Rubina Naushad Lakhani

Data Engineer

Senior Azure Data Engineer with 9.5 years of experience (5.5 years specialized) in architecting high-scale data systems. Expert in building end-to-end Python and Azure pipelines for batch and real-time processing using ADLS, Kafka, and Snowflake. Proven track record of leveraging ADF, Databricks, and SQL Database to deliver high-throughput, event-driven solutions for complex, multi-format datasets.

✉ rubinalakhani006@gmail.com 📱 9949007066 📍 Hyderabad in linkedin.com/in/rubinalakhani

WORK EXPERIENCE

Senior Systems Engineer
Chubb

07/2024 - Present

Achievements/Tasks

- Architected end-to-end Python pipelines to automate the ingestion, decryption, and processing of 350k+ documents from ADLS, integrating OCR and classification APIs to ensure 100% metadata accuracy.
- Engineered scalable batch and event-driven workflows to handle high-volume historical datasets and real-time document streams, optimizing system throughput and reducing manual processing time.
- Leveraged Snowflake and Kafka to manage complex table operations and build near real-time data streams, ensuring seamless synchronization between document classification and metadata storage.
- Modify existing workflows in Informatica Powercenter and IICS to resolve data issues.

Senior Consultant
Capgemini

04/2022 - 07/2024

Achievements/Tasks

- Designed and developed complex data pipelines using Azure Data Factory to process and transfer data sets from various sources/formats including Azure SQL, JSON, REST API, CSV, XLSX, Cosmos NOSQL DB, and Parquet to data lake.
- Developed complex data transformation logic using Scala, and performed data cleaning, data enrichment, and data validation using PySpark and SparkSQL through Databaricks.
- Experience in working with Delta Lake on Databricks platform, including configuring and optimizing Delta Lake tables.
- Experience with Azure DevOps for continuous integration and deployment of ADF and Databricks code.
- 1-year experience in managing a team of 12-15 members as a technical lead, assisting in solving challenges and reviewing code before production deployment.

Technology Analyst
Infosys

06/2016 - 04/2022

Achievements/Tasks

- Proficient in connecting to various data sources, including Azure Data Lake Storage, Azure Blob Storage , Azure SQL Database, API and CSV From Azure Databricks.
- Created advanced data manipulation techniques using Python, and cleaned, enhanced and validated data using PySpark and SparkSQL.
- Experience in creating and maintaining interactive dashboards and reports using Power BI.
- Worked as an Oracle Fusion Middleware developer with experience in Oracle Data Integrator, Oracle BI Publish and SQL.

SKILLS

Databricks   Snowflake   PySpark

Azure Data Factory   SparkSQL   Delta Lake

Azure SQL   Power BI

CERTIFICATIONS

Microsoft Certified: Azure Data Engineer Associate (DP-203) (11/2023 - 11/2025)

Databricks Certified Data Engineer Associate
 (07/2024 - 07/2026)

ACHIEVEMENTS

Databricks Teaching Assistant for a cohort of 1300 members (05/2024 - 05/2024)
Assist the Databricks instructor in handling a cohort of 1200-1300 members by answering queries posted by the attendees on the Databricks Data Engineer Associate course.

XtraMile Certificate (06/2022)
For oustanding performance.

Certificate of Excellence: Delivery Ninja (01/2022)
For consistent performance and outstanding commitment to work.

EDUCATION

Bachelor of Engineering in Information Technology
 (2012 - 2016)
Muffakham Jah College of Engineering and Technology