전자신문

CIO Summit 2026

CIO SUMMIT 2026

인간과 AI의 조화, 멀티 에이전트 성공의 열쇠

2026. 2. 26.(목)

그랜드 인터컨티넨탈 서울 파르나스

CIO Summit 2026

AI 가속의 시대

멀티 에이전트 성공을 위한 엔터프라이즈 AI 인프라 전략

HS 효성인포메이션시스템 | Hitachi Vantara

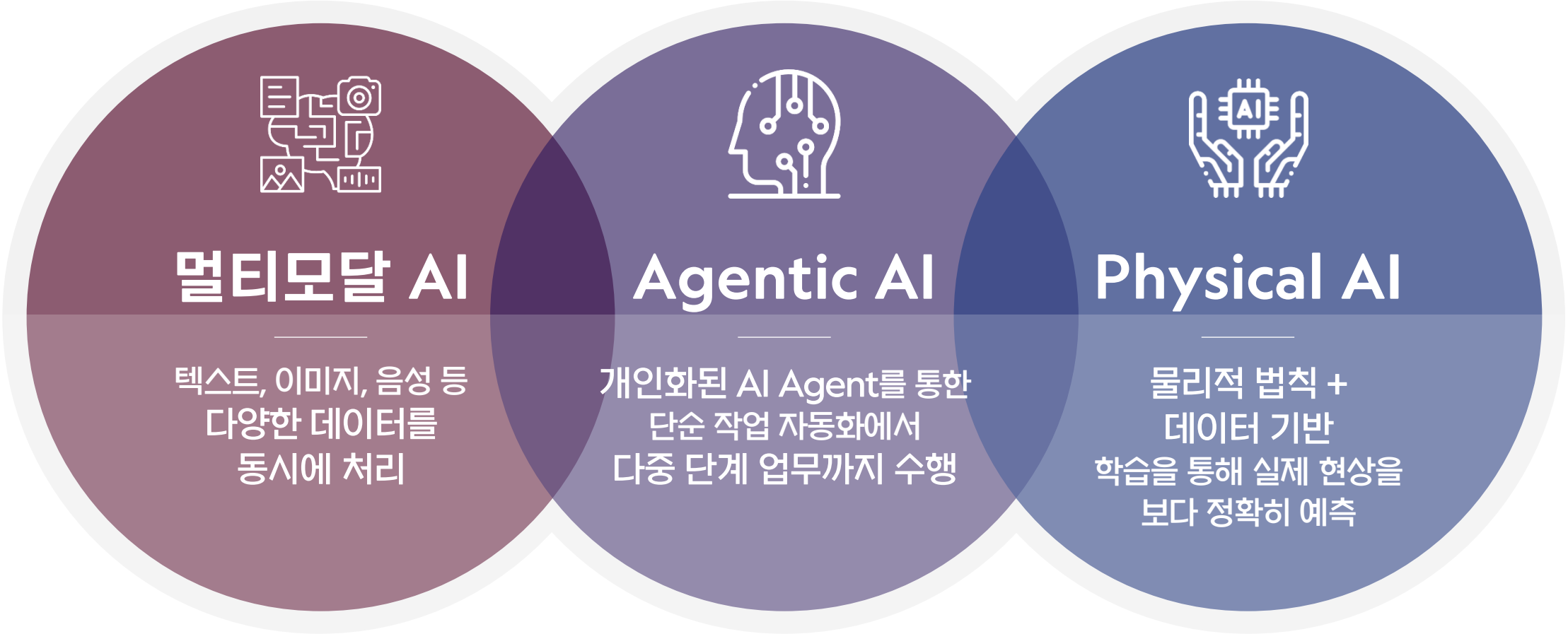
Agenda

I. AI 트렌드
II. 멀티에이전트 구축을 위한 HS효성 AI 플랫폼
III. AI 플랫폼 구성 및 사례

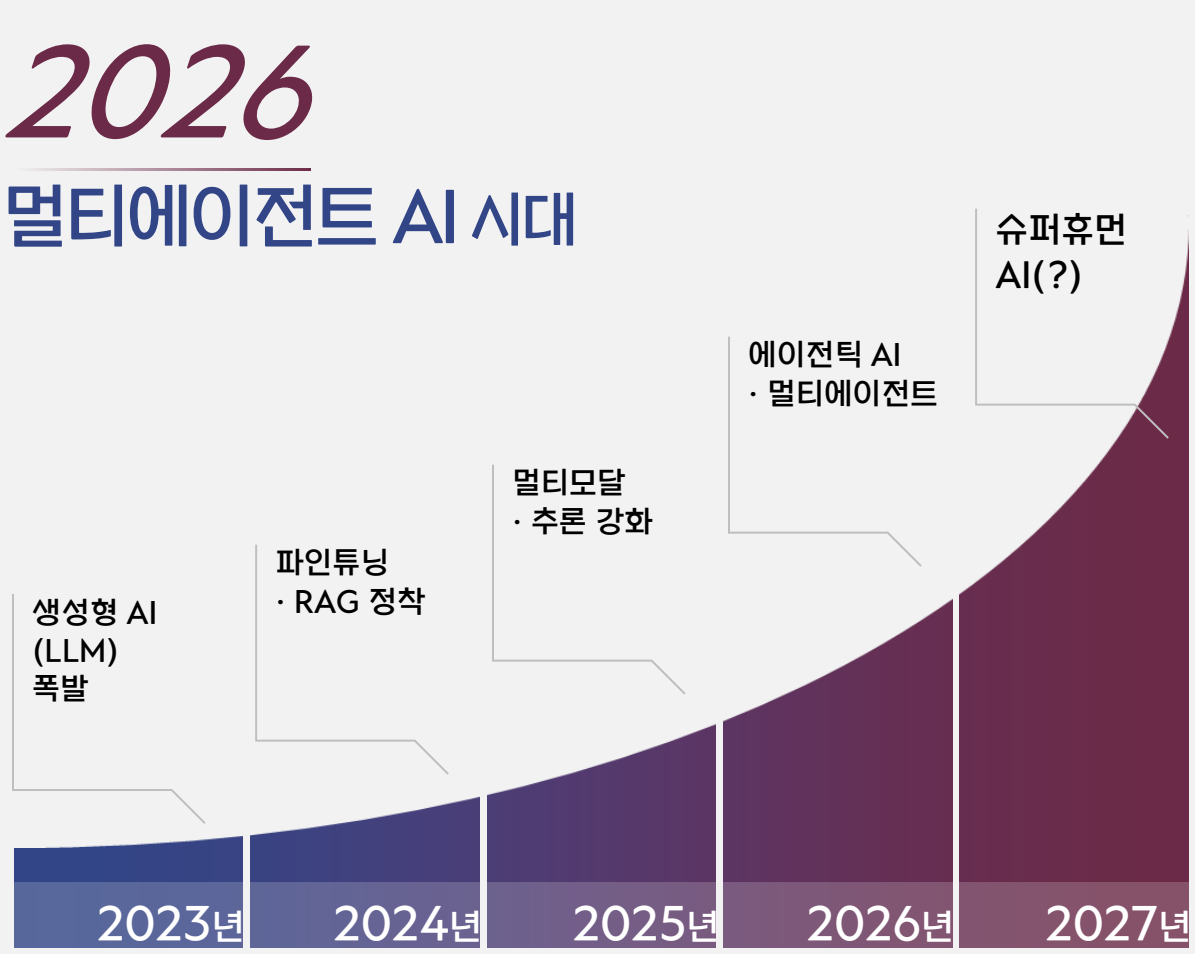
I. AI 트렌드

1. 지속가능한 미래를 위한 AI
2. AI 기술 발전 전망
3. 멀티에이전트 시스템의 진화
4. 에이전틱 AI 인프라 4대 필수 요건

1. 지속가능한 미래를 위한 AI



2. AI 기술 발전 전망

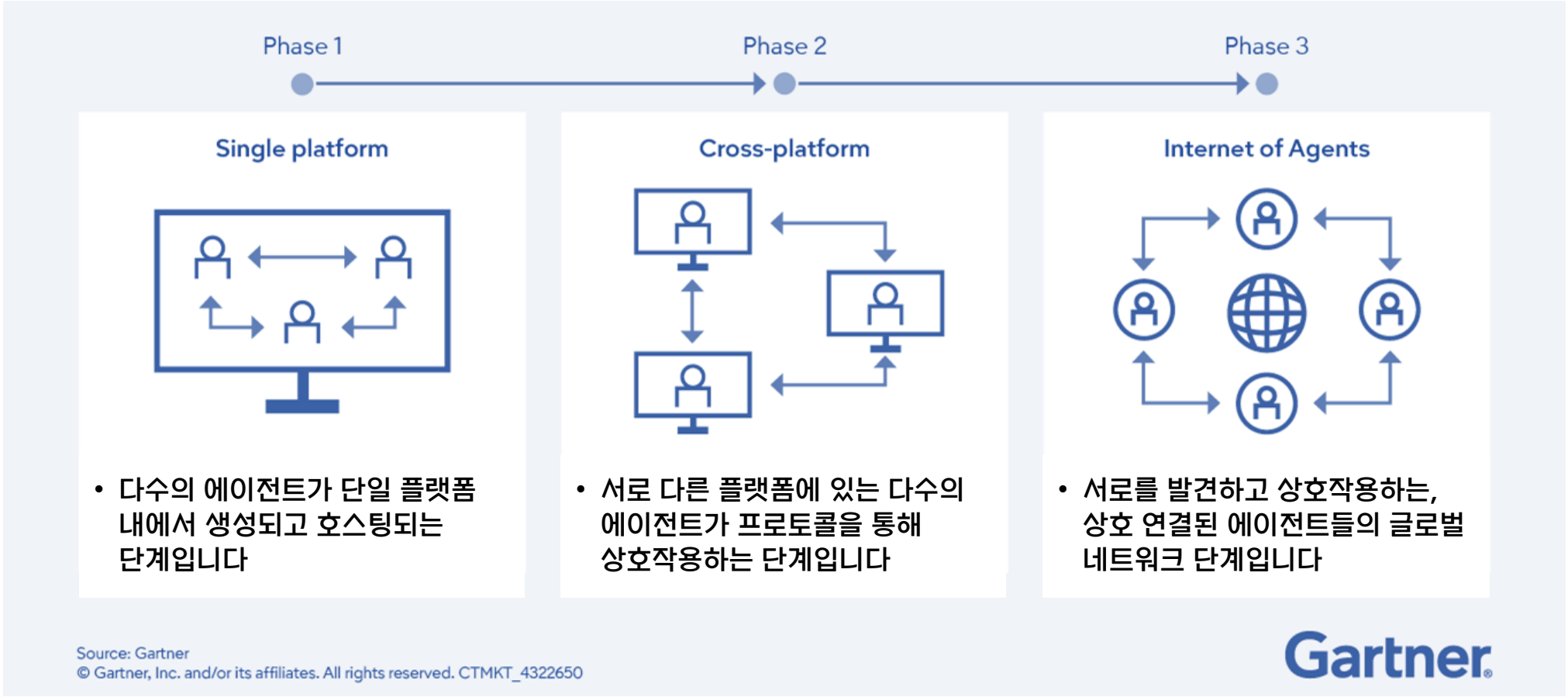


source:AI-2027·Gartner·Deloitte 전망

모델명	특징 요약
Llama 4.5 (Enhanced) 2026년 1월 발표	• 10M 토큰의 초거대 컨텍스트와 네이티브 멀티모달 능력 및 에이전트 성능 최적화 (MXFP4 양자화 포맷 지원)
