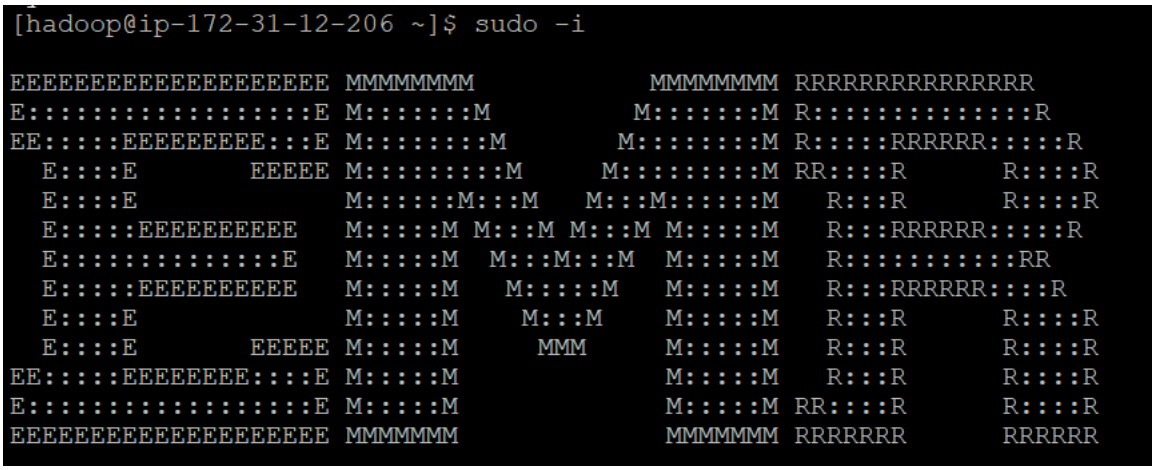
Task 2. Use Sqoop command to ingest the data from RDS into the HBase Table.

Solution: We followed the below steps to complete the above task:

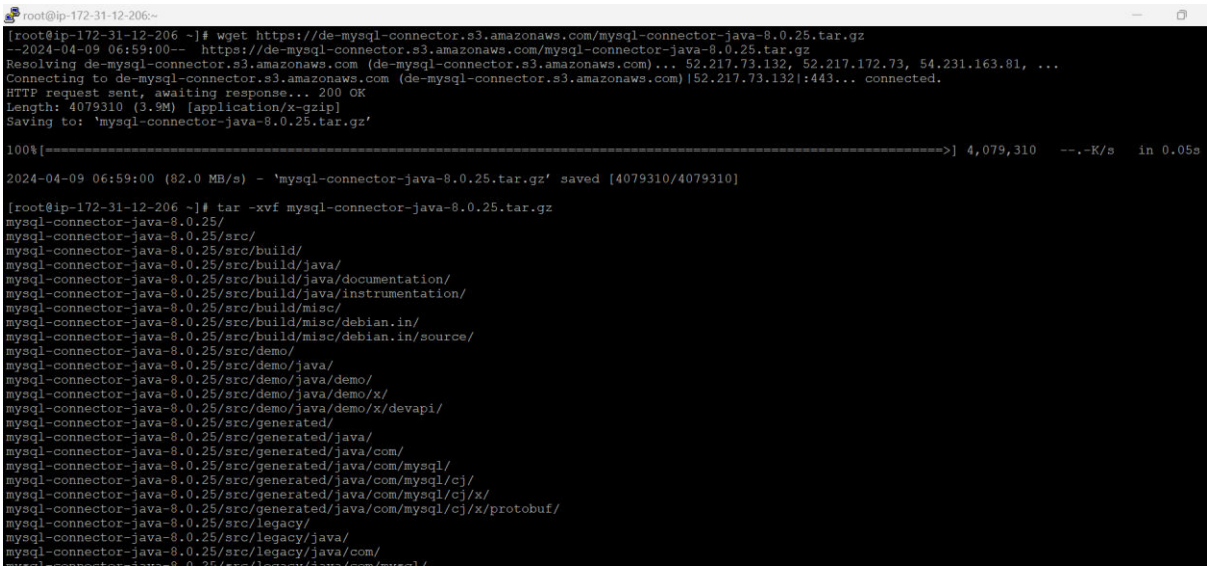
1. First, we logged into the EMR cluster and switch to root user by running **sudo -i** command.



2. Next, we need to run the below commands to install the MySQL connector jar file:

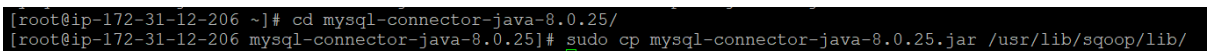
   wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz

