

UMAIR

Cloud Data Engineer

Phone: 923490219531
Location: Karachi, Pakistan

Email: umairnawaz7228@gmail.com
Portfolio Web: <https://umair-portfolio-web.vercel.app>

Linkedin: www.linkedin.com/in/umairnawaz7228
Github: <https://github.com/umair7228>

SUMMARY

Skilled Cloud Data Engineer with experience building, optimizing, and managing scalable data pipelines on AWS and Snowflake. Proficient in ETL processes, real-time data streaming, and data warehousing, with hands-on expertise in Python, SQL, Apache Kafka, and Airflow. Demonstrated success in ensuring data accessibility, reliability, and integration across systems, with a strong focus on performance and cloud architecture.

PROJECTS EXPERIENCE

Pandemic Insights: ETL Pipeline with AWS Glue, Athena & Redshift

Developed a scalable ETL pipeline using AWS Glue, Athena, and Redshift to process and analyze COVID-19 data from multiple sources. Designed fact and dimension tables, optimized data queries, and loaded the results into Redshift for data-driven insights. (Check out all the projects in my website link given above)

Serverless Data Lake Architecture on AWS

A fully automated serverless data lake solution using AWS services to ingest, process, and notify users about data transformations. This architecture leverages S3, Lambda, Glue Crawlers and Jobs, CloudWatch, and SNS to streamline ETL processes and deliver scalable data insights.

Event-Driven Data Processing System

Built an event-driven architecture using AWS SNS, SQS, Lambda, and S3 to enable real-time data processing. Events are triggered via SNS and managed with SQS, while Lambda functions execute tasks dynamically, with data stored in S3. This system ensures efficient handling of dynamic data flows with optimized response times.

Real-Time Stock Market Data Pipeline

Built a real-time data pipeline for stock market analysis using Apache Kafka for seamless data streaming and AWS for scalable data storage and querying real-time data with Athena, crawler and glue. The project enables timely access to market data for analysis, integrating multiple services to ensure high availability and efficient data handling across different components.

Weather and S3 Data Integration Pipeline

Developed an automated data pipeline using Apache Airflow on AWS EC2, integrating weather data from OpenWeather API and Amazon S3 into an RDS PostgreSQL database. Designed to process and join data in parallel, this pipeline outputs weather data to Amazon S3, ensuring structured storage and efficient data querying.

Real-Time Data Pipeline with SCD Implementation

Created an end-to-end data pipeline generating synthetic data with Python, leveraging Apache NiFi for data extraction and transfer to Amazon S3, and Snowpipe for ingestion into Snowflake. Incorporated SCD (Slowly Changing Dimensions) Type 1 and Type 2 tracking to manage data evolution, providing a comprehensive framework for data updates.

E-Commerce Sales Analysis

Conducted a comprehensive analysis of e-commerce sales data to uncover actionable insights for optimizing business strategies. Used SQL and Python for data cleaning, transformation, and visualization, supporting data-driven decision-making processes that enhance e-commerce operations.

EDUCATION

Cloud Data Engineering

S.M.I.T (Saylani Mass IT Training)

- Karachi, Pakistan
- Training in Cloud Data Engineering, focusing on cloud platforms, data pipelines, and SQL for efficient data management.

Agentic AI Engineering

GIAIC (Governor Sindh Initiative)

- Karachi, Pakistan
- Pursuing Agentic AI Engineering and Frontend Development at GIAIC, focusing on AI models, machine learning, automation, and building intuitive, user-centric web applications.

SKILLS

- AWS
- T-SQL, PostgreSQL
- Apache Kafka
- Docker
- Next.js
- TypeScript
- CSS (Tailwind CSS)
- Snowflake
- Python (Pandas, Numpy)
- Apache Airflow
- React.js
- JavaScript
- HTML

LANGUAGES

English

Proficient



Urdu

Native



Courses & Certifications

- Data EngineeringSnowflake
- Associate Data Engineering in SQLDataCamp
- Data WarehousingDataCamp
- SQLDataCamp
- PythonDataCamp