

서버 임차 규격서

1. 사업 개요

사업명	2026년도 AI 기반 엔드프로덕트 상용화 지원 Type1 사업
과제명	AI 기반 스마트 방수총 및 화점인식 진압 사족보행로봇 솔루션
목적	과제 수행을 위한 AI 모델 학습·추론용 GPU 서버 임차

2. 계약 개요

계약형태	임차(임대차)
수량	4대
임차기간	5개월
공급가액	180,000,000원(173,904,000원)
부가세액	18,000,000원(17,390,400원)
총액(VAT포함)	198,000,000원(191,294,400원)
납품(설치)기한	계약일로부터 60일 이내

3. 서버 하드웨어 규격(1대 기준, 이상 스펙)

폼팩터	4U 랙마운트형 GPU 서버
CPU	AMD EPYC(Turin 9005 계열) 2소켓, 소켓당 16코어/32스레드 이상, 3.0GHz 이상
메모리	DDR5 ECC Registered 512GB 이상 (5200MHz 이상)
스토리지	NVMe Enterprise SSD 3.84TB 이상, 2.5" 핫스왑 베이 8개 이상
GPU	NVIDIA 프로페셔널 GPU 2장 이상, 장당 메모리 96GB(GDDR7) 이상, CUDA 코어 24,000개 이상
네트워크	10Gbps BaseT 듀얼포트 이상 96GB(GDDR7) 이상, CUDA 코어 24,000개 이상
전원	3200W 이상 이중화(N+1), 80 PLUS TITANIUM 등급 이상
인증	NVIDIA 서버 인증(NVIDIA Qualified Server)
원격관리	Out-of-band 관리(BMC/iKVM) 지원
OS/드라이버	Ubuntu, CUDA Toolkit, cuDNN, GPU 컨테이너 런타임 설치 지원
보증	무상 2년 이상

4. GPU 관리·운영 소프트웨어 기능 요구사항

GPU 클러스터 모니터링	클러스터/서버/GPU 단위 사용률·온도·상태 실시간 확인
멀티테넌시/작업 스케줄링	다수 개발자 독립 컨테이너 환경, GPU 자원 자동/수동 스케줄링
멀티노드·멀티GPU 분산학습	복수 서버 GPU 동시 활용 분산학습 지원
GPU 가상 분할 지원	GPU 1개를 다중 인스턴스로 분할 활용 가능
개발환경 즉시 구축	사전 구축 컨테이너 이미지로 즉시 딥러닝 개발 가능
사용량 통계 리포트	GPU/서버/사용자별 자원 사용량 추출 및 리포트 제공

3. 계약방법

- 일반경쟁입찰(최저가+적격심사) 또는 협상에 의한 계약