

Architecture Diagram

Part 1

kaggle

Raw Data taken.



The raw dataset is unzipped in AWS EMR.



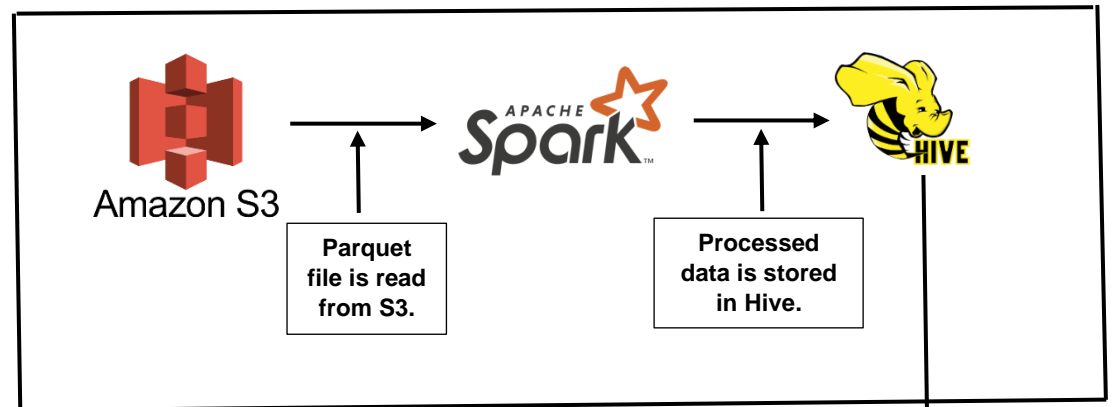
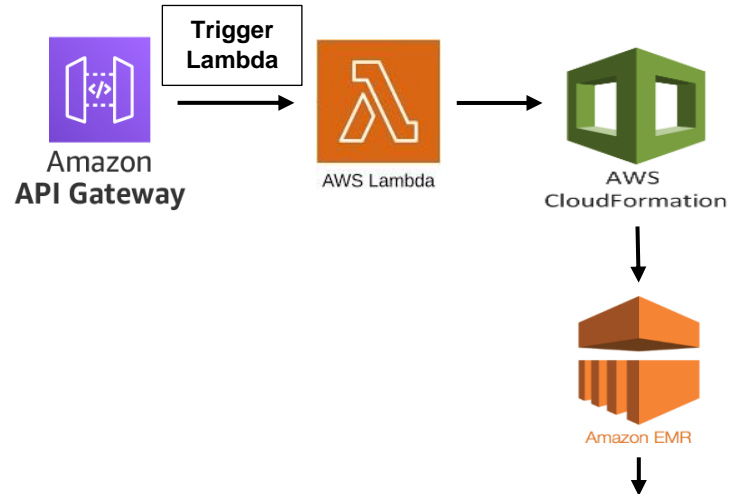
csv file is read in spark.



After importing and unzipping the data is stored in AWS S3.

csv file is formatted to parquet form and stored in S3.

Part 2 Automated



tableau

Tables stored in Hive are Visualized in Tableau by connecting To Amazon EMR Hadoop Hive.