

Jess | Key Projects

AI-Powered Data Analysis Automation System

2026.04 ~

Overview

Developed an AI-powered data analysis automation system that enables users to query and analyze data using natural language. Built a metadata retrieval and SQL generation architecture based on DataHub MCP, allowing users to perform data exploration and analysis without writing SQL manually.

Key Responsibilities

- Designed workflows for collecting additional context from natural language queries
- Developed a metadata retrieval architecture based on DataHub MCP
- Implemented automatic SQL generation using table and column metadata
- Built workflows for query execution result processing and post-processing
- Designed and implemented LLM-based data analysis automation workflows

Tech Stack

LLM, MCP, DataHub, Python

Real-Time Data Processing Platform

2026.02 ~ 2026.04

Overview

Built a real-time data processing platform using Spark Streaming in response to increasing real-time processing requirements in an existing batch-oriented Spark environment. Standardized the development, deployment, and operational processes for streaming jobs while establishing a monitoring system for stable real-time operations.

Key Responsibilities

- Built a real-time processing platform based on Spark Structured Streaming
- Designed automated pipelines for streaming job development, testing, and deployment
- Standardized job execution and operational environments
- Developed monitoring servers for processing delays, job status, and failures
- Created and shared internal guides for streaming development and operations

Tech Stack

Spark Streaming, Kubernetes, Python

AI Search Recommendation System

2026.01 ~ 2026.04

Overview

Developed a personalized search recommendation system using AWS Personalize Rerank. Built a recommendation architecture that reorders search results based on user behavior data and integrated the recommendation model with the search service in real time to provide personalized search experiences.

Key Responsibilities

- Designed and trained recommendation models using AWS Personalize Rerank
- Processed and prepared user behavior datasets for model training

- Implemented real-time API integration between the search service and recommendation model
- Developed personalized search result ranking logic
- Conducted recommendation quality improvement and model optimization experiments

Tech Stack

AWS Personalize, Spark, Python, AWS

CRM Recommendation System

2025.02 ~ 2025.06

Overview

Existing CRM campaigns relied heavily on manual SQL queries and operator experience for user targeting. To improve this process, developed an AWS Personalize-based CRM recommendation system that automatically extracts users with high purchase probability based on behavioral data.

Key Responsibilities

- Designed recommendation model training and operational architecture using AWS Personalize
- Built preprocessing pipelines for user behavior training datasets
- Developed automatic user targeting features for advertisements and promotional campaigns
- Developed internal tools for recommendation result management and retrieval
- Designed and implemented the end-to-end workflow for model training, inference, and service integration

Achievements

- Improved advertisement CTR by up to 2x
- Achieved approximately 2% CVR
- Automated manual CRM targeting processes

Tech Stack

AWS Personalize, Spark, Airflow, Python, S3

EKS Data Platform Integration & Modernization

2024.12 ~ 2025.12

Overview

Led an EKS-centered data platform integration project to resolve operational complexity and resource inefficiencies across EC2, EMR, and ECS environments. Migrated existing infrastructure to a dedicated AWS account, standardized operations with Kubernetes, migrated Spark workloads from EMR to EKS, and introduced DBT-based data quality management processes.

Key Responsibilities

- Designed and implemented AWS accounts and infrastructure using Terraform IaC
- Built Spark on Kubernetes environments and migrated existing Spark and batch infrastructures
- Standardized Kubernetes-based deployment and operational processes
- Introduced DBT and migrated existing SQL batch architectures
- Performed infrastructure monitoring and cost optimization

Achievements

- Reduced AWS infrastructure costs by approximately 30%
- Reduced batch operation time by approximately 50% through operational standardization

Tech Stack

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

Data Governance & Analytics Platform

2023.01 ~ 2025.01

Overview

Built a data governance and analytics platform to improve organization-wide data accessibility and utilization, which had previously been limited to the data team. Established integrated systems for data access, metadata management, and privacy protection to enable efficient data exploration and usage.

Key Responsibilities

- Migrated monolithic analytics servers to ECS-based architectures
- Built Hue/DP-based data access interfaces
- Implemented DataHub-based data catalog and metadata management systems
- Developed data masking and privacy protection logic for ISMS compliance
- Developed and operated an LLM-based RAG Slackbot using Confluence data

Achievements

- Passed ISMS compliance requirements
- Improved organization-wide data accessibility

Tech Stack

AWS ECS, DataHub, Spark, Python, LLM

Data Lake & Data Warehouse Modernization

2022.06 ~ 2024.06

Overview

Built an S3-based medallion architecture data lake and unified data pipelines to resolve fragmented ingestion structures and data consistency issues. Standardized core datasets and KPI definitions using a star-schema-based data warehouse architecture. Automated previously manual KPI aggregation processes and converted monthly update cycles into daily batch pipelines, enabling near real-time KPI visibility. Also established DW/Mart conventions and migrated legacy tables into the new data model structure.

Key Responsibilities

- Built an S3-based medallion architecture data lake
- Developed Airflow-based batch ingestion pipelines
- Improved data warehouse modeling and ingestion structures
- Standardized KPI metrics and automated dashboard pipelines
- Designed DW/Mart conventions and performed legacy migrations

Achievements

- Standardized KPI definitions and automated manual workflows
- Reduced KPI update cycles from monthly to daily

Tech Stack

Spark, Airflow, S3, SQL

MG Bank Big Data Platform Project

2021.08 ~ 2022.03

Overview

Participated in a Cloudera Hadoop-based big data platform project for building data analytics environments and integrating distributed data sources. Built Hadoop clusters across multiple servers and developed Sqoop and NiFi-based data ingestion pipelines to consolidate distributed data into a centralized data warehouse.

Key Responsibilities

- Built Cloudera Hadoop-based big data platforms
- Configured Kerberos and TLS-based Hadoop security environments
- Supported DW workflows using Sqoop, Hive, and Impala
- Developed external data ingestion pipelines using NiFi
- Performed Impala query tuning and statistical data generation

Achievements

- Improved query performance by approximately 20%

Tech Stack

Hadoop, Hive, Impala, NiFi, Kerberos

Hadoop Cluster Automation Module

2021.06 ~ 2021.08

Overview

Developed an Ansible-based automation module to streamline repetitive Hadoop cluster deployment and operational tasks. Modularized Ansible installation, hosts configuration, cluster deployment, and Kerberos security setup so that clusters could be automatically provisioned using only server information defined in configuration files.

Key Responsibilities

- Developed Hadoop cluster automation modules using Ansible and Shell Script
- Automated server environment setup and cluster deployment
- Configured Active Directory integration and account management automation
- Built centralized hosts-based server and environment variable management systems
- Automated repetitive operational procedures

Achievements

- Reduced cluster deployment time by more than 50%

Tech Stack

Ansible, Shell Script, Hadoop, Linux