



# 윤병우의 포트폴리오



## About Me

데이터 전문가로 성장중인 **윤병우**입니다. 초기에는 **Fulfilment** 업무를 맡으며 고객사 WMS 프로그램을 직접 개발하며 개발자로 첫발을 내딛었으며, **데이터 엔지니어링, Snowflake, AI챗봇** 및 다양한 AI 서비스등 구성 및 공공기관 AI 챗봇 프로젝트도 진행하고 있습니다.

어떤 기술을 사용할때는 정의된 문서를 먼저 살펴본 후 사용하려고 노력하고 있으며, **이런 습관으로 오픈소스 문서의 오타를 찾아 수정하여 기여하는 경험**을 하기도 하였습니다.

**Langchain PR 링크**

데이터를 흐르는 관점에서만 보는 것이 아닌, 실제 데이터로 무엇을 할 수 있는 지에 대해 많이 시도해보고 있습니다.

## Yoon ByeongWoo

Seoul, South Korea  
Tel. 010-4695-5394  
E. lms46784678@gmail.com  
Github.  
<https://github.com/Themath93>

## Experiences (1년 10개월 as DE)

- DHL SC(사원) 2021.04 ~ 2023.11
- Dothis 프로젝트(Data Engineer) 2023.09 ~ 2024.10
- (주)서울플랫폼(Data Engineer) 2023.11 ~ 2025.09
- (주)한국신용데이터(Data Engineer) 2025.09 ~ 재직중

## Navigator

## Education

- 계양고등학교 2008.03 - 2011.02
- 멀티캠퍼스 데이터 엔지니어 과정 2022.5 - 2022.11

## Job Experience

(주)서울플랫폼 (Data Engineer) 2023.11 ~ 2025.09 (1년 10개월)

데이터 엔지니어 역할

▼ 프로젝트 : 셀러키워드 1.7.2 고도화 - 링크 - 2023.12 ~ 2024.04

- 인원 : 데이터 엔지니어 2명, 프론트엔드 엔지니어 2명, 백엔드 엔지니어 1명
- 목표 : 네이버 쇼핑 데이터(Hidden API) 대량 실시간 스크래핑 로직 개발

▼ 개요

🗣️ 이커머스 셀러분들에게 키워드에 대한 경쟁정도, 시장성, 광고비등을 한눈에 제공하기 위해 동시에 다양한 쇼핑데이터를 비동기방식으로 한 번에 수집하여 인사이트를 제공하였습니다.

- 네이버 쇼핑 데이터 수집 및 스크래핑 방지 회피 로직 개발 (IP 프록시, Cookie,
- FastAPI 도입 및 로직 비동기화
- 팀 내 협업을 통해 스크래핑 로직 최적화 및 오류 수정

#### ▼ 📊 성과

- **AWS Lambda + 자체 개발한 Fakeheaders** 로 월간 최대 2000만건의 데이터 수집 안정화
- 사설 프록시 서버를 **AWS Lambda 프록시 서버로 전환하여 비용을 80% 이상 절감**, 이를 통해 연간 약 720만 원의 비용 절감 효과 달성
- 데이터 수집 안정화로 인해 서비스 제공 속도 개선 및 **추가 유료 서비스 개발(상품모니터링) 요구조건 달성**

#### ▼ 프로젝트 : 📦 셀러박스 2.3 고도화 - 🔗 링크 - 2024.03 ~ 2024.08

- 👥 **인원** : 데이터 엔지니어 1명, 프론트엔드 엔지니어 1명, 백엔드 엔지니어 1명
- 📖 **목표**: 수집 로직 비동기화 및 다양한 수집서버를 하나의 도메인으로 통합

#### ▼ 🗣️ 개요

🗣️ 보다 빠른 데이터수집 서버를 구축하기위해 **모든 로직을 비동기화** 및 서버를 EC2 → AWS Lambda 로 변경하여 **동시 대량 요청 대응 및 비용 절감** 하였으며, **API Gateway** 로 여러 수집 서버, 인증 서버를 통합하여 관리포인트를 최소화 하였습니다.

