Task 2: Using Sqoop commands to ingest the data from RDS into the HBase Table.

1: Logging in EMR instances and completing the initial steps of setup.

command to install the MySQL connector jar file

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

Already executed below command above to extract the MySQL connector

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

• The MySQL Connector directory created in the previous step and then we copy it to the Sqoop library to complete the installation.

"cd mysql-connector-java-8.0.25/"

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

" Mysql_secure_installation ": To setup MariaDB

```
The Address Part | Part
```

2: Ingesting data from mySQL RDS to HBase table

Note: --hbase-create-table: creates an HBase table if it does not exist

```
sqoop import \
--connect jdbc:mysql://database-1.cn4wwwocm81j.us-east-1.rds.amazonaws.com/taxi_yellow
\
--username admin \
--password 7087507642 \
--table trips \
--hbase-table trip_records_hbase_new \
--column-family col1 \
--hbase-create-table \
--hbase-create-table \
--hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime \
--hbase-bulkload \
--split-by payment_type
```

Code Explanation:

let's break down the SQoop command step by step:

1. <u>sqoop import</u>: This is the command to initiate the data import process using SQoop.

2. --connect

jdbc:mysql://database-1.cn4wwwocm81j.us-east-1.rds.amazonaws.c om/taxi_yellow: This specifies the JDBC connection string to connect to the MySQL database located at the provided endpoint "database-1.cn4wwwocm81j.us-east-1.rds.amazonaws.com" and selects the "taxi yellow" database.

- 3. <u>--username admin</u>: Specifies the username to authenticate with the MySQL database, in this case, "admin".
- 4. <u>--password 7087507642</u>: Specifies the password for the provided username, "admin".
- 5. <u>--table trips</u>: Specifies the name of the MySQL table from which data will be imported, in this case, "trips".

- 6. <u>--hbase-table trip_records_hbase_new</u>: Specifies the name of the HBase table where the data will be imported, in this case, "trip_records_hbase_new".
- 7. <u>--column-family col1</u>: Specifies the column family in the HBase table where the data will be stored, in this case, "col1".
- 8. <u>--hbase-create-table</u>: Indicates that the HBase table specified (--hbase-table) should be created if it does not exist.
- 9. <u>--hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime:</u>
 Specifies the columns from the MySQL table that will be used as the row key in the HBase table. In this case, the row key will be composed of values from the columns "tpep_pickup_datetime" and "tpep_dropoff_datetime".
- 10. <u>--hbase-bulkload</u>: Indicates that the data will be loaded into HBase using bulk loading, which can improve performance for large datasets.
- 11. <u>--split-by payment_type</u>: Specifies the column used to split the import into multiple parallel tasks. In this case, the import will be split based on the "payment_type" column.

This command essentially connects to a MySQL database, selects a specific table, and imports its data into an HBase table with specified configurations.

```
trip_records_hbase \
ly colfam1 \
                           nbase-table trip_records_indse \
column-family colfan1 \
hbase-create-table \
hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime \
hbase-bulkload \
- mbase-creats-table \
- mbase-creats-table \
- mbase-creats-table \
- mbase-creats-table \
- mbase-bulkload \
- mbase-bulkload
  24/03/03 13:34:00 INFO db.IntegerSplitter: Split size: 1; Num splits: 4 from: 1 to: 5 24/03/03 13:34:00 INFO mapreduce.JobSubmitter: number of splits:5
                                       // 3 13:33:00 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/247509e59f32cb435f3a0ff0bac30e7e/trip_records.jar
// 3 13:33:00 WARN manager.MySQLManager: It looks like you are importing from mysql.
// 3 13:33:00 WARN manager.MySQLManager: pthis transfer can be faster! Use the -direct
// 3 13:33:00 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
// 3 13:33:00 INFO MARN manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
// 3 13:33:00 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.maps
// 3 13:33:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.ma
    4/03/03 13:34:00 INFO db.IntegerSplitter: Split size: 1; Num splits: 4 from: 1 to: 5
4/03/03 13:34:00 INFO mapreduce.JobSubmitter: number of splits:5
4/03/03 13:34:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job 1709466620659 0002
4/03/03 13:34:00 INFO impl.YarnClientImpl: Submitted application application 1709466620659 0002
4/03/03 13:34:00 INFO impl.YarnClientImpl: Submitted application application 1709466620659 0002
4/03/03 13:34:00 INFO mapreduce.Job: The url to track the job: http://ip-172-31-6-200.ec2.Internal:20888/proxy/application_1709466620659_0002/4/03/03 13:34:00 INFO mapreduce.Job: Tob job 1709466620659_0002 urnning in uber mode: false
4/03/03 13:34:09 INFO mapreduce.Job: map 0% reduce 0%
```

```
May input records—1889595

May input records—1889595

May output records—1889595

May output lipse=198979891

Imput spit speces in put records—18895951

Cambine input records—18895951

Cambine copit records—18895951

Cambine input records—18895951

Redon: smittle system=19895951

Redon
```

Checking in hBase:

Commands:

1: To get into hbase	hbase shell
2: Tables in hbase:	List
3: Count data copied:	count 'trip_records_hbase

```
### Service | Se
```

```
Durient count: 3422000, cov. 2017-01-12 15:22:23 2017-01-13 15:28:3

Durient count: 3422000, cov. 2017-01-12 15:26:00 2017-01-21 15:28:28

Current count: 3425000, rov. 2017-01-12 15:28:20 2017-01-12 15:28:28

Current count: 3425000, rov. 2017-01-12 15:33:33:30 2017-01-21 15:33:33:30

Current count: 3425000, rov. 2017-01-12 15:33:33:30 2017-01-21 15:33:33:30

Current count: 342000, rov. 2017-01-12 15:44:40 2017-01-12 15:33:30

Current count: 342000, rov. 2017-01-12 15:44:40 2017-01-12 16:06:48

Current count: 342000, rov. 2017-01-12 16:02:10 2017-01-12 16:06:10

Current count: 342000, rov. 2017-01-12 16:02:07 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01-12 16:02:10 2017-01
```