Tahir Malik (Data Analyst / Engineer)

New Delhi, IN | tahir.malik296@gmail.com | +91 9971587221 | LinkedIn | GitHub | Portfolio

EDUCATION

Amity University (Gurugram, India)   **Jul 2017 – Apr 2021**
B. Tech in Information Technology (C.G.P.A – 8.45/10)
Relevant Coursework: Python, SQL, Big Data & Fundamentals, Economics (Minor)

PROFESSIONAL EXPERIENCE

Data Engineer Intern, Mactores, Mumbai   **Apr 2025 - Present**
- Gained hands-on experience in data engineering workflows on the Databricks Lakehouse Platform.
- Monitored, debugged, and optimized ETL pipelines built with Apache Spark and Delta Lake to ensure reliability and data quality.
- Worked on implementing real-time streaming pipelines using Kafka, Delta Lake, and Databricks Auto Loader.
- Achieved AWS Cloud Practitioner and Databricks Certified Data Engineer (Associate & Professional) credentials during this role.

Data Analyst, Capgemini, Gurugram   **Oct 2022 – Apr 2023**
- Assisted senior analysts in cleaning, preparing, and validating datasets to ensure data accuracy for client reporting.
- Supported the creation of dashboards and reports by preparing data, generating basic visualizations, and documenting insights.
- Helped troubleshoot minor data issues and collaborated with team members to maintain smooth and reliable analytics workflows.

Associate Software Engineer, DXC Technology, Noida   **May 2021 – May 2022**
- Assisted in maintaining ETL pipelines for a client project using Azure Data Factory, contributing to efficient data integration and transformation processes.
- Monitored and helped troubleshoot pipeline issues to reduce downtime and support reliable data delivery for key operations.
- Worked with senior engineers and cross-functional teams to improve workflows, implement minor enhancements, and maintain adherence to data standards.

PROJECTS

Fintech Data Migration Pipeline
- Developed scalable pipelines for migrating and transforming data of a fintech company from their traditional SQL database to Azure Data Lake, ensuring data consistency and minimal downtime.
- Implemented a medallion architecture within ADLS Gen2 using PySpark & Delta Lake to effectively manage and process data, improving data accessibility and reliability.

E-Commerce Data Streaming Analytics
- Developed a robust data pipeline leveraging Confluent Kafka to collect e-commerce clickstream and user activity data, enabling real-time analytics capabilities.
- Combined Azure Data Lake Storage Gen2 (ADLS Gen2) with Databricks to efficiently store, cleanse, and convert raw data into formats suitable for analysis, facilitating actionable insights.

TECHNICAL SKILLS

Languages: Python, SQL, C, C++
Big data & Fundamentals: Hadoop, Hive, Kafka, Apache Spark, Databricks, Airflow, MongoDB, Docker, Git
Cloud Services: Azure Data Factory, Azure Synapse Analytics, Azure Function App, Azure Data Lake Storage, Azure Cosmos DB, Azure SQL Database, GCP Dataproc, GCP Pub/Sub, GCP Cloud Storage, Snowflake, Dimension & Fact Table, BigQuery
Soft Skills: Communication, Analytical Thinking, Problem Solving, Team Player, Data-Driven Decision Making

CERTIFICATIONS

- Databricks Certified Data Engineer Professional
- Databricks Certified Data Engineer Associate
- AWS Certified Cloud Practitioner
- Data Engineering with Azure
- Introduction to Statistics
- Google Data Analytics Specialization