

Compute

Multi-node GPU Cluster

대규모 고성능 AI 연산을 위한 다량의 GPU 제공 서비스

Multi-node GPU Cluster는 대규모의 고성능 AI 연산을 위해 물리 GPU 서버를 가상화 없이 제공하는 서비스입니다. GPU가 장착된 Bare Metal Server 2대 이상을 사용하여 다수의 GPU를 클러스터링 할 수 있으며, Samsung Cloud Platform의 고성능 스토리지 및 네트워킹 상품과 연계하여 편리하게 GPU 서버를 사용할 수 있습니다.

손쉬운 GPU 아키텍처 구성

Samsung Cloud Platform의 Multi-node GPU Cluster는 고성능 NVIDIA SuperPOD 아키텍처를 적용한 Bare Metal Server를 제공하여, GPU를 활용한 다수의 사용자 또는 대규모 AI 모델 학습의 고성능 분산 워크로드 처리가 가능합니다.

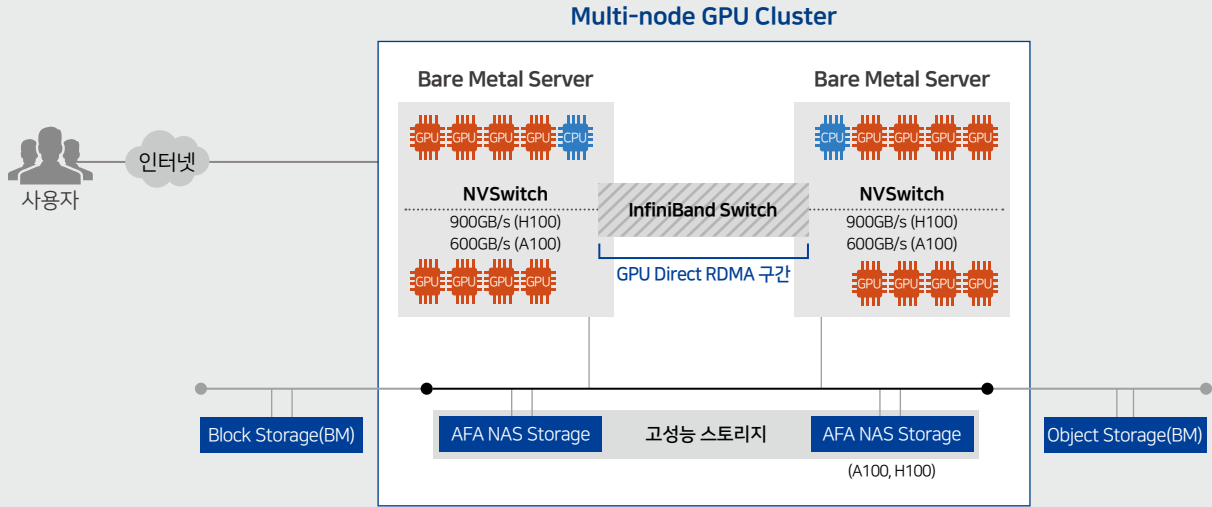
고성능 네트워크 연계

Multi-node GPU Cluster는 Samsung Cloud Platform의 네트워크 자원과 연동하여 고성능이 필요한 AI Job을 수행할 수 있습니다. InfiniBand 스위치를 이용하여 GPU Direct RDMA (Remote Direct Memory Address) 환경을 구성함으로써 GPU 메모리간 데이터 IO를 직접 처리하여 AI/Machine Learning 고속 연산이 가능합니다.

고성능 스토리지 연계

Multi-node GPU Cluster는 Samsung Cloud Platform의 다양한 스토리지 자원과 연계하여 사용할 수 있습니다. 고속 네트워크와 직접 연동한 고성능 SSD NAS File Storage를 사용 가능하며, Block Storage 및 Object Storage도 연계하여 사용 가능합니다.

서비스 구성도



주요 기능




- GPU Bare Metal Server 생성/관리
 - NVIDIA GPU 8장을 장착한 표준 GPU Bare Metal Server 제공
 - ※ Internal NVMe Disk, NVIDIA NVSwitch, NVIDIA NVLink
 - RDMA SW Stack의 OS 표준 이미지 제공 (OS : Ubuntu)
- 고성능 처리
 - InfiniBand 스위치를 이용하여 GPU Direct RDMA 환경 구성
 - 고성능 SSD File Storage 제공 (A100, H100)
- 스토리지 및 네트워크 연동
 - OS 디스크 외 추가 연결 스토리지 (Block, Object) 제공
 - 서버넷/IP 및 VPC Firewall 연계 설정

요금 기준

- 과금
 - 무약정 : 자원 신청 후 할당 시간 기준 시간 단위 과금
 - 약정 : 1년/3년 약정 조건에 따른 할인 적용. 신청 자원 기준 월 단위 정액 과금
 - ※ 약정 기간 내 해지 시 위약금 발생

FOR MORE INFORMATION

SAMSUNG SDS

 www.samsungsds.com / cloud.samsungsds.com
 contact.sds@samsung.com / scp_sales@samsung.com
 youtube.com/samsungsds

