

[금융 AI의 Next Step] 금융 혁신을 위한 AI 적용 및 확장 방안

권동수 전문위원 (HS효성인포메이션시스템)

최근영 CTO (지음AI)

2024년 12월 12일

Agenda

1. AI 트렌드
2. AI 기술 및 적용 사례
3. AI 확장형 구축 방안 및 사례

1. AI 트렌드

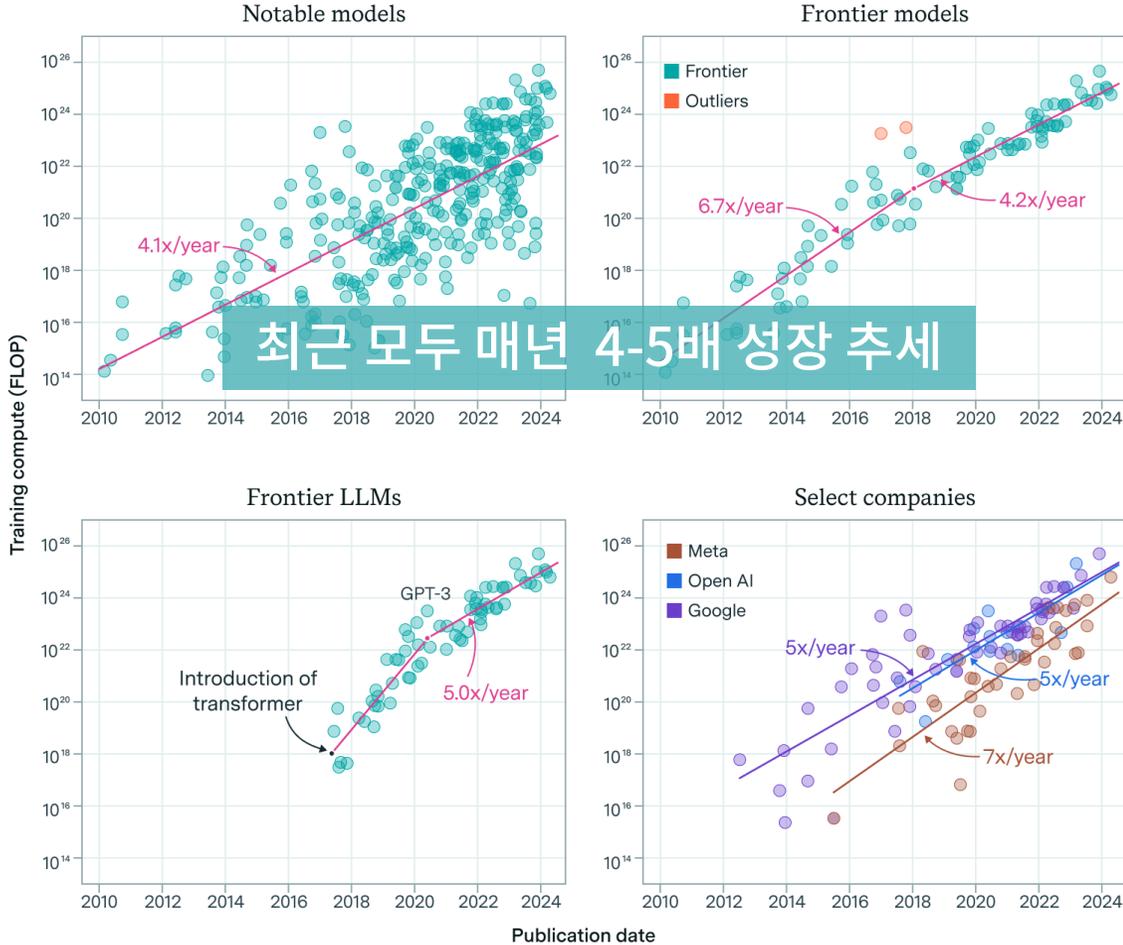
1. ANI(약인공지능), AGI(일반인공지능) and ASI(초인공지능)
2. 생성형AI 도입 시 고려사항
3. 금융 비즈니스 이해와 기술역량을 갖춘 AI 파트너



ANI(약인공지능), AGI(일반인공지능) and ASI(초인공지능)

Summary of compute trends in AI

EPOCH AI



AI Artificial Intelligence

ANI

2024년
(Artificial Narrow Intelligence)

- 음성 인식, 이미지 처리, 자연어 처리, 추천 시스템, 자율주행 등

AGI

(Artificial General Intelligence)

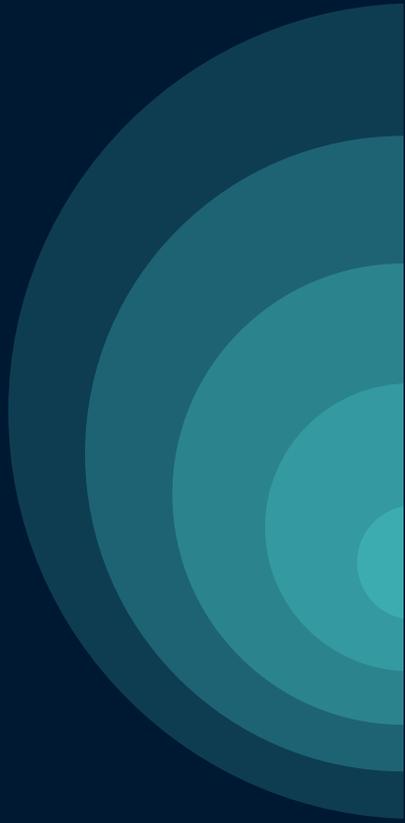
- 다양한 문제 해결, 과학적 발견, 창의적 작업 등

ASI

(Artificial Super Intelligence)

