

하이브리드 클라우드의 AI 모델용 개방형 플랫폼

주요 내용

입증된 앱 플랫폼 기반에서 AI 추론 및 에이전틱(Agentic) AI를 운영하고 확장합니다.

데이터 사이언티스트, 데이터 엔지니어, 애플리케이션 개발자, DevOps 팀의 역량을 강화하는 일관된 사용자 환경으로 팀 간 AI/ML 운영 효율성을 높입니다.

온프레미스, 클라우드 또는 엣지에서 AI/ML 워크로드를 구축, 학습, 배포, 모니터링함으로써 하이브리드 클라우드의 유연성을 제공합니다.

지능형 애플리케이션 및 생성형 AI 수용

AI(인공지능), ML(머신 러닝), DL(딥러닝)은 다양한 비즈니스와 산업 전반에서 애플리케이션 현대화 노력에 크게 영향을 미치고 있습니다. 데이터에서 전략적 가치와 새로운 인사이트를 이끌어내고 혁신해야 한다는 필요성으로 인해 AI 지원 클라우드 네이티브 애플리케이션, MLOps 및 GenAIOps 방법론 사용이 확장되고 있습니다. 이처럼 새로운 환경에는 프로덕션으로 이전 시 급격한 모델 비용 증가, 복잡한 사용자 정의, 경직된 배포 제약 조건, 혁신 속도를 따라잡기 위해 필요한 운영 등 복잡한 과제가 뒤따를 수 있습니다. 기업은 추론 비용을 절감하고, 확장 및 모니터링을 단순화하고, 지속적인 변화에 적응하는 솔루션을 요구합니다.

Red Hat® AI는 하이브리드 클라우드 환경 전반에서 엔터프라이즈 AI 솔루션 개발과 배포를 가속화합니다. 전체 AI/ML 라이프사이클을 관리하고 MLOps 및 GenAI Ops 기능을 제공하는 통합 플랫폼입니다. Red Hat AI는 특히 다음 4가지 주요 요소에 중점을 둡니다.

- ▶ 빠르고 유연하고 효율적인 추론으로 효율성 향상
- ▶ 모델과 데이터를 연결하는 경험 단순화
- ▶ 에이전틱(Agentic) AI 혁신 가속화
- ▶ 하이브리드 클라우드 전반에 AI 확장 시 유연성과 일관성 보장

Red Hat OpenShift® AI는 업계를 주도하는 하이브리드 클라우드 애플리케이션 플랫폼인 [Red Hat OpenShift](#)에 기반한 것으로, Red Hat AI 포트폴리오의 주력 제품입니다. 이 AI 플랫폼은 AI 엔지니어, 데이터 사이언티스트, 개발자가 생성형 및 예측 모델과 AI 기반 애플리케이션을 규모에 맞게 구축하고 배포할 수 있는 강력한 AI/ML 기반입니다. 기업은 하나의 공통된 플랫폼에서 원하는 툴을 실험하고, 협업하고, 시장 출시 시간을 단축할 수 있습니다. Red Hat OpenShift AI는 데이터 사이언티스트와 개발자가 원하는 셀프 서비스 환경과 엔터프라이즈 IT 팀이 요구하는 신뢰성을 결합합니다.

신속한 개발, 교육, 테스트, 배포

Red Hat OpenShift AI는 오픈소스 기술로 구축한 유연하고 확장 가능한 MLOps 플랫폼으로, 팀이 실험하고 모델을 서빙하고 혁신적인 애플리케이션을 제공할 수 있도록 일관된 운영을 지원하며 신뢰할 수 있는 기능을 제공합니다. OpenShift AI는 AI 기반 애플리케이션 제공을 가속화하므로, 기업은 초기 파일럿 단계를 넘어 더 빠르고 주도적으로 견고한 운영 배포 환경으로 이전할 수 있습니다.

또한 예측 AI 모델과 생성형 AI 모델을 빌드하고, 학습시키고, 튜닝, 배포, 모니터링하는 툴을 갖춘 통합 UI(사용자 인터페이스) 환경을 제공합니다. 소버린 및 프라이빗 AI를 지원하기 위해 통제되고 보호되는 풋프린트를 제공하는 데 특히 중점을 두어 하이브리드 클라우드 환경에 모델을 배포할 수 있습니다. 이러한 접근 방식은 민감한 데이터와 AI 모델을 지정된 지리적 경계나 조직 경계 내에 유지함으로써 규제 및 컴플라이언스 요구 사항을 충족하는 데 도움을 줍니다.

