

국내 최초 도심형 프리미엄 AI 인프라

마곡 B300

with NVIDIA HGX B300

국내도심GPU 클라우드의
새로운표준



마곡 B300 도입 안내

국내 최초 NVIDIA HGX B300 **Dedicated 클러스터**로 모든 AI 워크로드를 가속하고, **제로트러스트 · SPU 기반 스토리지 레벨 암호화로** 완벽한 데이터 보안을 보장합니다.

왜 마곡 B300인가?

- 01 실시간 추론 서비스에 최적화 H100 대비 추론 성능 약 10배
- 02 LLM / 멀티모달 AI 기반 서비스 토큰 비용 최적화
- 03 최고 수준의 데이터 보안 + 물리 격리 + 국내 데이터 주권

도입 대상 산업군

 AI · IT 테크	 바이오 · 신약	
 금융 · 공공	 모빌리티	 국방 · 보안

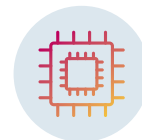
도입 상담 및 제안 요청

GPUaaS · NPUaaS · QCaaS

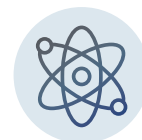
핵심 서비스 포트폴리오



GPUaaS HGX B300 클러스터
LLM 사전학습 · 파인튜닝 · RAG 추론
NVLink · Dedicated 단독점유



NPUaaS
Furiosa RNGD 기반 추론 특화 서비스
GPU 대비 1/4 수준전력효율



QCaaS
2026년 하반기 54큐비트급
양자컴퓨터 도입
하이브리드 워크로드 확장 옵션으로 제공


유연한 4단계 과금 모델

On-demand	Prepaid
Reserved	Dedicated

최대 23% 할인 · 자원 우선 배정 SLA 보장

완벽한 엔터프라이즈급 데이터 보안


제로트러스트


SPU 가속기


물리격리

국내 최초 도심형 NVIDIA HGX B300+ 전용 스토리지 프리미엄 AI 인프라

도심에서
가장 안전하게, 가장 빠르게,
가장 저렴하게 즉시 활용 가능한
프리미엄 GPU 인프라의 새로운 표준

- 01 실시간 추론 서비스 최적화 H100 대비 추론 성능 **약 10배**
- 02 SPU 기반 전용 스토리지 레벨 암호화
- 03 제로트러스트 기반 네트워크 보안
- 04 데이터 주권 + 즉시 활용

국내 도심 GPU 클라우드의
새로운 표준

GPU 시간당
₩10,000 부터
3년 약정 최저가

하이퍼스케일러 GPU 대비 최대
약 2-3배 저렴, AI API 토큰 비용 대비
최대 50-80% 절감

-23%
3년 약정 시 최대 할인율

OpEx 전환
수십억 CapEx 없이 즉시 활용

토큰비용 ↓
오픈 LLM 자체 운영으로 API 종량제 탈피

성능 우위에 더해, 합리적인 비용 구조와 명확한
가격 정책으로 예산을 보호합니다.

압도적비용경쟁력

하이퍼스케일러대비

약 2-3배
절감

AI API 토큰비용대비

최대
50-80%
절감

GPUaaS 시작가격 (per GPU·시간)

On-demand
₩13,000

필요할 때 즉시 사용,
약정 없음
단기 PoC · 검증용

Prepaid
₩11,700

선 결제 충전식,
약정 없음
-10% 할인 적용

1년약정
₩10,500

1년 예약 사용
-19% 할인 적용

3년약정
₩10,000

3년 예약 사용
(최저가)
-23% 할인 적용

* GPU 1장·시간기준/ 3년약정시최대23% 절감