

Task 2:

We are using the same RDS instance and EMR instance created for task 1.
Hbase has been selected in the options while creating EMR instance.

We need to perform the below steps for Hbase:

- ➔ Installing gcc
- ➔ Installing Happybase

Commands:

```
sudo yum update
```

```
yum install gcc
```

```
sudo yum install python3-devel
```

```
pip install happybase
```

Step 1: Create table in Hbase using python script, table can be created from hbase shell as well.

Create_table.py

```
import happybase
#create connection
connection = happybase.Connection('localhost', port=9090 , autoconnect=False)

#open connection to perform operations

def open_connection():
    connection.open()

#close the opened connection

def close_connection():
    connection.close()

#list all tables in Hbase

def list_tables():
    print("fetching all table")
    open_connection()
```

```

tables = connection.tables()
close_connection()
print("all tables fetched")
return tables

```

#create a table by passing name and column family as a parameter

```

def create_table(name,cf):
    print("creating table " + name)
    tables = list_tables()
    if name not in tables:
        open_connection()
        connection.create_table(name, cf)
        close_connection()
        print("table created")
    else:
        print("table already present")

```

```
create_table('yellow_tripdata', {'info' : dict() })
```

table yellow_tripdata has been created with column family as info.

Step2:

Scoop ingestion from RDS to Hbase

```

sqoop import --connect jdbc:mysql://vg-rds-nyc-tlc.c5dapn7s4k1r.us-east-
1.rds.amazonaws.com:3306/nyc_tlc \
--username admin --password admin123 \
--table yellow_tripdata \
--columns
"rowkey,VendorID,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,t
rip_distance,RatecodeID,store_and_fwd_flag,PULocationID,DOLocationID,payment_t
ype,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,t
otal_amount,congestion_surcharge,airport_fee" \
--hbase-table yellow_tripdata \
--column-family info \
--hbase-row-key rowkey -m 1

```