

지역특화형 초거대 AI 인프라 플랫폼을 위한 맞춤형 설계와 구현 전략

The Customized Design and Implementation Strategies for Region-Specific Hyper-Scale AI Infrastructure Platforms

Jihwan Kim, Daenyeon Kim, Seungman Chun

Gyeongbuk Institute of IT Convergence Industry Technology

Gyeongsan-si, Gyeongsangbuk-do, Republic of Korea

E-mail: jhkim2@gitc.or.kr

Abstract

This study emphasizes the need for region-specific hyper-scale AI infrastructure, tailored to each area's unique industrial needs, data regulations, and energy conditions. Analyzing cases from Gyeongsangbuk-do, Gyeongsangnam-do, and Gwangju, it confirms that such customized AI infrastructure enhances local industrial competitiveness and supports sustainable regional development, with potential for broader application.

Keywords : Large-scale AI, Region-Specific AI Platforms, AI Infra-Structure design

Introduction

지역별로 특화된 초거대 AI 인프라 플랫폼의 설계와 구축 필요성이 증가하고 있다. 지역특화형 초거대 AI 인프라 구축은 단순히 대규모 컴퓨팅 자원을 확보하는 것을 넘어, 각 지역의 특성을 반영한 맞춤형 전략이 필요하다. 따라서 본 연구는 이러한 배경을 바탕으로, 지역의 고유한 요구 사항을 반영한 맞춤형 초거대 AI 인프라 플랫폼의 설계 및 구현 전략을 제안하는 것을 목표로 한다. 지역별로 요구되는 산업적 필요, 데이터 처리 규제, 에너지 공급 환경 등을 고려한 맞춤형 인프라 구축 방안을 제시함으로써, 보다 효율적이고 효과적인 AI 인프라의 운영을 도모하고자 한다.

Implementation Strategies

○ 고성능 AI 연산 클러스터 설계
초거대 AI 모델 학습을 위해 고성능 GPU 서버 클러스터가 필요하다. 각 지역의 산업적 요구와 인프라에 맞춰 적합한 GPU 서버를 선택해야 한다. PCIe 기반 GPU 서버는 비용 효율적이고 중소형 데이터센터에 적합하며, NVIDIA DGX는 고성능 AI 학습에 최적화된 고가 솔루션으로 대규모 프로젝트에 적합하다. NVIDIA HGX는 대규모 병렬 연산이 필요한 데이터 집약적 산업과 대형 데이터센터에 적합하다.

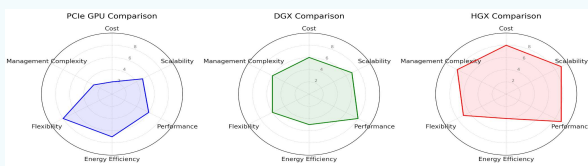


Fig. 1 Comparison of AI Compute Clusters

○ 네트워크 및 데이터 전송 최적화
고성능 AI 환경에서는 데이터 전송 속도와 네트워크 성능이 중요하다. Infiniband는 초고속 데이터 전송과 낮은 지연 시간으로 금융 및 의료 데이터 처리에 적합하지만 설치와 관리 비용이 높다. RoCE는 이더넷 인프라를 활용해 저비용으로 네트워크 성능을 최적화할 수 있어, 예산이 제한된 중소형 기업에 적합하다.

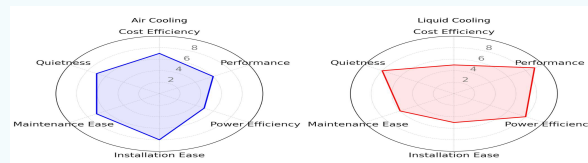


Fig. 2 Comparison of High-Performance Network

○ 스토리지 시스템 선택

AI 모델 학습에는 대량 데이터 처리를 위한 고성능 입출력(I/O)과 확장성이 필요한 스토리지 시스템이 필요하다. 병렬 파일시스템(Lustre, GPFS)은 실시간 데이터 처리에 유리하며 대규모 데이터센터에 적합하고, 클라우드 파일시스템(Ceph, MinIO)은 초기 비용이 낮고 유연성이 높아 클라우드 기반 인프라에 적합하지만 실시간 데이터 처리에는 한계가 있다.

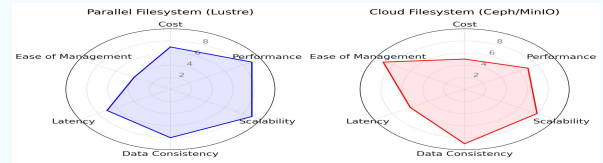


Fig. 3 Comparison of Storage File Systems

○ 냉각 시스템 설계

초거대 AI 연산 클러스터의 발열 관리를 위해 공랭식과 수랭식 냉각 시스템을 고려해야 한다. 공랭식은 설치가 간편하고 중소형 데이터센터에 적합하며, 수랭식은 고밀도 연산 환경의 대규모 데이터센터에 효과적이다. 에너지 여건에 따라 적합한 냉각 방식을 선택해 효율성을 극대화할 수 있다.

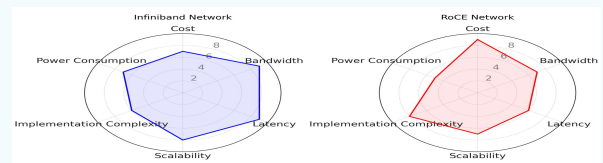


Fig. 4 Comparison of Cooling Systems

Case Study

○ 경상북도

258억 원을 투자하여 PCIe 기반 GPU 서버와 RoCE 네트워크를 활용한 AI 인프라를 구축하고, 농업 및 제조업 등 다양한 산업에 AI 기술을 적용하여 생산성과 효율성을 높인다.

○ 경상남도

227억 원을 투자하여 제조업에 특화된 초거대 AI 모델과 응용 서비스를 개발하며, 주요 제조기업에 AI 솔루션을 적용하여 효율성을 극대화한다. 퍼블릭 클라우드 기반 IaaS와 하이브리드 인프라를 구축하여 자원 관리의 유연성을 제공한다.

○ 광주광역시

4,061억 원을 투자하여 AI 중심 산업 융합 집적단지를 조성하고, HGX GPU 서버와 Infiniband 네트워크를 도입하여 고성능 AI 학습과 대규모 데이터 처리를 지원하는 인프라를 구축하여 산업 경쟁력을 강화한다.

Conclusion

이 연구는 각 지역의 산업적 요구를 반영한 맞춤형 초거대 AI 인프라 전략의 필요성을 강조한다. 경상북도는 농업 및 제조업, 경상남도는 제조업, 광주광역시시는 산업 경쟁력 강화를 목표로 각각 특화된 AI 인프라를 구축하였다. 연구는 이러한 맞춤형 AI 인프라가 지역 산업 경쟁력과 지속 가능한 발전에 중요한 역할을 하며, 다른 지역으로 확산 가능성이 있음을 제안한다.