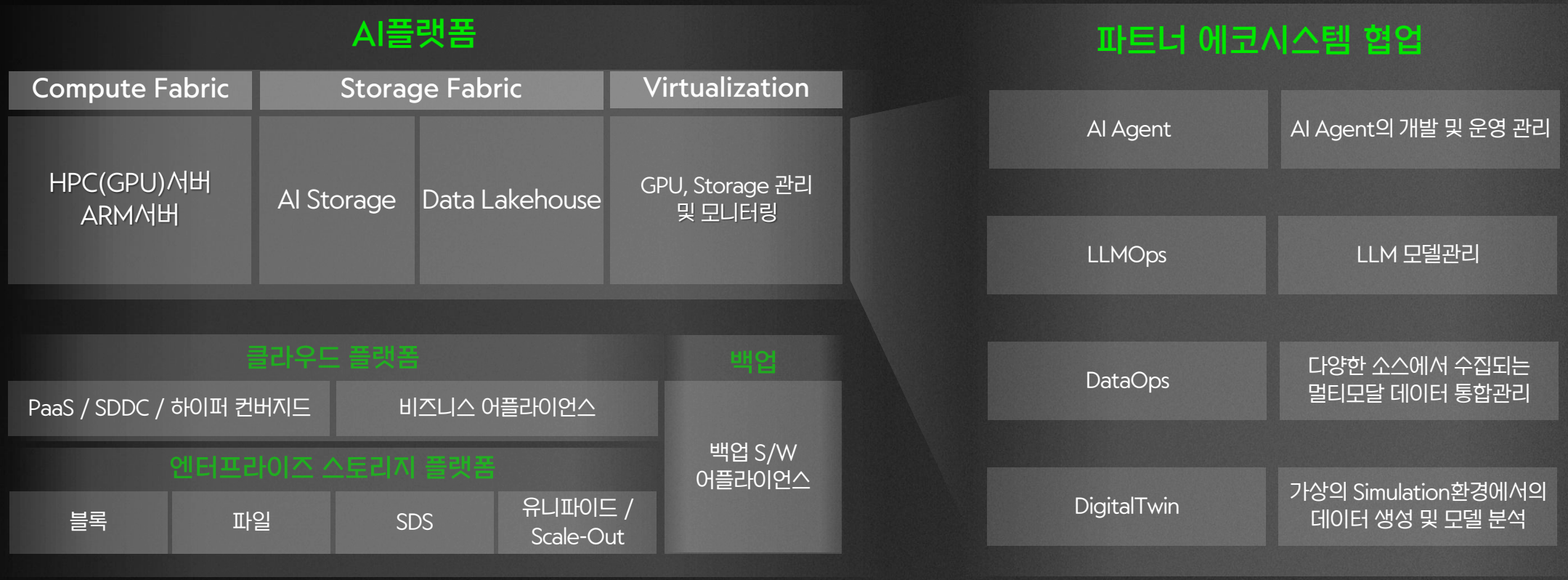
도입 후 활용은 ?

어떻게?