   tar -xvf mysql-connector-java-8.0.25.tar.gz



   cd mysql-connector-java-8.0.25/

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

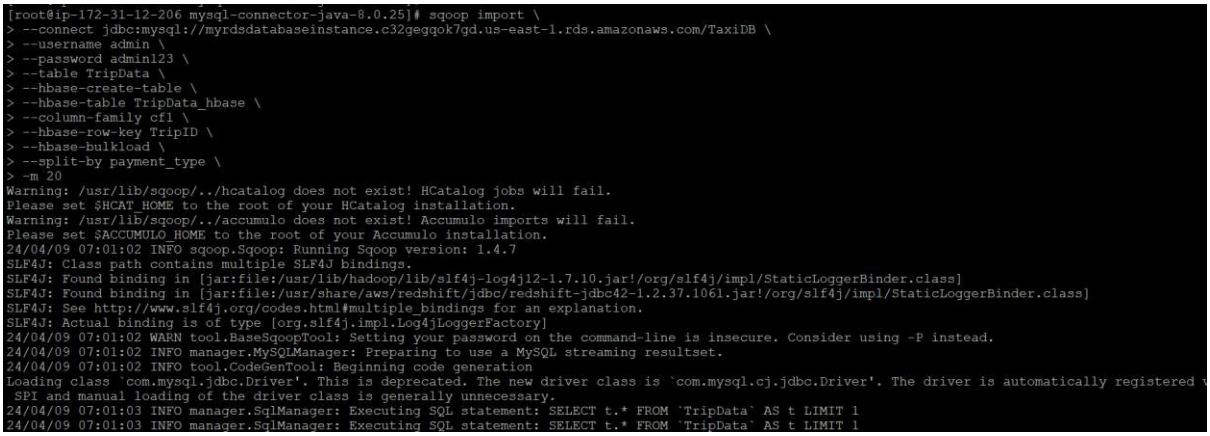


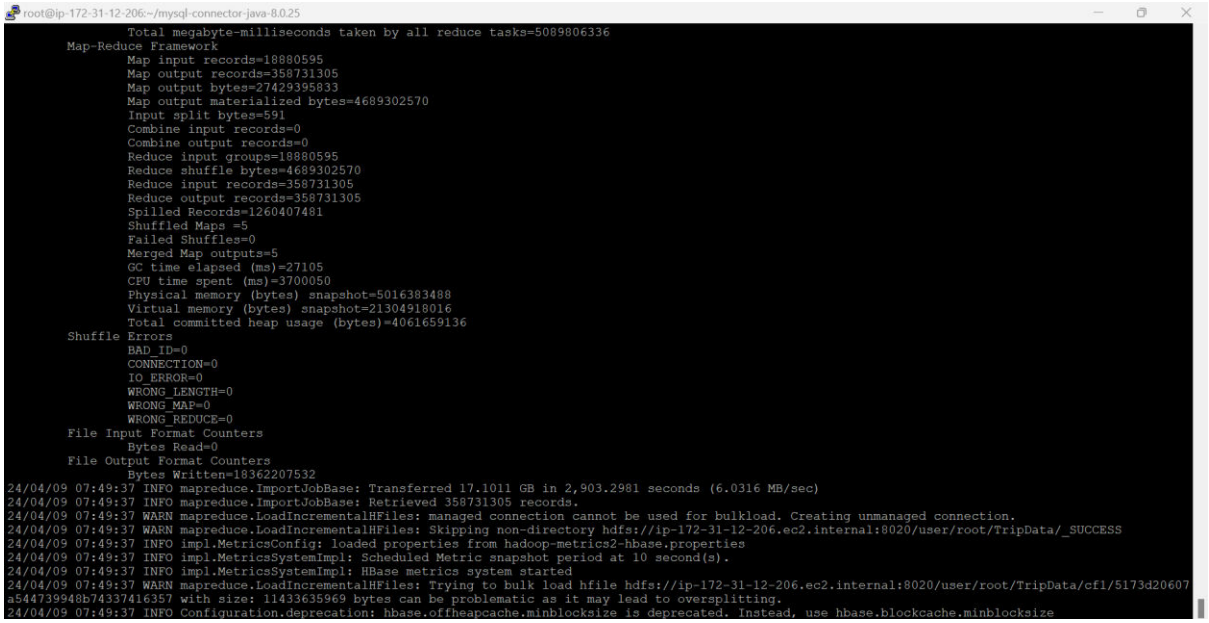
3. Next, for ingesting data from AWS RDS's MySQL database instance to HBase table, we ran the below command:

**sqoop import **
**--connect jdbc:mysql://myrdsdatabaseinstance.c32gegqok7gd.us-east-1.rds.amazonaws.com/TaxiDB **
**--username admin **
**--password admin123 **
**--table TripData **
**--hbase-create-table **
**--hbase-table TripData_hbase **
**--column-family cf1 **
**--hbase-row-key TripID **
**--hbase-bulkload **
**--split-by payment_type **
-m 20

Explanation for the above command is as follows:

- **sqoop import –** is a command used for importing data
- **--connect –** specifies the JDBC string of the MySQL database
- **--username –** specifies the username to connect to the MySQL database
- **--password** – specifies the password to connect to the MySQL database
- **--table –** specifies the MySQL table name from where the data will be imported
- **--hbase-create-table** – used to create the HBase table if not existing already
- **--hbase-table –** specifies the name of the HBase table into which the data will be imported to
- **--column-family cf1** – specifies the column family in HBase table
- **--hbase-row-key** – specifies the column which will the row key for the HBase table
- **--hbase-bulkload** – used for faster data loading when data volume is very large
- **--split-by** – specifies the column based on which the HBase regions will be created
- **-m 20** – specifies the number of mappers





4. Next, we entered the shell of HBase by entering **HBase shell** command.

```
[root@ip-172-31-12-206 mysql-connector-java-8.0.25]# hbase shell
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Fri Apr 17 15:18:24 UTC 2020
```

5. Next, we ran the command **list** to see the list of tables present (it should have the HBase table we created).

```
hbase(main):001:0> list
TABLE
TripData_hbase
1 row(s) in 0.2750 seconds

=> ["TripData_hbase"]
```

6. Next, we ran the command **describe 'TripData_hbase'** to see the table related details.

```
hbase(main):002:0> describe 'TripData_hbase'
Table TripData_hbase is ENABLED
TripData_hbase
COLUMN FAMILIES DESCRIPTION
{NAME => 'cf1', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER',
COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.1160 seconds
```

7. Next, we ran the command **count 'TripData_hbase'** to see how many rows are present based on the row key.