

## 데이블 / 정의근

# 온라인 광고 로그 파이프라인 및 AI 기반 데이터 분석 도우미 만들기

**2026.01**



# Content

## Ⅰ.프로젝트 개요

- 1. 프로젝트 주제명·····3
- 2. 프로젝트 수행 배경·····3
- 3. 프로젝트 목표 및 수행 내용·····3
- 4. 개발을 위한 환경 및 요구사항·····3
- 5. 활용되는 데이터 형태·····3

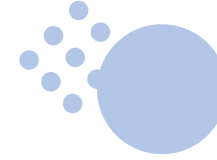
## Ⅱ.상세 구현/분석 방법 및 예시

- 1. 구현/분석 프로세스·····4
- 2. 구현/분석 상세 내용 및 예시·····4
- 3. 데이터 리소스 활용 상세·····5
- 4. 예상 계획 및 일정·····5

# I.프로젝트 개요

## 프로젝트 주제명

온라인 광고 로그 파이프라인 및 AI 기반 데이터 분석 도우미 만들기



## 프로젝트 수행 배경

- 온라인 광고는 인터넷 서비스 기반 기업들이 전통적으로 선택해 온 핵심 수익 모델 중 하나입니다. 데이터 엔지니어링 업무의 특성상, 단순히 기술 스택을 다루는 것을 넘어 처리하는 데이터가 어떤 비즈니스 맥락을 가지는지, 즉 도메인 지식을 이해하는 것이 매우 중요합니다. 본 프로젝트는 온라인 광고 도메인에 대한 이해를 바탕으로, 실제 데이터 엔지니어링 업무 흐름을 경험할 수 있는 기회를 제공하는 것을 목표로 합니다.
- 온라인 광고는 사용자의 요청에 따라 실시간으로 노출되며, 광고 노출 이후 클릭, 서비스 가입, 구매 등으로 이어지는 사용자의 모든 행동이 로그 형태로 실시간 수집됩니다. 이러한 특성으로 인해 온라인 광고 시스템은 대규모 실시간 로그 수집, 안정적인 데이터 적재, 그리고 분석을 위한 데이터 파이프라인 구성이 필수적인 영역입니다.

## 프로젝트 목표 및 수행 내용

- 온라인 광고에서의 사용자 행동 로그 생성 시뮬레이터 -> 실시간 로그 수집 파이프라인 -> 광고 성과 분석 데이터 ETL 파이프라인 -> SQL 기반 데이터 분석 환경 구축
- (Optional) AI 기반 데이터 분석 시스템 개발 : Text-to-SQL, SQL 실행 후 그래프 작성

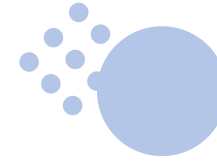
## 데이터 파이프라인 구축을 위한 환경 및 요구 사항

- 파이썬 개발 환경
- 도커(컨테이너 환경)
- 클라우드 환경 (AWS 선호)
- Github 환경

## 활용되는 데이터 출처 및 형태

- 온라인 광고 환경을 모사해서 자동 생성한 무작위 텍스트 데이터

## II.상세 구현/분석 방법 및 예시



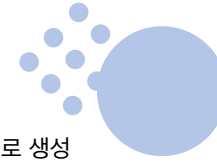
### 1. 구현/분석 프로세스

- 온라인 광고 환경을 모사한 무작위 로그 생성기 개발
  - 로그 생성기에서 메시지 큐 (e.g. AWS Kinesis) 로 실시간 로그 전송
- 실시간 로그 저장 및 로그 변환/적재를 위한 ETL 파이프라인 구축
- 데이터 대시보드 구축
- (Optional) AI를 활용한 Text-to-SQL 툴 개발

### 2. 구현/분석 상세 내용 및 예시

- 온라인 광고 환경을 모사한 무작위 로그 생성기 개발
  - 온라인 광고 데이터에 대한 이해 필수 (도메인 지식 쌓기)
  - 키워드 목록 : 광고 지면, 광고주, 캠페인, 크리에이티브, 노출, 클릭, 전환, 클릭 당 단가, CTR, CVR, RPM 등
  - 로그 생성기에서 메시지 큐 (e.g. AWS Kinesis) 로 실시간 로그 전송
- 실시간 로그 저장
  - 메시지 큐에 저장된 로그를 실시간 프로세서(e.g. AWS Lambda)를 통해 S3 에 저장
- 로그 변환/적재를 위한 ETL 파이프라인 구축
  - 스케줄러 프로그램(e.g. Airflow) 을 통해 주기적으로 분석용 데이터 생성
  - SQL 을 필수적으로 사용해야 하며 예외적으로 Pandas 또는 Polars 사용 가능
  - SQL 기반 데이터 스키마 구성
- 데이터 대시보드 구축
  - SQL 기반으로 데이터 분석할 수 있도록 환경 제공
  - 차트, 그래프 기능 필요
  - CTR, CVR, RPM 등 온라인 광고 세계에서 필수적으로 분석하는 수치들에 대한 대시보드 제공
- (Optional) AI를 활용한 Text-to-SQL 툴 개발
  - Vanna, Google ADK 와 같은 Agent 툴을 활용하여 자연어 입력을 SQL로 바꿔서 실행하고 그래프까지 그려서 제공하는 툴
  - 자연어 입력 예시 : "2026-01-23일 동안 캠페인들의 CTR, CVR, RPM 을 RPM 역순으로 보여줘."

## II. 상세 구현/분석 방법 및 예시

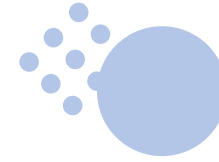


### 3. 데이터 리소스 활용 상세

- 온라인 광고 데이터에 대한 이해를 바탕으로 유사한 형태의 로그들을 무작위로 생성

### 4. 예상 계획 및 일정

- 2월
  - 1주차 : 온라인 광고 데이터에 대한 이해
  - 2주차 : 무작위 로그 생성기 개발 및 실시간 로그 저장
  - 3주차 : 로그 변환/적재를 위한 ETL 파이프라인 구축
    - 단, 3주차는 설연휴로 20~22일 진행 가능 예상
  - 4주차 : 로그 변환/적재를 위한 ETL 파이프라인 구축
- 3월
  - 1주차 : 로그 변환/적재를 위한 ETL 파이프라인 구축 + 데이터 대시보드 구축
  - 2주차 : 데이터 대시보드 구축, 앞서 개발한 시스템 개선
  - 3주차 ~ 4주차 : (일정이 미뤄지지 않는 경우) AI를 활용한 Text-to-SQL 툴 개발
    - 일정이 미뤄져서 진행이 어려운 경우 관련 기술 스터디 및 데이블에서 개발한 툴 설명/데모로 대체 예상



## 기업 연계 프로젝트 주제

Ver1.0

2025년 06월

집필자

펴낸곳

SK planet

책임편집

SK planet T아카데미

주소

서울시 마포구 마포대로 122

전자우편

tacademy@tacademy.skplanet.com

전화번호

02-6119-1166

6

본 교재의 판권 및 저작권은 집필자 및 펴낸 곳의 소유이므로 무단 전재나 복사를 금합니다.