2. HS효성 AI플랫폼

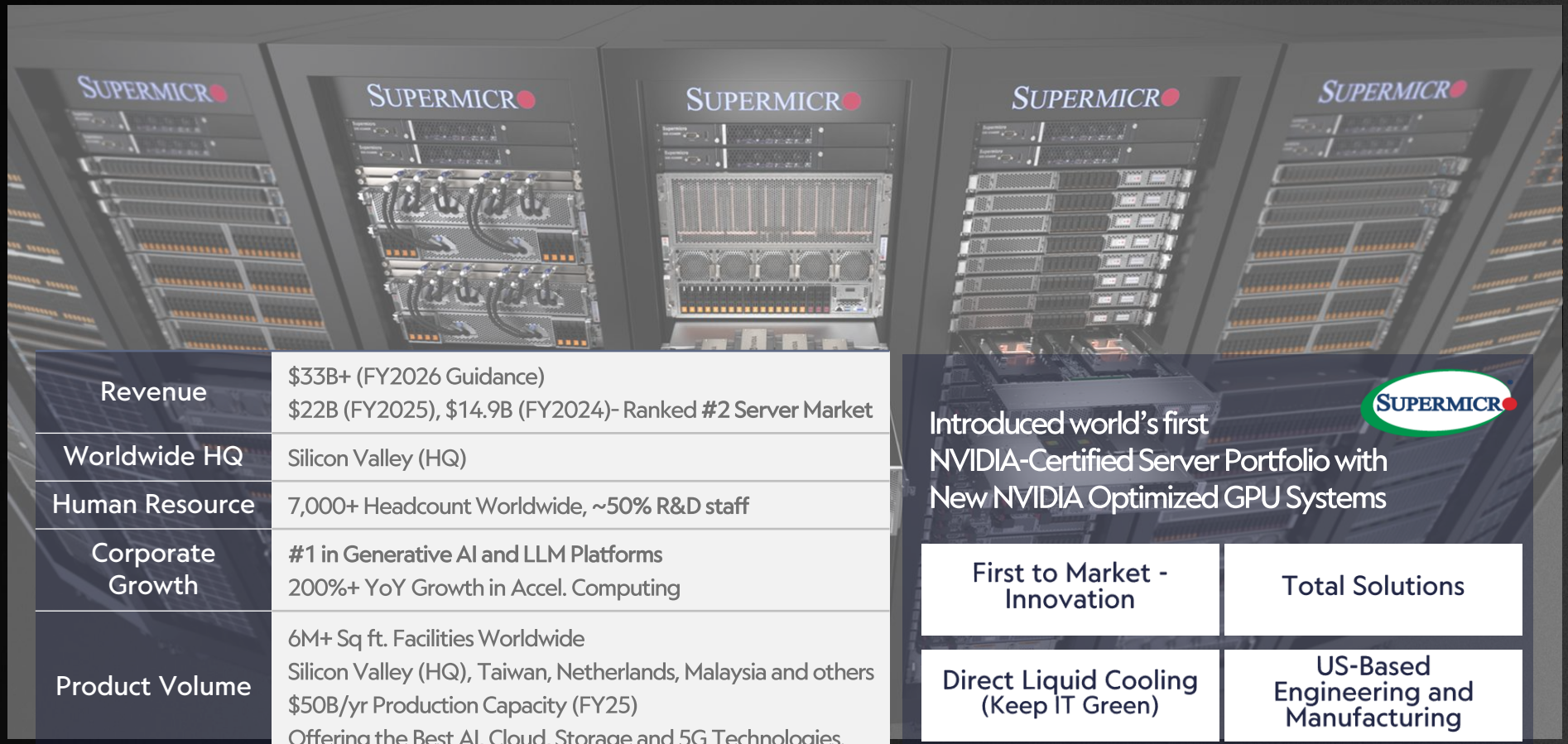
확장성과 유연성을 갖춘 AI플랫폼

파트너 에코시스템 시너지 강화



3. AI플랫폼 Component - GPU

- 고성능 GPU Compute Fabric - SUPERMICRO **HGX GPU** Server
- HS효성인포메이션시스템은 국내 SUPERMICRO의 **공식 총판**



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 NVIDIA Vera Rubin NVL 72 Super Cluster

3. AI플랫폼 Component - Arm

- 저전력 고 전성비 Compute Fabric - GreenCore **Arm CPU** Server
- HS효성인포메이션시스템 & 엑세스랩 공동 개발한 **국산 Arm** 서버

