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-00000-20429a3a-dc5a-4539-9557-abbea1bf7616-c000.csv | wc -l

<Screenshot of the file in HDFS>

```
[hadoop@ip-172-31-70-125 ~]$ hadoop fs -ls datewise_aggregated_data
Found 2 items
-rw-r--r--   1 hadoop hadoop          0 2024-04-26 19:36 datewise_aggregated_data/_SUCCESS
-rw-r--r--   1 hadoop hadoop       3758 2024-04-26 19:36 datewise_aggregated_data/part-00000-20429a3a-dc5a-4539-9557-abbea1bf7616-c000.csv
[hadoop@ip-172-31-70-125 ~]$ hadoop fs -cat datewise_aggregated_data/part-00000-20429a3a-dc5a-4539-9557-abbea1bf7616-c000.csv | wc -l
289
```