생성형 AI 애널리스트 예측

“조직들이 하이브리드 인프라 선택 시 워크로드 및 데이터 요구 사항을 일치시키기 위해 노력함에 따라 AI는 2026년 디지털 인프라 예산을 높이는 매우 중요한 요인이 될 것으로 예상됩니다. 의사 결정자의 90%는 AI가 2026년 디지털 인프라 예산 및 기술 선택을 주도하는 중요한 요소가 될 것으로 생각합니다.”¹

AI 도입 간소화

Red Hat OpenShift의 애드온인 OpenShift AI는 오픈소스 커뮤니티를 강력한 AI 에코시스템과 결합하여 AI 도입을 확대하고 AI 이니셔티브에 대한 신뢰를 향상하도록 설계한 플랫폼입니다. 기업에 적합한 AI/ML 기술을 더욱 유연하고 자유롭게 선택할 수 있습니다. 사용자는 자체 예측 모델을 빌드하거나, 외부 생성형 AI 모델로 시작한 후 플랫폼에 제공되는 모델 서버 중 하나를 사용해 검색 증강 생성(RAG) 기술로 모델을 강화할 수 있습니다. 이 플랫폼에서는 최적화 및 검증된 거친 Llama, Mistral, DeepSeek, Granite 등의 제3자 모델에 신속하게 액세스할 수 있습니다. 이들 모델은 Red Hat AI의 Hugging Face 기반 리포지토리에서 사용 가능하며 vLLM에서 효율적으로 실행됩니다. 카탈로그에서 사용자는 이러한 모델을 살펴보고 자체 모델을 추가할 수 있습니다.

팀 전반의 운영 일관성 향상

Red Hat OpenShift AI는 데이터 사이언티스트, AI 엔지니어, 개발자, DevOps 팀이 협업하여 AI 솔루션을 적시에 효과적으로 제공할 수 있도록 지원하는 일관된 사용자 환경을 제공합니다. 사용자는 셀프 서비스 액세스로 협업 워크플로우, 그래픽 처리 장치(GPU) 가속화, 간소한 운영을 활용하며 하이브리드 클라우드 환경 및 네트워크 엣지 전반에서 대규모로 AI 솔루션을 일관적으로 제공할 수 있습니다.

IT 운영 팀은 손쉽게 확장 또는 축소 가능하며 강화된 거버넌스와 보안을 제공하는 입증된 플랫폼에서 간소화된 구성과 더욱 자동화된 워크플로우의 이점을 누릴 수 있습니다.

하이브리드 클라우드의 유연성 확보

Red Hat OpenShift AI 플랫폼에서는 클라우드, 온프레미스 데이터센터, 에어 갭(air-gap) 환경 등 다양한 환경에서 AI/ML 워크로드의 학습, 배포 및 모니터링이 가능하여 규제, 보안 및 데이터 요구 사항을 충족할 수 있습니다. 이 플랫폼은 NVIDIA, AMD, Intel과 같은 벤더가 제공하는 다양한 AI 가속기와 호환됩니다. 이 기능을 확장하여 서비스로서의 GPU 환경을 구축할 수 있으며, 이 환경에서 조직은 GPU 리소스를 중앙에서 관리하고, 파티셔닝하고, 스케줄링하는 동시에 해당 리소스의 사용 현황을 자세히 관측할 수도 있습니다.

생성형 AI 및 에이전틱

생성형 AI 프로젝트의 경우, 플랫폼 엔지니어를 위한 대시보드 환경인 AI hub(개발자 미리 보기)와 같은 구성 요소를 통해 제공되는 전용 사용자 환경에서 모델 및 MCP 서버를 설정하고 배포하기 위한 카탈로그, 레지스트리 및 모델 배포가 통합 제공됩니다. Gen AI studio(개발자 미리 보기)는 AI 엔지니어와 애플리케이션 개발자가 배포된 모델 및 MCP 서버를 액세스하여 실험, 비교, 평가 및 테스트할 수 있는 AI 자산 엔드포인트이자 AI 플레이그라운드입니다.