GreenCore

강점1) H/W 보드 직접 설계, 개발

국내 유일 Arm 보드 일체 설계/개발 → 유지보수 및 기술자문 유연 대응

강점2) S/W 일체형 지원

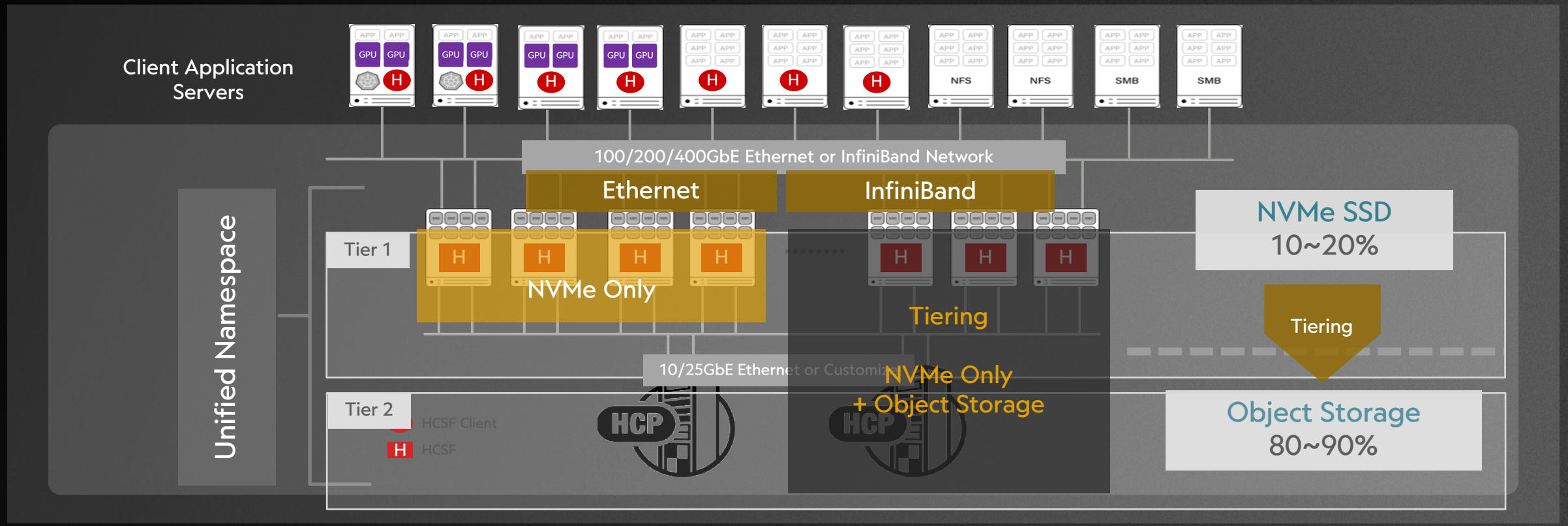
Arm 관련 OS 및 다양한 오픈소스에 대한 가이드 지원

강점3) 맞춤형 관리환경

Arm 서버 전용 BMC 제공 및 관리 GPU 화면 제공

3. AI플랫폼 Component - HCSF/WEKA

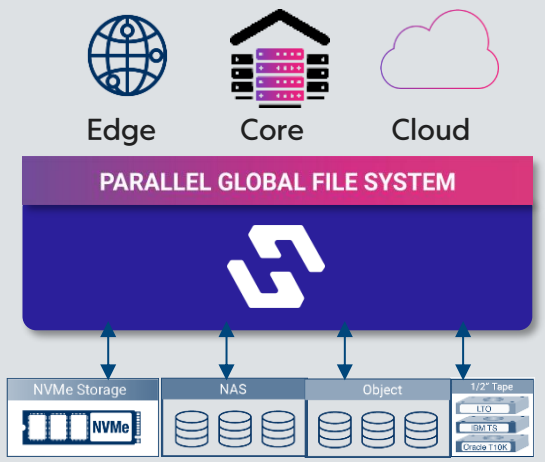
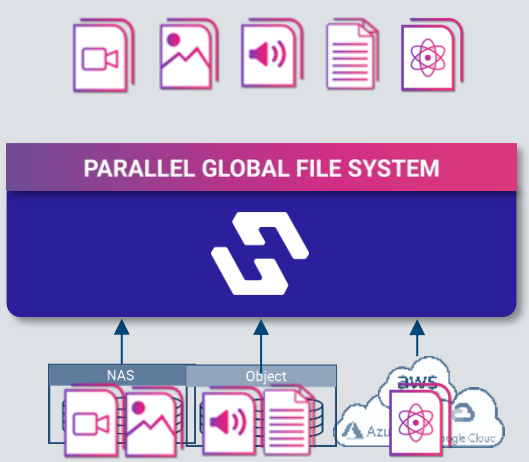
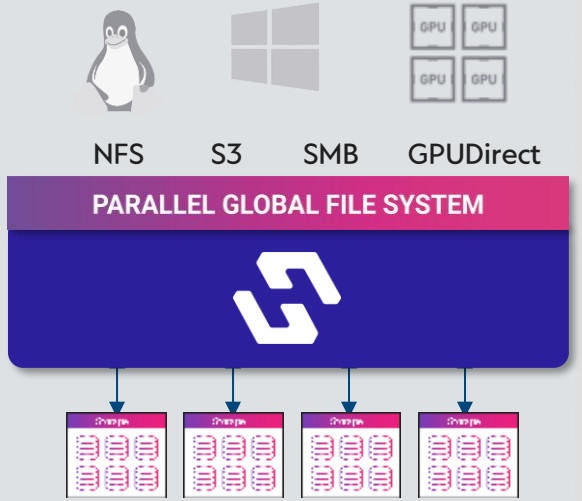
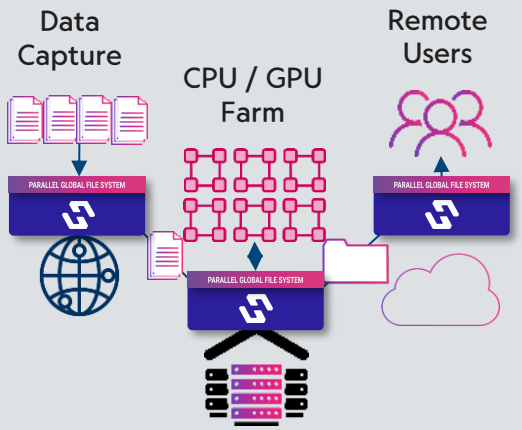
• 초고성능 Storage Fabric - 고성능 분산병렬 파일시스템 어플라이언스 **HCSF/wekaFS**



