

Employee_ID: 1360

Data Engineer

TECHNICAL SKILLS

Databases: MySQL, SQL Server, Azure SQL.

ETL/ELT Tools: Azure Databricks, Azure Data Factory, AWS Glue, DBT, Informatica.

Data Analytics Tools: Power BI, Thoughtspot, Power BI Services.

Data Warehousing

Tools: Snowflake, Databricks.

Cloud Computing: AWS, Microsoft Azure.

Data Storage tools: Azure Blob storage, AWS S3, Azure Data Lake storage.

Source Control: GitHub

Languages: Python, SQL.

CI/CD: Azure DevOps.

Gen AI: AWS Bedrock, Azure OpenAI

CERTIFICATION



EDUCATION

Bachelor of Engineering (B.E. Computer Science) - K.L.N. College of Engineering, Madurai, Tamil

SUMMARY STATEMENT

Forward-thinking Data Engineer known for driving business impact by building robust, scalable, and high-performing data pipelines. Expert in optimizing end-to-end data workflows to ensure timely, trusted data delivery for critical decision-making. Adept at applying automation and GenAI to simplify complex processes, eliminate inefficiencies, and enhance overall productivity. Committed to turning raw data into strategic insights that empower organizational growth.

EXPERIENCE

Data Engineer - 2 Years 4 months

Optisol Business Solutions Pvt Ltd, Madurai

A pivotal contributor to multiple data migration and analysis initiatives at Optisol Business Solutions, leveraging a diverse technology stack to deliver seamless integration, ensure data integrity, and enhance overall system performance.

PROJECTS

Google Sheets to Azure SQL Server Data Migration (ETL) - Jiffy

This project involved migrating data from Google Sheets-based CSV files into an Azure SQL Server database while addressing schema differences and structural inconsistencies. The work included detailed manual data profiling and field mapping to ensure precise alignment with the target data model. An automated ETL pipeline was developed using Azure Data Factory (ADF), with Azure Data Lake serving as secure and scalable staging for ingestion and transformation. Custom Python scripts were implemented for advanced data processing, including generating JSON outputs from CSV files and converting column values into HTML-formatted content required by the destination schema. The migration process was strengthened with validation rules, error handling, and quality checks to ensure accuracy, integrity, and seamless system integration.

On-Premises to Cloud Data Warehouse Migration - Republic Services

This project involved migrating data from US Ecology into Republic Services' cloud data ecosystem by integrating multiple on-premises databases - Azure SQL Server, Oracle, and MS SQL Server—into a Snowflake staging schema. Informatica was used as the primary ETL tool to build full and incremental load pipelines, while DBT handled data transformations from the staging layer to the ODS schema. AWS Step Functions automated the end-to-end orchestration across DEV, STG, QA, and PROD environments, ensuring consistent and reliable processing. Responsibilities included developing and maintaining Informatica workflows, monitoring daily loads, troubleshooting execution issues, and implementing data validation rules through DBT Core in VS Code, supported by custom macros for integrity checks. The project ensured high data quality, optimized Snowflake models for analytics performance, streamlined incremental load configurations via PuTTY, and aligned deliverables with business requirements through collaboration with cross-functional teams.

On-Premises Oracle to On-Premises SQL Server Legacy Database Migration - Republic Services

This project involved migrating data from an on-premises Oracle database to an on-premises SQL Server legacy system. Manual and GenAI-assisted schema mapping was performed to create a precise mapping file. Using the Python-based AutoSQLGen tool, raw data was extracted from Oracle into SQL Server, after which SQL scripts applied the mapping logic to transform and load the data into staging tables with required cross-walk validations. The processed data was then moved into mirror staging, a structural replica of the original Oracle tables, ensuring accurate alignment. AutoSQLGen's column analyzer and auto-insert modules were used to load data into the test environment for application and database-level unit testing before handing off to QA. Once validated, the mirror data was promoted to production, and the same workflow—with delta analysis—was followed for incremental loads. AutoSQLGen automated most of the extraction, transformation, and loading processes, significantly reducing manual effort and ensuring consistency.

E-commerce Store Performance Analysis

The E-commerce Store Performance Analysis project aimed to enhance business insights and decision-making by analyzing transaction data from an online store. The initiative encompassed data migration, modeling, transformation, visualization, and generating actionable insights. Python scripts were developed for extracting and transforming data from various sources before loading it into SQL Server for advanced processing. Azure Data Factory (ADF) was employed to orchestrate the seamless transfer of data from SQL Server to Azure Blob Storage. Data was then securely loaded into Snowflake using external staging, ensuring integrity throughout the process.

In Snowflake, data models were designed and implemented to optimize and structure the e-commerce data, making it suitable for in-depth analysis and reporting.