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 and debugged ETL pipelines with Apache Spark & Delta Lake to ensure reliability and data quality.
- Learned how to implement real-time streaming pipelines using Kafka, Delta Lake, and Databricks Auto Loader.
- Earned Databricks Certified Data Engineer Associate & AWS Cloud Practitioner credentials during this role.

Professional Development (Career Break)   **May 2022 - Apr 2025**
- Took a career break for family health responsibilities while consistently upskilling.

Associate Software Engineer, DXC Technology, Noida   **May 2021 - May 2022**
- Supported and maintained ETL pipelines for Carrefour using Azure Data Factory, ensuring efficient data integration and transformation processes.
- Monitored and troubleshooted pipeline issues to minimize downtime, enhance performance, and ensure accurate data delivery for business-critical operations.
- Collaborated with cross-functional teams to optimize workflows, implement enhancements, and ensure adherence to organizational 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 Associate**   - **Grow Data Skills: Data Engineering with Azure**

- **Google Data Analytics Specialization**   - **AWS Certified Cloud Practitioner**