• 티어링 시 구성 용량은 요구 성능 및 환경에 따라 달라질 수 있음

3. AI플랫폼 Component - HammerSpace

• Storage Fabric - Data Orchestration **HammerSpace**

모든 환경, 모든 스토리지 구축	물리적 데이터 이관 X	병렬 처리를 위한 빠른 데이터 액세스 제공	정책에 따른 손쉬운 데이터 자동 배치
			

3. AI플랫폼 Component - Hitachi IQ Studio

- 기업용 AI 에이전트 구축·운영을 간소화하는 통합 플랫폼 - 히타치 iQ 스튜디오
- 노코드 에이전트 빌더와 온프레미스 보안 환경 제공

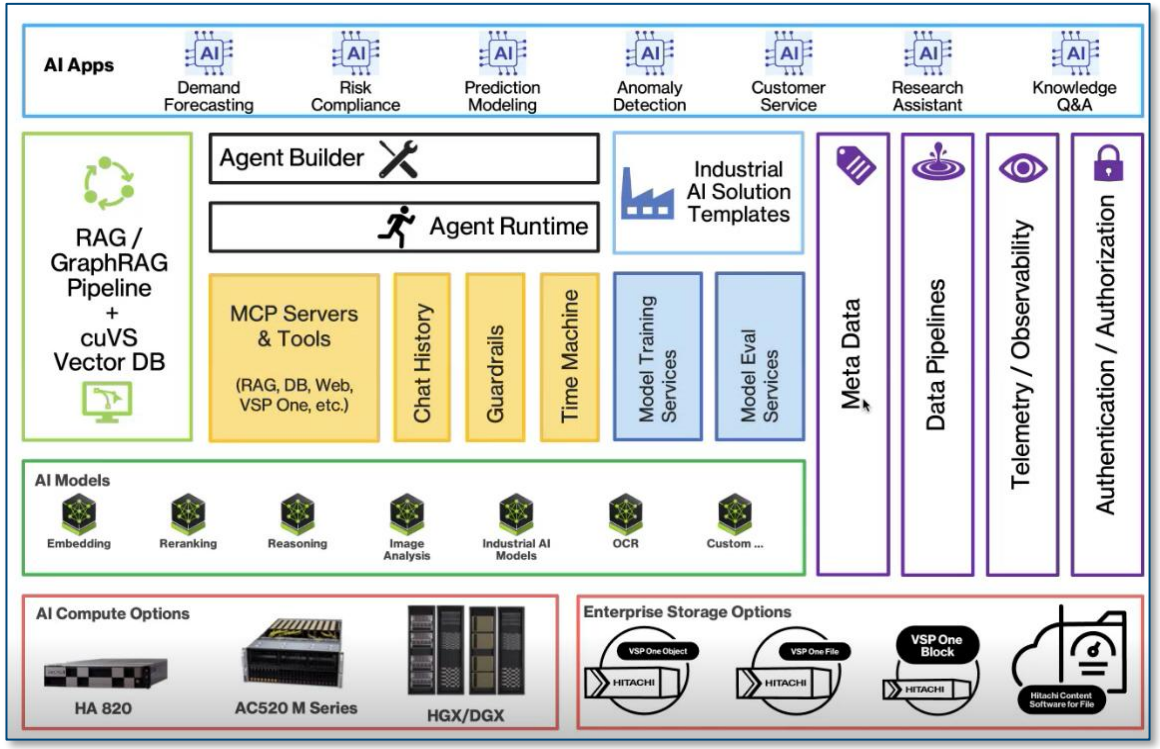
히타치 iQ 스튜디오

엔터프라이즈 AI 에이전트 구축/운영 간소화

'히타치 iQ 스튜디오'

- AI 에이전트 전 주기 통합 관리
- 엔비디아 AI 플랫폼 기반
- RAG/MCP 기반 AI-Ready 데이터제공
- AI 거버넌스 및 감사 추적 지원

히타치 iQ 스튜디오 아키텍처



4. AI 도입 이슈에 대한 고민 해결

1. AI 인프라 기술
- 통합 AI플랫폼 제공

- GPU 가상화, 고성능 스토리지, 네트워크, 컨테이너
- 슈퍼마이크로 GPU서버와 스토리지 조합으로 아키텍처 단순화

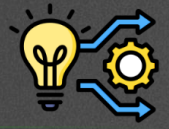
2. 비용효율적 구성
- 성능과 비용 효율 데이터 운영

- 고성능 데이터 처리 인프라 제공
- 초고성능 병렬 파일 스토리지 (Weka-HCSF)
- 고성능 파일 통합 스토리지(해머스페이스)
- 비용효율적 저장용 데이터레이크(오브젝트 스토리지)