gpt-oss-120B 2025년 8월 발표	• 강력한 추론 성능을 제공하는 전문가 혼합(MoE) 구조의 AI 모델(MXFP4 양자화 포맷 지원)
Qwen3-Max 2025년 9월 발표	• 중국 오픈소스 대형 모델, 최근 발표 및 규모 확대
Llama 4 Maverick 2025년 4월 발표	• 멀티모달·긴맥락 대응 강화, 메타의 주력 모델
Gemma 3 2025년 3월 발표	• 효율 중심 오픈모델, 비용/운용 측면에서 추천
Mixtral 8×22B 2024년 4월	• 오픈소스 전문가 혼합(MoE) 모델 중 성능·효율 균형 우수상용

3. 멀티에이전트 시스템의 진화

- 단순 챗봇을 넘어 스스로 판단하고 도구를 사용하는 'AI 에이전트'로의 진화
- 인프라의 한계: 기존 범용 서버로는 에이전트의 복잡한 추론과 실시간 데이터 접근 속도를 감당하기 어려움



4. 에이전틱 AI 인프라 4대 필수 요건

01		
	초저지연 추론 및 컴퓨팅 (Low-Latency & Bursty Compute)	◦ 가변적 리소스 할당: 에이전트의 활동은 일정하지 않고 특정 순간에 폭발적인 계산량을 요구(Burst)합니다. 이를 유연하게 지원하는 탄력적 인프라(Elastic Infra)가 필요합니다.
02		
	고성능 메모리 및 데이터 레이어 (Memory & Data Fabric)	◦ 실시간 데이터 공급: 에이전트가 외부 도구를 사용할 때 데이터를 실시간으로 읽고 쓸 수 있는 병렬 파일 시스템이 뒷받침되어야 GPU 병목 현상 없이 자율 판단을 내릴 수 있습니다.
