Rajas Akre

DATA ENGINEER|rajasakre06@gmail.com|9168085149

Professional Summary

Results-driven Data Engineer with 4 years of experience designing, building, and optimizing scalable data pipelines in the BFSI domain at Tata Consultancy Services (TCS). Proven expertise in Azure Data Factory, Azure Databricks, PySpark, and SQL-based solutions for enterprise-grade ETL systems. Led a team of engineers to deliver efficient data processing solutions for banking and financial services, improving performance and enabling real-time insights. Azure Databricks Certified Associate with strong knowledge of cloud data warehousing, data lakes, and distributed data processing.

Technical Skills

- Cloud & Big Data: Azure Data Factory, Azure Databricks, ADLS Gen2, Azure SQL, Azure Synapse
- Programming: Python, PySpark, SQL, HiveQL
- Databases: SQL Server, Hive, MongoDB, Cassandra
- ETL & Data Pipelines: ADF, Databricks, Spark, Hive, Delta Lake
- Data Warehousing: Star/Snowflake schema, Data Modeling
- Tools & Versioning: Git
- Certifications: Databricks Certified Data Engineer Associate

Professional Experience

Tata Consultancy Services (TCS) – Nagpur, April 2021 – Present

Responsibilities:

- Designed and developed scalable, secure, and efficient **end-to-end data pipelines** using **Azure Data Factory (ADF)**, **Azure Databricks**, and **ADLS Gen2**, handling large-scale BFSI datasets.
- Implemented **data transformation and cleansing logic using PySpark** in Databricks notebooks for structured and semi-structured financial data.
- Ingested and integrated data from diverse sources including **SQL Server, MongoDB, and Cassandra** using ADF's **copy activity**, **batch triggers**, and parameterized pipelines.
- **Led a team of data engineers**; conducted peer code reviews, mentored team members, and enforced coding standards and best practices.
- Optimized **Spark jobs** using **partitioning**, **broadcast joins**, **caching**, and **adaptive query execution (AQE)** to reduce shuffle operations and latency.
- Developed **parameterized, reusable stored procedures** and **T-SQL scripts** in **Azure SQL Database** for data quality checks, reconciliation, and business reporting.
- Automated pipeline monitoring, alerting, and logging using **ADF diagnostics**, **Azure Monitor**, **Log Analytics**, and **custom PowerShell scripts**.
- Built **CI/CD pipelines** for ADF and Databricks notebooks using **Azure DevOps**, ensuring seamless deployment and rollback strategies.
- Implemented **data validation frameworks** to ensure data consistency and integrity across environments (Dev, QA, Prod).
- Collaborated with business stakeholders and QA teams to define **data SLAs**, **error handling**, and **rollback mechanisms** for mission-critical loads.

Key Achievements:

- **Reduced end-to-end ETL execution time by ~47%**, from **6 hours to 3.2 hours**, by optimizing Spark cluster configurations and refactoring pipeline logic.
- **Achieved 99.9% SLA adherence** for daily ingestion of financial and transactional data feeds across 15+ source systems.
- Increased pipeline reusability by **implementing metadata-driven pipeline orchestration** with configuration-based control tables.
- Delivered **automated failure recovery and retry logic**, reducing manual intervention by 80%.
- Contributed to **cost optimization** by tuning cluster auto-scaling policies and job scheduling, reducing Databricks compute cost by 25%.

Education

Bachelor of Engineering (B.E.) – Computer Science

Government College of Engineering, Nagpur | 2016 – 2020

Certification

Databricks Certified Data Engineer Associate | Issued: 2025