- 인간의 능력을 초월하는 모든 분야 활동 영역

비즈니스에 성공적인 생성형 AI를 도입하고 궁극적으로 지속가능 성장을 위해 기업이 고려할 5가지 중요한 사항



#1 목표 파악 : 맞춤형, 확장성, 성과 지향적 생성형 AI

#2 핵심 엔터프라이즈 표준과 구체적인 기술 지침 결합

#3 하이브리드 환경에서 생성형 AI 지원

#4 고성능·로딩 최적화·AI 어플라이언스 도입

#5 금융 비즈니스의 이해와 기술 역량을 갖춘 AI 파트너 선택

[출처] Five actions enterprises can take to accelerate GenAI adoption and improve business outcomes (Hitachi Vantara)

금융 비즈니스 이해와 기술역량을 갖춘 AI 파트너

AI 인프라 부터 전문 AI 서비스까지! AI 에이전트를 위한 최상의 조합

HS 효성인포메이션시스템

- AI/HPC 인프라 전문업체
- GPU, 네트워크, 초고성능 저장소, K8S, HCI
- AI Platform 구축



iumAI 지음에이아이

- 금융 서비스 전문 AI 업체
- 100여개 이상 금융 부문 컨설팅, 아키텍처, 전략수립
- AI 프로젝트 수행



2. AI 기술 및 적용 사례

1. 현실 AI 프로젝트의 Success & Pain Point
2. ZIUM AI GNOS 기술 및 서비스
3. 금융분야 망분리 개선 준비 사항
4. AI 프로젝트 추진 방법