OpenShift AI는 API 계층과 유연한 확장형 기반을 통한 제공함으로써 에이전틱(Agentic) AI를 가속화합니다. Llama Stack API 및 MCP(기술 프리뷰) 지원은 Llama Stack API 엔터프라이즈급 구현을 포함하며 다양한 AI 기능을 위한 표준화된 단일 진입점을 제공합니다.

추가 틀에는 실제 추론 배포를 지원하는 LLM 평가(LM Eval) 및 LLM 벤치마킹이 포함됩니다. LLM compressor는 Red Hat AI의 Hugging Face 기반 리포지토리에서 최적화 및 검증된 모델을 구축하는 데 Red Hat이 사용하는 것과 동일한 방법으로 조직의 사용자 정의 모델 크기를 줄이는 알고리즘을 제공합니다.

¹ IDC Tech Supplier. “AI 요구사항으로 인한 온프레미스 인프라 배포 및 퍼블릭 클라우드와의 상호운용성 수요 급증, 2025년.” Doc #US53418426, 2025년 10월. (고객 로그인 필요)

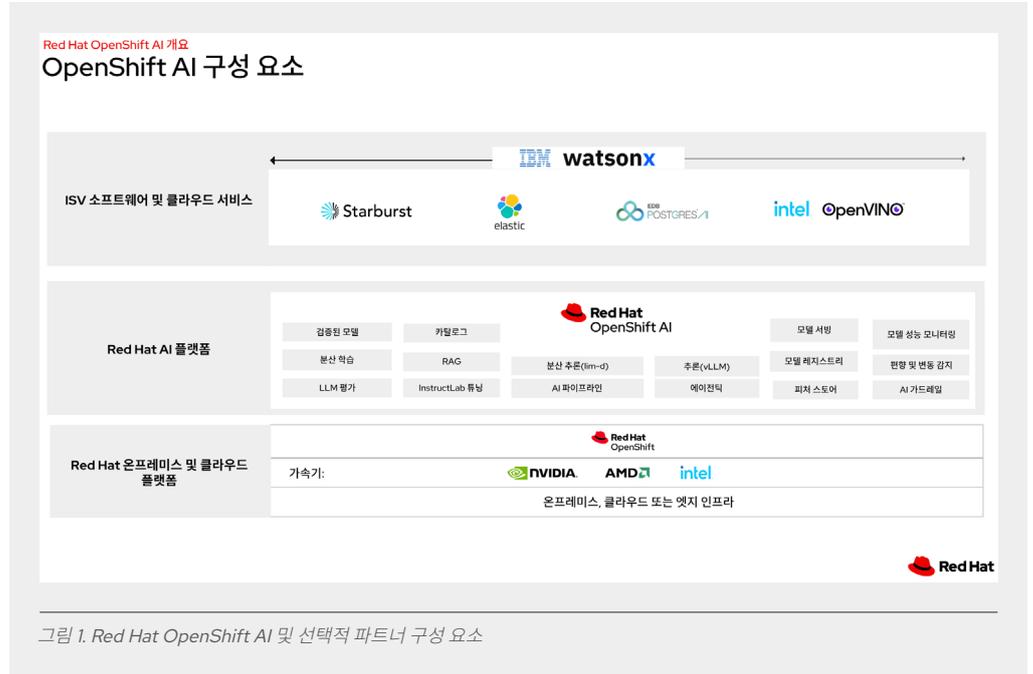


그림 1. Red Hat OpenShift AI 및 선택적 파트너 구성 요소

Red Hat OpenShift AI가 제공하는 다음과 같은 핵심 툴과 기능으로 탄탄한 기반을 구축할 수 있습니다.

