

# 윤병우

☎ 010-4695-5394   @lms46784678@gmail.com

## [핵심 역량]

- 대용량 데이터 수집 및 적재 파이프라인 구성 경험
- Snowflake 내부 데이터 웨어하우스 구축 (SQL, Spark[Snowpark])
- Python Flask, FastAPI 기반의 모델, 데이터 수집 서버, AI 챗봇 서버 구축 경험
- 데이터분석을 위한 Pandas, SQL, Spark 활용 경험
- AWS 기반의 EC2, Lambda 서버리스 서버 구성, RDS, DMS, S3, Kinesis 활용 프로젝트 경험
- 기획, 디자인, 프론트와의 다양한 프로젝트 경험

## [데이터 수집 파이프라인 구성 경험]

대량 수집 파이프라인을 운영해본 경험(일 800만 + 월 2000만 요청) 이 있습니다. 대량 데이터 수집을 빠짐 없이 진행하기 위해 Kafka 와 수동 커밋을 활용하고 있습니다.

데이터 수집 로직의 빠른 유지보수를 위해 FastAPI 기반의 수집서버를 구성 및 AWS Lambda 를 활용하여 요청이 필요없는 시기에는 비용 최소화 및 대량 수집 요청에도 대응 할 수 있도록 구성하고 있습니다.

## [데이터 웨어하우스 경험]

모 회사에서는 20개 이상의 프랜차이즈의 주문 데이터를 각각의 RDS(mysql) 에서 관리하고 있으며, 이를 통합으로 분석하기 위해 Snowflake 를 도입과정에 참여하며 멀티 계정의 RDS 를 통합하는 프로젝트를 진행하고 있으며, 수집된 데이터를 기반으로 한 서비스 구축을 위해 통합 SpeedLayer RDS 를 구축하여 준 실시간, 배치 ETL 작업을 하는 프로젝트를 직접 구성하고 고도화 하는 작업을 하고 있습니다.

## [AI + 데이터 엔지니어]

일을 하다보면 다양한 분들로 부터 데이터 요청을 받을 일이 많은데요. 간단한 쿼리작업 같은 일을 대신해주는 SQL Agent 를 직접 구성하는 스몰 프로젝트를 진행 한 경험이 있습니다. LangGraph 로 구성한 챗봇과 RAG 를 결합하여 자연어를 적절한 쿼리로 바꾸어 쿼리 생성 및 쿼리 결과를 보여주는 서비스를 만든 경험이 있습니다.

## [☑데이터 엔지니어로서]

시작은 데이터 수집으로 시작하며 대량의 데이터 수집, 방화 로직 회피등의 경험과 Snowflake 를 도입하며 여러 데이터 소스를 통합하는 경험을 하였으며,

## [팀원으로써]

저의 업무의 시작은 "왜?" 입니다. 업무 몰입에 중요한 것은 목적성이라고 생각합니다. 특히, 데이터 엔지니어로서 하는 역할이 서비스에 직접적인 부분이 적기 때문에 모든 일에는 목적성을 잃지 않으려 노력하고 있습니다. 그러기 위해 프로젝트에 몰입하기전에 저는 제가 하는 모든일이 "왜?" 필요한지 고민하려고 하고 있습니다.



## 주식회사 한국신용데이터

2025.09 - 재직중 (1개월) | 정규직 | 데이터엔지니어 (수집) | 데이터엔지니어 (수집)

### 재직중

2025.09 - 2025.09

재직중입니다.



## 서클플랫폼주식회사

2023.11 - 2025.09 (1년 11개월) | 정규직 | 데이터 엔지니어 | 데이터 웨어하우스 및 데이터 수집 인프라 구축

### AI Recipe

2025.03 - 2025.04

역할: 요리데이터 대량 수집 및 임베딩, 요리레시피 찾기 챗봇 구성

성과

- 풀 스트리밍 방식의 챗봇 구현
- 사용자의 자연어로 DB 에서 적절한 레시피 1개를 찾아가는 로직 검증 완료

### SQL Agent

2025.02 - 2025.03

역할: 셀러박스 셀러들을 위한 SQL Agent 챗봇 구성

성과

- 셀러박스 최초의 내부 데이터 기반 챗봇서비스
- 기본 통계이외의 다양한 데이터 정보를 자연어로 추출하는 로직 검증 성공

### Tips 진행

2024.11 - 2025.02

역할: 도매몰, 1688 상품데이터 실시간 수집

성과

- 3개의 도매몰, 1688 데이터 수집 과제 달성

### Snowflake DW 구축 고도화

역할 : SpeedLayer(mysql) 스키마 구성 및 인덱싱, Snowflake 파이프라인 고도화

성과

- Snowflake 파이프라인을 snowpipe 에서 스케줄방식의 Copy Into 방식으로 프랜차이즈당 월 180\$ 절감
- 점주앱 B/E 의 쿼리비용 감소 (쿼리당 비용 -> 요청량에 관계없는 시간당 비용방식으로 변경)

