

# 컨테이너 가상화 기반 AI 개발 운영 플랫폼 아키텍처 설계 및 구현 방안

이용구<sup>1</sup>, 김민규<sup>2</sup>

<sup>1,2</sup>한국수력원자력 중앙연구원

Yong-Ku Lee<sup>1</sup>, Min-Kyu Kim<sup>2</sup>

<sup>1,2</sup>Central Research Institute, Korea Hydro & Nuclear Power Co., Ltd.

lee.yongku@khnp.co.kr, kim.minkyu@khnp.co.kr

**요약** 본 논문은 현대 기업 환경에서 인공지능(AI) 시스템의 개발부터 배포, 운영에 이르는 전체 라이프사이클을 효율적으로 관리하기 위한 컨테이너 가상화 기반 AI 개발 운영 플랫폼 아키텍처를 제안한다. 기존의 베어메탈 서버 또는 서버 가상화 기반의 단일 환경 AI 개발 및 운영 방식으로는 급속히 증가하는 AI 워크로드의 복잡성과 다양성에 대응하기 어려워, 컨테이너 기술을 활용한 일관되고 유연한 개발 및 운영 환경의 필요성이 대두되고 있다. 이에 본 연구는 쿠버네티스(Kubernetes) 기반의 컨테이너 오케스트레이션 환경과 Kubeflow를 통합하여 AI 개발, 학습, 파인튜닝, 배포 등 전체 라이프사이클을 체계적으로 관리할 수 있는 아키텍처를 설계하고 구현 방안을 제시한다. 제안된 아키텍처는 다계층 구조로 설계되어 인프라 레이어부터 컨테이너 오케스트레이션, AI 플랫폼, 데이터 관리, AI 개발 파이프라인, 모니터링 및 운영 자동화, 보안 및 거버넌스 레이어에 이르는 종합적인 솔루션을 제공한다. 특히 본 연구는 AI 워크로드의 특성을 고려한 자원 관리 전략, GPU 등 고성능 하드웨어 활용 방안, 분산 학습 및 서빙 아키텍처, 그리고 MLOps 관점에서의 자동화된 CI/CD 파이프라인 구축 방안을 중점적으로 다룬다. 본 연구에서 제안하는 아키텍처는 AI 모델 개발 생산성 향상, 컴퓨팅 자원 활용 효율성 증대, 운영 안정성 확보, 그리고 확장성 측면에서 기존 방식 대비 유의미한 개선을 가져올 것으로 기대된다. 본 연구를 통해 기업이 제한된 네트워크 환경 및 내부 인프라를 활용하여 현대적인 AIOps 환경을 구축하고, AI 기술을 활용하여 기업 요구사항에 효과적으로 적용할 수 있는 실질적인 방안을 제시하고자 한다. 현재 AI 개발 운영 플랫폼 구축을 위한 인프라 도입이 추진되고 있는 단계이며, 구축 완료 후 효과에 대한 우수성 검증과 확대 적용 가능성도 함께 모색할 예정이다.

**핵심주제어:** 컨테이너 가상화, 쿠버네티스, Kubeflow, AIOps, MLOps, AI 라이프사이클 관리, 클라우드 네이티브 아키텍처, 엔터프라이즈 AI 플랫폼