Mayur Dudhe

+91- 9922417719

mayurdudhe111@gmail.com Chh. Sambhajinagar(MH)

OBJECTIVE

Recent graduate with a strong foundation in data analysis, programming, and database management. Skilled in SQL, Python, and data processing tools. Eager to apply these skills in data engineering, contribute to data-driven projects, and gain hands-on experience in big data and cloud technologies.

CAREER SUMMARY

1. Graduated with a strong foundation in data engineering concepts.
2. Proficient in SQL, Python, and basic data processing tools.
3. Familiar with Hadoop ecosystem components (HDFS, MapReduce, Spark).
4. Knowledgeable in cloud technologies (AWS Services)
5. Strong analytical and problem-solving skills.
6. Familiar with Agile methodologies and collaborative environments.
7. Motivated to contribute to impactful data-driven projects in the retail sector.

TECHNICAL SKILLS

Operating System: Windows, Unix

Hadoop Ecosystem: HDFS, Map-Reduce, Hive, Spark.

Relational Database: Oracle

Programming Language: Python, PySpark, SQL

Cloud Technologies:

AWS: (RDS, S3, Glue, Lambda,Athena,EMR,EC2,SNS,Step functions,EventBridge, cloudwatch)

Visualization Tool: PowerBi

Development IDE: Pycharm,Putty,SQL workbench,Databricks

ACADEMIC PROJECT : **ETL Pipeline with AWS Glue**

Objective:

To design and implement an automated ETL pipeline using AWS Glue, AWS Lambda, and Amazon RDS for efficient data processing, validation, and storage. The goal is to streamline the ingestion, transformation, and validation of product category data (inventory, price, and order), ensuring high-quality data is stored for downstream systems and reporting.

Roles & Responsibilities:

- **Data Ingestion (S3 - Source)**:
  - Ingested various product category data feeds (e.g., inventory, price, order) from upstream systems into S3.
  - Managed data uploads in CSV/JSON format for further processing by downstream systems.
- **Event-Driven Processing (AWS Lambda)**:
  - Utilized AWS Lambda to automate the triggering of AWS Glue ETL jobs based on data uploads in S3.
  - Ensured event-driven, dynamic processing of different file types (inventory, price, order).
- **Data Transformation & Validation (AWS Glue)**:
  - Cleaned and transformed incoming data, removing unwanted characters and correcting formatting inconsistencies.
  - Validated data according to business rules, ensuring correct prices and inventory records.
  - Stored valid records in the master schema and flagged invalid data for further analysis. ● **Parent-Child Record Validation (AWS Glue)**:
  - Performed validation on parent-child relationships (e.g., price linked to product records) using AWS Glue.
  - Managed unmatched records by storing them in an unprocessed schema for further review.
- **Data Storage (Amazon RDS - PostgreSQL)**:
  - Stored, cleaned and validated data in Amazon RDS (PostgreSQL) for structured storage and querying.
  - Ensured invalid records were placed in a separate unprocessed schema for manual intervention.
- **Data Extraction (AWS Glue ETL Jobs)**:
  - Extracted data from Amazon RDS (PostgreSQL) and performed further transformations.
  - Converted data into Parquet format to optimize storage and enable efficient querying. ● **Data Upload (S3 - Target Location)**:
  - Uploaded the transformed Parquet files to a target S3 bucket, making the data available for downstream systems such as SAP.
  - Enabled reporting and visualization for stakeholders using clean, well-structured data.

Key Technologies Used:
AWS Glue, AWS Lambda, Amazon RDS (PostgreSQL), Amazon S3, Parquet

SKILLS GAINED

Data ingestion, ETL pipeline development, cloud-based data processing, event-driven architecture, data transformation and validation, AWS services (Glue, Lambda, S3, RDS), data storage optimization (Parquet format).

EDUCATIONAL QUALIFICATION

Bachelor of Engineering in Computer Science from Swami Vivekanand Institute of Technology and Management, Sohna in Academic Year 2024

SELF DECLARATION

I hereby declare that the above mentioned are true to the best of my knowledge.

Mayur Bandu Dudhe

9922417719