### 점주앱(가칭)

2024.09 - 2024.11

역할 : AWS Aurora RDS -> Snowflake 데이터 이관 및 Data warehouse, Data mart Stream, Batch 파이프라인 구축

성과

- 기존 RDS 에서의 분석 DB -> Snowflake 로 이관 및 비용 최적화
- Mysql 의 Delete & Insert 방식을 Snowflake 를 활용하여 Stream + Merge Query 를 적용
- 기존 분석에서는 ETL 프로세스상 3일전의 데이터만 제대로 분석할 수 있는 문제를 준실시간(운영DB와 최소 10초 이내로 동기화 가능)으로 분석할 수 있는 환경 구축

### 셀러 키워드 1.7.2 고도화

2023.12 - 2024.04

역할 : 데이터 수집, 네이버 쇼핑 스크래핑 방지 회피 로직 개발, FastAPI 도입 및 로직 비동기화

성과

- 타사는 쇼핑의 데이터 수집을 포기할정도로 네이버의 보안이 강화 되었지만 AWS Lambda + 자체 개발한 Fakeheaders 로 월간 최대 2000만건의 데이터 수집 안정화
- 월 60만원이상의 프록시업체의 비용을 AWS Lambda 로 전환하여 비용 100% 절감

### 셀러박스 2.3 고도화

2024.03 - 2024.11

역할 : 데이터 수집 주기의 단축을 위한 모든 로직 비동기화, AWS Lambda 로 전환, 상품발송을 위한 오픈마켓 API 통합 구성

성과

- 데이터 수집의 모든 로직을 비동기화 하여 데이터 수집 시간을 평균 60% 이상 단축
  - asyncio semaphore 를 활용하여 고객사의 RPS(Requests Per Second) 최대 활용
  - 데이터 수집주기 10분 10분 주기 + 수집 버튼 도입
- EC2 Lambda 로 변경하여 비용 30 ~ 40% 절감
  - EC2 인스턴스 - 8 개 ( 메인 7개, 개발서버 1개) Lambda Serverless( 메인 1개, 개발 1개), Lightsail Login window Server 1개



(주)유니에스

2021.04 - 2023.11 (2년 8개월) | 정규직

### WMS 물류 자체 시스템 개발 및 유지 보수

2022.05 - 2023.08

- WMS 자체 개발 및 업무 프로세스 개선 경험
- 코로나로 인한 고객사(헬스케어)의 물량 상승으로 인한 관리 어려움으로 Excel 을 기반으로 하여 VBA 와 Python(xlwings)를 활용하여 자체 CI 구성
- 여러 물류창고의 재고를 통합관리해야 함에도 그 해 실물 전산 재고 100% 일치 달성
- 물류량 400% 상승에도 기존 업무시간내에 처리 가능하도록 프로세스 최적화

## 학력



계양고등학교

2008.03 - 2010.03 | 졸업 | 이과

## 스킬

Python

Docker

Apache Kafka

Apache Airflow

Apache Spark

MySQL

SQL

AWS

Linux

Notion

GitHub

Flask

FastAPI

Slack

JIRA

Git

데이터웨어 하우징

데이터베이스 관리

## 수상/자격증/기타



[멀티캠퍼스 K-Digital] 데이터 사이언스/엔지니어링 전문가 17,18회차

2022.11 | 기타

Python 을 기반으로 하여 데이터기반 의사결정을 위한 ETL Pipeline 시스템의 제작과 Data Science 와의 협업을 위주한 프로젝트 교육프로그램에 참여하여 우수상을 수상하였으며, 혼자 개발해오던 방식에서 Team 프로젝트를 진행하며 협업의 위대함을 깨닫게 되어 사람의 역량은 단순히 1+1=2 라는 수학적 결과로 이어지지 않다는 것을 깨닫게 되는 과정이었습니다.



제48회 SQL 개발자(SQLD)

2023.04 | 기타

SQL 을 심도있게 배우기위해 해당 자격증 취득했습니다.



Dothis 프로젝트 (Data Engineer)

2023.09 | 기타

최근의 유튜버들은 영상 편집보다 소재 찾기에 더욱 힘들어하고 있으며 이를 해소하기 위해 다양한 소스의 데이터를 기반으로 AI 를 학습시켜 트랜드에 적합한 소재 추천 및 스크립트 구성까지 해줄 수 있는 서비스를 개발 하는 프로젝트입니다.

- ☑MariaDB 의 Datawarehouse ETL 운영
- Airflow 를 활용한 ETL 파이프라인 Dag 구성
- 분산 크롤링 아키텍처 구성을 위해 Kafka 를 활용하여 Producer, Consumer 자체 구성하여 활용

## 링크

---

[🔗 https://github.com/Themath93](https://github.com/Themath93)

[🔗](https://lizard-dive-9a3.notion.site/37b80353afce44a19913c66614f23674?pvs=4) 경력기술서, 포트폴리오를 확인 하실 수 있는 노션링크 입니다.

<https://lizard-dive-9a3.notion.site/37b80353afce44a19913c66614f23674?pvs=4>