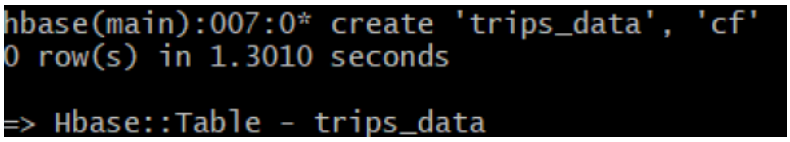
Task 2

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

1) Previous Task we have created the RDS Table and load data from csv files.

2) Now we will create HBase table and by using sqoop will import RDS table data to HBase table

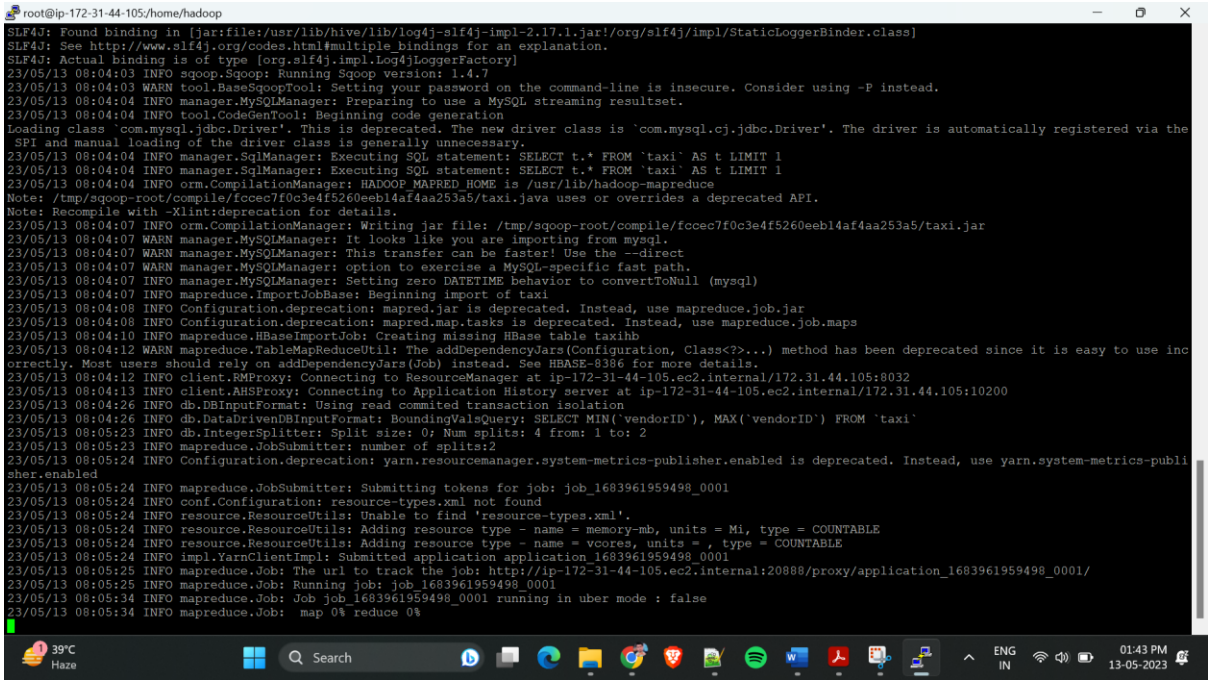
Create HBase Table – HBase table named trips_data' with a column family named 'cf'