현실 AI 프로젝트의 Success & Pain Point

Success Point

서비스 구현에 집중

양질의 데이터

단계적 혁신 로드맵

최신 기술 도입시기 결정

도메인 지식과 기술 보유 파트너

VS

Pain Point

기술 구현에 집중

다량의 데이터

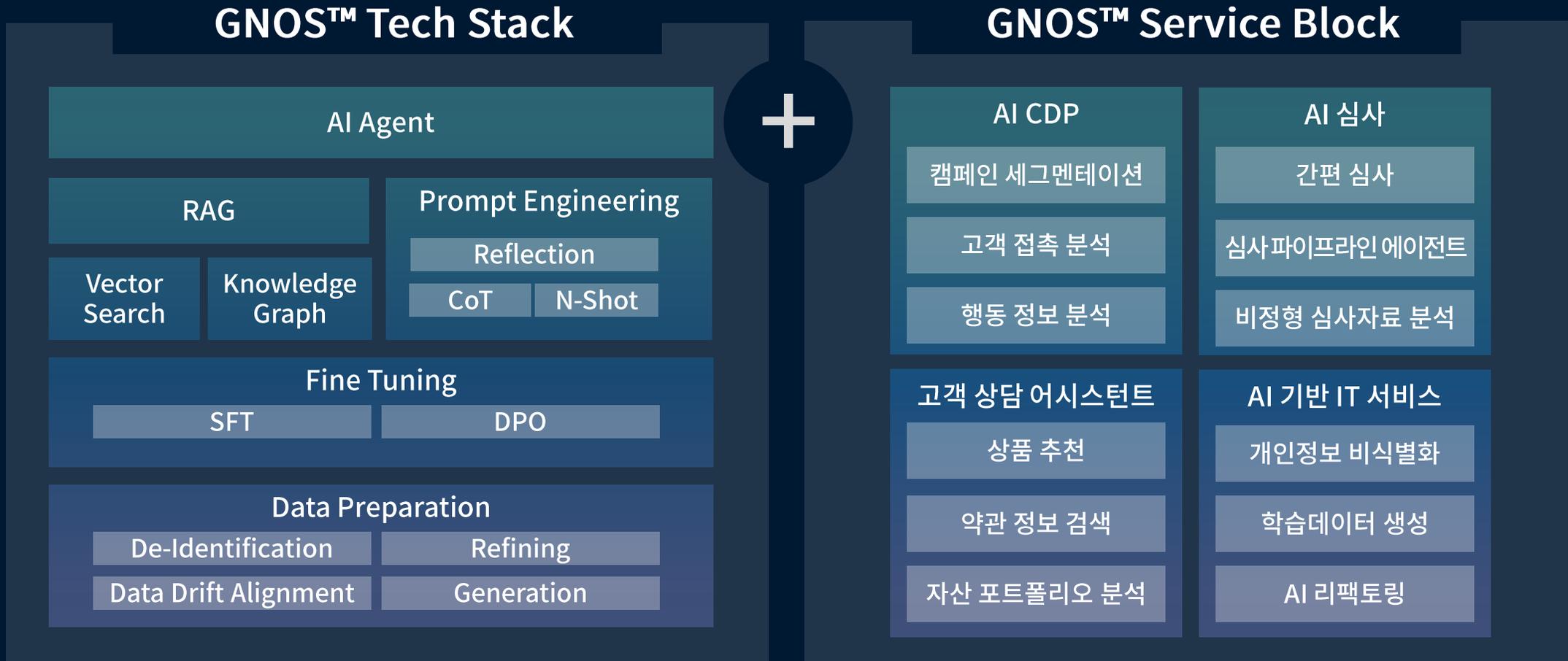
빅뱅 방식의 혁신

모든 최신 기술 적용

혁신의 전 과정을 자체 역량만으로

ZIUM AI GNOS 기술 및 서비스

금융 기업의 안정적인 AI 사업 추진을 위한 최적의 기술과 서비스



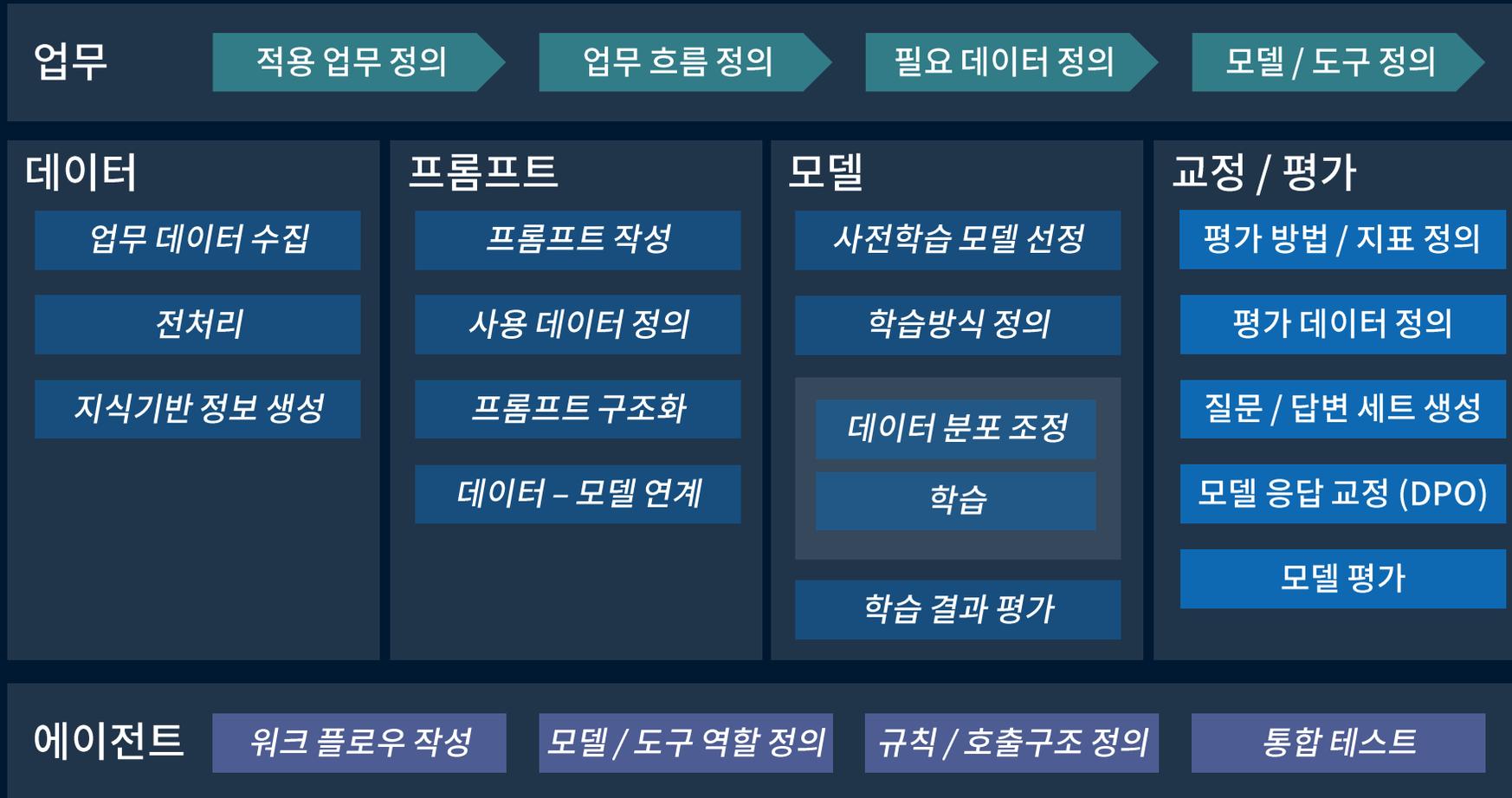
망분리 개선 내용



준비 사항

- 내부 보안 관리 체계 강화 (Strengthening Internal Security Management System)
- 제로 트러스트 보안 (Zero Trust Security)
- 모범 사례 기반 리스크 평가 (Risk Assessment Based on Best Practices)
- 가명 처리 개인신용정보 활용 (Utilization of Pseudonymized Personal Credit Information)
- 복원 정보 암호화 및 별도 관리 (Encryption and Separate Management of Recovery Information)
- 제3자 리스크 관리 강화 (Strengthening Third-party Risk Management)

