Gokulraj G

ASSOCIATE-PROJECTS AT COGNIZANT

+91 9715892390 | Gokulraj.Govindaraj@cognizant.com | LinkedIn

Professional Summary

Aspiring Azure Data Engineer with 3.5 years of experience in designing and implementing scalable data solutions using Azure cloud technologies. Expertise in ETL pipeline automation, data integration and transformation using Azure Data Factory, Databricks, PySpark and SQL. Strong focus on performance optimization and scalable data solutions.

Skills & Tools

- **Database:** SQL (CTE, Window Functions, Index), Oracle PLSQL, Spark SQL
- **Programming Languages:** Python, Shell Scripting, PySpark
- **Cloud Services** – Azure (Azure Data Factory, Azure Data Lake, Azure Databricks).
- **Tools** – GitHub, SQL Developer, Toad, Visual Studio, WinSCP, Putty, PyCharm

Professional Experience

HCL Tech

Merck & Co, pharmaceutical

Clinical Data Layer – ENCODING, SDTM (Data Engineer)   May 2022 – Mar 2025

- Built ingestion pipelines in Azure Data Factory with Databricks Notebooks to move and transform clinical data from VEEVA to Encoding Catalog.
- Built an event-driven ADF pipeline that was triggered upon file arrival in ADLS. The pipeline utilized Data Flows to perform transformations and loaded the processed data into a reporting container for visualization and business reporting.
- Automated IMMUTA API integration in Databricks Notebook to dynamically register new data sources whenever a new study was introduced or table/column-level changes were detected, supporting data masking and blinding for compliant data access.
- Possess strong working knowledge of Databricks jobs, dbutils, and performance optimization techniques including partitioning, Adaptive Query Execution (AQE), and bucketing, ensuring scalable and efficient data pipelines.
- Hands on experience in implementing data modelling practices, including SCD Types and Change Data Capture (CDC) using Delta Table Merge API across all the Delta tables to maintain accurate and current clinical datasets.
- Proficient in working with various file formats such as JSON, CSV and parquet with additional knowledge of Spark Structured Streaming for real time data processing.
- Familiarity with Azure Synapse Studio, data integration pipelines, dedicated and serverless SQL pools and Notebooks.
- Possess strong understanding of data partitioning strategies and data skew handling techniques to optimize distributed data processing and improve performance in PySpark and Azure Databricks workloads.
- Experience with Delta Lake features like versioning, schema evolution and time travel.

Clinical Data Repository (PLSQL Developer)   Dec 2021 – May 2022

- Designed, developed and optimized PL/SQL procedures, functions, packages and triggers to support healthcare data processing and patient data transformation.
- Achieved 70% performance improvement by refactoring a complex dynamic SQL query with over 100 left joins into a single PIVOT based join.
- Implemented a custom job scheduling workflow using DBMS_SCHEDULER and Oracle triggers by inserting job details into a temporary table and polling it with a procedure executed every 30 seconds.

- Built a dynamic masking package to generate Oracle views based on masking configuration, ensuring data privacy and compliance with healthcare standards.
- Designed and implemented a CI/CD pipeline for database deployment using Liquibase and GitHub Actions, automating version control, validation and rollout of PLSQL changes across environments.

Awards and recognitions:
- Start Innovator
- Top Performer Award
- .

Education & Certifications
- BE(EEE) – Government College of Engineering, Bodinayakkanur - 76 %.
- Higher Secondary Certificate (HSC) – GHSS, Pavakkal - 87 %
- **Microsoft certified** DP-203 Azure Data Engineer
- **Databricks certified** Data Engineer Associate

Profiles
- LeetCode   - https://leetcode.com/u/gokulsakthi/
- HackerRank   - https://www.hackerrank.com/profile/gpgokul67