

데이터 파이프라인 오케스트레이션

오케스트레이션이란?

- 여러개의 컴퓨터 시스템, 애플리케이션 또는 서비스를 조율하고 관리하는 것
- 복잡한 테스크와 워크플로우를 쉽게 관리할 수 있도록 도와준다.
- 한마디로 워크플로우 관리도구.

오케스트레이션 기능

- 스케줄링
- 테스크 의존관계 정의
- 실행결과 알림 및 보안
- 실패시 재실행(backfill: 일정시간동안 workflow를 재실행하는 과정)

테스크 의존관계 정의

- DAG(Directed Acyclic Graph)
 - 방향성 (Directed) : 간선에 방향이 존재
 - 비순환 (Acyclic) : 그래프의 한 정점에서 시작하여 다시 시작점으로 이어지는 간선이 존재X = 사이클 존재X