AI 프로젝트 추진 방법



- ✓ 할 일과 역할을 명확하게
- ✓ 프롬프트 정보 구조화
- ✓ 지식기반 시스템 구성
- ✓ 특화된 모델의 조합
- ✓ 학습 데이터 분포 조정
- ✓ 명확한 평가 지표

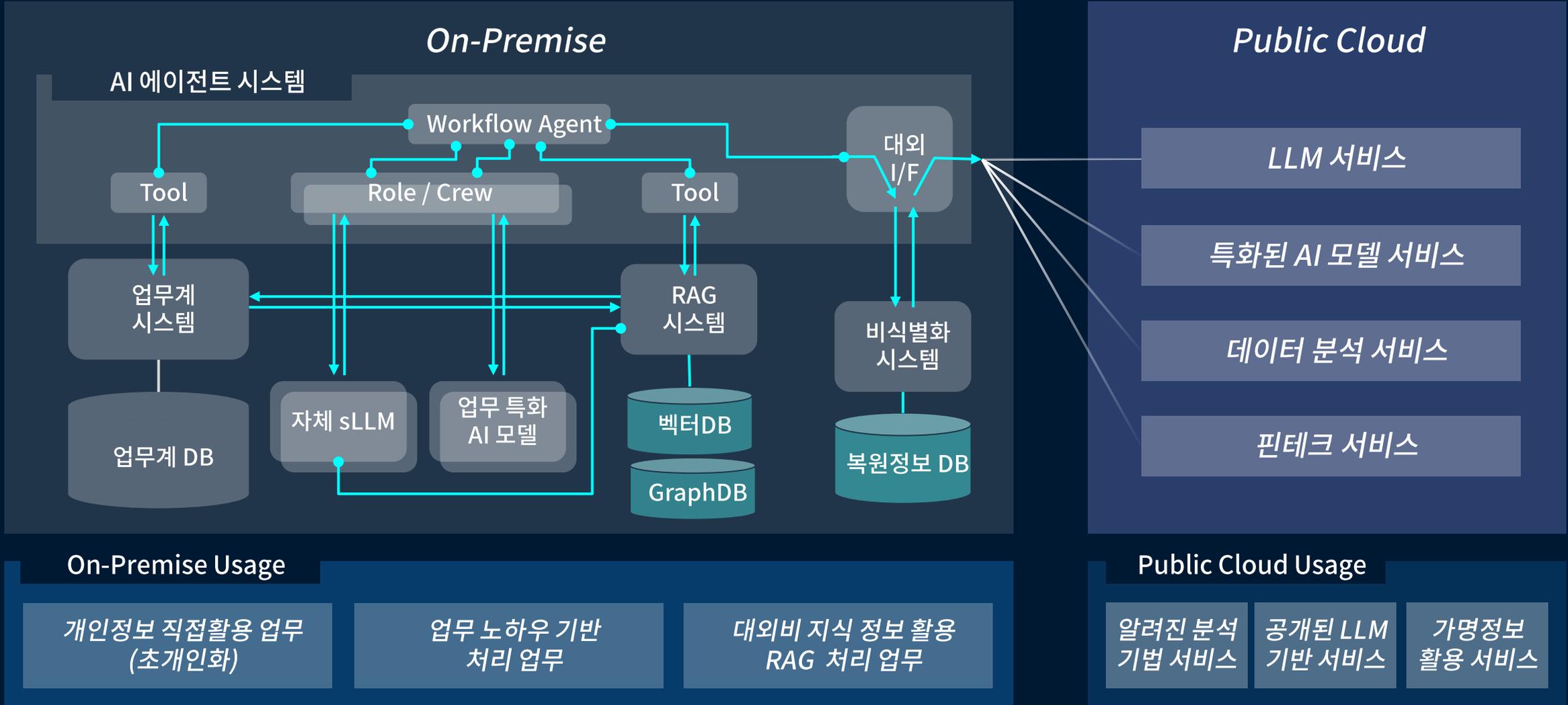
3. AI 확장형 구축 방안 및 사례

1. AI 확장형 하이브리드 구성
2. 금융 AI 플랫폼 구축
3. 통합 AI 플랫폼
4. 확장형 구축 사례
5. AI 플랫폼 구축을 위한 체크포인트



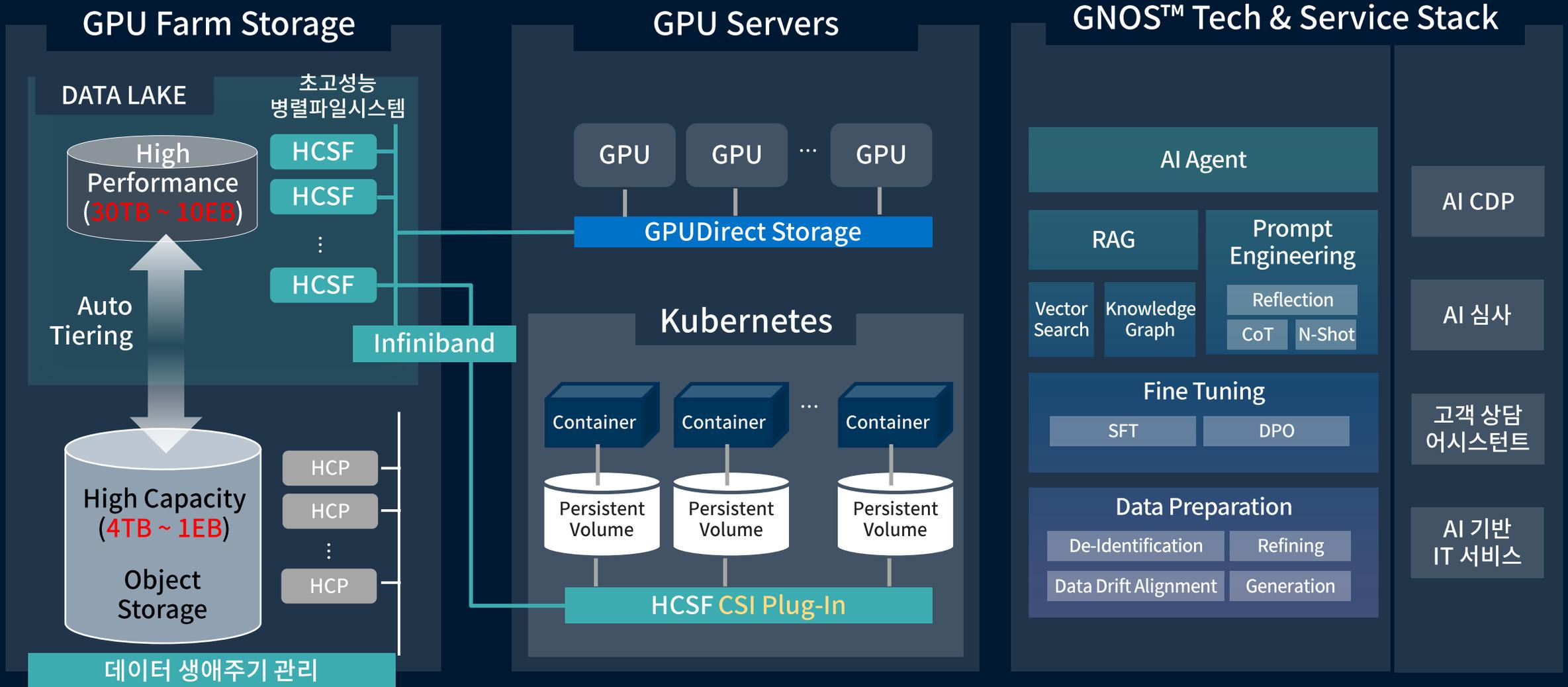
AI 확장형 하이브리드 구성

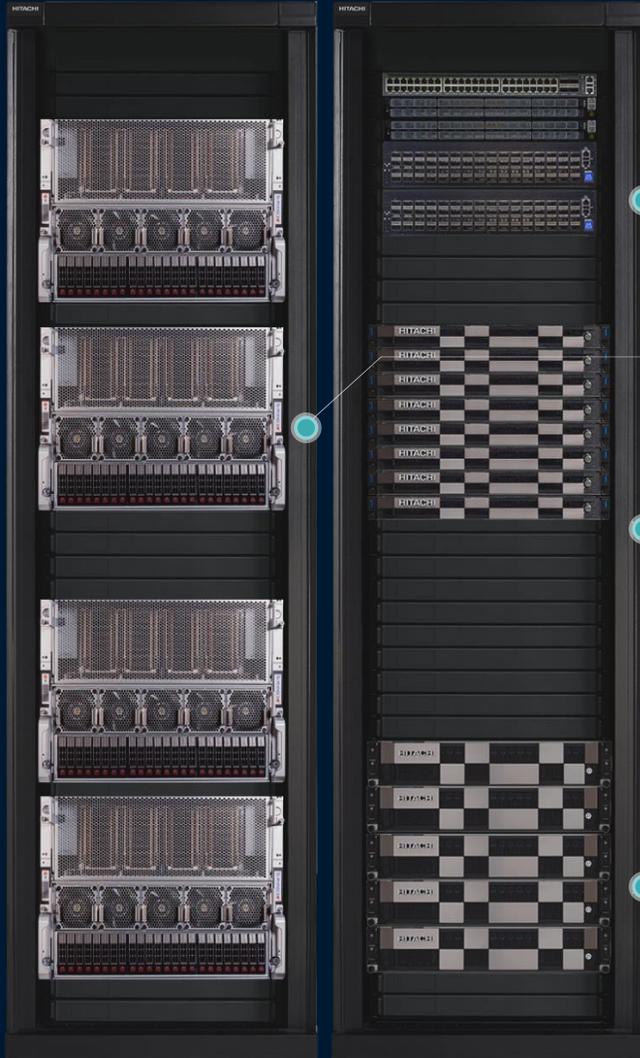