- ▶ **모델 구축 및 사용자 정의.** 데이터 사이언티스트는 [JupyterLab](#) UI에서 탐색적 데이터 사이언스를 수행할 수 있는데, 이 UI는 일반적인 Python 라이브러리를 통해 즉시 사용 가능하고 안전하게 빌드된 노트북 이미지를 제공합니다. 생성형 AI 프로젝트의 경우, OpenShift AI는 검색 증강 생성(RAG) 및 분산 InstructLab 학습을 지원하며 더 효율적으로 생성형 AI 모델에 기술과 지식을 기여할 수 있도록 모델 조정 툴링을 제공합니다.
- ▶ **모델 서빙.** Red Hat OpenShift AI는 모델 서빙용 코어 엔진으로 KServe를 사용하는 다양한 프레임워크를 제공하여 예측 머신 러닝이나 파운데이션 모델을 프로덕션 환경에 배포하는 작업을 간소화합니다. 최대 확장성이 요구되는 LLM의 경우, OpenShift AI는 vLLM 런타임을 통한 병렬화된 서빙을 제공합니다. Llm-d는 파이프라인을 모듈형 서비스로 분리하여 LLM 추론을 최적화하는 프레임워크를 제공하며, 이 프레임워크는 스마트 자동 스케일링 및 효율적인 요청 라우팅을 지원합니다.
- ▶ **AI 파이프라인.** Red Hat OpenShift AI는 그래픽 프론트엔드를 사용하여 AI 태스크를 파이프라인으로 오케스트레이션하고 파이프라인을 구축할 수 있는 파이프라인 구성 요소를 제공합니다. 기업은 데이터 준비, 모델 빌드, 모델 서빙과 같은 프로세스를 연결할 수 있습니다.
- ▶ **모델 모니터링.** Red Hat OpenShift AI는 운영 지향적인 사용자가 모델 서버 및 배포된 모델에 대한 운영 및 성능 메트릭을 모니터링하는 데 도움이 됩니다. 사용자는 성능 및 운영 메트릭에 즉시 사용 가능한 시각화 기능을 사용하거나 데이터를 기타 관측성 서비스와 통합할 수 있습니다.
- ▶ **분산 워크로드.** 팀은 분산 워크로드를 통해 모델 학습, 튜닝, 서빙, 데이터 처리 작업을 가속화할 수 있습니다. 이 기능은 작업 실행 우선순위를 지정하고 작업을 분산하며 노드를 최적의 방법으로 사용할 수 있도록 지원합니다. 고급 GPU 지원은 파운데이션 모델의 워크로드 요구 사항을 처리하는 데 도움이 됩니다.

- ▶ **AI 가드레일, 편향 및 변동 감지.** Red Hat OpenShift AI는 데이터 사이언티스트와 AI 엔지니어가 학습 데이터를 기반으로 모델이 공정하고 편향되지 않았는지 모니터링하고 나아가 실제 배포 과정에서도 공정성을 유지하도록 지원하는 툴을 제공합니다. AI 가드레일은 중요한 안전 제어를 구현하는 사용자 정의 가능한 프레임워크를 제공하여 프로덕션 용도로 모델의 투명성, 공정성, 신뢰성을 보장할 수 있습니다. 변동 감지 툴에는 배포된 ML 모델의 입력 데이터 분포 정보가 포함되어 모델 추론에 사용된 실시간 데이터와 모델 학습에 사용된 데이터 사이에 큰 차이가 있는 경우를 감지합니다.
- ▶ **카탈로그 및 레지스트리.** Red Hat OpenShift AI는 플랫폼 엔지니어가 최적화된 생성형 AI 모델을 검색, 비교, 평가할 수 있는 내부 모델 카탈로그 및 큐레이션된 카탈로그를 제공합니다. 데이터 사이언티스트와 AI 엔지니어가 예측 및 생성형 AI 모델, 메타데이터, 모델 아티팩트를 공유, 수정, 배포, 추적할 수 있도록 지원하는 중앙 레지스트리도 제공합니다.
- ▶ **피처 스토어.** 성능 향상 및 워크플로우 가속화를 위해 ML 모델에 사용할 잘 정의된 정제 데이터 피처를 관리합니다.

전체 AI 라이프사이클을 위한 툴

Red Hat OpenShift는 기업이 성공적으로 모델을 학습시켜 배포하고 이를 프로덕션 단계로 전환할 수 있도록 지원하는 서비스와 소프트웨어를 제공합니다(그림 2 참조).

