Lucky

Phone: +91-9319073062 Email: luckymishra310118@gmail.com LinkedIn GitHub Leetcode GeeksForGeeks

Professional Summary

Results-driven Data Engineer with 1+ year of experience architecting scalable ETL pipelines in Azure cloud ecosystem. Specialized in big data processing using PySpark and Azure Databricks, successfully processed around 10 TB daily data volumes. Achieved 67% reduction in processing time and 10% cost savings through performance optimization and resource management.

Experience

Accenture Solutions Pvt. Ltd. **08/2024 – Present**
Associate Software Engineer *Gurugram, India*

- Developed and maintained 7+ ETL pipelines using **PySpark** in Azure Databricks, processing over **10 TB** of data daily to support batch analytics and business reporting.
- Automated 10+ data workflows in **Azure Data Factory**, orchestrating Databricks notebooks and Spark jobs via trigger-based and scheduled executions, reducing manual intervention by **80%**.
- Delivered near real-time insights by integrating 10+ processed datasets into **Azure Synapse Analytics**, enabling data-driven dashboards and cross-functional reporting.
- Optimized PySpark job performance by reducing runtime from **18 to 6 hours** and cutting resource usage by 40% through caching, partitioning, and Spark tuning—resulting in **10%** lower storage costs.
- Ensured data pipeline reliability and SLA adherence by working with cross-functional teams and leveraging ADF activity logs and **Databricks** monitoring tools.

Technical Skills

Cloud & Big Data: Azure Data Factory, Databricks, Azure Synapse Analytics, Azure Data Lake Gen2, Delta Lake
Programming Languages: Java, Python, SQL, Data Structures and Algorithms
Developer Tools: Git, Visual Studio, Jupyter, Oracle SQL Developer
Python Libraries: Pandas, NumPy, Matplotlib, Seaborn, BeautifulSoup, Scikit-learn
Big Data Tools: Hadoop, Apache Spark, Hive, ETL

Personal Projects

Flight Analytics Pipeline with Delta Live Tables and dbt

- Implemented Auto Loader with Delta Live Tables for streamlined data ingestion.
- Created parameterized PySpark functions to dynamically generate dimension tables, enhancing reusability and reducing duplication across ETL logic.
- Embedded data quality checks using Delta Live Tables expectations with automated error handling and data validation.
- Built a Lakehouse architecture on Databricks using the medallion model and Delta Live Tables, with Unity Catalog for governance.
- Utilized dbt to create analytical models and materialized views, delivering actionable insights for stakeholder dashboards.
- Tech Stack Used: **PySpark, Databricks, Auto Loader, DLT, dbt, SQL**.

End-to-End Data Lakehouse Pipeline

- Developed a complete ETL pipeline using ADF to extract data from a SQL database into ADLS raw layer.
- Established incremental data loading via ADF Lookup activities and stored procedures, eliminating redundant data pulls and improving pipeline efficiency.
- Architected scalable data transformation workflows in Azure Databricks, simulating the conversion of bronze layer data to the silver layer for downstream consumption.
- Modeled dimensional and fact tables in the gold layer using Delta format, enabling efficient analytics and reporting.
- Tech Stack Used: **Azure Data Factory, Azure Databricks, ADLS Gen2, Azure Synapse Analytics**.

Achievements and Certifications

- Secured **AIR 5** in IGDTUW-CET.
- Microsoft Certified: **Azure Data Fundamentals (DP-900)**

Education

Indira Gandhi Delhi Technical University for Women **08/2022 – 05/2024**
Master's in Computer Applications, CGPA: 8.4 *Delhi, India*