3. AI 확장형 구축 방안 및 사례



금융 AI 플랫폼 구축

3. AI 확장형 구축 방안 및 사례





고속 네트워크

200Gb/400Gb Ethernet & InfiniBand
연산 및 저장자원 간 성능 최적화를 위한 초고속 네트워크

GPU 서버

HGX, H100, H200, B200 등
NVIDIA가 공인한 슈퍼마이크로의 최신 GPU서버

초고성능 병렬 파일시스템

Hitachi Content Software for File
GPU환경, AI워크로드에 최적화된 NVMe 기반 초고성능 분산 병렬파일시스템

오브젝트 스토리지

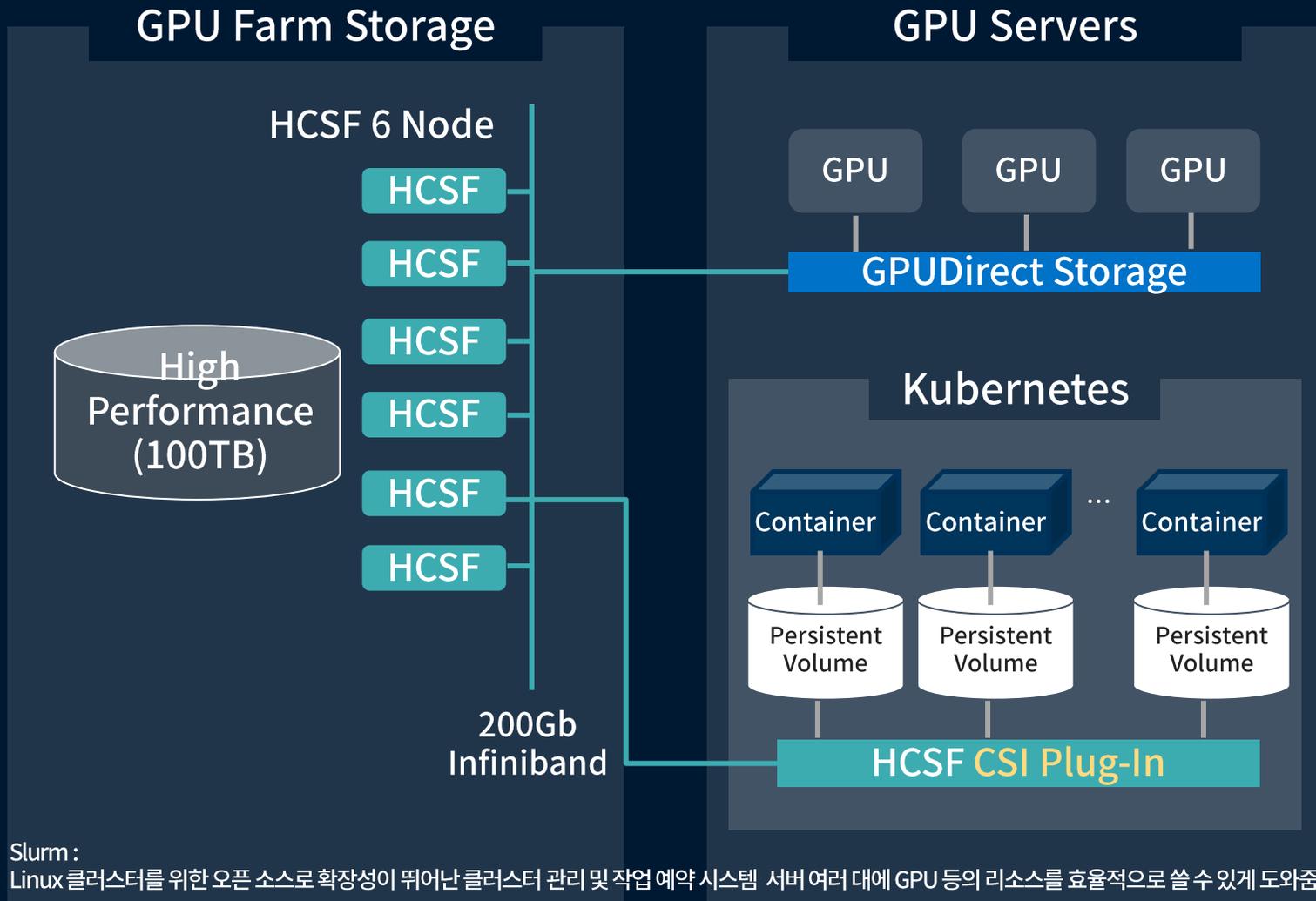
Hitachi Content Platform
확장성이 뛰어난 오브젝트 스토리지와 정책기반 자동화 티어링으로 효율성 향상

AI 인프라는 HS효성에 맡기고,
‘비즈니스 혁신’에만 집중하세요.

<p>01</p> <p>사전 설계·검증된 통합 AI 인프라 플랫폼 제공</p>	<p>02</p> <p>데이터 생애주기를 위해 데이터레이크 구성</p>	<p>03</p> <p>컨설팅부터 설계, 구축, 운영까지 통합 수행</p>
<p>04</p> <p>IDC 공인 국내 하이엔드/고성능 스토리지 시장 10년 연속 1위 기업</p>	<p>05</p> <p>파트너 에코시스템과 긴밀한 협업으로 최적의 AI Ops 제공</p>	<p>06</p> <p>LLM, AI/HPC 프로젝트 다수 수행</p>

사례 1. 국내 그룹 SI 기업 (100TB)

