Employee_ID:1362

Data Engineer

TECHNICAL SKILLS

Databases: MySQL, SQL Server, Azure SQL.

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

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, C

CI/CD: Azure DevOps.

Gen AI: AWS Bedrock, Azure AI

CERTIFICATION

EDUCATION

B. Tech – Electrical and Electronics Engineering

SUMMARY STATEMENT

Innovative Data Engineer with a strong track record of accelerating business outcomes by designing and implementing efficient, scalable data pipelines. Experienced in optimizing data workflows, enabling faster decision-making through clean, reliable data delivery. Adept at leveraging automation and Gen AI to streamline complex data engineering tasks, reduce manual effort, and improve operational efficiency. Passionate about transforming raw data into actionable insights that drive strategic growth.

EXPERIENCE

Data Engineer - 2.4 years

Optisol Business Solutions Pvt Ltd, Madurai

In my current role, I specialize in designing and implementing advanced data workflows, utilizing Snowflake for scalable cloud data warehousing solutions. I have extensive experience in developing SQL-based stored procedures, functions, and views, along with integrating Snowflake with Databricks to enable high-performance data analysis. Additionally, I excel at creating dynamic, interactive dashboards and reports in Power BI, transforming raw data into actionable insights for business decision-making.

PROJECTS

#1 E-commerce Store Performance Analytics-Snowflake Developer

- Implemented Python scripts to extract and transform data from various online store sources, loading it into SQL Server for further processing.
- Utilized Azure Data Factory (ADF) to orchestrate the movement of data from SQL Server to Azure container storage, ensuring seamless data transfer and synchronization.
- Loaded data from Azure container storage to Snowflake using external staging, maintaining data integrity and security during the transfer process.
- Designed and implemented data models within Snowflake to structure and optimize the e-commerce store data for analytical purposes.
- Developed Power BI dashboards and reports to visualize the transformed data, providing stakeholders with actionable insights into online store performance and trends.

OPEN-SOURCE CONTRIBUTIONS

- **Ora2Pg Code Refactoring Tool:** ORA2PG is an advanced database migration tool that streamlines the complex process of migrating from Oracle to PostgreSQL databases. Our AI-powered solution makes what is typically a challenging technical endeavor accessible, efficient, and reliable

- **ERDynamo:** This Python script processes SQL stored procedures, extracting table relationships and join conditions using Azure OpenAI (GPT-4o-mini). It analyzes SQL files to identify stored procedure names, tables, join types, and conditions, converting the results into a structured format. The extracted insights are normalized and stored in an Excel file for easy review. After extracting the details, the script generates a Mermaid.js script to visualize the ERD.\

- **Impact Analysis Tool:** This automated impact analysis tool swiftly identifies potential risks from database schema changes. By leveraging AI, it analyzes stored procedures, pinpointing impacted areas like datatype modifications or column removals. The tool then generates comprehensive reports, providing actionable insights for database modernization and minimizing disruption.

- **Query Performance Analyzer:** A specialized tool designed to optimize database performance by analyzing SQL queries for bottlenecks, execution patterns, and potential optimizations. The analyzer evaluates indexing strategies, query structure, and resource consumption to recommend improvements that reduce execution time and enhance system efficiency.

- Leveraged advanced analytics techniques within Power BI to uncover meaningful insights related to sales trends, customer behavior, product performance, and more.

- Collaborated with stakeholders to understand business requirements and translate them into actionable data analysis solutions, driving strategic decision-making processes

#2 Deesha - Azure ETL Developer & Power BI Developer

- Leveraged PySpark in Databricks for efficient extraction and processing of data from the Data Lake.
- Implemented Spark transformations for data cleansing, filtering, aggregation, and enrichment.
- Defined and enforced schema consistency for transformed data, ensuring data integrity within the Data Lake Gold Layer.
- Orchestrated the integration between Databricks and Azure SQL, orchestrating data extraction, transformation, and loading (ETL) processes.
- Utilized Azure SQL Database as the target destination for processed data, ensuring seamless integration with existing SQL-based systems.
- Contributed to the development of Power BI dashboards by providing high-quality, transformed data from Azure SQL.
- Collaborated with stakeholders to design and implement interactive visualizations and reports in Power BI, facilitating data-driven decision-making.

#3 Snowflake-Powered Market Research Analytic - Snowflake Developer

- Designed an efficient ELT pipeline leveraging AWS S3, Snowflake, and DBT for data processing and analytics. Designed an effective
- Implemented automated and efficient data loading from AWS S3 into Snowflake tables using Snow pipe, ensuring real-time data availability.
- Utilized DBT to construct comprehensive and efficient transformations, facilitating seamless migration from source to target databases within Snowflake.
- Engineered an efficient incremental model in DBT for processing incremental data, enhancing data processing capabilities and reducing latency.
- Created efficient visualizations using Power BI after the data was transformed, providing actionable insights and enhancing data-driven decision-making.

BLOGS & PUBLICATIONS

- **Streamlining Data Processing with AWS, Snowflake and DBT**, offering comprehensive insights into ELT methodologies, showcasing expertise in cloud computing, data warehousing, and transformation tools.

- **Unlocking Data Potential: Leveraging DBT on Snowflake with Docker for Dynamic Transformation Pipelines,** demonstrating proficiency in Docker container deployment for dynamic data transformation pipelines using DBT on Snowflake.

- **ERDynamo — The Ultimate Snowflake ERD Accelerator** - An in-depth exploration of automating entity-relationship diagram creation for Snowflake databases, showcasing how ERDynamo leverages AI to analyze complex stored procedures and visualize data relationships for improved database management.

- **From REST to MCP: Standardizing the Way AI Agents Think and Act** - A comprehensive analysis of the evolution from REST APIs to Message Control Protocol (MCP), discussing how standardizing AI agent communication enhances interoperability, efficiency, and reliability in multi-agent systems.

- **Accelerating Data Migration with Microsoft Fabric & Copilot: AI-Driven Simplicity** - Detailed insights into leveraging Microsoft Fabric and Copilot technologies to streamline data migration processes, demonstrating how AI-powered tools significantly reduce migration time while improving accuracy and providing actionable insights.

OptiSol
Solutions | Services

#4 Republic Services :US Ecology Data Migration – Data Engineer

- Develop and deploy Informatica IICS pipelines for both incremental and full loads to ingest data from source systems into the Snowflake staging schema
- Manage deployment of table DDLs and IICS job exports to STG and QA environments via ServiceNow tickets.
- Design DBT models for full and incremental loads, incorporating column naming conventions, deduplication, and data transformation rules.
- Implement DBT test cases for not null and uniqueness validations in YML files and create GitHub pull requests for code integration across branches.
- Collaborate with QA to address defects, ensuring high data quality and compliance with project standard