

합리적인 GPU 선택

SKT Enterprise GPUaaS

Powered by Lambda

Basic Infra For AI
GPU는 SKT와 함께하세요



SKT Enterprise GPUaaS

Powered by Lambda

• NVIDIA H100 SXM5

- ① 32 ~ 2,040 GPU Cluster with 3,200Gbps InfiniBand
- ② GPU 간 NVLink(7.2TB/s) 통신
→ PCIe 대비 7배 뛰어난 성능

• High-End 병렬 스토리지

- ① 200Gbps 연결로 초고속 환경 지원
- ② 최상의 퍼포먼스 제공
(90GB/s read, 65GB/s write Peak)

• 완벽하게 격리된 Private Cloud

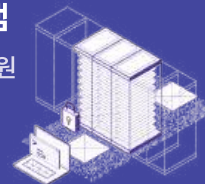
- ① 사용자만 접근 가능한 시스템 구조
- ② 고객별 방화벽 제공으로 추가 보안 설정

• 유연한 네트워크 호환성

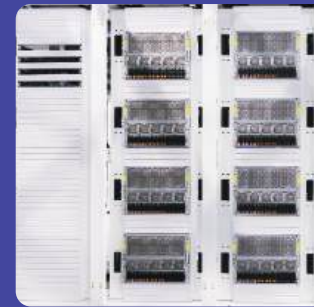
- ① Internet Access
- ② 전용회선
- ③ Public Cloud 연동

• 다양한 프로모션 프로그램

- ① 고객 지원 · 교육 프로그램 지원
- ② 약정 기간 & Upfront 할인
- ③ 첫구매·재구매 할인 혜택



(GPUaaS : GPU as a Service)



AI 데이터 센터 (Lambda - US 리전 전경)

- ✓ 처음으로 GPU 를 사용한다면
- ✓ 최신 GPU 로 바꾸고 싶다면
- ✓ 검증된 파트너를 찾고 있다면

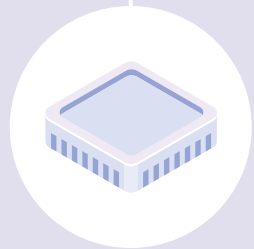
합리적인 GPU 선택

SKT Enterprise
gpu-sales@sk.com

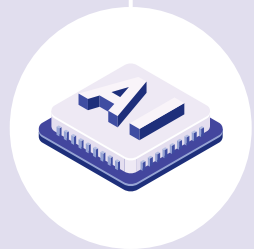


Global TOP 수준 슈퍼 컴퓨터 'TITAN' 운용

아키텍처 최적화로
성능 효율성* 87% 달성
(슈퍼 컴퓨터 Top 500 평균 67%)
* 이론 성능 대비 실제 성능 비율
국내 제 1금융권 구축·운영 경험



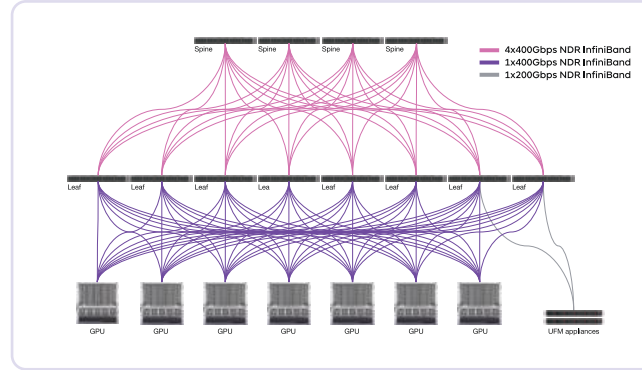
미래형 데이터 센터
국내 최초 44kW 랙 도입
(랙당 H100 서버 4대 탑재)
Global Top Player와
AI DC 기술 공동 개발
(수냉식, 액침 냉각)



NVIDIA 전략적 파트너
AI DC 최적화 기술
서울 리전 공동 구축
SK브로드밴드
가산 데이터센터 운용

① 최고 성능을 위한 Architecture

Infiniband 기반의 SXM5 Cluster



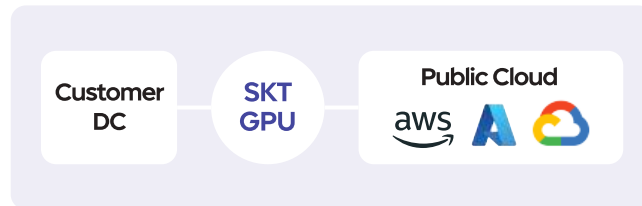
- 8 x 400Gbps (NDR) Infiniband 구성
- Full mesh Non-blocking 구조

② SKT·람다 기술로 검증된 구성·성능

- 국내 최고 수준의 비용 대비 성능 실현 (PFLOPS/원)
- PCIe 기반 상품 대비 약 7배 빠른 네트워크 속도
- High-End 병렬 스토리지 지원

③ 최적의 Hybrid 연동 호환성

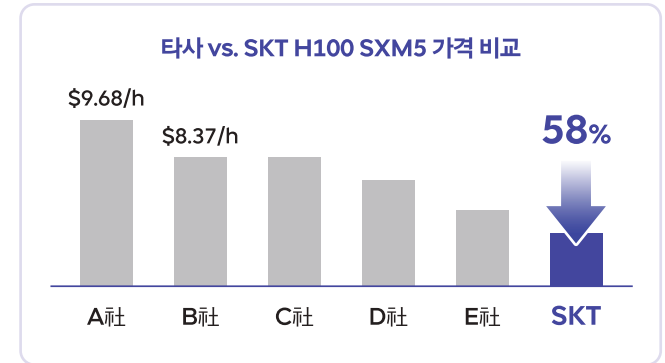
Public Cloud·전용 회선 연동 지원



- SKT 통합 бил링을 통한 일원화 관리
- SK텔레콤·SK브로드밴드 회선 기반 가격 경쟁력 확보

④ 국내·외 CSP 대비 최대 58% 가격 인하

AI Cloud Infra Architecture & 성능 최적화 집중
AI 모델 학습·추론에 적합한 비용 효율적 솔루션 제공
· 최대 2,040개 GPU 단일 Cluster 구성 지원



⑤ 추가 비용 없는 과금 체계

TCO 관점의 비용 부담 완화

미국 A사	7,700만원(월)	9.2억원(연) 추가 비용 발생
한국 B사	1억원(월), 12억원(연)	추가 비용 발생

SKT - DTO 과금 할인

[1PB/월 DTO 비용 비교]

- DTO^(*) 과금 할인으로 Foundation model, serving model size에 따른 추가 비용 우려 없음
- GPU/Storage 등 필수 요소에 한정된 과금 체계

^(*) DTO: Data transfer out