03		
	안전한 실행 환경 (Sandboxed Execution Environment)	◦ 샌드박싱 (Sandboxing): 에이전트가 생성한 코드가 메인 시스템에 해를 끼치지 않도록 격리된 안전한 실행 환경이 필수입니다.
04		
	통합 관측성 및 거버넌스 (Observability & Governance)	◦ 에이전틱 MLOps (AgentOps): 에이전트의 추론 단계별 로그를 기록하고, 오류 발생 시 즉시 개입할 수 있는 모니터링 시스템이 필요합니다.

II. 멀티에이전트 구축을 위한 HS효성 AI 플랫폼

1. 멀티에이전트 인프라 구성의 복잡성
2. HS효성 AI 플랫폼
3. AIOps 플랫폼 - Backend.AI, AIPub / MoAI
4. AI 오케스트레이션 플랫폼 - Hitachi IQ Studio
5. AI 도입 이슈에 대한 고민 해결

1. 멀티에이전트 인프라 구성의 복잡성

- 멀티에이전트 인프라 설계 시 HPC 클러스터부터 고성능 스토리지·GPU 활용도까지,
  복합적인 HW 및 솔루션 구성에 대한 검증 필요 → Reference 기반 **최적의 구성안 설계** 필요!

이슈 1.

AI 솔루션 기술 부족

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

이슈 2.