3. 에코시스템 구축
- 다양한 솔루션 접목

- AI 적용을 위해 필요한 다양한 솔루션 접목
- 기존의 방식과 다른 접근 체계 가능
- AI Ops, LLM 기반 챗봇 등 서비스 전문 파트너와 연계

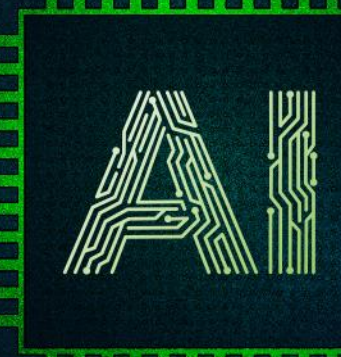
4. 운영 효율화
- 통합 제안 및 운영 지원



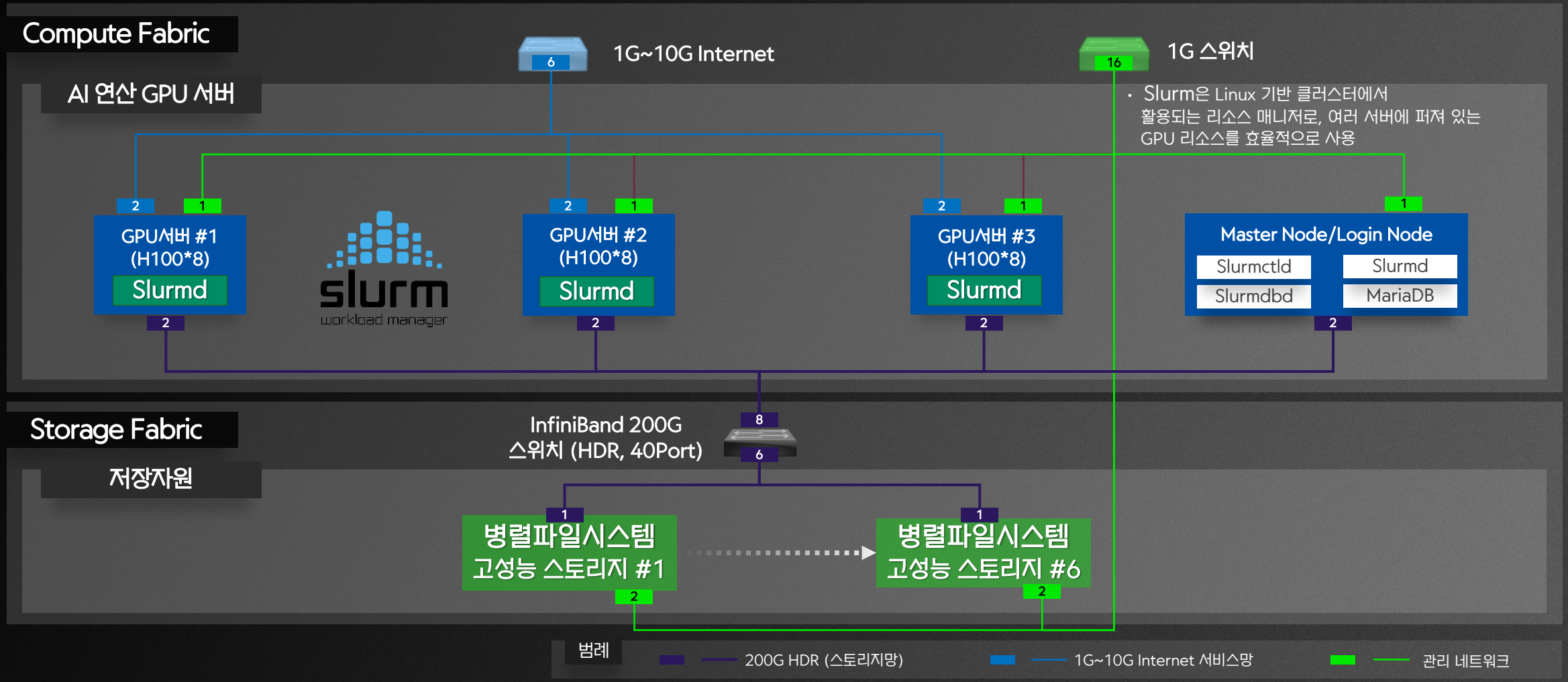
- AI 인프라에서 필수적인 연산자원과 (슈퍼마이크로 GPU서버) 네트워크, 저장자원 (SAN/NAS 및 HCSF, 해머스페이스, 오브젝트 스토리지 등)을 통합 구성
- 다양한 연계 솔루션을 통합 구축을 통해 운영 효율성 확보

