**Batch Processing with AWS**

Prepared by Nikunj Rathod

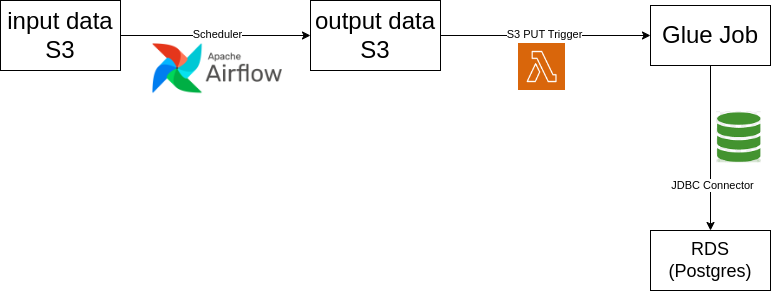
# **Overview**

This document provides a detailed workflow for implementing batch processing with AWS over a 6-week period. The workflow covers data ingestion, loading data into an SQL database, and using AWS services such as Apache Airflow, AWS Glue, and Amazon RDS. Different Machine Learning models and their Comparision, Trained LLM to answer the simple questions about the dataset as well as Visualization of the data with PowerBI and Python Matplotlib.

# Table of Contents

|  |  |
| --- | --- |
| 1 | High Level Diagram |
| 2 | Steps to implement Workflow |
| 3 | Data Exploration |
| 4 | Machine Learning modelling |
| 5 | User Interface and LLM Development |
| 6 | Finding and Visualizations |
| 7 | Conclusion |
| 8 | References |

## High Level Diagram



## Steps to Implement Workflow

**Data Ingestion:** I have uploaded the raw data (synthetic\_logdata\_v2.csv) to my **Amazon S3** bucket (input directory). I have created an **Apache Airflow Job** that will run at scheduled specific time every day which will read the file from input directory and process the file in batch of 10000 lines and put it in the output directory of the same S3 bucket.

**Loading data into SQL Database:** I have made **AWS RDS (Postgres DB)** Instance for storing the transformed data. When any batch file comes into the output directory of the S3 bucket, The **Lambda Function** will trigger and the same will trigger the Glue Job which will perform the transformations on it and sends it to the Postgres DB table which is already made inside the Database. I have used the **JDBC connector** for connecting GLUE to Postgres.

**Transformations:**

This is how a link would look like

* List item 1
* List item 2
* List item 3

*“Quote - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.”*

## **Heading 2**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

### **Heading 3**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna.

#### **Heading 4**

Strong Body Text – 11px, regularDuis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat.



##### Heading 5

**Body Text – 11px:** Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

###### Heading 6

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.