

# 하이브리드 클라우드 전반에서 ML 개발, 훈련, 테스트 및 배포

## 주요 내용

온프레미스 또는 퍼블릭 클라우드에서 지능형 애플리케이션을 신속하게 제공합니다.

비즈니스의 AI 도입을 간소화하고, AI 도입을 늘리고, 유연성을 제공합니다.

데이터 과학자, 데이터 엔지니어, DevOps 팀이 효과적으로 협업할 수 있도록 지원하는 일관된 사용자 환경으로 팀 간 AI/ML 운영 일관성을 확립합니다.

클라우드, 온프레미스 또는 엣지 등 데이터 위치와 가까운 곳에서 AI/ML 워크로드를 빌드, 훈련, 배포 및 모니터링할 수 있는 기능으로 하이브리드 클라우드 유연성을 확보합니다.

Open Data Hub는 Red Hat OpenShift 상에 서비스로서의 AI(AlaaS) 플랫폼을 구축하기 위한 청사진과도 같습니다. 따라서 Jupyter 및 Kubeflow와 같은 업스트림 작업에서 상속받으며 Red Hat OpenShift AI의 기반이 됩니다.

## 인공지능 및 머신 러닝 배포 가속화

생성형 AI(gen AI) 도입이 빠르게 늘어남에 따라 인공지능, 머신 러닝 및 딥러닝(AI/ML/DL)이 기업과 조직에 중요한 요소로 빠르게 자리 잡았습니다. Gartner는 현재, 기업의 85%가 2026년경에는 생성형 AI API(애플리케이션 프로그래밍 인터페이스, application programming interface)를 사용하거나 생성형 AI 지원 애플리케이션을 배포할 것으로 내다보고 있습니다.<sup>1</sup> 그러나 이러한 기술을 배포하는 것은 복잡한 작업이 될 수 있습니다. 데이터 과학자는 AI/ML/DL 모델을 개발하기 위해 노력하는 과정에서 빠르게 진화하는 틀 간의 연계 부족을 자주 경험합니다. 이러한 격차는 데이터 과학자, 소프트웨어 개발자, IT 운영 팀의 생산성과 협업에 부정적인 영향을 미칠 수 있습니다. AI/ML 배포 확장은 리소스의 제약을 받고 관리가 복잡할 수 있으며, 하드웨어 가속화를 위한 값비싼 GPU(그래픽 처리 장치, Graphics Processing Unit) 리소스와 생성형 AI를 위한 분산 워크로드가 필요합니다. 많이 사용되는 클라우드 플랫폼은 확장성을 비롯하여 매력적인 툴셋을 제공하는 하지만, 오히려 이러한 툴로 인해 사용자가 종속되는 경우가 많아 아키텍처 및 배포 측면에서 선택의 폭이 제한됩니다.

오픈소스 Open Data Hub 프로젝트에 따라 Red Hat® OpenShift® AI<sup>2</sup>는 데이터 과학자가 생성형 AI가 포함된 ML/DL 모델을 신속하게 훈련, 테스트, 서빙 및 모니터링하도록 지원합니다. 사용자는 인프라 프로비저닝을 기다릴 필요 없이 모델링 및 애플리케이션 개발에 즉시 집중할 수 있습니다. Red Hat OpenShift의 애드온으로서 완전 매니지드 클라우드 서비스 또는 자체 관리형 소프트웨어 제품의 형태로 제공되는 OpenShift AI는 Red Hat 구성 요소, 오픈소스 소프트웨어, 기술 파트너 오퍼링을 온프레미스, 클라우드 또는 엣지 인프라에서 모델을 개발 및 서빙하는 유연성과 결합합니다.

## Red Hat OpenShift AI

OpenShift AI는 조직이 AI/ML 모델링을 수행할 수 있도록 공동 오픈소스 및 제 3사 툴의 통합 세트를 배포하기 위한 효율적인 방법을 제공합니다. 이 플랫폼은 개별 클라우드 공급업체가 제공하는 규범적이고 독자적인 AI/ML 제품군에 대한 대안이 됩니다. 플랫폼을 도입하는 사용자는 실험적인 모델을 구축할 수 있는 협업적인 오픈소스 툴셋과 플랫폼을 확보할 수 있어 인프라나 퍼블릭 클라우드 전용 툴에 대한 종속성(Lock-In)에 대해 걱정하지 않아도 됩니다. 또한 파트너 툴을 활용하여 기본 플랫폼을 확장하고 역량을 높일 수 있습니다. 모델은 하이브리드 클라우드 및 엣지 환경 전반에서 컨테이너 지원 형식으로 일관되게 프로덕션 단계로 제공할 수 있습니다. OpenShift AI는 입증되고 확장 가능한 보안 중심의 플랫폼에서 간편한 구성으로 관리하기 쉬운 환경을 IT 운영 팀에 제공합니다.