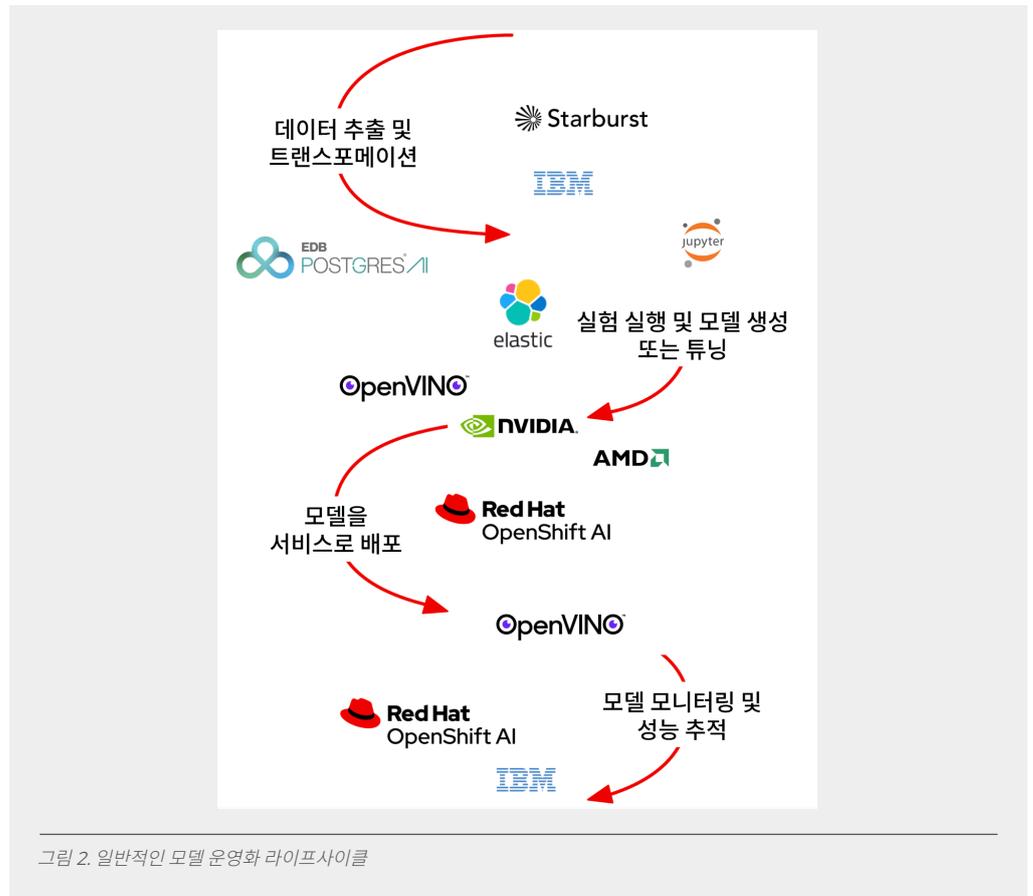


그림 2. 일반적인 모델 운영화 라이프사이클

Red Hat OpenShift AI 대시보드는 원활한 도입을 위해 모든 애플리케이션과 도큐멘테이션을 찾고 액세스할 수 있는 중앙 위치를 제공합니다. 시작하는 방법에 관한 스마트 튜토리얼은 공통 구성 요소와 통합 파트너 소프트웨어에 대한 모범 사례 가이드를 제공하며, 이는 대시보드에서 직접 이용할 수 있어 데이터 사이언티스트가 더욱 빠르게 학습하고 시작할 수 있도록 도와줍니다. 다음 섹션에서는 Red Hat OpenShift AI에 통합된 기술 파트너 톨에 대해 설명합니다. 일부 톨에는 기술 파트너의 추가 라이선스가 필요합니다.



Starburst

Starburst는 팀이 데이터를 활용하여 쉽고 빠르게 비즈니스 기능을 개선하도록 하여 분석을 가속화합니다. 셀프 관리형 제품 또는 전체 관리형 서비스로 제공되는 Starburst는 데이터 액세스를 대중화하여 데이터 소비자에게 포괄적인 인사이트를 제공합니다. Starburst는 주요 대량 병렬 처리(Massively Parallel Processing, MPP) SQL(Structured Query Language) 엔진인 오픈소스 Trino(이전 명칭: PrestoSQL)를 기반으로 구축되었습니다. Trino 전문가들이 구축하고 운영하는 Starburst는 데이터를 이동하지 않고도 위치와 관계없이 다양한 데이터세트를 검토할 수 있는 자유를 제공합니다.