초기 투자 비용 부족

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

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 AI 플랫폼

확장성과 유연성을 갖춘 AI 플랫폼

파트너 에코시스템 시너지 강화

AI 플랫폼

Compute Fabric	Storage Fabric		AIOps
HPC(GPU)서버 ARM서버	AI Storage	Data Lakehouse	GPU 가상화 관리

클라우드 플랫폼		백업
PaaS / SDDC / 하이퍼 컨버지드	Private Cloud AI	백업 S/W 어플라이언스

엔터프라이즈 스토리지 플랫폼			
블록	파일	SDS	유니파이드 / Scale-Out

파트너 에코시스템 협업

AI Agent	AI Agent의 개발 및 운영 관리
LLMOps	LLM 모델관리
MLOps	다양한 ML모델 관리
DataOps	다양한 소스에서 수집 되는 멀티모달 데이터 통합관리

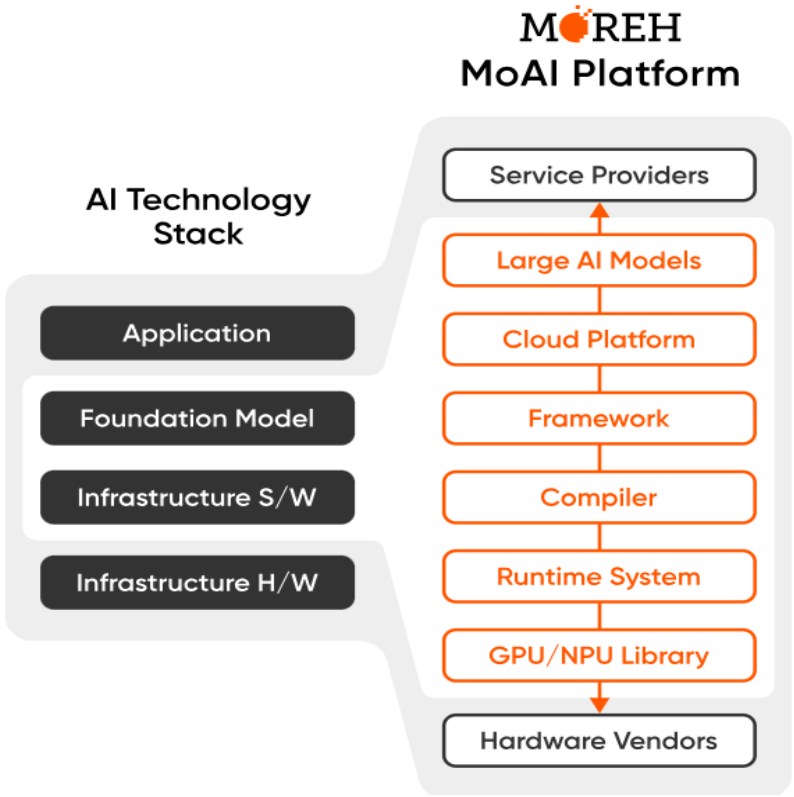
3. AIOps 플랫폼 - Backend.AI, AIPub / MoAI

NVIDIA
NVIDIA 인증 SW - 국내 기업

