WireFrame: Sentiment Analysis Pipeline Using HDFS and Spark

HDFS

S3-Bucket

Dump data from s3

CSV file

Jsonfile/

Json file

Read data from HDFS

Spark

Lambda Function

Token,Pre-Processing

Runs every 1 hour

HDFS

Sentiment-Analysis

Model- Training

Event Bridge

Write output to hdfs