3. AI 확장형 구축 방안 및 사례



사업 목적

- sLLM 모델 개발 업무를 위한 AI/HPC 인프라 구축

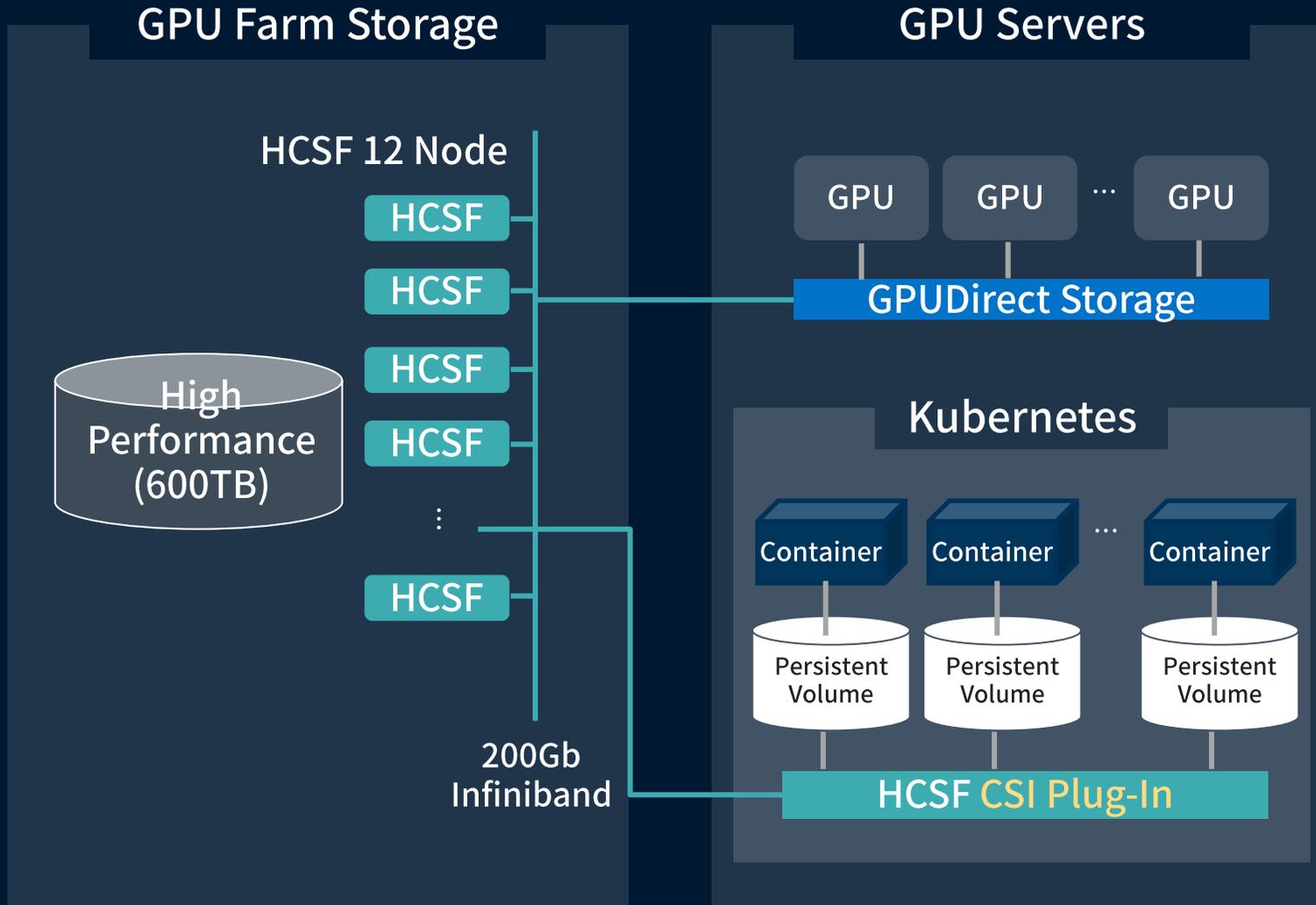
구축 내용

- 최신 Supermicro GPU 서버, 고성능 스토리지 HCSF, 고속 네트워크, Slurm AI HPC 인프라 구축
- Kubernetes 연계(CSI) 기능을 제공하여 볼륨 운영 최적화/효율화 지원

도입 효과

- MLPerf 테스트 검증된 최고 성능 GPU 서버와 GPUDirect Storage 적용으로 GPU 연산 성능 최적화
- 확장에 따른 성능 저하 이슈 없는 무제한 급증 설 지원 저장 자원 제안
- AI 플랫폼 구축 경험 있는 전문가로 구성된 전담 조직으로 AI 플랫폼 연계 기술 지원 보장

사례 2. 국내 AI 서비스 기업 (600TB)



