Contact

+91-7075088458

navyasrimushini24@gmail.com

www.linkedin.com/in/navyasri24

Technical Skills

- SQL
- Python
- PySpark
- Snowflake
- AWS(IAM,S3,EC2,GLUE, ATHENA,REDSHIFT)

Soft Skills

- Team Building
- Communication
- Adaptability
- Active Listening

Certifications

- SnowPro Core Certification
- Aws Certified Cloud practitioner

Navyasri Mushini

Professional Summary

Detail-Oriented Data Engineer with one year of strong experience in designing, implementing, and optimizing data pipelines for robust data solutions. Proficient in utilizing various data engineering tools and technologies to streamline data processes and enhance data accessibility and reliability. Demonstrated ability to work collaboratively with cross-functional teams to deliver high-quality data solutions that support business objectives.

Experience

Designation   : Junior Associate data Engineer

Company   : Kasmo Technologies, Hyderabad

Experience   : 1 Year

Responsibilities

- Develop and maintain robust data pipelines to ingest, process, and transform large volumes of data from diverse sources,ensuring scalability, reliability, and efficiency.
- Design and implement ETL processes to facilitate the seamless transfer of data between on premises and cloud-based storage environments, leveraging your expertise in cloud services such as AWS Cloud Platform.
- Continuously evaluate and adopt emerging technologies and tools in the field of data engineering to enhance productivity, efficiency, and innovation.
- Gain proficiency in data engineering tools and technologies such as SQL databases, Apache Spark, Snowflake, python through practical application and training

Project

Title   : Data migration
Duration   : July -2023 – Jan -2024
Technologies used   : sql server,redshift,Amazon s3,shellscript,pyspark, Snowflake

Project Description:

This project involved migrating data from various local databases to Snowflake, utilizing a multi-stage approach for data extraction, transfer, and loading. The process ensured data integrity and consistency through rigorous checksum and reconciliation checks at each stage.

Roles & Responsbilities

- Connected to various source databases using native connectors to ensure secure and efficient data extraction.
- Extracted data from source databases and saved it in Parquet file format to maintain data structure and enable efficient storage and transfer.
- Automated the extraction process using scripts to handle different database sources and ensured regular extraction schedules.
- Developed shell scripts to transfer Parquet files from local storage to Amazon S3.
- Implemented file reconciliation processes to verify file integrity during transfer by comparing local and S3 file sizes and checksums.
- Created and managed S3 folders to organize and store incoming files, ensuring that metadata files were moved to designated metadata folders.
- Created and ran scripts to load data from S3 into Snowflake, utilizing Snowflake's external table features.
- Designed and implemented data reconciliation processes to compare checksums between source files and target tables in Snowflake, ensuring data integrity before final loading.
- Generated YAML files to automate and document the data load process.
- Managed the creation of views in Snowflake to facilitate data access, incorporating MD5 hash key values for data validation.
- Automated the deletion of the MD5 column from views before loading data into target tables to maintain a clean data schema.

Project

Title   : Manjushree Technopack Limited (MTL)
Duration   : Feb -2024 – may-2024
Technologies used : sql server, pyspark ,snowflake

Project Description   :

As a part of Manjushree Technopack Limited (MTL), a leading rigid plastic packaging company, I contributed to the development of   a comprehensive reporting system that encompassed Sales, Budget, Operations, and Financial reports. The project aimed to provide stakeholders with real-time insights and enable data-driven decision-making across various business areas. One of the project's unique challenges was the regular acquisition of new companies by MTL. This required integrating the acquired company's data with MTL's existing data to create a Unified solution. To address this, we utilized Snowflake and developed a data model.

Roles & Responsbilities

- Worked on the company's Sales module that was maintained in SAP systems.
- Understood that the data architecture involved the various source systems, planned zones of ETL processing, transformations, and reports required by the end-user
- Implemented PySpark pipelines to extract data from diverse sources, such as CSV files and databases.
- Utilized PySpark capabilities to write DataFrames to load transformed data into the Snowflake data warehouse system.
- Utilized in-built Python packages such as datetime for handling date and time operations, to manipulate DataFrames.
- Created Streams to capture inserts/updates/deletes on Source data.
- Used Snowpipe to pull source files from external cloud storage such as AWS S3 buckets.
- After data cleansing, created the data warehouse with Dimension and Fact table in Star Schema

- Created Stored Procedures to implement Slowly Changing Dimension logic (Full load, SCD1, SCD2) for Dimension tables in SnowSight.
- Performed Incremental load for Fact tables.
- Created Tasks to perform ETL operations based on a schedule
- Hands-on understanding of data masking policies and row-level security materialized view for security
- Shared secure and relevant data required by the Marketing team with the use of data sharing.
- Received recognition for the extensive analysis of Budget data and building the entire flow to come up with the Fact table for Budget data.

Education

- **Graduation**

  Kakinada Institute of Engineering and Technology – II

  Computer Science and Engineering (2019-2023) - 70%

- **Intermediate**

  GRC Modern Junior college – MPC (2017-2019) -88%

- **SSC**

  ZPHS Gachibowli (2017) – 88%