Employee ID : 1271

LEAD DATA ENGINEER - SNOWFLAKE

TECHNICAL SKILLS

Databases: SQL, Document & Graph Databases like PostgreSQL, Oracle, MySQL, MongoDB, DynamoDB, Neo4j.

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

Data Analytics Tools: Power BI, Quicksight.

Data Warehousing Tools: Snowflake, Azure Synapse Analytics, AWS Redshift, Azure Databricks.

Cloud Technologies: AWS, Microsoft Azure, Google Cloud Platform.

Web Frameworks and technologies: REST API, HTML, CSS.

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

Data Governance & Quality: Data Quality Controls, Data Validation Frameworks, Data Profiling, Monitoring & Alerting, SLA Management

Workflow Scheduling tools: Airflow, AWS StepFunctions

Source Controls: GitHub, AWS Code Commit, GitLab.

Languages: Python, SQL, JavaScript.

Deployment tools : Jupyter Notebooks, Visual Studio Code, Databricks.

Tracking & documentation Tools: JIRA, Confluence, Rally, SharePoint, Notion

PROFESSIONAL SUMMARY

Results-driven Data Engineer with 6+ years of experience in designing, developing, and optimizing scalable data pipelines and analytical solutions, specializing in Snowflake Cloud Data Platform. Proficient in end-to-end ETL/ELT development, data modeling, and performance tuning to support data-driven decision-making across enterprise environments.

Hands-on expertise in cloud data integration using tools such as Azure Data Factory, AWS Glue, dbt, and Informatica, along with strong proficiency in SQL, Python, and Snowflake's advanced features including micro-partitioning, Time Travel, Clustering, and Streams & Tasks. Adept at building star and snowflake schemas, optimizing query performance, and implementing data governance and security best practices.

Collaborative and detail-oriented, with a strong track record of delivering cost-efficient, high-performing, and automated data workflows. Passionate about leveraging modern data architectures, cloud platforms, and data engineering best practices to enable robust analytics and reporting solutions..

EXPERIENCE

- Lead Data Engineer in OptiSol Business Solutions Pvt Ltd – June 2023 - Present
- Worked as a Data Engineer At OneIntegral Technologies– May 2019 – June 2023

PROJECTS

#1 Data Migration to Enterprise Data Warehouse – Republic Services

Technology: Snowflake, Informatica, Generative AI, dbt , pyspark, python, AWS step functions
Duration: 16 months
Team Size: 8
Description:
The project's objective was to migrate on-prem SQL and oracle servers to enterprise warehouse to generate business insights from the underlying data.
Role:

- Build 30+ integrations of source objects from 4 different database systems and did Initial & Incremental loads using Informatica mapping task flow data integration services.
- Build 30+ dbt models to do ELT transformations between staging data schema to ODS schema within snowflake.
- Automated DDL scripts for creating tables in staging

CERTIFICATIONS

EDUCATION

Bachelor Of Technology (2013-2017) (Leather Technology) – AC.Tech, Anna University.

schema using python and Generative AI.
- Data analysis to validate data load via informatica using pyspark.
- Automated data modelling to build star schema using Generative AI.

#2 Warehouse Optimization-Bricz

Technology Stack:
AWS S3, AWS Lambda, AWS Glue, Amazon SQS, AWS CloudWatch, Snowflake, Snowpipe, Python, PySpark, dbt, GitHub Actions

Duration: 8 months
Team Size: 4

Project Description:
The goal of this project was to optimize warehouse and supply chain operations by building a cloud-native, event-driven data pipeline that automated ingestion, transformation, and reconciliation of data for real-time analytics and machine learning use cases. The solution integrated AWS services with Snowflake to deliver a highly scalable and automated data ecosystem.

Role & Responsibilities:

- Designed an event-driven ingestion architecture using AWS S3 as the data landing zone, triggering AWS Lambda functions for downstream processing.
- Developed ETL pipelines in AWS Glue (PySpark) to cleanse, validate, and transform raw data before storing it in curated S3 buckets.
- Configured Amazon SQS to trigger data refresh workflows and communicate status updates between AWS services and Snowflake.
- Implemented Snowpipe to continuously load transformed data from S3 external staging tables into Snowflake actual tables, enabling near real-time data availability.
- Built and maintained dbt models within Snowflake for data modeling and business rule transformations.
- Developed an automated reconciliation framework integrated with BotQ to validate data consistency between raw, transformed, and Snowflake layers.
- Implemented monitoring and alerting through AWS CloudWatch for data pipeline health checks and failure notifications.
- Integrated CI/CD pipelines using GitHub Actions for automated deployment of AWS Glue scripts and dbt models.
- Supported data analytics and machine learning teams by ensuring high-quality, timely, and accessible data through Snowflake’s secure and scalable environment.

