Sqoop Import command used for importing a table from RDS to HDFS:

- Instal MysSQl connector jar to run Scoop

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

- Before importing the RDS dataset, a directory is created using the following command:

```
hadoop fs -mkdir -p /user/root/UPGRAD/ETL
```

- Using the scoop, Data is ingested from RDS to HDfs using the following:

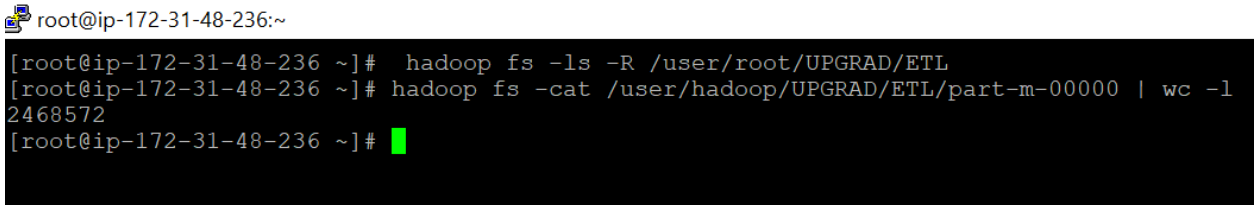
```
Sqoop import--connectjdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com:3306/testdatabase
--username student
--password STUDENT123
--table SRC_ATM_TRANS
--target-dir /user/hadoop/UPGRAD/ETL -m 1
```

The command used to see the list of imported data in HDFS:

- To count the files imported in the hdfs use the following command

```
<  hadoop fs -cat /user/hadoop/UPGRAD/ETL/part-m-00000 | wc -l >
```

- To know the top 3 records following command is used:



< hadoop fs -cat /user/hadoop/UPGRAD/ETL/part-m-00000 | head -n 3 >

```
root@ip-172-31-48-236:~
[root@ip-172-31-48-236 ~]# hadoop fs -cat /user/hadoop/UPGRAD/ETL/part-m-00000 | head -n 3
2017,January,1,Sunday,0,Active,1,NCR,NÃfÂ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,MasterCard,5643,Withdrawal,,,55.230,11.761,2616038,Naestved,281.150,1014
,87,7,260,0.215,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,MasterCard,1764,Withdrawal,,,57.048,9.935,2616235,NÃfÂ¸rresundby,280.640,
1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,VISA,1891,Withdrawal,,,57.048,9.935,2616235,NÃfÂ¸rresundby,280.640,1020,9
3,9,250,0.590,92,500,Rain,light rain
cat: Unable to write to output stream.
[root@ip-172-31-48-236 ~]#
```

Screenshot of the imported data:

```
root@ip-172-31-48-236:~
2024-01-22 16:40:39,371 INFO mapreduce.Job:  map 100% reduce 0%
2024-01-22 16:40:39,380 INFO mapreduce.Job: Job job_1705940468658_0001 completed successfully
2024-01-22 16:40:39,521 INFO mapreduce.Job: Counters: 33
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=298029
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=85
                HDFS: Number of bytes written=531214815
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=1212912
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=25269
                Total vcore-milliseconds taken by all map tasks=25269
                Total megabyte-milliseconds taken by all map tasks=38813184
        Map-Reduce Framework
                Map input records=2468572
                Map output records=2468572
                Input split bytes=85
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=229
                CPU time spent (ms)=28430
                Physical memory (bytes) snapshot=631095296
                Virtual memory (bytes) snapshot=3085852672
                Total committed heap usage (bytes)=595066880
                Peak Map Physical memory (bytes)=640131072
                Peak Map Virtual memory (bytes)=3094335488
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
2024-01-22 16:40:39,533 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 43.1468 seconds (11.7415 MB/sec)
2024-01-22 16:40:39,536 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-48-236 ~]#
```