Sanskar Dikshit

Bangaluru, India | sanskar.dikshit11@gmail.com | +91 8527873619 | LinkedIn | GitHub | Personal Website

Technical Skills

Primary Skills: SQL (Advanced) – CTEs, Window Functions, Indexing, Query Optimization, Python, Pyspark
Cloud Data Engineering: Microsoft Azure, SQL Server Management Studio, Azure SQL Database, Extract Transform Load (ETL), Azure Data Factory, Azure Synapse Analytics, Azure Data Lake Storage, Databricks, Power BI, Data Warehouse, Qlik CDC
Project Management: Agile, Azure DevOps, Jira, Git, CI/CD methodologies

Experience

Data Engineer, TEKsystems Global Services – Bengaluru, India 09/2022 – Present

- Built and deployed 30+ ETL data pipelines using ADF inside Azure Synapse Analytics, handling complex workflows like SCD Type 1, SCD Type 2, etc for processing datasets exceeding 30 million rows across multiple sources and targets.
- Integrated data from diverse sources including FTP, MS Excel, REST API, SQL Server, SharePoint, File System, and HTTP, seamlessly processing over 100 GB of data monthly.
- Developed a parameterized ADF pipeline utilizing a configuration table to process multiple tables, streamlining data pipeline automation and significantly reducing the number of pipelines required.
- Utilized dedicated SQL pool and server-less SQL pool in Azure Synapse Notebooks for data transformation and seamlessly integrated these notebooks in the ADF pipelines.
- Performed user testing and documented all development workflows and best practices in detail, enabling smooth knowledge transfer to different teams and stakeholders.
- Designed Change Data Capture (CDC) tasks in Qlik, enabling real-time data processing of over 100,000 rows daily from Oracle databases to Azure SQL with near-zero latency.

Data Engineer Internship, TEKsystems Global Services – Bengaluru, India 02/2022 – 06/2022

- Utilized Talend Open Studio for developing an efficient ETL solution to load resource utilization data from SharePoint to SQLite database, enabling the stakeholders to query the data, increasing the efficiency of the resource allocation process by improving the billable utilization from 78% to 91% in the month following our solution.
- Created a python script for extraction of skills of employees from employee profiles (.PPT files) using spaCy and python-pptx library with an accuracy of 0.96. This script helped in profile based tags generation to be added to the resource database.

Projects

Scalable ETL Pipeline & BI Dashboard for Lok Sabha Attendance Analysis [GitHub Link]

- Built a scalable end-to-end ETL pipeline in Azure Data Factory to ingest Lok Sabha attendance data from the sansad.in REST API, automating daily data ingestion and storage in Azure Data Lake (ADLS).
- Optimized data processing by consolidating 1,700+ CSV files into a single Parquet file using Databricks (PySpark), reducing storage footprint and query latency.
- DDeveloped an interactive Power BI dashboard with 25+ dynamic visuals, including drill-through pages, navigation panels, and advanced cross-filtering, enabling real-time analysis of MP attendance trends.
- Tech Stack: Azure Data Factory (ADF), Azure Databricks, ADLS, Power BI, REST API

Education

Manipal University Jaipur 2018 – 2022
B.Tech in Computer Science **CGPA:** 8.34/10

Certifications/Achievements

LeetCode SQL 50, HackerRank SQL Gold

Azure Data Engineer Associate Microsoft Certified
Power BI Data Analyst Associate Microsoft Certified