- 여러 오픈마켓 API 분석 및 통합 작업을 수행하여 API 간 호환성 향상
- 기존 코드 리팩토링을 통해 성능 개선 및 유지 보수 효율성 증가
- 데이터 수집서버 아키텍처 개선
- 배포 자동화 프로세스 구축을 주도하여 팀의 작업 효율성을 높임

#### ▼ 📊 성과

- 모든 데이터 수집 로직을 **비동기화** 후 수집 시간을 평균 **60% 이상 단축**, 데이터 수집 주기를 30분에서 10분으로 개선 및 수동 수집 버튼 도입
- **asyncio semaphore**를 활용하여 오픈마켓의 RPS(Requests Per Second)를 최대한 활용, **API 호출 효율성을 극대화**
- 8개 이상의 수집 API 서버를 하나의 도메인(API Gateway)으로 통합하여 관리 비용 절감 및 안정성 향상
- Github Actions와 Serverless(Lambda) 구성을 통해 배포 시간을 **90% 이상 단축**

#### ▼ 프로젝트 : ❄️ 스노우플레이크 DW(Data warehouse) 구축[소복소복] - 🔗 링크 - 2024.11 ~ 2025.09

- 👥 **인원** : 데이터 엔지니어 1명, 프론트엔드 엔지니어 1명, 백엔드 엔지니어 1명, PM 1명, 앱 디자이너 1명
- 📖 **목표** : 10개 이상의 카페 브랜드 점주분들이 브랜드에 상관없이 실시간으로 주문을 확인하고 고객에게 개인적인 마케팅을 할 수 있는 브랜드 통합앱(점주앱) 프로젝트

#### ▼ 🗣️ 역할

- 각 브랜드마다 데이터베이스가 달라 이를 통합할 수 있는 데이터베이스로 이관하는 작업 수행
  - Snowflake, AWS DMS 활용
  - RDS(binlog) → DMS(CDC) → S3(Parquet) → Snowflake(Snowpipe)
  - 각 브랜드마다 RDS 인스턴스가 다르기때문에 이관작업순서를 매뉴얼로 정리

#### ▼ 📊 성과

- 기존 분석에서는 ETL 프로세스상 3일전의 데이터만 제대로 분석할 수 있는 문제를 준실시간(운영DB와 최소 10초 이내로 동기화 가능)으로 분석할 수 있는 환경 구축
- 기존 MSTR 과 연결된 DW(RDS)에서는 사용할 수 없던 복잡한 Query 를 손쉽게 수행

▼ 프로젝트 : ❄️ 스노우플레이크 DW(Data warehouse) 고도화 - [링크](#) - 2025.02 ~ 2025.09

- 👥인원 : 데이터 엔지니어 3명
- 📖목표 : 점주앱의 Snowflake Query 비용 절감 및 데이터 파이프라인 비용 절감
- ▼ 🧑역할
  - 점주앱에 사용되는 B/E API 쿼리 확인 및 SpeedLayer RDS 의 스키마구성 및 인덱싱
  - 스노우플레이크 내부 파이프라인 구성 변경

▼ 프로젝트 : 📦 셀러박스3.0 수집 로직 개편 - [링크](#) 2025.06 ~ 2025.07

- 👥인원 : 데이터 엔지니어 1명, 프론트 엔지니어 1명, 기획자 1명
- 📖목표 : 수집 로직 서비스 신뢰도 제고를 위한 수집 로직 고도화
- ▼ 🧑역할
  - Flask 로직 → FastAPI 이전
  - 대량의 수집요청을 막힘 없이 처리하기 위한 중앙 RPS 관리 서버 도입
  - 주문건별 데이터 수집 로직 생성
  - 다양한 수집 인스턴스 개편
- ▼ 📈성과
  - 주문, 문의, 리뷰, 정산등의 데이터 수집, 처리 로직의 실패가 없도록 조정
  - 개별 주문 수집 API 를 구성하여 미 완결된 주문건을 완결상태로 보낼 수 있도록 하여 데이터 신뢰성 증가
  - 코드 관리와 유지보수방식을 개편하여 디버깅, 테스트, 배포 시간 단축