OneIntegral Technologies, Chennai May-2019 to June-2023

#3 ATM Downtime Reduction
Technology: Databricks, Azure Storage, Python, Pyspark, Spark ML.
Duration: 1 year and 3 months
Team Size: 3
Description:
The project's objective was to train machine learning model using data generated using sensors, failure codes, maintenance code and spare replacement codes from ATM machine from its history and predict probable future failures and recommend spare part that needs to be replaced in case of machine failure.
Role:

- Loaded huge volume of historical data from EJ logs, sensor logs, failed component images, spare replaced history into Azure Storage and onto Databricks.
- Built error patterns and classified them into fatal and non-fatal patterns using pyspark.
- Mapped the spare parts replaced to fix the fatal patterns using Python script.
- Prepared training data from error patterns after data cleaning and feature engineering.
- Implemented Auto ML to select the best model and hyperparameters.
- Created pipeline for periodic training and model evaluation.

#4 Cash on Delivery Tracker
Technology: Azure Storage, Azure Functions, Python, Pandas, Pyspark, Azure SQL Database, Azure Data Factory, Power BI.
Duration: 6 months
Team Size: 3
Description:
The main scope of the project was to build a system to track station wise COD remittance versus deposit posted status. The scope was further extended to aggregate reports at district, state, region levels to track business performance at every level.
Role:

- Streamlined data intake using Azure Functions and Azure Storage.
- Built scripts to transform remittance and deposit details into a standard reconciliation structure using Azure Data Factory and Pyspark.
- Built reports using pyspark and SQL queries.
- Extracted reports aggregated at various levels of business using Pyspark.
- Built dashboards using Power BI to enable the management team to derive insights from data after reconciliation.

#5 MIS Dashboard
Technology Stack:
Snowflake, AWS Glue, AWS Lambda, Amazon S3, Snowpipe, dbt, AWS Step Functions, AWS QuickSight, PySpark, Python, GitHub Actions

Duration: 2 years 3 months
Team Size: 4

Project Description:
The objective of this project was to build a Management Information System (MIS) Dashboard providing real-time business insights across multiple e-commerce channels. The solution automated data ingestion, preprocessing, and computation of cost and revenue at each business layer to detect revenue leakages, optimize operational costs, and measure performance KPIs. The data platform was designed around Snowflake for centralized storage, transformation, and analytics.

Role & Responsibilities:

- Automated data extraction from multiple e-commerce portals using Python (Selenium) and integrated with AWS S3 as the raw data landing zone.
- Developed AWS Glue ETL jobs (PySpark) to standardize schema, validate mandatory fields, and perform data cleansing and enrichment before loading into Snowflake staging tables.
- Implemented Snowpipe for near real-time ingestion of preprocessed data from S3 into Snowflake.
- Created dbt models on Snowflake to build business logic transformations, compute cost and revenue metrics across business hierarchies, and generate aggregated datasets for reporting.
- Designed stored procedures and task automation within Snowflake to orchestrate daily data refreshes and reconciliation checks.
- Used AWS Lambda functions to trigger Snowpipe and dbt model refresh workflows orchestrated via AWS Step Functions.
- Built interactive AWS QuickSight dashboards leveraging Snowflake datasets to visualize trends in cost, revenue, and performance KPIs across time and business segments.
- Implemented CI/CD automation using GitHub Actions for version control and deployment of Glue jobs and dbt transformations.
- Collaborated with business stakeholders to define KPIs, improve data accuracy, and ensure consistent visibility across business layers.

OPEN-SOURCE CONTRIBUTIONS

- Detecting Moving Objects Using Gaussian Mixed Model: In this work cars passing through a highway are being separated from their environment using Gaussian Mixed Model using OpenCV.
- https://github.com/chandrakanthkarunakaran/GaussianMixedModel-Moving-Object-Detection.git

- Graph Anonymization Using Network Obfuscation: Graph anonymization algorithm is developed using networkx to protect privacy of users in social networks when the network is plotted as a dataset, by obfuscating the said network.
  https://github.com/chandrakanthkarunakaran/Graph_Anonymization.git

- Memory Game: This is an interactive game built using Pygame to build memory strength of the player.
- https://github.com/chandrakanthkarunakaran/MemoryGame.git

- Zoo Management Application Using Python/Django: This is a complete web application built using python and Django to manage visitors, create shows, manage bookings, and enroll animals for an imaginary zoo.
- https://github.com/chandrakanthkarunakaran/ZooManagement.git

.