

Data Analytics

Data Ops



데이터 처리 작업을 위한 워크플로우를 작성하고 작업 수행을 자동화하는 서비스

Data Ops는 주기적, 반복적으로 발생하는 데이터 처리 작업에 대해 워크플로우를 작성하고 작업 스케줄링을 자동화하는 Apache Airflow 기반의 관리형 워크플로우 오케스트레이션 서비스입니다. Samsung Cloud Platform의 Kubernetes Engine 클러스터 환경에서 단독으로 사용하거나, 다른 애플리케이션 SW와 함께 사용할 수 있습니다.

편리한 설치 및 관리

Data Ops는 표준 Kubernetes 클러스터 환경에서 웹 기반 콘솔을 통해 손쉽게 설치가 가능합니다. Apache Airflow와 관리 모듈이 자동 설치되며, 통합 대시보드를 통해 웹 서버 및 스케줄러의 실행 상태에 대한 통합 모니터링이 가능합니다.

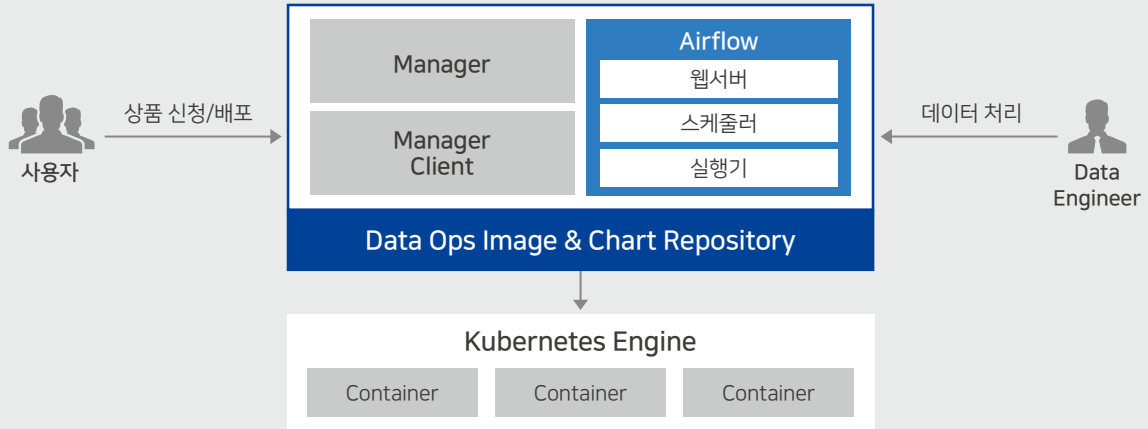
동적 파이프라인 구성

Python 코드를 기반으로 데이터 작업에 대한 파이프라인 구성이 가능합니다. 데이터 작업 예약과 연동하여 동적으로 작업을 생성하기 때문에 원하는 워크플로우 형태와 스케줄링을 자유롭게 구성할 수 있습니다.

편리한 워크플로우 관리

DAG (Direct Acyclic Graph : 방향성 비순환 그래프) 구성을 웹 기반의 UI를 통해 시각화하여 관리하기 때문에 데이터 흐름의 전후 및 병렬 관계를 쉽게 이해할 수 있습니다. 또한 각 작업의 타임아웃, 재시도 횟수, 우선순위 정의 등을 손쉽게 관리할 수 있습니다.

서비스 구성도



주요 기능




- 간편한 설치
 - 컨테이너 환경의 오픈소스 Airflow 설치
- GUI 기반의 편리한 관리
 - Airflow 설정을 컨테이너 환경에서 손쉽게 관리
 - Airflow 플러그인 배포
 - Airflow 서비스에 대한 상태 모니터링
- 워크플로우 작성 및 스케줄링
 - Python 기반의 워크플로우 작성으로 확장성 용이
 - 스케줄러를 통한 작업 수행 자동화
 - DAG 작업별 리소스 관리
 - 데이터 처리 작업에 대한 이슈 및 실패에 대한 재처리 플랜
- Airflow 구성 요소
 - 웹서버 : DAG의 구성 요소와 상태를 시각화. Airflow 설정에 대한 구성 관리
 - 스케줄러 : 다양한 DAG와 해당 작업을 오케스트레이션. DAG 예약/실행
 - 실행기 : Kubernetes 기반 동적 Executor인 KubernetesExecutor 제공
 - 메타데이터 DB : DAG, 해당 실행 및 사용자, 역할 및 연결 등 Airflow 구성에 대한 메타데이터 저장

요금 기준

- 과금
 - Data Ops가 사용하는 Kubernetes Engine의 Pod들의 CPU 사용시간

FOR MORE INFORMATION

SAMSUNG SDS

 www.samsungsds.com / cloud.samsungsds.com
 contact.sds@samsung.com / scp_sales@samsung.com
 youtube.com/samsungsds

