Scripts Execution
Explanation of the solution to the batch layer problem

- In this project our goal is to detect the credit card fraud in
  the real time streaming transaction.

- This whole project is divided into two part. In this document
  we will focus on first half of the project i.e

  - Task 1: Load the transactions history data
    (card_transactions.csv) in a NoSQL database.

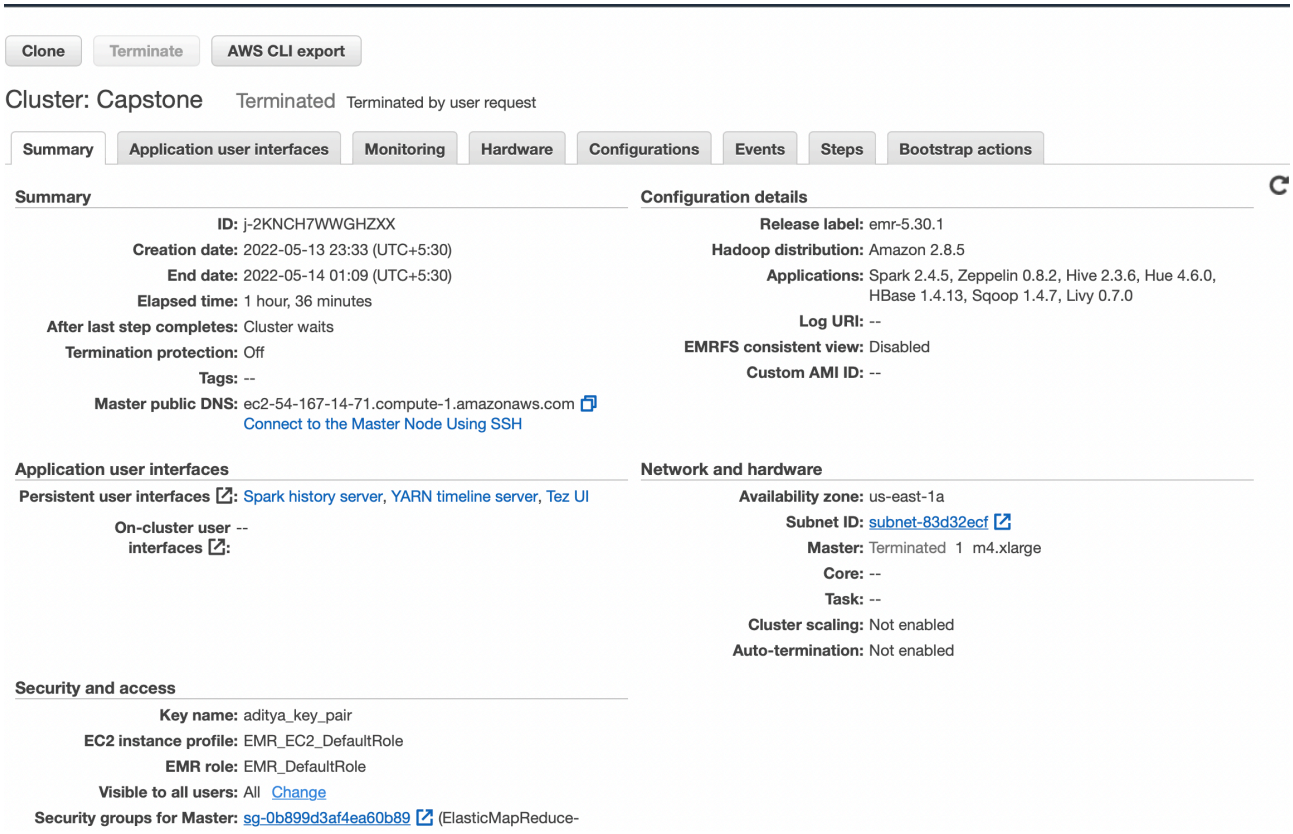
  - Task 2: Ingest the relevant data from AWS RDS to
    Hadoop.

  - Task 3: Create a look-up table with columns specified

earlier in the problem statement.

- Task 4: After creating the table, you need to load the
  relevant data in the lookup table.

- In order to perform the aforementioned tasks, I have created
  a single node EMR cluster with the below configuration.



• List of documents submitted in the zip: • Load NoSQL.pdf - Task 1
• Sqoop Data Ingestion.pdf - Task 2 • Create NoSQL.pdf - Task 3

• PreAnalysis.pdf - Task 4

• **TASK 1:**

- To load the data from csv file to Hbase (NoSQL database). I have uploaded the csv file to S3 bucket. Loc: s3://capstone-aditya/input/card_transactions.csv

- Then from S3, I have imported the file into hdfs using distcp command.

- After that I have opened the hive shell and created a table to hold the data from csv file and then created hive-hbase integrated table so that all data must be reflected in Hbase also.

- After importing data to hive-hbase integrated table, I used random UUID for

row key creation.

- Finally verified all the records imported successfully into table.

- All the commands and screenshot attached in the **Load NoSQL.pdf.**

- **TASK 2:**
  - In order to ingest data from AWS RDS to hdfs, I have
    used sqoop to import the data.

  - After importing the data into hdfs, I used hive to create
    and load the imported data into hive table.

  - Finally verified all the records imported successfully into

table.

- All the commands and screenshot attached in the
  Sqoop Data Ingestion.pdf.
  • **TASK 3:**
  • In order to create Lookup table I have used hive-base
  integrated table.
  • All the commands and screenshot attached in the
  Create NoSQL.pdf.
  • **TASK 4:**

- To load the data into Lookup table first we need to
  calculate some fields.

- I have created two intermediate table in hive, first one to
  hold the last 10 transactions and second one to
  calculated the UCL of last 10 transactions.

- Finally load the data into lookup table.

- Finally verified all the records imported successfully into
  table.

- All the commands and screenshot attached in the
  PreAnalysis.pdf.

• All the command and screenshot are consolidated in

Scripts Execution.pdf.

• This is the logic behind my way to complete task (till task 4).