Load data from AWS RDS to Hadoop

<Command to run the python file>

1. CREATING AN EDITABLE PYTHON FILE

   vi datewise_bookings_aggregates_spark.py

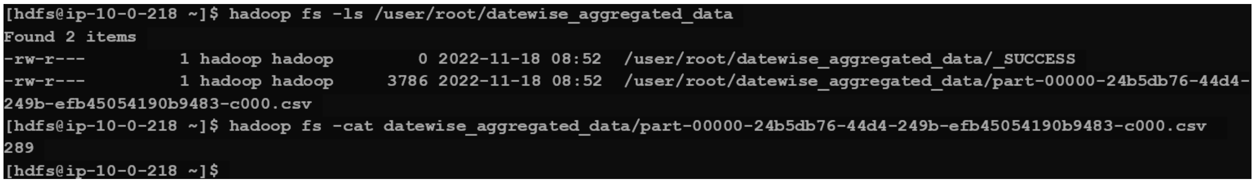
2. SPARK SUBMIT

   spark2-submit --jars "spark-sql-kafka-0-10_2.11-2.3.0.jar"
   datewise_bookings_aggregates_spark.py

<Command to move the csv file to HDFS>

hdfs dfs -put home/datewise_aggregated_data/user/root/clickstream_data_flatten

<Screenshot of the file in HDFS>



```
[hdfs@ip-10-0-218 ~]$ hadoop fs -ls /user/root/datewise_aggregated_data
Found 2 items
-rw-r---         1 hadoop hadoop        0 2022-11-18 08:52  /user/root/datewise_aggregated_data/_SUCCESS
-rw-r---         1 hadoop hadoop     3786 2022-11-18 08:52  /user/root/datewise_aggregated_data/part-00000-24b5db76-44d4-
249b-efb45054190b9483-c000.csv
[hdfs@ip-10-0-218 ~]$ hadoop fs -cat datewise_aggregated_data/part-00000-24b5db76-44d4-249b-efb45054190b9483-c000.csv
289
[hdfs@ip-10-0-218 ~]$
```