

류현지 | 주요 프로젝트

데이터 분석 자동화 시스템 개발

2026.04 ~

프로젝트 개요

자연어 기반으로 데이터를 조회하고 분석할 수 있는 분석 자동화 시스템 개발 프로젝트를 수행했습니다. 사용자의 질의를 기반으로 추가 정보를 수집하고, DataHub MCP 기반 메타데이터 검색 및 SQL 자동 생성 구조를 구축하여 자연어만으로 데이터 조회와 분석이 가능하도록 시스템을 개발했습니다.

주요 수행 내용

- 자연어 질의를 기반으로 추가 정보 수집 및 분석 워크플로우 설계
- DataHub MCP 기반 메타데이터 검색 구조 개발
- 테이블·컬럼 메타데이터 기반 SQL 자동 생성 기능 개발
- 데이터 조회 결과 반환 및 후처리 워크플로우 구성
- LLM 기반 데이터 분석 자동화 워크플로우 설계 및 개발

사용 기술

LLM, MCP, DataHub, Python

실시간 데이터 플랫폼 구축

2026.02 ~ 2026.04

프로젝트 개요

기존 Spark 배치 중심 데이터 처리 환경에서 실시간 처리 요구사항이 증가함에 따라 Spark Streaming 기반 실시간 데이터 처리 플랫폼을 구축했습니다. Streaming Job 개발·배포·운영 체계를 표준화하고 모니터링 환경을 함께 구축하여 안정적인 실시간 데이터 처리 및 운영 환경을 마련했습니다.

주요 수행 내용

- Spark Structured Streaming 기반 실시간 처리 플랫폼 구축
- Streaming Job 개발·테스트·배포 자동화 파이프라인 설계
- Job 실행 및 운영 환경 표준화
- 처리 지연·상태·장애 여부 모니터링 서버 구축
- 팀 내 Streaming 개발 및 운영 가이드 문서 작성·공유

사용 기술

Spark Streaming, Kubernetes, Python

AI 검색 추천 프로젝트

2026.01 ~ 2026.04

프로젝트 개요

검색 결과 개인화를 위해 AWS Personalize Rerank 기반 검색 추천 시스템 개발 프로젝트를 수행했습니다. 사용자 행동 데이터를 기반으로 검색 결과를 유저별로 재정렬하는 개인화 추천 구조를 구축했으며, 기존 검색 엔진 결과와 추천 모델을 실시간으로 연동하여 사용자별 검색 경험을 제공할 수 있도록 개선했습니다.

주요 수행 내용

- AWS Personalize Rerank 기반 추천 모델 설계 및 학습
- 사용자 행동 데이터 기반 학습 데이터 구성 및 전처리

- 검색 서비스와 추천 모델 간 실시간 API 연동 구성
- 검색 결과 개인화 반영 구조 개발
- 추천 품질 개선 및 모델 고도화 실험 수행

사용 기술

AWS Personalize, Spark, Python, AWS

CRM 추천 시스템 개발

2025.02 ~ 2025.06

프로젝트 개요

기존 CRM 캠페인은 마케터가 직접 SQL 조건을 작성하여 타겟 유저를 추출하는 방식으로 운영되어 수작업과 운영자 경험에 크게 의존하고 있었습니다. 이를 개선하기 위해 사용자 행동 데이터를 기반으로 구매 가능성이 높은 유저를 자동 추출하는 AWS Personalize 기반 CRM 추천 시스템을 개발했습니다.

주요 수행 내용

- AWS Personalize 기반 추천 모델 학습 및 운영 구조 설계
- 사용자 행동 데이터 기반 학습 데이터 전처리 파이프라인 구축
- 특정 광고 및 기획전별 구매 가능성이 높은 사용자 자동 추출 기능 개발
- 추천 결과 조회를 위한 내부 운영 도구 개발
- 모델 학습·추론·서비스 연동 전 과정 설계 및 개발

성과

- 광고 CTR 최대 2배 개선
- CVR 약 2% 달성
- 수작업 기반 CRM 타겟팅 프로세스 자동화

사용 기술

AWS Personalize, Spark, Airflow, Python, S3

EKS 데이터 플랫폼 통합 및 현대화

2024.12 ~ 2025.12

프로젝트 개요

EC2-EMR-ECS로 분산 운영되던 데이터 플랫폼의 운영 복잡도와 리소스 비효율 문제를 해결하기 위해 EKS 중심 데이터 플랫폼 통합 프로젝트를 수행했습니다. 기존 데이터 인프라를 데이터팀 전용 AWS 계정으로 이전하고 Kubernetes 기반 운영 구조로 표준화했습니다. 기존 EMR 기반 Spark 환경을 EKS 기반 오픈소스 Spark로 전환하고 오토스케일링 기반 운영 체계를 구축했습니다. 또한 DBT를 도입하여 데이터 품질 검증 및 데이터 모델 관리 체계를 개선했습니다.