Starburst는 Red Hat OpenShift가 제공하는 확장 가능한 클라우드 스토리지와 컴퓨팅 서비스와 통합되어 안정적이고 보안에 중점을 둔 효율적이고 경제적인 방법으로 모든 엔터프라이즈 데이터를 쿼리합니다. 장점은 다음과 같습니다.

- ▶ **자동화.** Starburst와 Red Hat OpenShift 오퍼레이터는 클러스터의 자동 구성, 자동 튜닝, 자동 관리를 제공합니다.
- ▶ **고가용성과 점진적 스케일 다운.** Red Hat OpenShift 로드 밸런서는 Trino 코디네이터와 같은 서비스를 상시 가동 상태로 유지할 수 있습니다.
- ▶ **유연한 확장성.** Red Hat OpenShift는 Trino 작업자 클러스터를 쿼리 로드와 따라 자동으로 확장할 수 있습니다.

Hewlett Packard Enterprise

HPE Machine Learning Data Management 소프트웨어

기업은 노트북 실험부터 중요한 엔터프라이즈 배포까지 모든 것을 용이하게 지원하는 데이터 관리 솔루션을 필요로 합니다. HPE Machine Learning Data Management 소프트웨어(이전의 Pachyderm)를 사용하면 데이터 사이언스 팀이 자동 데이터 버전 관리를 통해 보장된 데이터 계보를 기반으로 컨테이너화된 데이터 중심의 ML 파이프라인을 구축하고 확장할 수 있습니다. 실제 데이터 사이언스 문제를 해결하도록 엔지니어링된 HPE Machine Learning Data Management 소프트웨어는 ML 라이프사이클을 자동화하고 확장할 수 있도록 데이터 기반을 제공하는 동시에 재현 가능성을 보장합니다. HPE Machine Learning Data Management 소프트웨어는 비정형 데이터에서 데이터 웨어하우스, 자연어 처리, 동영상 및 이미지 ETL(extract, transform, load), 금융 서비스, 생명과학에 이르는 광범위한 활용 사례를 통해 다음 기능을 제공합니다.

- ▶ 팀이 데이터 변경 사항을 높은 성능으로 추적할 수 있는 방법을 제공하는 자동화된 데이터 버전 관리
- ▶ 데이터 처리 속도를 높이는 동시에 컴퓨팅 비용을 낮추는 데이터 중심의 컨테이너화된 파이프라인
- ▶ ML 라이프사이클 내 활동과 자산에 대한 고정된 기록을 제공하는 변경 불가능한 데이터 계보
- ▶ DAG(방향성 비순환 그래프, Directed Acyclic Graph)의 직관적인 시각화를 제공하고 디버깅과 재현 가능성을 지원하는 콘솔
- ▶ JupyterLab Mount Extension을 통해 버전 관리 데이터에 대한 포인트 앤 클릭(point-and-click) 인터페이스를 제공하는 Jupyter Notebook 지원

- ▶ 기업 내 다양한 팀 전반에서 HPE Machine Learning Data Management 소프트웨어를 규모에 맞게 배포하고 관리하는 강력한 툴을 통해 엔터프라이즈 관리



AI 솔루션 배포를 가속화한 NVIDIA

비즈니스 성공에서 AI/ML 애플리케이션이 수행하는 역할이 점점 더 커짐에 따라, 기업은 복잡한 워크로드를 처리하고, 하드웨어 사용을 최적화하며, 확장성을 제공할 수 있는 플랫폼이 필요하게 되었습니다. 확장 가능한 데이터 프로세싱, 데이터 분석, ML 학습, 추론은 모두 고도로 리소스 집약적인 컴퓨팅 태스크를 나타냅니다. NVIDIA 소프트웨어는 GPU의 병렬 처리 기능을 활용하여 엔드 투 엔드 데이터 사이언스의 모든 측면을 가속화할 수 있습니다.

