**TULASI**

Data Engineer | +1(223) 356-3459|

**Professional Summary:**

Experienced Data Engineer with 10 years of expertise in building, optimizing, and managing data pipelines, ETL and ELT processes, and scalable data solutions across banking, retail, and healthcare domains. Proficient in Python, Snowflake, SQL, Airflow and PySpark for data processing and transformation, with hands-on experience in AWS and Azure cloud platforms. Skilled in streaming technologies like Kafka and batch processing using Spark. Knowledge in relational databases ensuring data integrity through data quality frameworks and governance. Strong in data modelling, version control, Agile methodologies, and pipeline architecture, collaborating effectively with teams for secure and efficient solutions.

**Technical skills:**

**Data Engineering:** Python, Snowflake, PySpark, SQL, AWS, Azure, Databricks, ETL-Processing, PostgreSQL, Data Warehousing, Jupyter Notebook, Data Analysis, Alation Cloud, DataVault 2.0 Modeling

**Tools:** Aspire, VS-Code, PyCharm, AWS Glue, ADF, Airflow, JIRA, Jenkins, Tableau, PowerBI, Collibra

**Version Control & CI/CD:** GitHub, Azure DevOps, ClearCase

**API Technologies:** Java, JSON, REST, Postman

**Machine Learning:** SciKit-Learn, Pandas, NumPy, EDA, RFE, NLP, Seaborn

**Methodologies:** Agile, Scrum

**Professional Experience:**

**Data Engineer - Hitachi Digital Services Mar 2023 - Aug 2024**

**Client: Texas Capital Bank, TX.**

All Banks need to streamline data from various sources for analysis and reports. Data is extracted from different sources, transformed as per the business requirements using AWS Glue jobs and Stored Procedures to move to Snowflake Data Warehouse and processed in Data Lake and Data Marts.

* Developed scripts for data processing in Snowflake, Python, AWS Glue, and stored procedures, leveraging Databricks for scalable data processing and transformation tasks.
* Created AWS Glue and Azure Data Factory jobs to fetch data from S3 and Azure Blob Storage, using Azure Databricks for processing, transforming it into external tables and materialized views, and performing ETL transformations.
* Designed and developed robust ETL pipelines using PySpark in Databricks, as well as on AWS platforms, enabling the extraction, transformation, and loading of data from various sources into data lakes and data warehouses.
* Developed PySpark scripts in Azure Databricks for processing data from multiple sources to destination buckets on AWS S3 and Azure Blob Storage. Implemented complex data transformation logic in PySpark to clean, aggregate, and enrich datasets, ensuring data integrity and accuracy for downstream analytics.
* Integrated multiple data sources, including databases and file types such as CSV, XLS, XML, and Text, into the ETL pipeline using PySpark in Azure Databricks, streamlining data ingestion processes and improving accessibility across AWS and Azure environments.
* Possess knowledge of streaming platforms such as Kafka, Kinesis, and Azure Event Hubs, and understanding of how they facilitate real-time data processing and ingestion.
* Collaborated with cross-functional teams to gather and translate functional and technical requirements into effective reporting solutions, ensuring data quality and accuracy through the development of aggregate data models, dimension views, and fact tables across cloud platforms.
* Worked with Azure DevOps boards for managing tasks, tracking progress, and collaborating with team members to ensure efficient and timely delivery of project milestones.
* Orchestrated and automated end-to-end ETL workflows using Apache Airflow to trigger and manage AWS Glue jobs, enabling streamlined data processing and enhancing pipeline reliability across AWS services.
* Developed Airflow DAGs to monitor and manage Glue job execution with custom retry logic, status tracking, and alerting, ensuring high availability and fault tolerance in data pipelines.

**Data Engineer - VHS Consulting Mar 2022 - Feb 2023**

**Client: Sainsbury’s, London, UK.**

**Project #1: Customer Complaints**

Customer Complaints is the different stream that handles all the complaints received from the customers on various products. This stream contains twelve different objects with different types of complaints data. Performed data engineering tasks of these streams from scratch. The DataVault model has three layers and moves data from Stage to RDV, BDV and then to Presentation layers. Have created the Satellite, Link & Hub objects for every stream.

* Worked extensively on ETL, data transformations, Snowflake Data Warehouse, SQL and Python.
* Built data pipelines specific to each Complaints stream.
* Performed data extraction from different sources and loaded into Stage using Snowpipe, data is transformed & loaded as per business requirements using the MWAA tool.
* Snowpipe ETL processes then ingests data from AWS Landing buckets to the Stage.
* The DDL’s are built & deployed with Jenkins pipelines while the DML’ are managed by Airflow jobs.
* Implemented the MWAA Airflow Jobs for ingesting data across various buckets to make pipeline work.
* All the database changes are specified in Liquibase change logs.
* Data profiling is done to maintain the quality of the data in Alation cloud.
* Deployed yaml files for all Snowflake objects with classification data else Aspire standards will mask it.
* Involved in Training and KT sessions for newly onboarded Teammates.
* Collaborated with cross-functional teams to gather and translate functional and technical requirements into effective reporting solutions, ensuring data quality and accuracy through the development of aggregate data models, dimension views, and fact tables.
* Followed Agile methodology managing Sprint meetings like planning and refinement on JIRA board.
* Created GitHub Repos by replicating the common repos along with Jenkins pipeline.
* Familiarity with data governance, privacy regulations like GDPR,CCPA,HIPAA and security best practices.

