

용도설명 및 구매사양서

1. 품 명	(국문) 인공지능 융합 분석 연산 시스템 (영문) Integrated AI Analytics and Computation System		
2. 수량/단위		3. 금 액	원
4. 연구과제명	가계 정(구매)	5. 품 목 번 호 (H S K No.)	
6. 제작회사명 (국 명)			
7. 국내제작 가 능 여 부		8. 기기의 중복 보유여부	
9. 사용용도/ 내구년수			
10. 품명 및 세부사양		단위	수량
A.기본구성사양 1.인공지능 융합 분석 연산 시스템 - NVIDIA DGX B200 1440GB (8-GPU NVIDIA B200 180GB System), Full Support 3 Years B.기타사양 1. 딥러닝 학습용 고성능 GPU 슈퍼컴퓨터 (NVIDIA DGX B200 1440GB) - CPU : Dual Intel Xeon Platinum 8570 Processors (Total 112 Cores, 2.1 GHz, MAX 4 GHz) - Memory : 2TB - GPU : 8x NVIDIA B200 180GB (1440GB Total) - Storage1(OS) : 2x 1.92TB M.2 NVME SSDs - Storage2(Cache) : 8x 3.84TB U.2 NVME SSDs - NVLink : 18x NVLinks Per GPU (1800GB/s Per GPU to GPU Bi-directional Bandwidth) - NVSwitch : 2x NVIDIA NVSwitches (Total 14.4TB/s Bi-directional Bandwidth) - Networking1(Cluster) : 4x OSFP port serving 8x Single-port ConnectX-7 VPI		SET	1.

- Networking2(Storage) : 2x Dual-port QSFP112 BlueField-3 DPU
(Up to 400Gb/s Infiniband/Ethernet)
- TDP: 14.3kW max
- Dimension : 10U / 444mm(H) x 482.2mm(W) x 897.1mm(L)
- DGX Software Stack Installed :
 - . Ubuntu Linux Host OS / Driver / NV Docker / Accelerated Solution
 - . NVIDIA GPU 사용 프로세스 병목 탐색 및 가이드 필수 제공
(클러스터 구성시)
 - . Docker 기반 GPU Distribute 환경 구축 및 성능 Test (MPI Horovod)
 - . CUDA Parallel Programming 교육 지원
 - . NVIDIA KOREA 정품 (제조사 정품공급확약서 제출 必)
- Software & Support :
 - . 8x NVIDIA AI Enterprise Essentials Subscription per GPU, 3 Years
 - . Ubuntu Linux OS / Driver / CUDA / 기본 DL Framework 설치Nvidia AI enterprise를 통해 가상환경 및 최신 DL framework가 사용자 선택에 따라 설치 가능해야 함.
 - . Mandatory Installation DGX B200, Partner Installed
- 워런티 3년
- 서버제조사 정품공급 및 기술지원확약서 제출

2. 딥러닝 추론용 고성능 GPU 서버 (NVIDIA L40S 384GB)

- CPU : 2x AMD EPYC 9354 (3.25G/256M/32C/280W)
- Memory : 16x 64GB DDR5 PC5 ECC/REG
- Storage1(OS) : 2x SSD 960GB SATA3 2.5"
- Stprage2(Data) : 2x SSD 1.92TB SATA3 2.5"
- NIC1 : 1x 10G 2Port ETH LAN (SFP+)
- NIC2 : 1x 10G 2Port ETH LAN (RJ45)
- RAID Card : 1x 8-CHANNEL SATA/SAS RAID Card
- PCIe Slot1 : 1x PCIe 5.0 Slot (x16 link) for HBA/RAID cards
- PCIe Slot2 : 8x FHFL PCIe 5.0 x16 (in x16 Link) slot for Dual Slot GPU
- PCIe Slot3 : 1x PCIe 5.0 x16 (x16 link) + 1x PCIe 5.0 (x8 link)
- PSU : 4x 3000W 80Plus Titanium PSU
- Dimension : 4U / 439(W) x 175.6(H) x 789(D) mm
- GPU Option : Up to 8x GPU (Dual Slot GPU)
- GPU : 4x NVIDIA L40S 48GB GDDR6 PCIe (Passive)
- Installed :
 - . Ubuntu Linux OS / Driver / CUDA / 기본 DL Framework 설치가장 최신의 안정화버전의 OS 및 모든 주변장치들이 정상적으로 동작할 수 있는 드라이버가 설치 되어야 함. 설치된 GPU 호환의 가장 최신 CUDA 및 DL framework 설치 되어야 함. 의존성 문제 없어야 함. 세부 버전은 납품시 담당자와 협의하여 해당 버전 설치 지원 되어야 함.
 - . NVIDIA GPU 사용 프로세스 병목 탐색 및 가이드 필수 제공
(클러스터 구성시)
 - . Docker 기반 GPU Distribute 환경 구축 및 성능 Test (MPI Horovod)
 - . CUDA Parallel Programming 교육 지원
- 워런티 3년
- 서버제조사 정품공급 및 기술지원확약서 제출

① 현장 납품(O) , ② 현장 설치(O) , ③ A/S필요() , ④ 기타()