3. AI플랫폼 사례 및 구성

1. A사 사례-IT 대기업 AI플랫폼 인프라(자체 LLM 개발)
2. B사 사례-대학병원 AI 분석 플랫폼 GPU FARM 구축
3. C사 사례-대기업 DX GPU AI 인프라 구축 (연구 개발)
4. D사 사례-은행 데이터레이크 GPU AI 인프라 구축
5. HS효성 AI플랫폼 구성
6. AI 인프라 도입 시 고려 사항

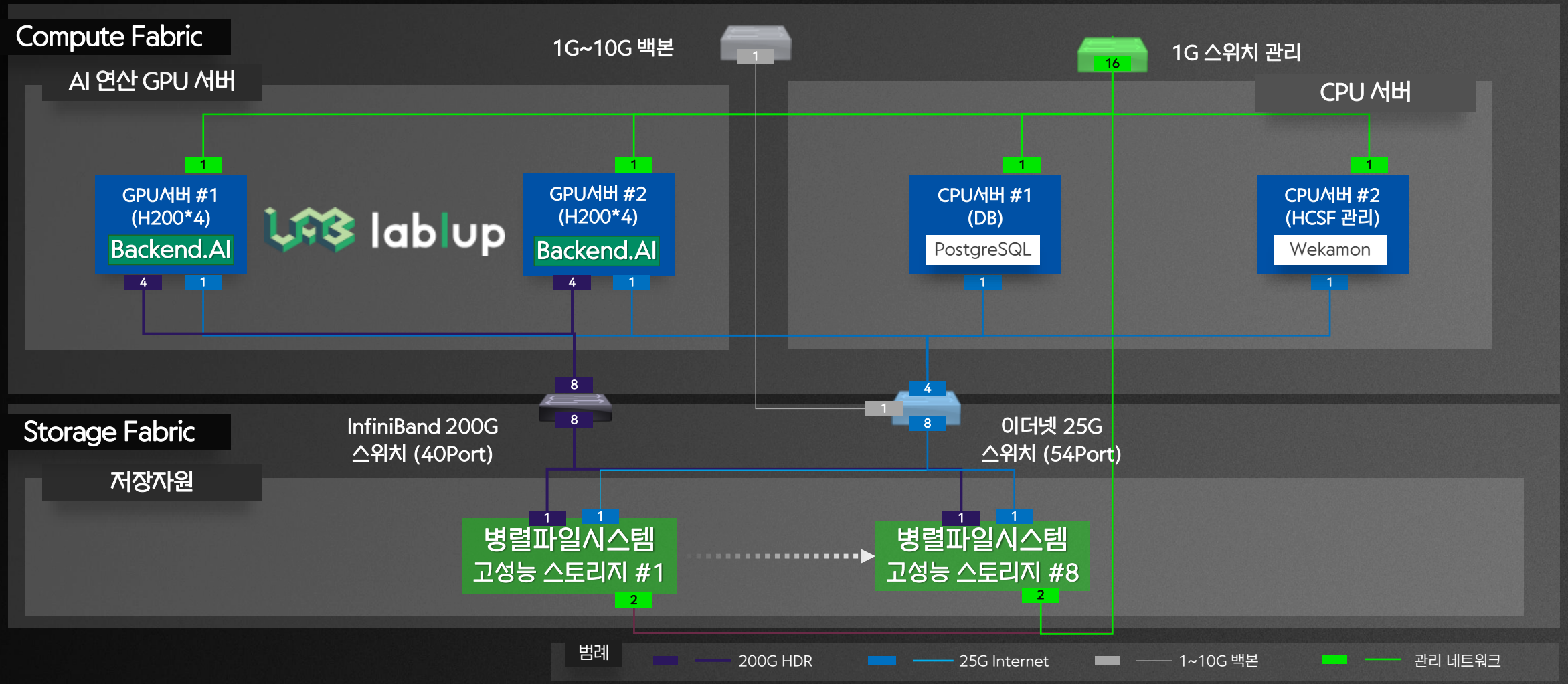


1. A사 사례-IT 대기업 AI플랫폼 인프라(자체 LLM 개발)

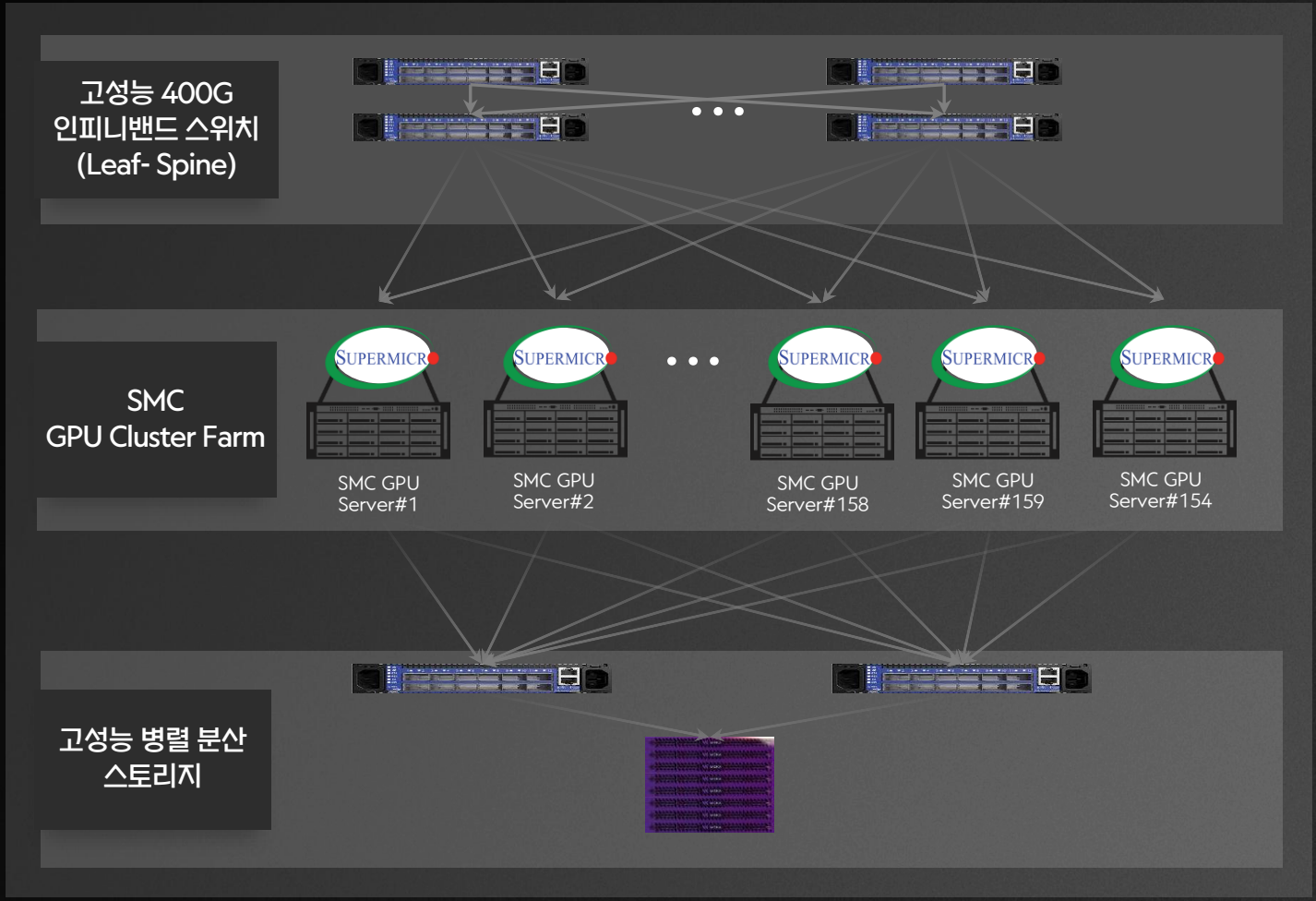


2. B사 사례-AI 분석 플랫폼 GPU FARM 구축

HPC 클러스터 (HW & SW) & HCSF(분석특화 저장소) & GPU 가상화 분할 및 클러스터 관리 솔루션 구축 사례



3. C사 사례-대기업 DX GPU AI 인프라 구축 (연구 개발)



사업 목적

- 고객사 DX GPU AI 인프라 구축 목적의 고성능 AMD GPU Cluster Farm 인프라 도입 및 구성
- 운영 154 GPU Cluster Farm / 개발 6 GPU Cluster Farm 구축 : 총합 160대 도입

