# Tahir Malk (Data Engineer)

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

## EDUCATION

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

**Amity University** **Jul 2018 – Apr 2021**
**Bachelors in economics (C.G.P.A - 8/10)** **Gurugram, India**
Relevant coursework: Indian Economy, Micro & Macro Economics (Minor Degree)

## PROJECTS

**Adventure Works Data Analytics** **Nov 2024 – Dec 2024**

- Developed an end-to-end ETL pipeline utilizing Azure Data Factory (ADF), Azure Data Lake Storage Gen2 (ADLS Gen2), Databricks, and Azure Synapse Analytics, enhancing data processing efficiency.
- Implemented a medallion architecture within ADLS Gen2 to effectively manage and process data, improving data accessibility and reliability.
- Integrated Azure Function App for dynamic record updates in response to new files added to the data source, streamlining data management processes.

**E-Commerce Data Streaming Analytics** **Dec 2024 – Jan 2025**

- 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.
- Engineered a production-grade workflow featuring Kafka producers for managing user-profiles, delivery-statuses, orders, and clickstreams showcasing proficiency in practical streaming applications.

**Incremental Data Analytics** **Jan 2025 – Feb 2025**

- Engineered a comprehensive end-to-end data pipeline utilizing Azure Data Factory (ADF), Azure Data Lake Storage Gen2 (ADLS Gen2), Azure SQL Database, and Azure Databricks to initially retrieve historical data followed by incremental updates from the source.
- Applied a medallion architecture within ADLS Gen2 to enhance data management and processing efficiency.
- Constructed a star schema by establishing dimension and fact tables, implementing Slowly Changing Dimension (SCD) Type-1 methodology for effective record upserting.

## TECHINCAL SKILLS

**Languages**: Python, SQL
**Big data & Fundamentals**: Hadoop, Hive, Kafka, Apache Spark, Databricks, Airflow, MySQL Database, 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
**Software Tools**: MS PowerPoint, MS Excel, Premiere Pro, Photoshop, Illustrator

## CERTIFICATIONS

- Grow Data Skills: Data Engineering with Azure
- Codebasics Excel: Mother of Business Intelligence
- Google Data Analytics Specialization
- Introduction To Statistics