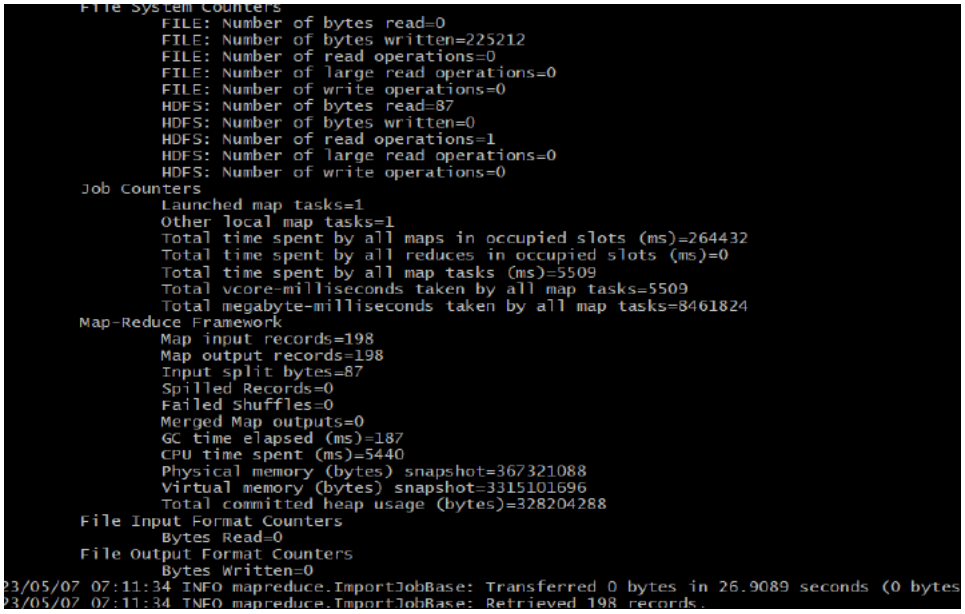


Sqoop import –

We provided RDS server connection details here to get Trips table records. And passed RDS/HBase table Name, Column-family and Row-key to insert the data into HBase table by using 4 mappers.

```
sqoop import --connect jdbc:mysql://case-study-dbb.ck4jzoqb1yn7.us-east-1.rds.amazonaws.com:3306/YellowTaxi \

--username admin \

--password user1234 \

--table taxi \

--hbase-table trips_data \

--column-family cf \

--split-by vendorID \

--num-mappers 4;
```





HBase Table Records After Sqoop import –

```
base(main):002:0> scan 'trips_data'
ROW                                                  COLUMN+CELL
1                                                    column=cf:Airport_fee, timestamp=1683443492125, value=0.0
1                                                    column=cf:DOLocationID, timestamp=1683443492125, value=236
1                                                    column=cf:PULocationID, timestamp=1683443492125, value=140
1                                                    column=cf:RatecodeID, timestamp=1683443492125, value=1
1                                                    column=cf:VendorID, timestamp=1683443492125, value=1
1                                                    column=cf:congestion_surcharge, timestamp=1683443492125, value=0.0
1                                                    column=cf:extra, timestamp=1683443492125, value=0.5
1                                                    column=cf:fare_amount, timestamp=1683443492125, value=6.5
1                                                    column=cf:improvement_surcharge, timestamp=1683443492125, value=0.3
1                                                    column=cf:mta_tax, timestamp=1683443492125, value=0.5
1                                                    column=cf:passenger_count, timestamp=1683443492125, value=1
1                                                    column=cf:payment_type, timestamp=1683443492125, value=2
1                                                    column=cf:store_and_fwd_flag, timestamp=1683443492125, value=N
1                                                    column=cf:tip_amount, timestamp=1683443492125, value=0.0
1                                                    column=cf:tolls_amount, timestamp=1683443492125, value=0.0
1                                                    column=cf:total_amount, timestamp=1683443492125, value=7.8
1                                                    column=cf:trip_distance, timestamp=1683443492125, value=1.2
10                                                   column=cf:Airport_fee, timestamp=1683443492125, value=0.0
10                                                   column=cf:DOLocationID, timestamp=1683443492125, value=48
10                                                   column=cf:PULocationID, timestamp=1683443492125, value=246
10                                                   column=cf:RatecodeID, timestamp=1683443492125, value=1
10                                                   column=cf:VendorID, timestamp=1683443492125, value=1
10                                                   column=cf:congestion_surcharge, timestamp=1683443492125, value=0.0
10                                                   column=cf:extra, timestamp=1683443492125, value=0.5
10                                                   column=cf:fare_amount, timestamp=1683443492125, value=12.0
10                                                   column=cf:improvement_surcharge, timestamp=1683443492125, value=0.3
10                                                   column=cf:mta_tax, timestamp=1683443492125, value=0.5
10                                                   column=cf:passenger_count, timestamp=1683443492125, value=1
10                                                   column=cf:payment_type, timestamp=1683443492125, value=2
10                                                   column=cf:store_and_fwd_flag, timestamp=1683443492125, value=N
10                                                   column=cf:tip_amount, timestamp=1683443492125, value=0.0
10                                                   column=cf:tolls_amount, timestamp=1683443492125, value=0.0
10                                                   column=cf:total_amount, timestamp=1683443492125, value=13.3
10                                                   column=cf:trip_distance, timestamp=1683443492125, value=1.0
100                                                  column=cf:Airport_fee, timestamp=1683443492125, value=0.0
100                                                  column=cf:DOLocationID, timestamp=1683443492125, value=162
100                                                  column=cf:PULocationID, timestamp=1683443492125, value=75
100                                                  column=cf:RatecodeID, timestamp=1683443492125, value=1
100                                                  column=cf:VendorID, timestamp=1683443492125, value=1
100                                                  column=cf:congestion_surcharge, timestamp=1683443492125, value=0.0
100                                                  column=cf:extra, timestamp=1683443492125, value=0.5
100                                                  column=cf:fare_amount, timestamp=1683443492125, value=9.5
100                                                  column=cf:improvement_surcharge, timestamp=1683443492125, value=0.3
100                                                  column=cf:mta_tax, timestamp=1683443492125, value=0.5
100                                                  column=cf:passenger_count, timestamp=1683443492125, value=1
```