BACKEND.AI

AIPub

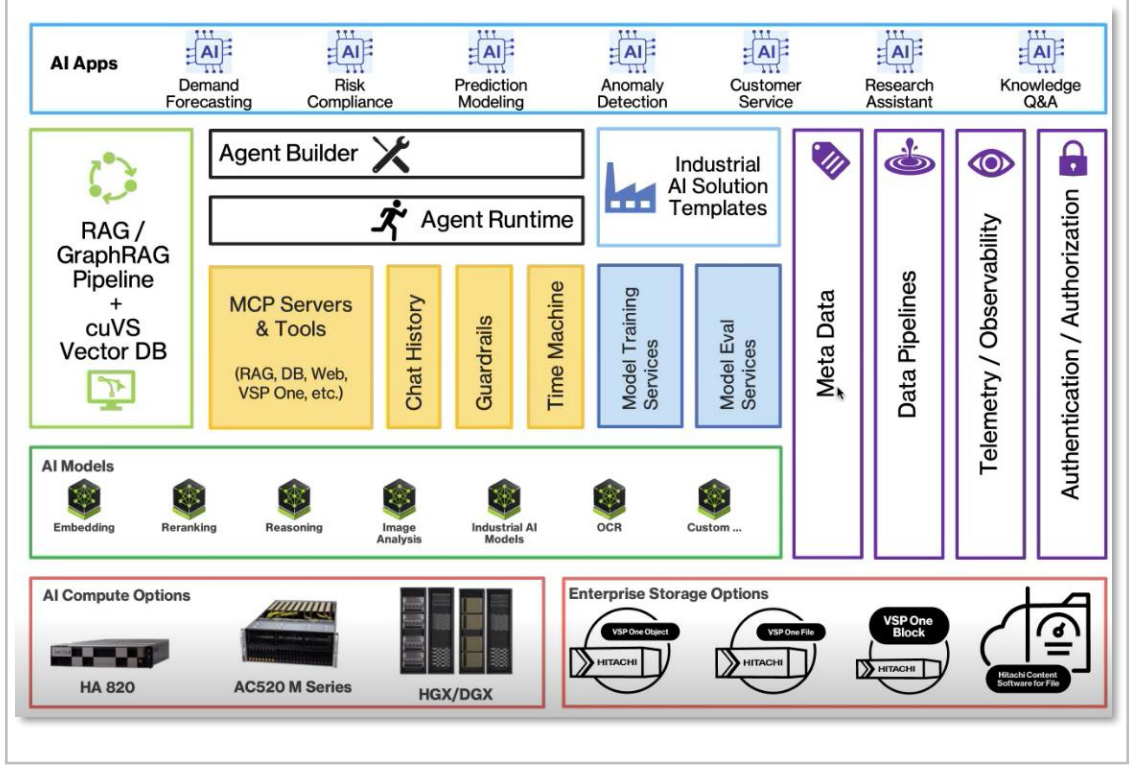
AMD
AMD 시리즈 B 투자 - 국내 기업



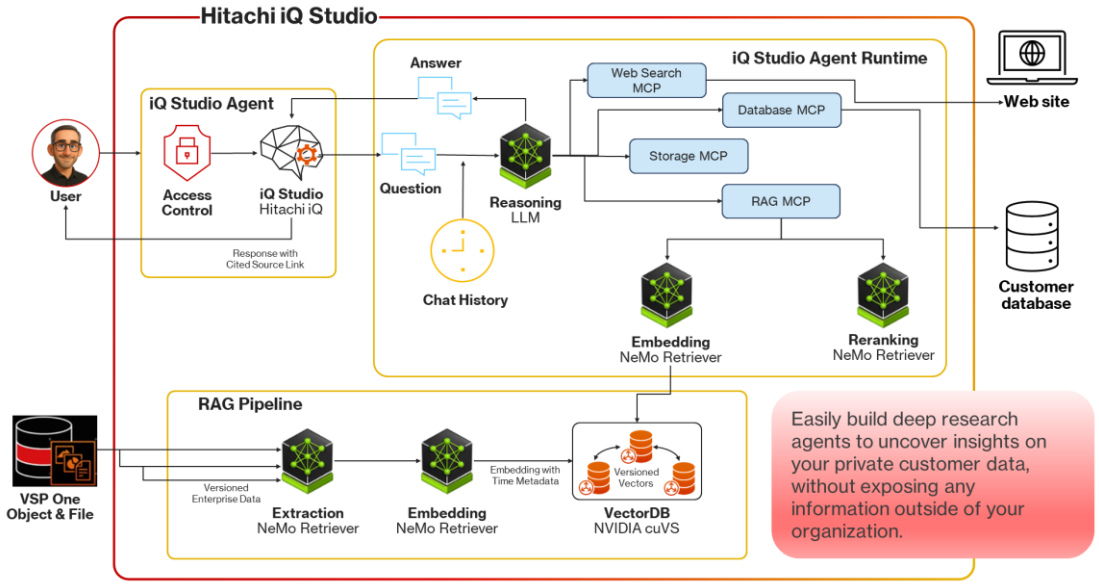
4. AI 오케스트레이션 플랫폼 - Hitachi iQ Studio

- Hitachi iQ Studio는 기업의 데이터를 기반으로 자율적인 판단과 실행이 가능한 에이전틱 AI를 노코드로 신속하게 구축하고, NVIDIA와의 협업을 통해 온프레미스 환경에서 인프라 최적화와 강력한 거버넌스를 동시에 제공하는 **통합 AI 오케스트레이션 플랫폼**

Hitachi iQ Studio Architecture



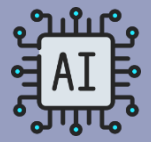
Hitachi iQ Studio Sample Workflow



5. AI 도입 이슈에 대한 고민 해결

01

AI 인프라 기술
- 통합 AI 플랫폼 제공



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

02

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)

