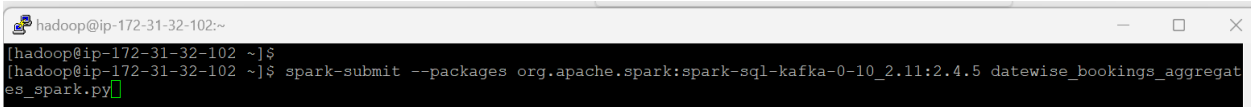
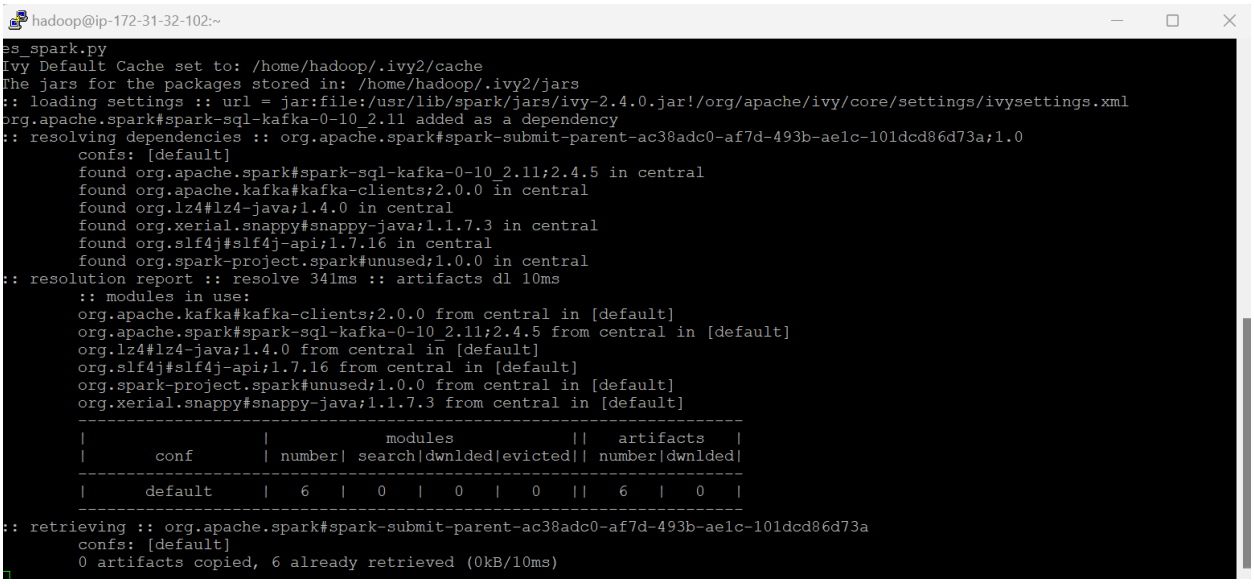
Load aggregate bookings data to Hadoop

1. Creating aggregates for finding date-wise total bookings using the Spark script 'datewise_bookings_aggregates_spark.py'

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





2. Command to move the csv file to HDFS

agg_df.coalesce(1).write.format('csv').mode('overwrite').save('/user/root/datewise_bookings_agg',header = 'true')

3. Screenshot of the files containing the aggregated bookings data in HDFS

hadoop fs -ls /user/root/datewise_bookings_agg/

hadoop fs -cat /user/root/datewise_bookings_agg/part-00000-18cdabab-83e6-4637-8841-cf5ccb3fae86-c000.csv| head -n 5