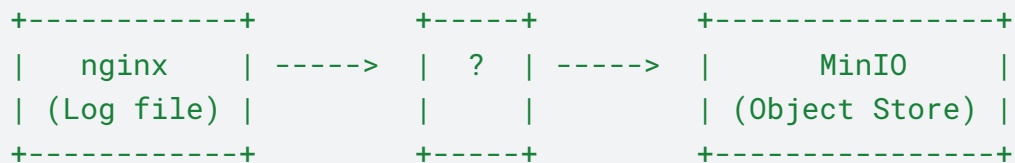


데이터 엔지니어 기술 과제 안내

본 과제는 지원자의 경력을 바탕으로 데이터 파이프라인 설계 및 구현 역량을 평가하기 위해 마련되었습니다. 아래 내용을 확인하신 후, 주어진 환경과 요구사항에 따라 파이프라인을 설계하고 구현하여 제출해 주시기 바랍니다.

과제 개요: Nginx Access Log 적재 파이프라인 구현

Unset



Nginx에서 발생하는 **access log**를 MinIO에 적재하는 파이프라인을 구축해 주세요. 로그는 파일 형태로 저장되며, 이를 주기적으로 수집하여 **Parquet** 포맷으로 변환 후 MinIO에 업로드합니다.

제공 환경

아래와 같이 과제 실행에 필요한 환경이 구성되어 있습니다.

Shell

```
.
├── airflow
│   └── dags
│       └── hello_world.py
├── minio
│   └── data
├── nginx
│   ├── logs
│   └── nginx.conf
├── docker-compose.yaml
├── docker-compose-example.yaml
├── docker-compose-airflow.yaml
└── README.md
```

컴포넌트 설명

1. Nginx
 - **docker-compose.yaml**에서 정의됩니다.

- **./nginx/logs** 디렉토리에 아래와 같은 JSON 형식의 **access** 로그가 저장됩니다.

JSON

```
{
  "remote_addr": "172.18.0.1",
  "remote_user": "",
  "http_user_agent": "Mozilla/5.0 (Macintosh; Intel Mac OS
X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/136.0.0.0 Safari/537.36",
  "host": "localhost",
  "hostname": "605701da20f0",
  "request": "GET / HTTP/1.1",
  "request_method": "GET",
  "request_uri": "/",
  "status": "200",
  "time_iso8601": "2025-05-28T21:47:42+00:00",
  "time_local": "28/May/2025:21:47:42 +0000",
  "uri": "/index.html",
  "http_referer": "",
  "body_bytes_sent": "615"
}
```

2. MinIO

- **docker-compose.yaml**에서 정의됩니다.
- 접속 정보:
 - 사용자명: minio_admin
 - 비밀번호: minio_password
 - 포트: 9001

3. Airflow(선택사항)

- **docker-compose-airflow.yaml**에서 정의됩니다.
- 접속 정보:
 - 사용자명: airflow
 - 비밀번호: airflow
 - 포트: 8080

실행 방법

Shell

```
# Nginx, MinIO 실행
docker-compose -f docker-compose.yaml up

# Nginx, MinIO, Airflow 포함 실행
```

```
docker-compose -f docker-compose.yaml -f
docker-compose-airflow.yaml up -d
```

요구사항

- **Nginx access** 로그 데이터를 **1시간** 간격으로 읽어, MinIO에 **Parquet** 포맷으로 저장하는 파이프라인을 구현하세요.
- 구현한 파이프라인은 **docker-compose** 기반으로 실행 가능해야 합니다.
- **docker-compose.yaml**을 수정하거나 새로운 파일로 작성해도 됩니다.
- 저장된 데이터는 **Apache Spark**를 사용하여 조회 가능한 구조여야 합니다.
- 파이프라인 구현 방식 및 사용 도구에 제한은 없습니다.
 - 제공된 **Airflow** 역시 사용하지 않아도 무방합니다.
- 단, **Nginx** 로그의 포맷은 변경되지 않아야 합니다.

제출 방법

아래 항목을 제출해 주세요.

- 파이프라인 소스 코드
- (선택) 설계 및 구현 문서

제출 방식:

- 이메일(recruit.kx@kakaoenterprise.com) 첨부 제출
- 또는 그리팅 영입시스템 과제 서버를 통해 제출
- 또는 개인 **GitHub Repository**에 업로드 후 **URL** 제출

평가 기준

- 요구사항 충족 여부 및 동작 정확성
- 데이터 파이프라인의 설계 완성도 및 안정성
- 코드 품질 및 유지보수 가능성
- (선택) 설계/구현 문서의 명확성 및 구조

본 과제는 지원자의 데이터 엔지니어링 실무 능력을 종합적으로 평가하기 위한 것입니다. 제출해 주신 결과물은 실무 환경과 유사한 문제 해결 능력을 판단하는 주요 참고자료로 활용됩니다.

지원자님의 훌륭한 결과물을 기대합니다. 감사합니다!