**Project #2: Own Brand-Healthy Living**

Own Brand are products being labeled and sold as their own products by the Clients handling all Nutritional & Ingredients data of the products. Processed data for health ratings of products based on ingredients and nutrition involved in each product, so customers are fully product aware and decide to opt-in or opt-out. Built a model with a decision tree to classify the products into Five categories - Healthy, Healthier Choice, Better For You, Occasional Treat and Pure Indulgence. Different sets of attributes were used for categories of food like General Foods, Beverages & Ready Meals.

* Performed Exploratory Data Analysis after extracting data from Kafka topics and transformed per requirements with Snowflake flattening of JSON data and loaded data using MWAA tool.
* Assisted in model development, optimizing the model for better accuracy, testing and validation.
* Performed data analysis with standard statistical methods, interpreted & provided results as summary.
* Retrieved desired attributes through flattening and created Satellite, Link and Hub objects.
* Data received from the Evolve Kafka topic onto the stage table is moved through three layers of Data Vault model - RDV, BDV and presentation layers.
* Worked on client‘s proprietary tools to load static files data into stage by creating yaml file with Data Quality rules and created Barbossa roles which deal with the roles and policies of Snowflake users.
* Utilized Agile methodologies and tools (e.g., JIRA) to manage MicroStrategy development projects, ensuring efficient delivery and alignment with business needs.
* Designed and implemented robust backend and frontend structures for business visualization and reporting, leveraging industry-leading integration and analytics tools to build a modern, high-performing data warehouse that meets business requirements.

**Senior Developer/Data Engineer - Silmaril Software Aug 2019 - Feb 2022**

**Client: Cars24**

**ETL for Sales Data**

Involved in Data Engineering activities to Pull the data from source which is Azure blob storage and make the

data available in our stage which is snowflake. Further data science team visualizes the data.

The sales data has different dimensions like manufacturer, model, engine, year of manufacturing,

registration year etc for which the data will be visualized, and proper steps would be taken by the marketing and business team to increase sales.

* Extensively performed data engineering by extracting data from Azure Blob Storage and creating ETL pipelines using Azure Databricks with PySpark.
* Transformed data based on business requirements and loaded it into Snowflake for further analytics and reporting.
* Implemented ETL pipelines using PySpark in Azure Databricks to process large datasets, enabling efficient data transformation and integration from multiple sources into Snowflake.
* Designed and optimized PySpark jobs, improving ETL execution time by 40% through effective data partitioning and caching strategies in Azure Databricks.
* Utilized DataFrame and RDD operations for data manipulation, including filtering, aggregating, and joining datasets to support analytics and reporting requirements.
* Enabled business and data science teams to visualize data and gain insights into sales performance using Snowflake as the data warehouse.
* Deployed code to higher environments, executing enhancements and managing Change Requests as required.
* Followed Agile methodology, managing sprint planning and refinement meetings using the Jira board.

**Data Engineer -** **Vijaya Bank May 2013 – Jul 2015**

Portal Development

* Developed the Banking portal, Performed Unit testing and smoke testing Analysis and fixed defects.
* Prepared Quality Documents for MMIS Modules.
* Created Presentations & Business Tier using JSF & JSR168 Portlets, JavaBeans, DAO & Data Access Layer.
* Development and integration of project applications - WebSphere Portal presentation and WAS Layers.
* Used ClearCase Control Version System for source code. Executed Enhancements and Change Requests.

**Data Engineer - Infinite Computer Solutions Aug 2010 – Apr 2013**

**Client: ACS Xerox**

**Medicaid Management Information System (MMIS)**

Medicaid is a State administered health insurance program financed and run jointly by the Federal and State Governments for low-income people of all ages without money or insurance to pay for health care. The goal of the program is to provide medical and other health care services to eligible individuals making them self-sufficient. Each State sets its own guidelines, subject to federal rules and guidelines. Certain services must be covered by the States to receive Federal funds. Other services are optional and are elected by States. The Model Medicaid program used by each state is a basis for computer processing. The model MMIS is divided into four functional areas / subsystems - Member, Claims Processing, Provider, & Program.

* Development and integration of project applications - WebSphere Portal presentation and WAS Layers.
* Performed Unit testing and smoke testing Analysis and fixed defects.
* Prepared Quality Documents for MMIS Modules.
* Created Presentations & Business Tier using JSF & JSR168 Portlets, JavaBeans, DAO & Data Access Layer.
* Used ClearCase Control Version System for source code. Executed Enhancements and Change Requests.

**Education & Certifications:**

Certified - Skillsoft Snowflake SnowPro - 2023

Master of Computer Applications - Andhra University College of Engineering, India. - 2009