NVIDIA NIM을 사용하면 Red Hat OpenShift 환경 내에서 NVIDIA GPU의 관리 기능과 성능이 향상되어 AI 애플리케이션이 NVIDIA AI 소프트웨어 및 하드웨어의 잠재력을 최대한 활용할 수 있습니다. NVIDIA NIM과 Red Hat OpenShift AI의 통합을 통해 리소스 할당을 개선하고, 효율성을 높이며, AI 워크로드 실행 생산성을 높일 수 있습니다.



Intel OpenVINO 툴킷

Intel OpenVINO 툴킷은 Intel 플랫폼상에서 고성능 DL 추론 애플리케이션의 개발과 배포를 가속화합니다. 이 툴킷을 사용하면 가상으로 신경망 모델을 도입, 최적화, 튜닝하고 OpenVINO 에코시스템의 개발 툴을 사용하여 포괄적인 AI 추론을 실행할 수 있습니다.

- ▶ **모델.** 소프트웨어 개발자는 자체 DL 모델을 유연하게 사용할 수 있습니다. 또한 시장 출시 시간을 단축하기 위해 Intel이 [Hugging Face와 협업하여 제공하는 OpenVINO 툴킷](#)을 통해 사전 학습 및 최적화된 모델을 사용할 수도 있습니다. OpenVINO는 Pytorch, ONNX, TensorFlow 및 많이 사용되는 다른 모델 형식을 지원합니다.
- ▶ **최적화.** OpenVINO 툴킷은 모델을 변환하여 편의성과 성능을 향상할 수 있는 다양한 방법을 제공하여 소프트웨어 개발자가 AI 모델을 더 빠르고 효율적으로 실행할 수 있도록 돕습니다. 개발자는 모델 변환 단계를 건너뛰고 PyTorch, ONNX, TensorFlow, TensorFlow Lite, JAX 또는 PaddlePaddle 형식에서 직접 추론을 실행할 수 있습니다. OpenVINO IR로 전환하면 최적의 성능을 제공하며, OpenVINO의 Neural Network Compression Framework에서 사용할 수 있는 가중치 압축 및 양자화 기능을 사용하여 추가로 최적화할 수 있습니다. 같은 기능이 스토리지 및 런타임 풋프린트도 감소시킵니다.
- ▶ **배포.** OpenVINO Runtime Inference Engine은 추론 프로세스를 가속화하기 위해 애플리케이션에 통합되도록 설계된 API(애플리케이션 프로그래밍 인터페이스)입니다. '한 번 작성하여 어디에나 배포'하는 방식을 통해 CPU(중앙 처리 장치), GPU, NPU, FPGA 등 다양한 Intel 하드웨어에서 추론 태스크를 효율적으로 실행할 수 있습니다. OpenVINO 생성형 AI 확장 라이브러리는 생성형 AI 워크로드 배포를 간소화하고 많은 경우, 필요한 코드를 3~5개 행으로 줄입니다. OpenVINO Model Server는 에이전틱 및 모델 서빙 시나리오를 지원하는 여러 기능을 제공하여 개발 작업을 훨씬 더 줄입니다.



EDB

EDB Postgres AI는 거래, 분석 및 AI 워크로드를 처리하도록 설계한 강력한 지능형 플랫폼으로, 온프레미스나 클라우드 환경 등 데이터가 존재하는 위치에 관계없이 탁월한 유연성을 제공합니다. 엔터프라이즈 Postgres 데이터베이스 솔루션 분야 글로벌 리더인 EDB는 최대 3배 더 빨리 AI 프로젝트를 프로덕션 환경에 배치할 수 있는 개방형 엔터프라이즈급 소버린 데이터 및 AI 플랫폼을 제공합니다. Red Hat OpenShift AI와 EDB Postgres AI의 통합을 통해 사용자는 검색 증강 생성(RAG)에 대한 강력한 AI 기술 자료를 구축하고 AI 데이터,

모델, 애플리케이션을 어디서나 배포 가능한 풀스택 소버린 AI 플랫폼으로 통합할 수 있습니다. 이렇게 핵심 운영 데이터를 AI 레디 자산으로 트랜스포메이션함으로써 **효율성을 최대 30% 향상**할 수 있으며, 비정형 데이터를 포함한 비공개 데이터의 사용을 간소화하여 조직의 기술 자료에 근거한 모델 출력을 보장할 수 있습니다.



