SUMAN DEWAN

| **Data Engineer** |+91 6290515886 |suman.dewan1999@gmail.com | Linkedin |

SUMMARY

Data Engineer with over 3 years and 8 months of experience at Cognizant, specializing in delivering robust data pipelines and analytical solutions for top Pharma clients. Proficient in data integration, transformation, and orchestration, with a strong ability to write optimized code. Adept at prioritizing tasks, resolving technical issues, and delivering timely solutions. Eager to embrace new opportunities.

SKILLS

Programming Languages: SQL, PySpark, Spark SQL, Python
Tools and Technologies: Azure Data Factory, Azure Databricks, Dataiku, Azure Key Vault, Azure DevOps, Azure Logic Apps
Collaboration Tools: Jira, Confluence

EXPERIENCE

Data Engineer   **August 2021 - Present**
Cognizant   **Kolkata, India**

- **Build** robust data pipelines using ADF to ingest On-Prem &REST API data into the ADLS bronze layer, supporting both full and incremental loads.
- Hosted SHIR to securely copy the On-Prem tables to Azure Data Lake Gen2 bronze layer.
- Utilized Databricks to apply business logic transformations post ingesting data into raw bronze layer, cleaning the data & storing it into silver layer (Delta tables).
- Mounted ADLS with Databricks and created PySpark scripts to transform the necessary data and store them in the silver layer (Delta Tables).
- **Optimized** PySpark scripts and used AQE, Broadcast join, Repartition, Coalesce, Cache, Persists. Used Partitioning & Bucketing on the Delta Tables
- **Processed** the refined & cleaned silver layer data in Databricks to perform complex transformations & aggregations, with the final data being stored in the gold layer (Delta Tables).
- **Orchestrated & automated** the data pipelines with email alerts for success & failure notifications for Business users.
- **Created scripts** to execute optimize, vacuum, & compaction processes on Silver and Gold Layer Delta tables, and scheduled triggers to run these scripts weekly.
- **Led project** for client MDM team to device an Email ranking & scoring system utilizing ADF & Databricks that analyze each emails metadata and its user engagement metrices and generate scores, which has improved outreach rates by 25%.
- **Designed** & implemented 15+ end-to-end data pipelines from scratch for various client stakeholders to process weekly & monthly data, and compute over 30+ business-critical KPIs for brand performance, prescription adherence rates, customer effectiveness etc. at different granularities.
- **Delivered** data to cross-functional pharma teams for their dashboards, enabling them to monitor and forecast representative activities & customer engagement trends.
- Handled 1TB,100GB csv data by creating end-to-end pipelines in Databricks & Dataiku, and optimized PySpark code to load into Snowflake Tables with 30% reduced execution time.
- **Validated** end-to-end pipeline functionality through UAT, ensuring accuracy and reliability before production deployment.
- Collaborated with Data Platform team to move the pipelines from DEV to QA and QA to PROD using CI/CD pipeline in Azure DevOps.
- **Monitored** Weekly & Monthly Data Pipelines in Production Environment, debugged code and fixed issues.

EDUCATION

Bachelor of Technology (Computer Science & Engineering)   **2017 - 2021**
Future Institute of Technology   (CGPA 8.87)

CERTIFICATION

- Databricks Certified Data Engineer Associate
- Dataiku Core Designer
- Dataiku Advanced Designer