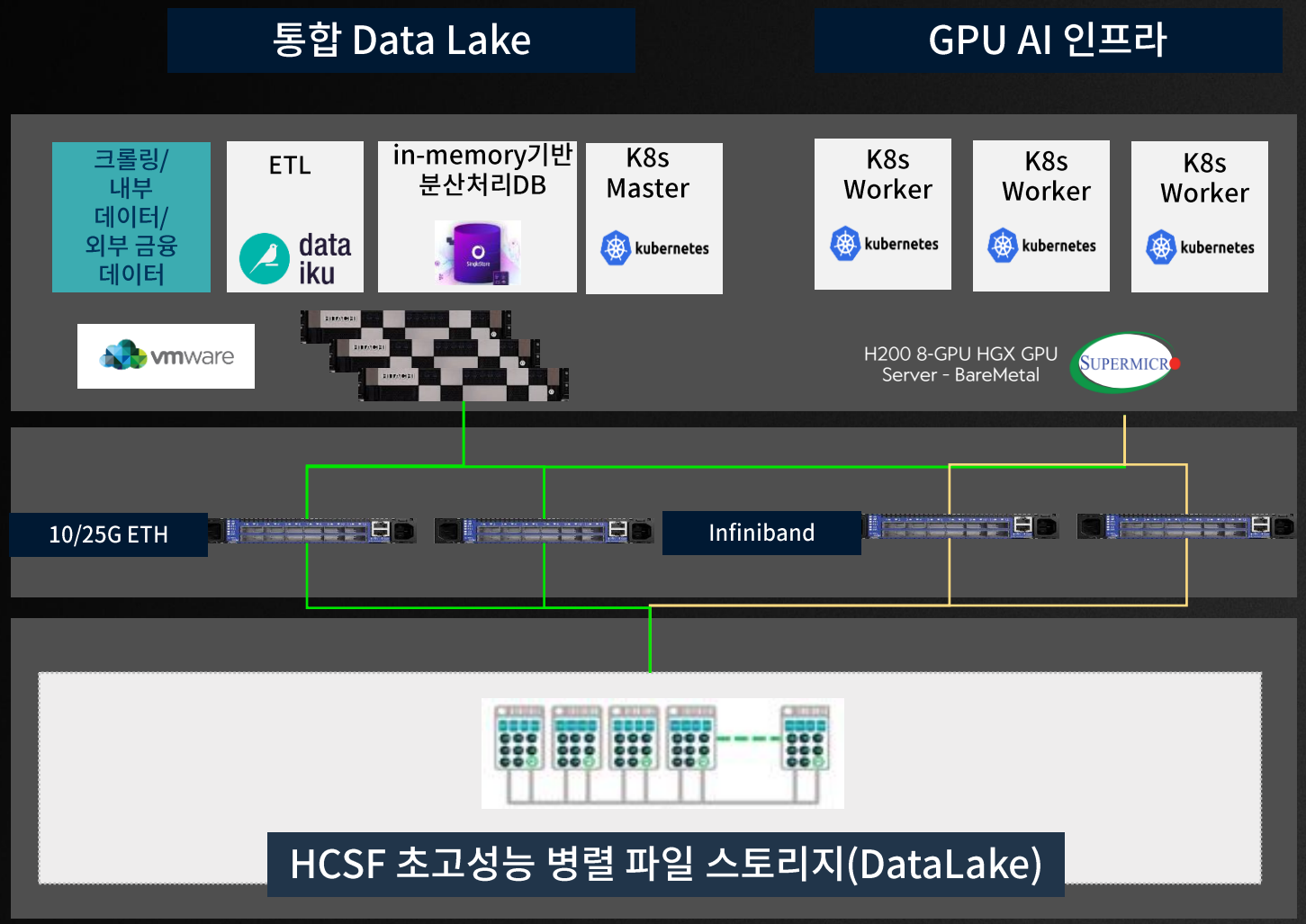
구축 내용

- 제조 연구 개발 및 분석을 위한 AMD GPU Farm 구축
- ROCm 라이브러리 활용을 통한 분석 환경 및 병렬분산스토리지 연계

도입 효과

- One Vendor GPU (Nvidia) 종속성 탈피 및 cost saving을 위한 고성능 AMD GPU가 장착된 안정적인 고성능 SMC GPU Server 도입

4. D사 사례-금융권 데이터레이크 GPU AI 인프라 구축



사업 목적

- 외부기관 데이터/국내금융기관 데이터 통합으로 **데이터 분석 및 AI 분석 /LLM 환경** 확립

구축 내용

- 정형/반정형/비정형 데이터 분석을 위한 **데이터 레이크** 구축
- GPU기반 AI 분석환경 및 LLM환경으로 **Scale-out 아키텍처 수립**

도입 효과

- Row기반/Column 기반 분석이 모두 가능한 **고성능 쿼리 분석 엔진** 도입
- 통합 플랫폼 (CPU -> + GPU) 을 이용해 **HW관리 포인트 최소화 및 확장 유연성 제공**
- **고성능 단일 데이터 레이크 저장소** 구축 운영

5. AI 인프라 도입 시 고려 사항

01

다양한 에코 파트너 협업 체계 구축 확인

빠르게 변화하는 AI시대 대응

02

AI플랫폼 구축 경험 확인

Compute Fabric,
Storage Fabric, AIOps Stack

03

국내외 실 사례를 통한 국내 기술력(인력) 여부 확인

장애 지원, 신규 AI 솔루션 연계 지원

감사합니다.