

물리적 망 분리 환경에서의 웹 기반 데이터 분석 플랫폼 구축 방안

이용구
한국수력원자력 중앙연구원
e-mail:lee.yongku@khnp.co.kr

How to build a web-based data analysis platform in a physical network separation environment

Yong-Ku Lee
Central Research Institute, Korea Hydro & Nuclear Power Co., Ltd.

요 약

망 분리 업무 환경이 적용된 기업은 외부 소프트웨어 패키지나 클라우드 기반 데이터 분석 서비스를 활용하는 데에 제약을 받는다. 이에 따라 클라우드 기반 Jupyter Notebook의 주요 서비스 특징에 대해 조사하고 분석하여 이러한 서비스를 사내 망 분리 환경에 적용하기 위한 구축 방안을 도출하고자 한다. 본 연구를 통해 사내 데이터의 외부 유출 가능성을 원천 차단함과 동시에 증가하는 데이터 분석 및 AI 모델 개발 수요에 능동적으 대응 가능한 환경을 구축하여 관련 연구개발 업무의 효율성을 제고하고자 한다.

1. 서론

데이터 보안의 중요성이 증가함에 따라, 많은 기업은 업무 망과 인터넷망을 분리하여 운영하고 있다. 그러나 이러한 망 분리 환경은 외부 소프트웨어 패키지나 클라우드 기반 데이터 분석 서비스의 활용을 제한하여, 내부 데이터를 활용한 인공지능(AI) 서비스나 Kaggle, Colab과 같은 웹 기반 데이터 분석 플랫폼과의 연계를 어렵게 만든다. 그럼에도 불구하고 사내 데이터 분석 및 AI 모델 개발에 대한 수요는 지속적으로 증가하고 있다. 본 연구는 이러한 망 분리 환경에서 웹 기반 데이터 분석 및 AI 모델 개발을 지원할 수 있는 효율적인 시스템 구축 방안을 제시하고자 한다.

2. 본론

2.1 주요 서비스별 특징

- Amazon SageMaker : Amazon Web Services(AWS)에서 제공하는 SageMaker는 대규모 데이터 셋 처리와 머신러닝 모델 구축, 학습, 배포를 위한 포괄적인 플랫폼을 제공한다. 다양한 프레임워크 지원, 자동화된 학습 및 배포 기능을 통해 사용자는 복잡하느없이도 고성능 AI 모델을 개발할 수 있다. 특히, 최근 LLM 공유 및 활용 플랫폼으로 각광을 받고 있

는 HuggingFace에서 다양한 연동 설정을 지원하고 있다.

- Google Colab : Google이 제공하는 무료 클라우드 기반 Jupyter Notebook 환경으로, 별도의 설치나 환경 설정 없이 바로 Python 코드를 실행할 수 있는 편리함을 제공한다. Google Drive와 연동되어 데이터 저장 및 공유가 용이하며, GPU 및 TPU를 무료로 사용할 수 있어 머신러닝 모델 학습에 최적화된 환경을 제공한다. 특히, 비용 부담 없이 AI 및 데이터 분석 실험을 진행하려는 사용자에게 적합하다.

- Azure Notebooks : Microsoft Azure에서 제공하는 클라우드 기반 Jupyter Notebook 서비스로, Python, R, F# 등의 언어를 지원하며 Azure의 강력한 클라우드 인프라와 통합되어 있다. Azure Machine Learning과 연계하여 대규모 데이터 분석과 머신러닝 프로젝트를 효율적으로 관리할 수 있는 기능을 제공하며, 기업 내 복잡한 데이터 분석 작업을 클라우드에서 수행할 수 있는 확장성을 갖추고 있다.

- Kaggle : 데이터 분석 대회 플랫폼으로 널리 알려진 Kaggle은 무료 Jupyter Notebook 환경을 제공하며, 방대한 공개 데이터셋과 다양한 머신러닝 대회 참여를 통해 학습할 수 있는 기회를 제공한다. 데이터 과학 커뮤니티의 활발한 참여가 이루어져 있으며, 초보자부터 전문가까지 폭넓은 사용자가 서로의 코드를 공유하고 학습할 수 있는 열린 환경을 제공한다. 특히, 실습 기반의 학습 및 협업을 중시하는 데이터

과학자들에게 적합한 환경을 제공한다.

2.2 구성 요소

웹 기반 데이터 분석 플랫폼을 구축하기 위해서는 다음과 같은 주요 구성 요소가 필요하다. 첫째, Python Jupyter Notebook의 컨테이너 서비스를 효율적으로 관리하기 위한 Kubernetes 운영 환경이 필수적이다. 이를 통해 다수의 노트북 컨테이너 인스턴스를 오케스트레이션하고, 자원 할당을 최적화할 수 있다. 둘째, 컨테이너 이미지의 배포 및 버전 관리를 위한 컨테이너 저장소가 필요하다. 이를 통해 다양한 버전의 환경을 일관되게 제공하고 관리할 수 있다. 마지막으로, Python 패키지 관리를 위한 내부 Repository 환경을 구축하여 외부 네트워크와의 연결 없이도 필요한 라이브러리 및 패키지들을 내부적으로 설치하고 유지할 수 있어야 한다. 이러한 요소들은 망 분리된 환경에서 데이터 분석과 AI 모델 개발을 원활하게 지원하는 데 필수적인 기반을 제공한다.

2.3 구성도

웹 기반 데이터 분석 플랫폼의 각 구성 요소들은 전체 시스템 내에서 유기적으로 결합되어, 사용자가 필요로 하는 분석 환경을 신속하게 배포, 활용, 그리고 반납할 수 있는 서비스를 제공한다. 이를 통해 사용자는 별도의 복잡한 설정 없이 원하는 분석 환경을 즉시 사용할 수 있으며, 플랫폼은 자원의 효율적인 관리와 확장성을 보장한다. 이러한 구조는 데이터 분석 및 AI 모델 개발 작업의 효율성을 크게 향상시킨다.



[그림 1] 웹 기반 데이터 분석 플랫폼 구성도

3. 결론

본 연구에서는 외부망에서 제공되는 서비스에 상응하는 웹 기반 데이터 분석 플랫폼의 구축 방안을 제시하였다. 제안된 플랫폼은 데이터 보안을 강화하면서도, 내부 개발자들이 안전한 환경에서 데이터 분석 및 AI 모델 개발 역량을 향상시킬 수 있는 기반을 제공한다. 이를 통해 내부 인프라만으로도 높은 수준의 데이터 분석 환경을 구현할 수 있으며, 기업의 데

이터 활용 및 AI 기술 개발의 효율성을 크게 증대시킬 수 있을 것으로 예상된다.

참고문헌

- [1] Kubernetes Documentation[웹사이트]. (2024.10.14). URL:https://kubernetes.io/docs/home
- [2] Project Jupyter Documentation[웹사이트]. (2024.10.14). URL:https://jupyter.org/documentation