Load data from AWS RDS to Hadoop

<Command to run the python file>

1. Create a python file to consume data from kafka

   vi datewise_bookings_aggregates_spark.py

2. Run spark submit command

   spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 datewise_bookings_aggregates_spark.py

<Command to move the csv file to HDFS>

1. Make a directory using mkdir command

   hadoop fs -mkdir datewise_aggregated_data

2. Loading the data from local file system to hadoop file system

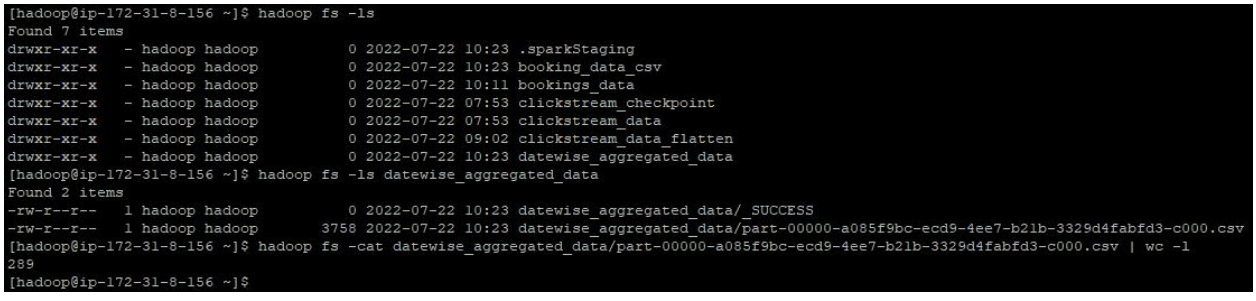
   hadoop fs- put ~/ datewise_aggregated_data datewise_aggregated_data

3. Checking the data file in hadoop

   hadoop fs -ls datewise_aggregated_data

   hadoop fs -cat datewise_aggregated_data / part-00000-a085f9bc-ecd9-4ee7-b21b-3329d4fabfd3-c000.csv | wc -l

<Screenshot of the file in HDFS>



<Screenshot of the booking_data in csv format in HDFS>