Ahmad Uzzam Masood

Portfolio | LinkedIn

3+ years of experience designing and orchestrating scalable, cloud-native ETL and streaming data platforms. Skilled in building automated, metadata-driven pipelines and real-time processing solutions across Azure and AWS. Proficient in Databricks, dbt, Apache Spark, Kafka, Airflow, Azure Data Factory, AWS Glue, Redshift, Snowflake, and Terraform. Experienced in optimizing ingestion latency, ensuring pipeline reliability, and deploying production-grade CI/CD workflows.

Experience

Data Engineer
01/2025 - Ongoing

Nayatel

- Built and operated dbt models on BigQuery and Snowflake, optimized incremental pipelines using partitioning & clustering.
- Implemented Terraform to automate infrastructure provisioning for BigQuery datasets, Composer environments, and service accounts.
- Designed and deployed metadata-driven ETL pipelines in Azure Data Factory and Databricks to ingest data from Oracle and REST APIs into PostgreSQL for Churn Analytics.
- Created parameterized Airflow DAGs using YAML-based configurations for dynamic mapping of pipeline parameters and validation rules.
- Set up automated validation, logging, alerting, and fully integrated CI/CD deployment workflows.

Data Engineer
07/2023 - 01/2025

PLC Group

- Developed large-scale ETL pipelines using PySpark for batch and streaming transformations across telemetry and energy metrics.
- Implemented Kafka-based ingestion to stream data from 2,500+ telecom sites into AWS S3 and Redshift.
- Built AWS Glue jobs for schema-drift-resilient transformations and automated data cataloging.
- Designed Redshift warehouse tables for analytical workloads and optimized distribution/sort keys to reduce query latency.
- Containerized Spark, Kafka consumers, and service components using Docker for reproducible deployments.

Core Skills

ETL & Data Integration: dbt, Airflow, Spark, Kafka, ADF, Glue, Lambda
Data Warehousing: BigQuery, Redshift, Snowflake, PostgreSQL, Azure SQL, Data Lake Gen2
DevOps & Automation: CI/CD (DevOps, GitHub Actions), Docker, Terraform
Monitoring: Composer, CloudWatch, Azure Monitor, Logging & Alerting

Projects

Covid Reporting Data Platform (Azure Data Factory, Data Lake, SQL, Power BI)

Built an end-to-end data platform in Azure to automate COVID-19 trend reporting. Designed metadata-driven ADF pipelines to ingest and transform data from APIs and Blob Storage into Data Lake Gen2 and Azure SQL Database for Power BI analytics. Implemented transformations using Mapping Data Flows, Databricks, and HDInsight, with integrated monitoring using Azure Monitor and automated CI/CD deployments through Azure DevOps.

Metadata-Driven ETL Framework (Airflow, YAML)

Designed a configurable ETL system using YAML metadata for dynamic parameter mapping, validation and logging in Airflow. Enabled source-to-destination mapping without code changes and automated validation rules. The system achieved reusable and scalable design supporting onboarding of new data sources within hours.

Predictive Site Outage Monitoring System (Spark, PostgreSQL, Kafka, FastAPI)

Designed and implemented a real-time outage monitoring system for distributed sites using PostgreSQL, Spark, Kafka, and FastAPI. This project handles large-scale site data, processes streaming energy and environmental metrics, and predicts outages with accurate alerts. It includes parallel processing with Spark, data ingestion from Kafka, and serves data via FastAPI endpoints.

Cloud Data Pipeline for Service Analytics (Redshift)

Designed a serverless data pipeline using AWS S3, Glue, Lambda, and Redshift to process and analyze large-scale service usage logs. Built orchestration and monitoring with Airflow and CloudWatch, improving data refresh frequency from daily to hourly and enabling near-real-time analytics.

Education

BSc Computer Engineering • COMSATS Islamabad • August 2022 • CGPA 3.09

Certifications

IBM Data Engineering
Coursera - 2025