UMAIR

Cloud Data Engineer | ETL & Real-Time Pipelines | AWS, Snowflake, Kafka, Airflow

Karachi, Pakistan • +92 349 4804814 • umairnawaz7228@gmail.com • linkedin.com/in/umairnawaz7228 • github.com/umair7228 • umair-portfolio-web.vercel.app

PROFESSIONAL SUMMARY

Cloud Data Engineer with hands-on experience building scalable ETL pipelines, real-time streaming systems and cloud-native architectures using AWS, Snowflake, Kafka, Spark and Airflow. Skilled at optimizing big data workflows and delivering automated, high-performance pipelines. Also a Cloud Data Engineering instructor who has mentored 60+ learners and designed project-based curriculum. Placed 2nd out of 1,000+ participants in a Cloud Data Engineering Hackathon.

CORE COMPETENCIES

ETL & Data Pipeline Development • Real-Time Streaming • Cloud-Native Architecture • Data Warehousing & Modeling • Big Data Processing • Workflow Automation • Performance Optimization • Monitoring & Alerting • Cross-Functional Collaboration • Mentoring & Technical Training

PROFESSIONAL EXPERIENCE

Lead Instructor, Cloud Data Engineering (Part Time) | Saylani Mass IT Training (SMIT) **Aug 2025 – Present**

- Design and deliver a comprehensive curriculum covering Python, EDA, web scraping, SQL, Kafka, AWS, Airflow and Snowflake.
- Mentor 60+ students, many from non-IT backgrounds, simplifying complex data engineering concepts through hands-on projects.
- Guide learners through real-world case studies, including building ETL pipelines, cloud-native architectures and real-time data flows.
- Develop project-based learning modules that strengthen problem-solving skills and industry readiness.
- Organize and run hackathons where learners apply their skills under real-world conditions.

Data Engineer | Curve Digital Solutions **May 2025 – Jul 2026**

- Optimized a large-scale ETL pipeline, cutting execution time by 80% (2.5 min to 30 sec) using efficient SQL and in-memory Python transformations.
- Designed and deployed real-time streaming pipelines with Apache Kafka, delivering event-driven data with sub-second latency.
- Automated recurring workflows with Apache Airflow, achieving 100% scheduling reliability with built-in monitoring and alerting.
- Partnered with data scientists and devops engineers to productionize ML models, improving delivery speed and scalability for predictive analytics.

Cloud Data Engineer | Saylani Tech **Jan 2024 – May 2025**

- Built ETL pipelines with Spark and Airflow, reducing manual workloads and enabling scalable, automated analytics.
- Designed optimized Snowflake data models, cutting query time and powering real-time dashboards.
- Developed cloud-native solutions on AWS (S3, Lambda, Glue, SNS, SQS, RDS, Redshift) for high-throughput data ingestion and transformation.

PROJECTS

Data Warehousing & ETL Pipeline for an Online Learning Platform (Detailed Github Repo)

- Built a Snowflake data warehouse with AWS S3 and Apache Spark for student analytics.
- Automated the ETL pipeline with Airflow and Snowpipe for real-time data ingestion.
- Created Power BI dashboards to track enrollments, completion rates and engagement.

Pandemic Insights: ETL Pipeline with AWS Glue, Athena & Redshift (Detailed Github Repo)

- Developed a scalable ETL pipeline to process and analyze COVID-19 data from multiple sources.
- Designed fact and dimension tables for optimized data querying in Amazon Redshift.
- Used AWS Glue and Athena for data transformation and insights generation.

Event-Driven Data Processing System (Detailed Github Repo)

- Designed an event-driven architecture using AWS SNS, SQS, Lambda and S3.
- Enabled real-time data processing through event triggers and message queues.
- Optimized data handling and response times for dynamic event flows.

Real-Time Stock Market Data Pipeline (Detailed Github Repo)

- Developed a real-time stock data pipeline with Apache Kafka and AWS.
- Integrated Glue, Athena and Crawlers for scalable storage and queries.

- Ensured timely access to market data for analytics.

Real-Time Data Pipeline with SCD Implementation **(Detailed Github Repo)**

- Created a synthetic data pipeline using Python and Apache NiFi for Amazon S3.
- Implemented Slowly Changing Dimensions (SCD) Type 1 and 2 for data evolution tracking.
- Used Snowpipe for seamless data ingestion into Snowflake for efficient updates.

TECHNICAL SKILLS

Languages: Python (Pandas, NumPy), SQL, TypeScript, JavaScript, Bash

Cloud & Warehousing: AWS (S3, Lambda, Glue, Athena, Redshift, SNS, SQS, RDS), Snowflake

Processing & Orchestration: Apache Kafka, Apache Airflow, Apache NiFi, Snowpipe

Other: Docker, Git, Power BI, Web Scraping (Beautiful Soup, Selenium), React.js, Next.js

EDUCATION

Bachelor of Science, Data Science and Business Analytics **Sep 2026 – Present**

Denning Institute of Technology and Entrepreneurship

Diploma, Cloud Data Engineering **Nov 2023 – Dec 2024**

Saylani Mass IT Training (SMIT), among the top 5 students

CERTIFICATIONS

20+ certifications across AWS, Snowflake, DataCamp and Udemy, including AWS Certified Data Engineer – Associate, Data Warehousing, Data Engineering in SQL and Python.