OpenShift AI는 자주 사용되는 생성형 AI 기반 모델을 지원합니다. 따라서 사용자는 고유의 활용 사례를 위해 자체 데이터를 사용하여 이렇게 사전 학습된 모델을 프롬프트 튜닝, 미세 조정 및 제공할 수 있습니다. 또한 워크로드를 위치에 상관없이 여러 Red Hat OpenShift 클러스터 전반에 분배할 수도 있습니다. 그리고 이 플랫폼은 Nvidia GPU와 Intel XPU 등 CPU(중앙 처리 장치) 및 GPU(그래픽 처리 장치) 기반 하드웨어 인프라를 지원하므로 조직은 자체 데이터 사이언스 플랫폼을 직접 관리할 필요 없이 AI 하드웨어 가속화를 더 간단하게 활용할 수 있습니다.

<sup>1</sup> Gartner 보도 자료. "Gartner, 기업의 80% 이상이 2026년경에는 생성형 AI API를 사용하거나 생성형 AI 지원 애플리케이션을 배포할 것으로 내다봄" 2023년 10월 11일.

<sup>2</sup> Red Hat OpenShift Data Science

Red Hat Consulting은 OpenShift AI 파일럿 인계이치먼트를 제공하여 조직이 OpenShift AI 여정을 시작하고 기존 엔터프라이즈에 통합할 수 있도록 지원합니다.

모델을 프로덕션으로 배포하기 위한 전략을 개발하기 위해 모델 실험 이상의 단계를 진행하고자 하는 조직을 위해 Red Hat Consulting은 MLOps Foundation 컨설팅 서비스도 제공합니다.

## 업스트림 오픈소스 및 상용 기술 파트너들

Red Hat OpenShift AI는 업스트림 Open Data Hub 프로젝트에서 찾을 수 있는 하위 툴셋(표 1)을 제공합니다. 조직은 전체 관리형 및 자체 관리형 Red Hat OpenShift와 같이 어떤 클라우드 환경에서도 모델을 개발, 테스트, 배포할 수 있으며 중앙에서 성능을 모니터링할 수 있습니다. Red Hat은 오픈소스 툴(예: Jupyter, Pytorch, Tensorflow)에 대한 정기적인 업데이트를 제공하여 통합, 테스트, 유지관리로 인한 업무 부담을 줄여줍니다. 또한 이 오퍼링은 여러 AI/ML 기술 파트너 오퍼링을 통합합니다(표 1). 게다가 Red Hat OpenShift에서 제품을 인증한 30곳 이상의 AI 기술 파트너로부터 추가적인 상용 기술 파트너 오퍼링을 추가할 수 있습니다.

표 1. Red Hat OpenShift AI 에코시스템

AI/ML 모델링 및 시각화 툴	사전 구축된 Notebook 이미지를 제공하는 <a href="#">JupyterLab UI</a> 와 공통 Python 라이브러리 및 패키지, <a href="#">TensorFlow</a> , <a href="#">PyTorch</a> , CUDA, 다수의 Notebook 세션을 관리하기 위한 <a href="#">Kubeflow Notebook 컨트롤러</a> , <a href="#">Anaconda</a> (전문가용 선택 가능), <a href="#">Intel의 AI Tools</a>
데이터 엔지니어링	<a href="#">Starburst</a> (Galaxy 및 Enterprise 선택 가능), <a href="#">Pachyderm</a> (선택 사항)
데이터 수집 및 스토리지	<a href="#">Red Hat AMQ</a> (애드온 선택 가능), <a href="#">Amazon Simple Storage Service(S3)</a>
GPU 지원	<a href="#">NVIDIA</a> (GPU 오퍼레이터 포함), Intel XPU( <a href="#">Intel Xeon 프로세서</a> 포함), <a href="#">Habana Gaudi</a> , <a href="#">Intel Data Center GPU Flex Series</a> )
모델 서빙 및 모니터링	모델 서빙(사용자 인터페이스를 갖춘 <a href="#">KServe</a> ), 모델 모니터링, OpenShift <a href="#">Source-to-Image(S2I)</a> , Red Hat OpenShift API Management(애드온 선택 가능), <a href="#">Intel Distribution of the OpenVINO 툴킷</a>
데이터 사이언스 파이프라인	<a href="#">데이터 사이언스 파이프라인</a> ( <a href="#">Kubeflow Pipelines</a> )은 데이터 준비와 같은 프로세스들을 하나로 연결하고, 모델을 구축하며, 모델을 서빙합니다

한국레드햇 홈페이지 <https://www.redhat.com/ko>



## Red Hat 소개

Red Hat은 권위 있는 어워드를 수상한 지원, 교육, 컨설팅 서비스로 고객이 여러 환경에서 표준화를 진행하고, 클라우드 네이티브 애플리케이션을 개발하고, 복잡한 환경을 통합, 자동화, 보안, 관리할 수 있도록 지원합니다.

f [www.facebook.com/redhatkorea](https://www.facebook.com/redhatkorea)  
구매문의 02-6105-4390  
[buy-kr@redhat.com](mailto:buy-kr@redhat.com)