Vishwa Chaitanya Dodla

vishwachaitanya892@gmail.com 945-213-3135 Frisco, TX

linkedin.com/in/vishwachaitanya

PROFESSIONAL SUMMARY

- Accomplished Data Engineer with **2+ years** of experience focused on executing **data engineering** tasks and designing efficient **data pipelines**.
- Experienced at managing and analysing large datasets using data processing frameworks such as **Databricks** and **Apache Spark**.
- Proficient in database management using **MySQL, OracleDB, SQL, and PLSQL**, assuring data integrity, retrieval, and optimisation.
- Expertise in leveraging **AWS cloud technologies** to optimise data processing, storage, and migration activities, resulting in efficient and scalable solutions.
- Extensive knowledge in using **AWS Glue, AB Initio, an ETL solution**, to automate data extraction, transformation, and loading operations. This knowledge guarantees that data is seamlessly integrated across a wide range of systems and platforms.
- Utilized **Snowflake** cloud-native data warehousing capabilities, leveraging AWS infrastructure to design and implement scalable data storage solutions.

SKILLS

Programming Languages
Python, SQL, PySpark, SparkSQL, Pl/SQL, NoSQL

ETL Tools
Ab initio, Informatica, AWS Glue.

Libraries
Pandas, NumPy, Plotly

Cloud
Azure, Aws

Database Tools
MySQL, Oracle, SQL, PostgreSQL, Teradata.

Big Data Frameworks
Databricks, Apache Spark

EDUCATION

Master's 12/2022 | University of Memphis, Tn
Information Systems

Bachelor's 03/2017 | JNTUH, India
Materials Science & Metallurgy engineering

WORK HISTORY

DATA ENGINEER 06/2023 – present | USA
UHG

- Implemented AWS DataSync to create seamless connections between on-premises data sources and Amazon S3, including Self-Hosted Integration for secure data transmission. Data was successfully moved from the sources to the destination Amazon S3 buckets.

- Utilized PySpark's parallel processing capabilities to load and transform 100GB of CSV files into Delta Lake Tables, resulting in a 95% boost in data processing efficiency.
- Strong knowledge of PeopleSoft modules including HRMS, Financials, Campus Solutions, etc.
- Developed and maintained PeopleSoft integrations using Integration Broker, ensuring seamless data flow between different systems.
- Implemented bulk data transfers, achieving a seamless 95% success rate, from an external AWS S3 stage to Snowflake internal stage, seamlessly executing the process using the powerful COPY command.
- Implemented a thorough quality assurance plan for verifying and processing data from source systems, which resulted in 99% accuracy of integrated data into the data warehouse.
- Developed Abinitio graphs to fetch data from DB2, Oracle, Teradata, Excel, flat files, and XML Files.
- Created PySpark notebooks in Databricks are used to read external JSON data and send it to Data Lake Tables.
- Designed and executed a data pipeline on AWS using Amazon Step Functions and AWS Glue Python Shell activity to schedule notebooks, leading to a notable 30% decrease in manual data processing efforts.
- Specific Databricks notebooks were scheduled as Databricks jobs and executed on job clusters and Databricks identified the best clusters for cost and resource allocation.
- Debugged the data pipeline, investigated anomalies, and resolved faults to guarantee seamless operation.

JUNIOR DATA ENGINEER 06/2020 – 08/2021 | India

Softkey Sol

- Developed data cleaning and manipulation procedures using Python libraries Pandas and NumPy for large datasets.
- Designed, developed, and maintained scalable data pipelines and ETL workflows in conjunction with cross-functional teams to guarantee effective data extraction, transformation, and loading procedures.
- Developed and optimized data models, schemas, and database architecture using Oracle, SQL Server, and MySQL technologies, resulting in a 20% increase in data retrieval efficiency.
- Loaded data from Snowflake internal stage into its tables using SnowSQL, leveraging a combination of COPY, LIST, PUT, and GET commands to meticulously validate internal stage files.
- Utilized big data processing frameworks such as Databricks, Apache Spark, and Hadoop to handle large-scale data processing, allowing for efficient data handling.
- Designed and developed AbInitio graphs, created psets and user defined functions using GDE as per the client requirement.
- Developed Spark applications using SparkSQL in AWS Databricks to perform data extraction, transformation, and aggregation from various file formats. Successfully loaded the processed data into target destinations
- Implemented Spark tuning and optimization techniques to increase data processing efficiency by70%.
- Optimized data pre-processing procedures to improve data quality and consistency, enhancing the accuracy and reliability of subsequent analyses and reporting.

ACADEMIC PROJECT

- Foot Traffic Data Analysis Project (Skills/Tools: Python, APIs, Pandas library, Data transformation, Data filtering, Data aggregation, Data frame manipulation)