## 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**

1. 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-b21b3329d4fabfd3-c000.csv | wc -l**

**<Screenshot of the file in HDFS>**