🤖 AI 엔지니어 역할

▼ 프로젝트 : 💬 SQL Agent [셀러박스] - [링크](#) 2025.02 ~ 2025.03

- 👥인원 : 데이터 엔지니어 1명, 프론트 엔지니어 1명, 기획자 1명
- 📖목표 : 셀러박스 셀러분들의 여러 오픈마켓의 데이터들을 모아 해당 데이터와 직접 대화할 수 있는 챗봇 구성
- ▼ 🧑역할
  - LangGraph를 활용하여 SQL Agent 챗봇 서비스 구성
  - DB → SQL Agent 로의 실시간 데이터파이프라인 구성 (민감정보 제거 전처리 포함)
  - FastAPI 를 사용하여 챗봇 서비스 API 서버 구성
- ▼ 📈성과
  - 서클플랫폼 최초의 챗봇 서비스 오픈
  - 주문확인, 기본통계 확인외에 챗봇서비스로 인해 앱 체류 시간 증가

▼ 프로젝트 : 💬 Chat My Excel - [링크](#) 2025.04 ~ 2025.04

- 👥인원 : 데이터 엔지니어 1명, 프론트 엔지니어 1명, 기획자 1명
- 📖목표 : 사용자로부터 Excel 파일을 제공받아 파일과 대화하는 챗봇 및 인터랙티브한 파트 생성
- ▼ 🧑역할
  - Excel 을 챗봇이 적은 토큰과 빠른 속도로 읽고 답할 수 있는 로직 구성
  - 사용자 요청에 따라 차트를 생성할 수 있는 로직 구성
  - LangGraph 를 통한 RAG 챗봇 구성
- ▼ 📈성과
  - 대량의 Excel 파일도 채팅을 통해 빠르게 확인하고 집계, 시계열 데이터를 빠르게 확인할 수 있는 서비스로 인해 마케터, 영업 담당자들의 업무시간 단축
  - Excel 을 내부적으로 DB 형태로 가공 후 SQL 질의 방식으로 진행하도록 하여 차트 생성에 용이하도록 구성, SQL 을 사용하기 때문에 Self-RAG 방식이 자연스럽게 도입되어 할루시네이션이 적게 일어나고 토큰을 적게 사용할 수 있음

▼ 프로젝트 : 💬 공공기관 지방세 답변 챗봇 2025.07 ~ 2025.09

- 👤인원 : AI 엔지니어 1명(본인), PM 1명, B/E 1명, F/E 1명
- 📌목표 : 공공기관 지방세 문의답변 채팅에 지방세 문의 답변 챗봇 구성
- ▼ 🧑역할
  - 폐쇄망에 FastAPI + LangChain 을 활용한 챗봇 서버 구성
  - 멀티 LLM 사용 Llama 4 17B, Llama 3 8B, Gemma 모델을 유동적으로 사용할 수 있는 로직 구성
  - CPU, GPU 인프라모니터링 API 구성
  - RAGAS 를 활용한 RAG 파이프라인 검증로직도입
- ▼ 📈성과
  - 진행 중인 프로젝트 입니다.
  - 예상성과) 공무원의 단순 반복적인 응대업무를 챗봇으로 대체

## (주)한국신용데이터-KCD (Data Engineer) 2025.09 ~

## Projects

### ▼ 외부프로젝트요약 펼치기

Project	Date	Role	Output	Link
고속도로 전기차 충전 현황 파악	22.08 ~ 22.08	데이터를 통한 의사결정 과정 결정, 이상 충전 신호 분석, 추천 충전소 시각화	우수상 수상	<a href="https://www.notion.so/70e8fb113d6b4bbb84d7ad41a31a80pvs=4">https://www.notion.so/70e8fb113d6b4bbb84d7ad41a31a80pvs=4</a>
증상에 따른 병원 추천 알고리즘	22.10 ~ 22.11	ETL 파이프라인 구성 및 데이터 크롤링	우수상 수상	<a href="https://www.notion.so/Hospital-Recommend-ETL-Crawling65f9b33d64eb4ca5bfe7a43b0c41ac9f?pvs=4">https://www.notion.so/Hospital-Recommend-ETL-Crawling65f9b33d64eb4ca5bfe7a43b0c41ac9f?pvs=4</a>
창고 재고관리 시스템	23.01 ~ 23.5	웹서비스, 프로그램화된 엑셀 파일 배포 및 유지보수	기존 매일 2시간 O.T 있던 업무환경 9 to 6 및 고객사의 빠른 재고 파악 가능해짐	<a href="https://www.notion.so/WMS-Warehouse-Management-System-9fdc127a9bfb41dc88e67bdecc81fb3?pvs=4">https://www.notion.so/WMS-Warehouse-Management-System-9fdc127a9bfb41dc88e67bdecc81fb3?pvs=4</a>
유튜브 소재추천 AI	23.09 ~ 진행 중	데이터엔지니어 (데이터 크롤링 병렬 구성, Airflow DAG 관리, AWS EC2 AMI 관리 및 S3 데이터 적재)	880만개 데이터 크롤링 122 시간에서 3.5 시간으로 단축 및 Airflow 도입 <b>계속 진행중</b>	<a href="https://dothis.kr/">https://dothis.kr/</a>













