Tecblic
HIGH ON TECH

Aditya
Sr. Data Engineer

OBJECTIVE:

Motivated Data Engineer with **5.2+ years** of experience in building scalable data pipelines and automating ETL workflows in cloud environments. Passionate about unlocking business insights through high-quality data architecture, while actively enhancing capabilities in notebooks, cluster management, and governance frameworks. Proficient in Azure-based data engineering with a strong foundation in PySpark, DBT, SQL, and Databricks.

EDUCATION:

- Bachelor of Engineering (B.E.)

TECHNICAL SKILLS:

Category	Technologies & Tools
Languages	Python, PySpark, SQL
Databases	Oracle SQL, Azure SQL Database
Cloud Platforms	Azure (Primary), AWS (Basic Exposure)
Azure Services	ADF, Azure Databricks, Synapse Analytics, Functions, Key Vault, Data Lake Storage, Storage
ETL Tools	Azure Data Factory, SQL
Data Architecture	Medallion Architecture (Basic Implementation), Structured Layering, Scalable Pipelines

Tecblic
HIGH ON TECH

Category	Technologies & Tools
Governance	Data Quality Checks, Basic Data Governance (RLS, masking not yet implemented)
BI Tools	Power BI (Basic usage, lacks DAX/RLS experience)
In Progress	Improving skills in Databricks Notebooks, Cluster Management, Real-time Dashboards, RLS
Certifications	Azure, Databricks, and Power BI certifications *(Currently preparing)*
Data Engineering	Spark, Hadoop, Airflow, DBT (Data Build Tool)

PROJECT EXPERIENCE:

E-Commerce Reward System – Data Pipeline for Fraud Detection and BI

Objective: Engineered a scalable and efficient data pipeline to optimize the analysis of the reward system, minimizing fraud detection errors. Ingested data from multiple sources, applying thorough data cleansing, validation, and business logic, with incremental loading to ensure up-to-date information. The processed data was used not only for training machine learning models aimed at advanced fraud detection but also provided critical insights to the Power BI team for visual reporting and business intelligence purposes.

Responsibilities:

- Developed and optimized **PySpark** scripts within **Azure Databricks** for ETL transformations.
- Created a **Medallion Architecture** pipeline with a structured data staging layer.
- Integrated transformed data into **Snowflake** for BI accessibility.

- Worked with ML teams (external to Databricks ML) to supply clean data.
- Collaborated with Power BI developers to deliver consistent, quality data.
- Did not use advanced notebooks or ML workflows in Databricks.

Tech Stack: Azure Data Factory, Azure Databricks, Snowflake, PySpark, Azure Key Vault

Telecom – Performance and Utilization Analysis of Cell Towers

Developed PySpark scripts to transform and convert raw data into a structured format, enabling business analysts to predict future company expansion by analyzing performance and utilization metrics of cell towers. This data enabled the company to optimize tower locations and predict performance trends for network planning.

Responsibilities:

- Transformed large datasets using PySpark in **Azure Databricks**.
- Implemented data validation routines for pipeline integrity.
- Integrated **Azure Data Lake Storage** for input/output stages.
- Supported business analysts with structured data for expansion planning.
- Lacked advanced notebook usage or detailed cluster tuning.

Tech Stack: Azure Databricks, PySpark, Azure Data Factory

End-to-End Data Engineering Workflow Optimization (Albertsons)

Environment: Snowflake, SQL, DBT, Tableau, Airflow

Description:

This project centered on optimizing data processing and engineering workflows using modern tools and methodologies. As the lead, I spearheaded the adoption of DBT (Data Build Tool) to streamline the end-to-end data engineering process. Overall, the project aimed to establish efficient and reliable data pipelines, effectively managing data processing initiatives from inception to completion.

Responsibility:

- Led data processing initiatives using DBT, overseeing the complete data engineering workflow.
- Wrote code for ETL processes in Snowflake, ensuring seamless data flow from raw to transformed to aggregated layers.
- Created and managed DBT jobs orchestrated using Airflow, optimizing data integration and enhancing workflow automation.

- Utilized DBT test module for comprehensive data testing in Snowflake.
- Led data processing initiatives using DBT, overseeing end-to-end data engineering workflow

SQL Query Development and Optimization

Developed and optimized complex SQL queries to extract and process data, ensuring accurate and efficient retrieval. Focused on enhancing performance by analyzing query factors and conducting extensive testing to verify the accuracy and reliability of transformed data.

Responsibilities:

- Wrote complex **T-SQL** queries within Azure SQL and Data Studio.
- Optimized performance by identifying and addressing bottlenecks.
- Conducted testing to verify query accuracy and business logic compliance.
- No integration with Power BI dashboards; focused purely on SQL-layer performance.

Tech Stack: Azure SQL Database, Azure Data Factory, T-SQL, Azure Data Studio

Healthcare Data Warehousing – ACA Compliance & Reporting

Objective: Built a unified data warehouse to enable compliance reporting under the Affordable Care Act (ACA) by integrating disparate healthcare data sources securely.

Responsibilities:

- Designed and deployed **ETL pipelines** in **Azure Data Factory** to ingest claims and enrollment data from multiple external sources.
- Created **dimension and fact tables** using **star schema modeling** to organize healthcare metrics.
- Developed PySpark jobs in **Azure Databricks** for cleansing and standardizing health plan data.
- Ensured **HIPAA-compliant** data handling via **Key Vault**, masking, and access management (RLS planned but not implemented).
- Loaded transformed data into **Azure Synapse Analytics** for downstream BI and regulatory teams.

Tech Stack: Azure Data Factory, Azure Databricks, PySpark, Azure SQL, Azure Synapse, Key Vault

Sales Forecasting & Reporting – Manufacturing Analytics Platform

Objective: Enabled better inventory and procurement decisions by building a forecasting data platform for the manufacturing division.

Responsibilities:

- Created batch ingestion pipelines using **ADF** and **Azure Storage** to pull daily sales, inventory, and shipment records.
- Used **PySpark in Databricks** to join and clean data for historical trend analysis.
- Implemented **basic forecasting logic** and feature engineering for ML teams (not within Databricks ML yet).
- Delivered curated datasets to **Power BI** team; no advanced DAX or real-time dashboards used.
- Applied **basic governance rules**, including naming conventions and validation steps; data masking and RLS are roadmap items.

Tech Stack: Azure Data Factory, Azure Databricks, Azure Data Lake, PySpark, Power BI