03

에코시스템 구축
- 다양한 솔루션 접목

- AI 적용을 위해 필요한 다양한 솔루션 접목
- 기존의 방식과 다른 접근 체계 가능
- AI Ops, SingleStore 등

04

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 구축하여 운영 효율성 확보

III. AI 플랫폼 구성 및 사례

1. AI 플랫폼 구성 HW
2. IT 대기업 AI 플랫폼 인프라(자체 LLM 개발)
3. 대학병원 AI 분석 플랫폼 구축
4. 대기업 DX GPU AI 인프라 구축(연구 개발)
5. AI 인프라 도입 시 고려 사항

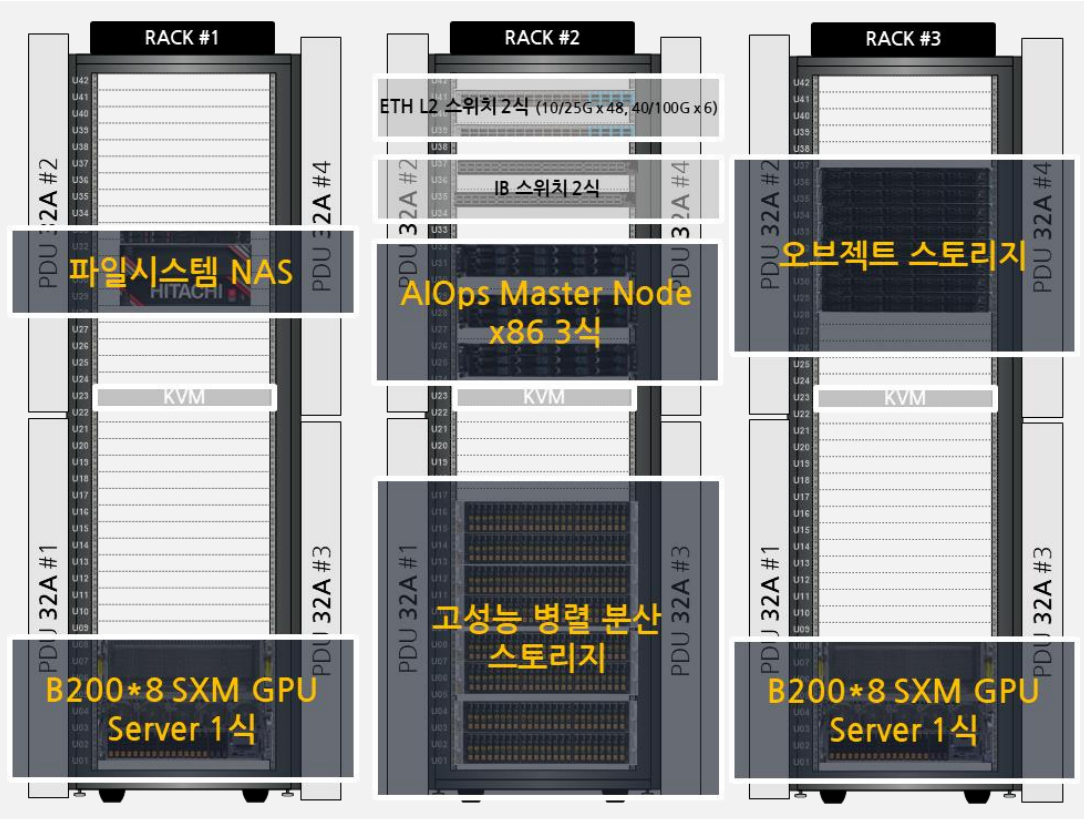
1. AI 플랫폼 구성 HW

GPU Infra + 고성능 Storage + 고성능 NW + AIOps SW

소규모 추론용 AI 플랫폼

고성능 학습/추론용 AI 플랫폼

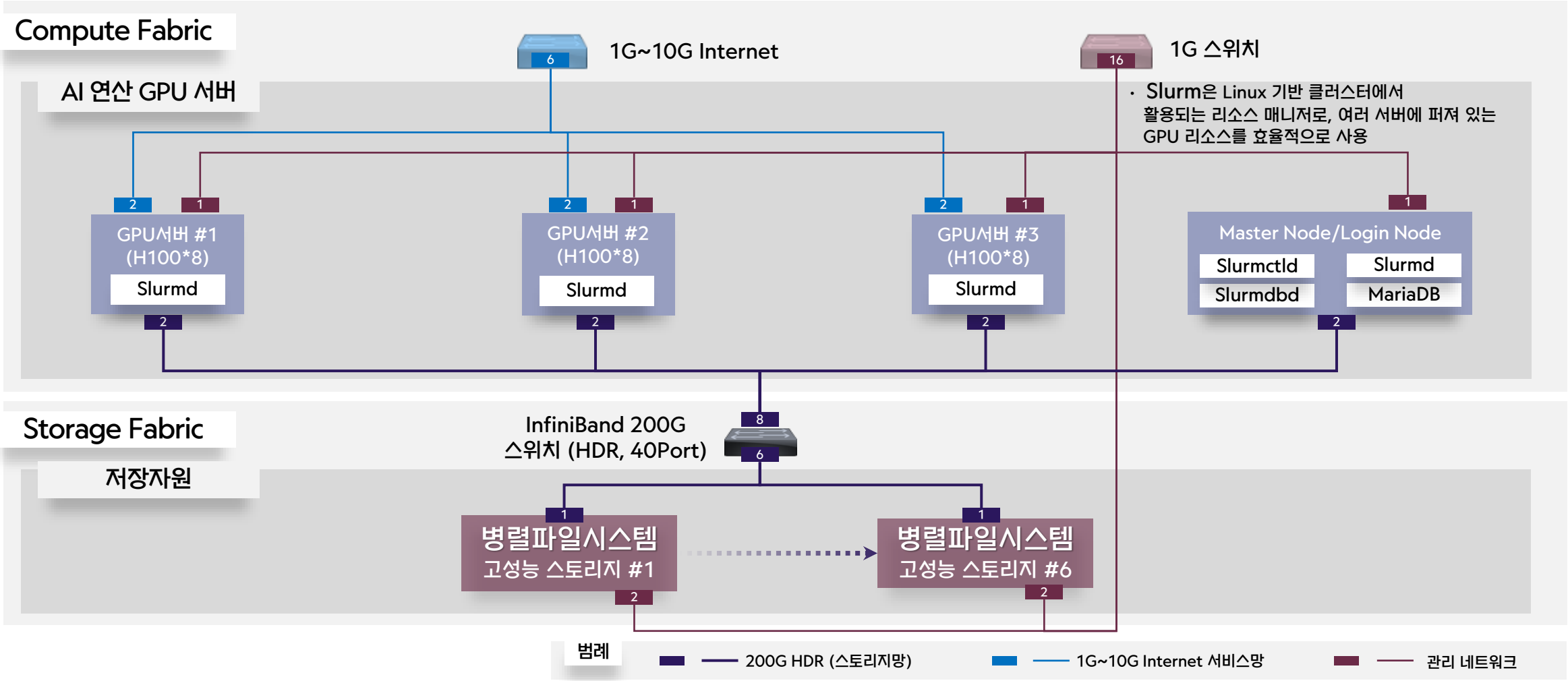
초고성능 학습/추론용 AI 플랫폼



*상기 랙 실장도의 서버 이미지 및 수치, mount 크기(ex:2U → 1U)는 변경될 수 있습니다.

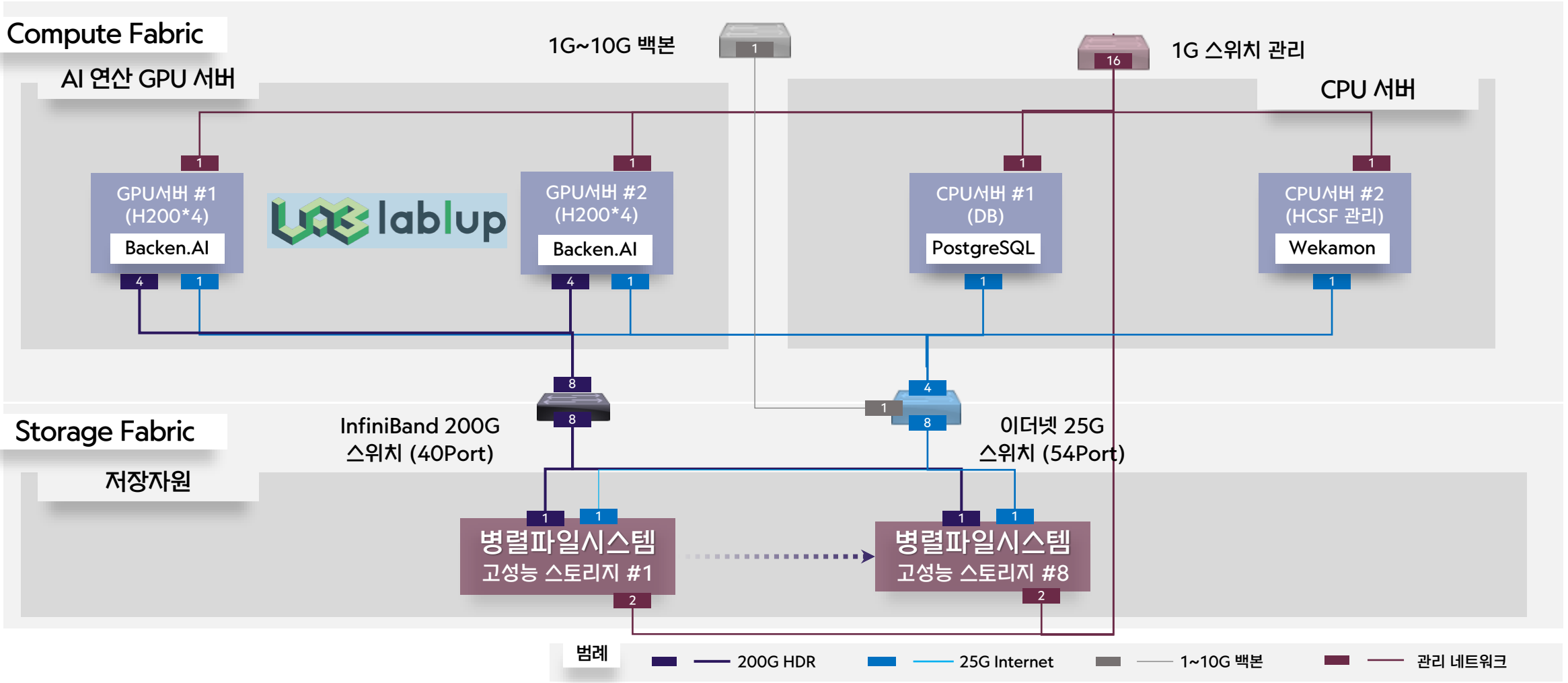
2. IT 대기업 AI 플랫폼 인프라(자체 LLM 개발)

