

---

# ETL\_Pipeline

AI16 – 배경태

# 목차

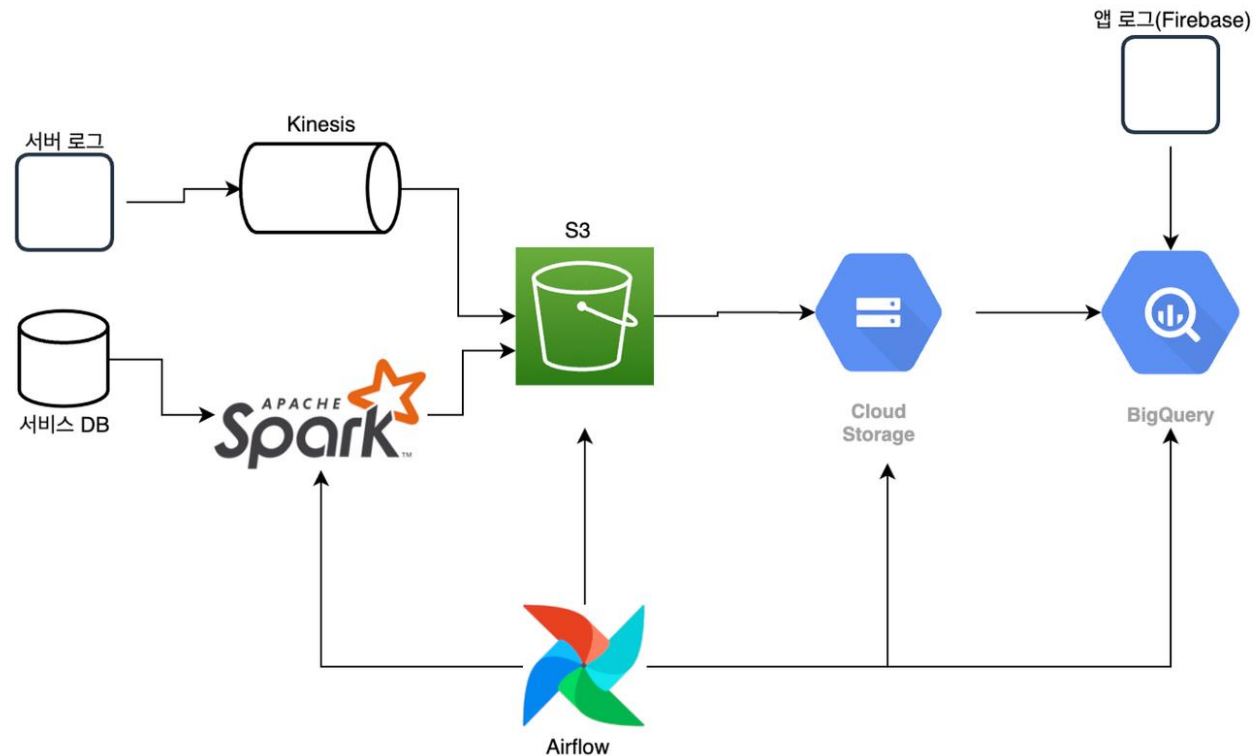
01. 프로젝트 개요

02. 프로젝트 수행

03. 프로젝트 수행 결과

04. 자체 평가 의견

# 1. 프로젝트 개요



1. Extraction (데이터 추출)

2. Transform (데이터 변환)

3. Load(데이터 적재)

데이터 추출 , 압축 변환을 진행하여 AWS S3객체로 저장하는게 목표

## 02. 프로젝트 수행

---



### 1. Extraction ( 데이터 추출 )

- url 을 통한 api 요청
- Json 형식의 데이터
- > 암호화, 복호화 진행

```
data = requests.get(url).json()
```

```
fernet = Fernet(key) # 키생성

#복호화
for i in data:
    decrypted_data = fernet.decrypt(log['data'])
    final_data = decrypt_to_json(decrypted_data_data)
```

## 02. 프로젝트 수행

---



### 2. Transform ( 데이터 변환 )

- 64자로 구성된 user\_id를 B64uuid 모듈 활용하여 44자로 축소
- Method : {'GET': 1, 'POST': 2} 또는 {1: 'GET', 2: 'POST'} -> 1,2로 구분
- inDate : 데이터 문자열 형식으로 압축

## 02. 프로젝트 수행

---



### 3. Load ( 데이터 적재 )

- AWS S3버킷 생성 -> 데이터 변환이 완료된 데이터를 적재
- 압축 알고리즘 (gzip 라이브러리) 적용 후 , 압축률 비교
- 압축된 데이터를 적재

## 04. 프로젝트 수행 결과

---

### ▶ [프로젝트 수행 결과]

- 데이터 적재를 위한 AWS 키 생성까지 진행
- 화재로 인한 추가작업 불가....

## 05. 자체 평가 의견

---

- ▶ 1. Git 을 이용한 실시간 작업, 백업 필수로 진행해야 할 것
- 2. 스케줄링에 대한 추가적인 학습 및 이해 필요
- 3. 파이프라인 전체 구조에 대한 개념 확실화 필요
- 4. ReadMe 매 작업 갱신마다 기입