

단일규격 구매사유서

□ 사업명 : 2026년도 AI 기반 경제정책정보 제공 서비스 고도화를 위한 AI 운용 인프라 증설

□ 구매목적

- 본 물품은 「AI 기반 경제정책정보 제공 서비스 고도화」 사업의 2차년도 추진을 위한 AI 운용 인프라 증설을 목적으로 도입하고자 하는 물품임.
- 1차년도에 구축된 GPU 서버에 GPU 및 메모리를 증설하여, 대규모 언어모델(LLM) 및 벡터기반 검색처리를 위한 RAG 기반 경제정책정보 서비스의 연산 처리능력을 고도화하고자 함.

□ 사유

- NVIDIA H200 NVL GPU 동일 모델 증설의 필요성
 - 1차년도 도입 GPU 서버에는 NVIDIA H200 NVL GPU 2장이 NVLink Bridge로 연결되어 있으며, 본 사업에서는 이를 4-way 구성으로 확장하고자 함.
 - NVLink Bridge를 통한 다중 GPU 병렬처리 구성은 물리적·소프트웨어적으로 완전히 동일한 GPU 모델에서만 가능하며, 이종 GPU 혼용 시 NVLink 구성이 불가능하여 LLM 추론 가속 효과를 확보할 수 없음.
- 기존 메모리(32GB Dual Rank x8 DDR5-6400 Memory Kit (HPE))와의 동일 사양 유지 필요성
 - 서버 메모리는 동일 시스템 내에 상이한 사양·인증 체계의 모듈이 혼재할 경우 전체 메모리 동작 속도가 최저 속도로 하

향(다운클럭)되거나 부팅 시 비인증 모듈 경고가 발생할 수 있어, 1차년도에 장착된 메모리와 동일한 DDR5-6400 사양으로 증설하여야 함.

- 또한 비인증 메모리 장착 시 메모리 관련 장애에 대한 HPE 정품 보증(Warranty) 적용에 제약이 발생할 수 있어, 1차년도 도입 서버의 자산 가치 보호를 위해 HPE 정품 인증(HPE SmartMemory) 메모리로 단일규격 지정이 불가피함.

□ 대상 물품

- 물 품 : NVIDIA H200 NVL GPU, 32GB Dual Rank x8
DDR5-6400 Memory Kit (HPE)