

# 온라인 광고 로그 처리 파이프 라인 및 분석 플랫폼 만들기

---

2026. 01. 27. 화  
ML Engineer  
정의근

# 01 프로젝트 목표 및 수행 내용

## 프로젝트 목표

- 데이터와 비즈니스 모델에 대한 이해 중요성 알기
- 현업 데이터 엔지니어 업무 맛보기

## 수행 내용

- 온라인 광고 데이터 이해, 비즈니스 모델 이해
- 온라인 광고 사용자 행동 로그 생성 시뮬레이터 작성
- 실시간 로그 수집 파이프라인 구축
- 광고 성과 분석 데이터 ETL 파이프라인 구축
- SQL 기반 데이터 분석 환경 구축
- (Optional) AI 기반 데이터 분석 시스템 개발
- 참여 인원 전원 필수
  - 온라인 광고 데이터 이해, 비즈니스 모델 이해
  - 구현 기술들에 대한 이해
- 업무 분장 & 협업
  - 나머지 작업들

**중요: 하는 만큼 배우게 될 것**

## 02 수행 내용 이해를 위한 키워드 또는 기술 스택

---

- 온라인 광고 데이터 이해, 비즈니스 모델 이해 (도메인 지식 쌓기)
  - 광고 지면, 광고주, 캠페인, 크리에이티브 등
  - 노출, 클릭, 전환, 클릭 당 단가(CPC), CTR, CVR, RPM 등
- 온라인 광고 사용자 행동 로그 생성 시뮬레이터
  - 랜덤하게 광고 로그를 생성하는 **python programs**
  - **AWS Kinesis Producer** 역할
- 실시간 로그 수집 파이프라인 (**실시간**)
  - **AWS Lambda -> S3** 파일 저장
  - **AWS Kinesis Consumer** 역할
- 광고 성과 분석 데이터 ETL 파이프라인 (**배치**)
  - **Airflow** = 배치 스케줄러
  - **python + SQL to AWS Athena -> S3** 파일 저장
  - **Data Schema** = a set of AWS Glue Tables
- SQL 기반 데이터 분석 환경 (데이터 대시보드)
  - **Redash + AWS Athena + Glue Table <- S3** 파일 로드

## 02 수행 내용 이해를 위한 키워드 또는 기술 스택

---

- **(Optional) AI 기반 데이터 분석 시스템**
  - Vanna, Pydantic AI, Google ADK 와 같은 Agent SDK 활용
  - 1. 자연어 입력을 SQL로 변환 (Text-to-SQL)
    - 자연어 입력 예시 : "2026-01-26 기준 1주 동안 캠페인들의 CTR 변화를 분석해줘."
  - 2. SQL을 실행해서 데이터 해석 및 그래프 제공