SOBAN MUNIR

JUNIOR DATA ENGINEER

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

LinkedIn: https://www.linkedin.com/in/soban-munir-89b64b333 | GitHub: https://github.com/soban-munir

PROFESSIONAL SUMMARY

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.

RELEVANT EXPERIENCE (PROJECT BASED)

Azure Data Engineering Projects

- Built an end-to-end Azure ETL pipeline using ADF + Databricks (PySpark) + Delta Lake + Synapse, implementing Bronze–Silver–Gold architecture and Power BI reporting.
- Developed an enterprise-style data pipeline for AdventureWorks using ADF + Databricks + Delta Live Tables + Synapse, including incremental loads, data validation, and star-schema modeling for analytics.

SKILLS

Cloud & Data Platforms: Azure Data Factory(ADF),Azure Databricks , Azure Data Lake/Blob storage, Azure Synapse Analytics
Languages: Python, PySpark, SQL
Data Engineering & Architecture: ETL / ELT Pipelines, Big data , Delta Lake , Medallion Architecture, Data Warehousing, Data Modeling, SCDs
Version Control & Collaboration: Git, GitHub, CI/CD Pipelines

PROJECTS

Netflix Azure Data Engineering Project | Azure Data Factory, Databricks, Delta Lake, Synapse

- Designed and implemented an end-to-end ETL pipeline on Azure integrating data from GitHub through Azure Data Factory into Delta Lake using Databricks.
- Applied data validation, transformation, and incremental loading through Delta Live Tables ensuring data accuracy and performance.
- Created a Gold layer schema for optimized analytics and reporting.
- Developed interactive Power BI dashboards for visual insights on Netflix datasets, improving data-driven decision-making.

Azure End-to-End Data Pipeline for AdventureWorks | ADF, Databricks, Delta Live Tables, Synapse

- Built a multi-layered Bronze–Silver–Gold architecture by ingesting 10 CSV source files from GitHub into Azure Data Lake using Azure Data Factory.
- Built PySpark pipelines in Databricks to clean and transform data into the Silver layer.
- Engineered a Dimensional Data Warehouse (Star Schema) in Azure Synapse Analytics for enterprise-grade analytical workloads.
- Delivered a production-ready Gold layer supporting low-latency reporting and BI consumption.

EDUCATION

BS Computer Science
Riphah International University —*Currently Pursuing*

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