- HPC 클러스터 (HW & SW) & HCSF(분석특화 저장소) & 오픈소스 클러스터 관리 솔루션 구축 사례



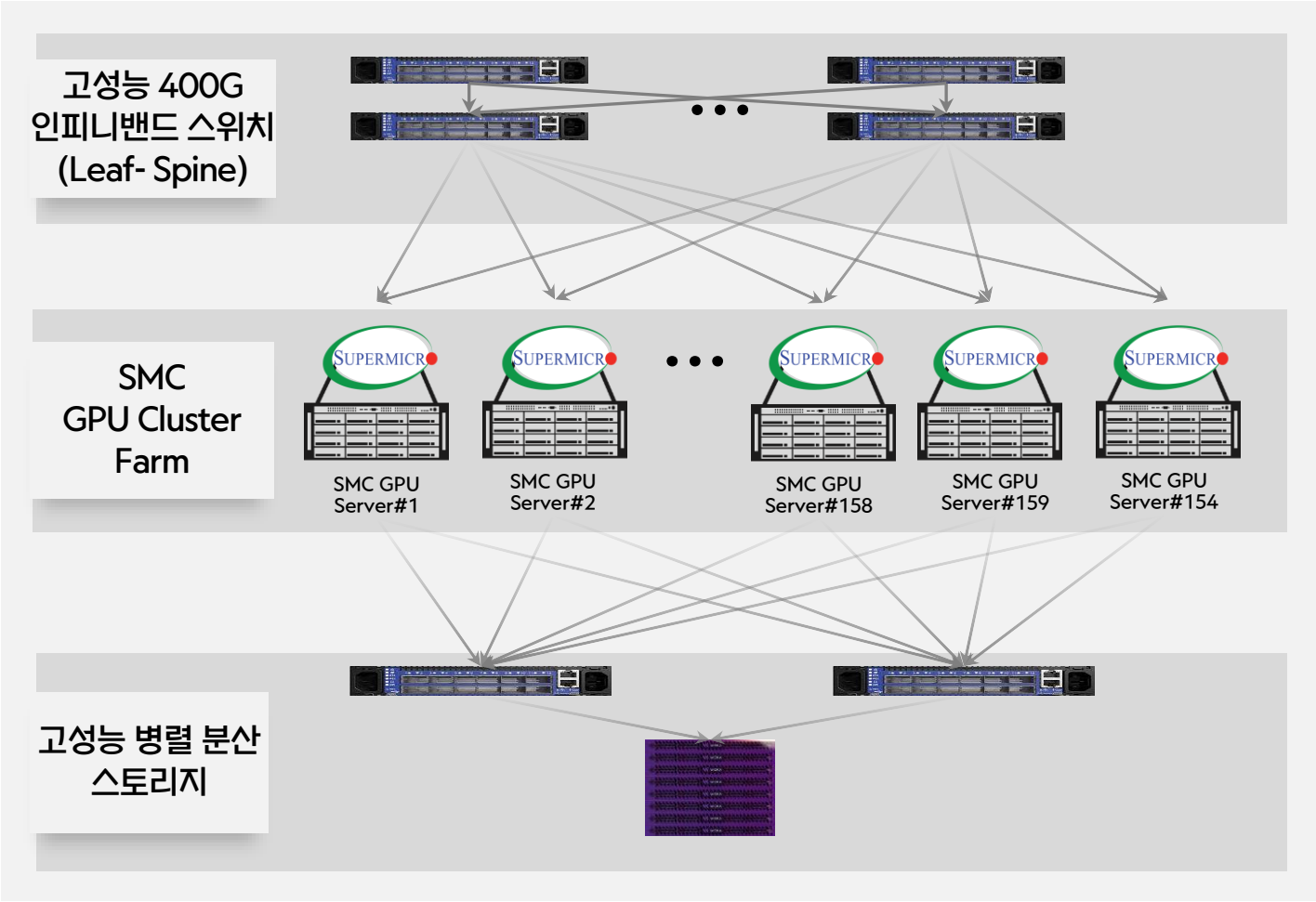
3. 대학병원 AI 분석 플랫폼 구축

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



4. 대기업 DX GPU AI 인프라 구축(연구 개발)

- 고성능 AMD GPU Server Cluster Farm 구축 사례



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

SuperMicro + HIS 선정 이유

- SuperMicro Server의 모듈식 설계에 따른 유연성과 확장성, 서버 성능을 보장하는 안정성
- HIS의 전문 기술지원 인력을 통한 최고의 직접 기술지원 서비스와 HPC/고성능스토리지/AI 구축 레퍼런스

AMD GPU 선정 이유

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

5. AI 인프라 도입 시 고려 사항

01

다양한 에코 파트너 협업 체계 구축 확인

빠르게 변화하는 AI시대 대응

02

AI 플랫폼 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.

HS효성인포메이션시스템
권동수 전문위원 (his-dskwon@hshyosung.com)