Elastic

Elastic Search AI Platform(ELK Stack 기반²)은 정확한 검색 기능과 지능적인 AI를 결합하여 사용자가 LLM을 더욱 빠르게 프로토타입화하고 통합하며, 생성형 AI를 활용하여 확장 가능하고 비용 효율적인 애플리케이션을 빌드할 수 있도록 지원합니다. Elastic Search AI Platform을 사용하면 혁신적인 RAG(검색 증강 생성, Retrieval Augmented Generation) 애플리케이션을 빌드하고, 관측성 문제를 사전에 해결하며, 복잡한 보안 위협에 대처할 수 있습니다. Elasticsearch는 온프레미스, 선택한 클라우드 제공업체 또는 에어 갭 환경 등 애플리케이션이 있는 곳에 배포할 수 있습니다.

Elastic은 Red Hat OpenShift AI, Hugging Face, Cohere, OpenAI 등을 포함한 에코시스템의 임베딩 모델과 단 한 번의 간단한 API 호출을 통해 통합됩니다. 이러한 접근 방식은 RAG 워크로드에 대한 하이브리드 추론 관리 코드를 명확하게 유지하며, 다음과 같은 기능을 제공합니다.

- ▶ **칭킹, 커넥터, 웹 크롤러**: 다양한 데이터셋을 검색 계층으로 수집
- ▶ **시맨틱 검색**: 기본 제공 ML 모델인 ELSER(Elastic Learned Sparse Encoder)와 **E5 임베딩 모델**을 통해 다국어 벡터 검색 지원
- ▶ **문서 및 필드 수준 보안**: 기업의 역할 기반 액세스 제어(RBAC)에 매핑되는 권한 및 자격 부여 구현

Elastic Search AI Platform을 사용하면 전 세계 개발자가 함께하는 커뮤니티의 구성원이 되어 항상 가까이에서 영감과 지원을 얻을 수 있습니다. [Slack](#), 토론 [포럼](#) 또는 소셜 미디어에서 Elastic 커뮤니티를 확인해 보세요.

결론

Red Hat OpenShift AI를 사용하는 기업은 AI 기반 애플리케이션 여정에서 실험하고 협업하며 궁극적으로 여정을 가속화할 수 있습니다. 데이터 사이언티스트와 AI 엔지니어는 Red Hat OpenShift AI를 사용하여 하이브리드 클라우드 전반에 모델을 빌드하고 배포하는 유연성을 확보하게 됩니다. IT 운영 팀 및 플랫폼 엔지니어는 MLOps 및 GenAIOps 기능의 이점을 활용하여 모델을 프로덕션 환경에 더욱 신속하게 배포할 수 있습니다. GPU에 대한 액세스를 포함하여 개발자와 AI 엔지니어 및 데이터 사이언티스트를 위한 셀프 서비스는 이미 엔터프라이즈 IT에서 사용하고 완전히 신뢰하는 애플리케이션 플랫폼에서의 혁신 속도를 높입니다.

Red Hat OpenShift AI는 강력한 파트너 에코시스템을 기반으로 신뢰할 수 있고 일관된 통합 플랫폼을 지속적으로 제공하며 효율적인 추론, 에이전틱(Agentic) AI, 확장 가능한 하이브리드 클라우드 운영에서 고유한 차별화 요소를 제공하고 있습니다.

자세히 알아보기

[Red Hat OpenShift AI](#)를 참조하여 지금 시작하세요.



Red Hat 소개

Red Hat은 전 세계 고객에게 높은 수준의 지원과 교육 및 컨설팅 서비스를 제공하여 **권위 있는 어워드를 수상**한 바 있으며 이를 통해 고객 환경 전반의 표준화, 클라우드 네이티브 애플리케이션 개발, 복잡한 환경의 통합, 자동화, 보안 및 관리를 지원합니다.

² ELK 스택은 Elasticsearch, Kibana, Beats, Logstash로 구성됩니다.