주요 수행 내용

- Terraform 기반 AWS 계정 및 인프라 IaC 코드 설계·구축
- Spark on Kubernetes 환경 구축 및 기존 Spark-배치 인프라 마이그레이션
- Kubernetes 기반 배포 및 운영 구조 표준화
- DBT 도입 및 기존 SQL 배치 구조 마이그레이션
- 운영 리소스 모니터링 및 비용 최적화 작업 수행

성과

- AWS 인프라 비용 약 30% 절감

- 배치 운영 구조 일원화 및 소요시간 약 50% 단축

사용 기술

AWS EKS, EMR, Terraform, Kubernetes, Spark, DBT

데이터 활용 및 거버넌스 플랫폼 구축

2023.01 ~ 2025.01

프로젝트 개요

데이터팀 중심으로 제한되어 있던 데이터 접근 환경을 개선하고 전사 데이터 활용성을 높이기 위해 데이터 활용 및 거버넌스 플랫폼 구축 프로젝트를 수행했습니다. 데이터 접근 인터페이스, 메타데이터 관리 체계, 개인정보 보호 체계를 통합적으로 구축하여 데이터 탐색과 활용이 가능한 환경을 마련했습니다.

주요 수행 내용

- 모놀리식 구조 분석 서버를 ECS 기반 구조로 분리 및 마이그레이션
- Hue·DP 기반 데이터 접근 인터페이스 구축
- DataHub 기반 데이터 카탈로그 및 메타데이터 관리 체계 구축
- 데이터 마스킹 및 개인정보 처리 로직 개발을 통한 ISMS 대응
- Confluence 데이터 기반 LLM RAG Slackbot 개발 및 운영

성과

- ISMS 통과
- 임직원 데이터 접근성 확대

사용 기술

AWS ECS, DataHub, Spark, Python, LLM

데이터레이크 및 데이터 웨어하우스 현대화

2022.06 ~ 2024.06

프로젝트 개요

분산되어 있던 데이터 적재 구조와 데이터 정합성 문제를 해결하기 위해 S3 기반 메달리온 아키텍처 데이터레이크를 구축하고 데이터 파이프라인을 일원화했습니다. 또한 사람마다 다르게 사용하던 지표 기준을 통일하기 위해 스타 스키마 기반 데이터 웨어하우스를 구성하여 핵심 데이터를 표준화하고 KPI 지표 체계를 정립했습니다. 기존 수동 입력 기반으로 관리되던 KPI 집계 프로세스를 자동화하였으며, 월 단위로 갱신되던 데이터를 일 배치 기반으로 전환하여 주요 데이터를 매일 확인할 수 있는 환경을 구축했습니다. 이를 위해 DW 및 Mart 컨벤션을 설계하고 기존 레거시 테이블 구조를 신규 데이터 모델로 마이그레이션했습니다.

주요 수행 내용

- S3 기반 메달리온 아키텍처 데이터레이크 구축
- Airflow 기반 데이터 배치 및 적재 파이프라인 구축
- 데이터 웨어하우스 모델링 및 적재 구조 개선
- KPI 주요 지표 일원화 및 대시보드 자동화 체계 구축
- DW 및 Mart 컨벤션 설계 및 마이그레이션 수행

성과

- 지표 일원화 및 수작업 자동화
- 월 → 일 KPI 갱신 주기 단축

사용 기술

Spark, Airflow, S3, SQL

새마을금고 빅데이터 시스템 구축

2021.08 ~ 2022.03

프로젝트 개요

데이터 분석 환경 구축 및 데이터 통합을 위해 Cloudera Hadoop 기반 빅데이터 플랫폼 구축 프로젝트에 참여했습니다. 여러 서버로 Hadoop 클러스터를 구성하고, 분산되어 있던 데이터를 통합 데이터 웨어하우스로 적재하기 위한 Sqoop, NiFi 기반 데이터 수집 및 적재 파이프라인을 개발했습니다.

주요 수행 내용

- Cloudera Hadoop 기반 빅데이터 플랫폼 구축
- Kerberos-TLS 기반 Hadoop 보안 환경 설정
- Sqoop-Hive-Impala 기반 DW 워크플로우 구축 지원
- NiFi 기반 외부 데이터 수집 파이프라인 개발
- Impala 쿼리 튜닝 및 통계 데이터 생성 작업 수행

성과

- 쿼리 조회 성능 약 20% 개선

사용 기술

Hadoop, Hive, Impala, NiFi, Kerberos

Hadoop 클러스터 구축 자동화 모듈 개발

2021.06 ~ 2021.08

프로젝트 개요

반복적인 Hadoop 클러스터 구축 및 운영 작업을 자동화하기 위해 Ansible 기반 클러스터 구축 자동화 모듈을 개발했습니다. Ansible 설치, hosts 구성, 클러스터 환경 배포, Kerberos 보안 설정 등을 모듈화하여 ini 파일에 서버 정보만 입력하면 자동으로 클러스터 구성이 완료되도록 구축했습니다. 이를 통해 서버별 수작업 환경 설정 과정을 제거하고, hosts 기반으로 클러스터 서버 정보를 일원화하여 운영 효율성을 개선했습니다.

주요 수행 내용

- Ansible-Shell Script 기반 클러스터 구축 자동화 모듈 개발
- 서버 환경 설정 및 클러스터 배포 자동화
- Active Directory 연동 및 계정 관리 자동화 환경 구성
- hosts 기반 클러스터 서버 정보 및 환경 변수 관리 체계 구성
- 반복 수작업 기반 운영 절차 자동화 수행

성과

- 구축 시간 50% 이상 단축

사용 기술

Ansible, Shell Script, Hadoop, Linux