### Projects

이름	태그	날짜
📦 셀러박스 주문 수집고도화	인프라 AWS ETL 데이터엔지니어 서클플랫폼	@2025년 6월 10일 → 2025년 7월 15일
💬 Chat My Excel	AI 데이터분석 서클플랫폼 진행중	@2025년 4월 1일 → 2025년 5월 1일
❄️ 통합 데이터 웨어하우스 [멀티 계정]	AWS ETL 데이터엔지니어 진행중	@2025년 2월 1일
🤖 SQL Agent	AI 데이터엔지니어 서클플랫폼 자연어처리	@2025년 2월 1일 → 2025년 3월 25일
❄️ Snowflake 데이터 이관(브랜드 점주 앱)	AWS ETL 데이터엔지니어 서클플랫폼	@2024년 11월 1일
📦 셀러박스 고도화	AWS Crawling 서클플랫폼 업무개선	@2024년 6월 1일

이름	태그	날짜
 <a href="#">셀러키워드(셀키) 1.7 고도화</a>	AWS Crawling 데이터엔지니어 서클플랫폼 자체개발	@2023년 12월 1일 → 2024년 3월 1일
 <a href="#">Dothis 유튜브 소재추천 AI</a>	AI AWS 데이터엔지니어	@2023년 9월 28일 → 2024년 11월 15일
 <a href="#">WMS(Warehouse Management System)</a>	Service 업무개선 자체개발	@2023년 1월 5일 → 2023년 5월 5일
 <a href="#">Hospital Recommend(ETL &amp; Crawling)</a>	Crawling ETL 자연어처리	@2022년 10월 11일 → 2022년 11월 18일
 <a href="#">고속도로 전기차 충전 현황 파악</a>	공공데이터 데이터분석 데이터시각화	@2022년 8월 12일 → 2022년 8월 25일

## Skills

### Skills

Skill	Tags	Experience Level
 <a href="#">Python</a>	AI Data Engineering Language	중
 <a href="#">RDBMS</a>	Data Engineering Database SQL	중
 <a href="#">Kafka</a>	Data Engineering	중
 <a href="#">Airflow</a>	Data Engineering Scheduler	중
 <a href="#">Crawling</a>	Data Engineering	상
 <a href="#">AWS</a>	Cloud Data Engineering	중
 <a href="#">FastAPI</a>	API Framework	중
 <a href="#">Git</a>	Configuration Management Tool	중
 <a href="#">Databricks</a>	Cloud Data Engineering SQL Spark	하
 <a href="#">Snowflake</a>	Cloud Data Engineering SQL	중
 <a href="#">Grafana</a>	BI Monitoring SQL	하
 <a href="#">LangChain</a>	AI Data Engineering	상

## Thank You!

최근들어 AI 서비스가 많이 생겨나면서 많은 변화가 생기고 있습니다. 저는 그중에서도 사람만이 해야할 수 있는 것은 반드시 있다고 생각하고 있습니다.

많은 AI 서비스가 출시되고 있지만, 품질 높은 AI 서비스는 AI의 작동 방식을 잘 이해하고 서비스하려는 도메인을 제대로 파악한 사람이 작성한 프롬프트와 AI 제어 방식에서 비롯된다고 생각합니다.

저는 챗봇과 AI 서비스 프로젝트를 진행하며 원하는 답변을 얻기가 굉장히 어렵다는 것을 많이 깨닫고 있습니다. 이는 풍부한 경험과 도메인에 대한 깊은 이해를 가진 사람만이 해결할 수 있는 과제라고 생각합니다.

저는 AI 자체보다 서비스에 대한 이해를 깊게 가져가며 품질 높은 AI 서비스를 만드는 개발자가 되고 싶습니다.

감사합니다.

To the Top [Contact](#)©Yoon ByeongWoo 2025