사업 목적

- AI 교육 및 자사 AI 학습 모델 테스트 환경을 위한 프라이빗 클라우드 기반 HPC 인프라 구축

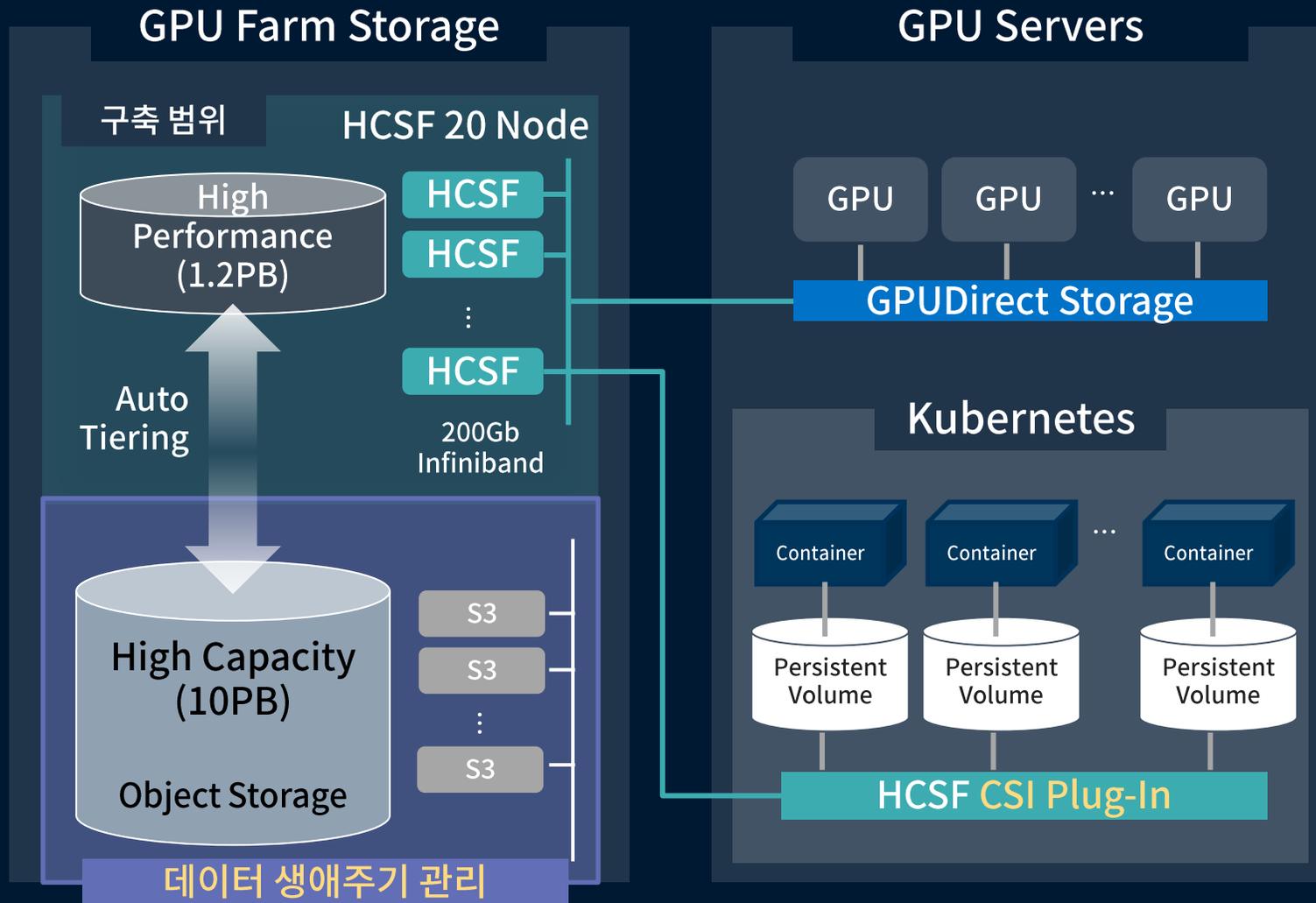
구축 내용

- GPU 환경 구현을 위한 고성능 스토리지 HCSF 구축
- GPU 성능 최적화 GDS 지원 스토리지 구축
- Kubernetes 연계(CSI) 기능을 제공하여 불륨 운영 최적화/효율화 지원

도입 효과

- 다수 개발자/분석가/교육생 공동 운영을 위한 GPU 전용 스토리지 인프라 확보
- 향후 요구 사항 변화(데이터 증가, 사용자 증가 등)에 유연한 대응 확보
- GPU 서버 인프라 고려한 스토리지 최적화

사례 3. 국내 R&D기업 (1.2PB)



사업 목적

- 자연어처리알고리즘 학습 성능 개선을 위한 GPU 서버와 전용 스토리지를 도입

구축 내용

- GPU 환경 구현을 위한 고성능 스토리지 HCSF 구축
- GPU 성능 최적화 GDS 지원 스토리지 구축
- Kubernetes 연계(CSI) 기능을 제공하여 볼륨 운영 최적화/효율화 지원
- 고객사 S3 스토리지와 Tiering 구성

도입 효과

- 자연어처리 기계 학습 연산 성능 향상
- 서버 인프라 구성 고려한 스토리지 최적화
- 자체 S3 스토리지 활용성 확보를 통한 비용 효율적인 구성



AI/HPC 요구 사항 확인 및 PoC/BMT 필수!!!
IOPS, 처리량 요구 성능 검증



AI/HPC 시스템 데이터 레이크 구축 경험 필수!!!
대용량의 파일 처리, 안정적인 확장



국내외 실 사례를 통한 국내 기술력(인력) 필수!!!
장애 지원, 신규 AI 솔루션 연계 지원, 금융전문가

감사합니다.

