Soban Munir

Lahore, Pakistan | sobanmunir.official@gmail.com | Linkedin | +92-335-1444470 | Github

Data Engineer

PROFILE

Azure Data Engineer with strong hands on experience designing end-to-end ETL/ELT pipelines using Azure Data Factory, Databricks (PySpark), Delta Lake (Medallion Architecture), and Azure Synapse Analytics for Big data processing. Skilled in Data Warehousing, data ingestion and transformation, incremental loads, and SQL optimization, delivering analytics-ready datasets for Power BI reporting and business insights.

EDUCATION

Intermediate in Pre Engineering
DPS College Okara — 2023

BS Computer Science
Riphah International University — 2028

EXPERIENCE (Project-based)

Netflix Azure Data Engineering Project

- End-to-End ETL: Designed an Azure pipeline using ADF and Databricks to ingest GitHub data into a multi-layered Data Lake.
- Transformation & Loading: Developed PySpark scripts for data cleaning and implemented incremental loading to ensure data accuracy.
- BI Insights: Created an optimized Gold layer and built Power BI dashboard to visualize streaming trends and metrics

AdventureWorks Data Pipeline

- Medallion Architecture: Built a Bronze-Silver-Gold framework by ingesting 10+ CSV sources from GitHub into Azure Data Lake via ADF.
- Data Engineering: Developed robust PySpark pipelines in Databricks to clean and transform raw data into a refined Silver layer.
- Enterprise Warehouse: Engineered a Star Schema in Azure Synapse to deliver a production-ready environment for high-performance analytics.

SKILLS

Cloud & Data Platforms: Microsoft Azure (Data Factory, Azure Databricks, Data Lake, Synapse
Languages: Python, PySpark, SQL
Data Engineering & Architecture: ETL Pipelines, Delta Lake, Data Warehousing, SCDs
Version Control & Collaboration: Git, GitHub, CI/CD Pipelines

CERTIFICATIONS

- Microsoft DP-203 (Azure Data Engineer Associate) — In Progress
- Pyspark Essential Training: Building Data Pipelines – LinkedIn Learning
- Complete Guide to SQL for Data Engineering – LinkedIn Learning