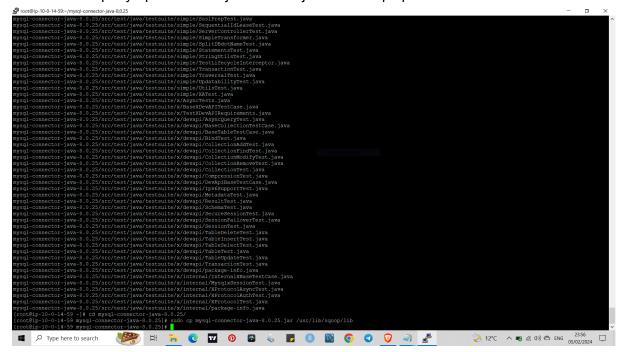
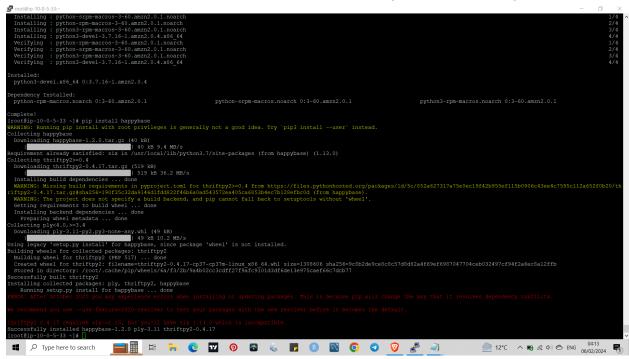
- 1. Login to our EMR instance, switch to the root user via **sudo -i** and after that connect the mysql connector using the commands which are given below:
- wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
- tar -xvf mysql-connector-java-8.0.25.tar.gz

```
| Monopolis | Notice | Notice
```

- cd mysql-connector-java-8.0.25/
- sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib



2. Switch to root user (sudo -i) and Install HBase using: pip install happybase



- 3. Go to HBase Shell using: hbase shell command.
- Create a table in HBase using: create 'trip_2017_hbase', 'cf1'
- 5. Use **list** command to check whether the table is created or not.
- Use describe command to describe the table.
- 7. exit from hbase shell
- 8. Switch to root user (**sudo -i**)
- 9. Using **sqoop import** command to ingest data from RDS into the HBase table.

```
sqoop import \
```

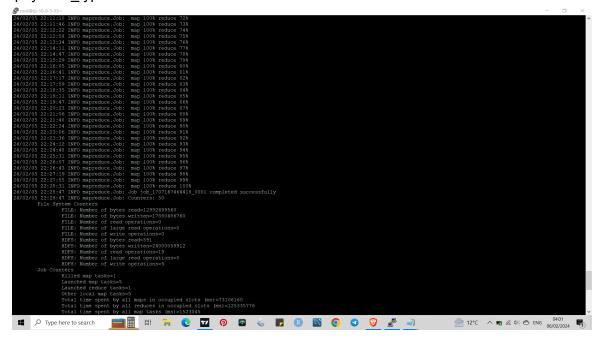
- --connect jdbc:mysql://mapreducedb.c5ekieu6m7o3.us-east-1.rds.amazonaws.com:3306/trip \
- --username harshuser \
- --password harsh123 \
- --table trip_2017 \
- --hbase-table trip_2017_hbase \
- --column-family cf1 \
- --hbase-create-table \
- --hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime \

- --hbase-bulkload \
- --split-by payment_type

Explanation of sqoop import command:

It is a "sqoop" command that imports data from a MySQL database table "trip_2017" into an HBase table "trip_2017 hbase".

- **--connect**: specifies the JDBC connection for the MySQL database.
- --username: specifies the username to use when connecting to the MySQL database.
- --password: specifies the password to use when connecting to the MySQL database.
- --table: specifies the name of the MySQL table to import data from.
- --hbase-table: specifies the name of the HBase table to import data into.
- **--column-family**: specifies the name of the column family in HBase where the imported data will be stored.
- --hbase-create-table: creates an HBase table if it does not exist.
- **--hbase-row-key**: specifies one or more columns from the MySQL table that will be used as the row key in HBase.
- --hbase-bulkload: uses HBase bulk load feature for faster data loading.
- --split-by: specifies a column from the MySQL table that will be used to split data into multiple HBase regions. In this command, data is imported from the MySQL table "trip_2017" into an HBase table named "trip_2017_hbase". The column family in HBase where the imported data will be stored is named "cf1". The row key in HBase is composed of two columns from the MySQL table: "tpep_pickup_datetime" and "tpep_dropoff_datetime". The data is loaded into HBase using a bulk load feature for faster loading. The data is split into multiple regions based on the column "payment_type".



Total time spent by all reduce tasks (ms)=195581

Total vorte: millisecond tasks (ms)=195581

Total vorte: millisecond tasks (ms)=195581

Total vorte: millisecond tasks (ms)=195581

Total supplyre: millisecond tasks (ms) all medical millisecond tasks (ms) all millisecond tasks (