Data Ingestion from RDS to HDFS using Sqoop

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

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student --password STUDENT123 \
--target-dir /user/root/bank_atm_project \
-m 1
```

```
[root@ip-172-31-88-50 ~]# sqoop import \
> --connect jdbc:mysql://upgraddetest.cyaielc9bmnf.us-east-1.rds.amazonaws.com/t
estdatabase \
> --table SRC_ATM_TRANS \
> --username student --password STUDENT123 \
> --target-dir /user/root/bank_atm_project \
> -m 1
```

```
22/11/19 03:10:13 INFO mapreduce.Job:  map 0% reduce 0%
22/11/19 03:10:40 INFO mapreduce.Job:  map 100% reduce 0%
22/11/19 03:10:40 INFO mapreduce.Job: Job job_1668826472141_0001 completed succe
ssfully
22/11/19 03:10:40 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=189689
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=531214815
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=1144992
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=23854
                Total vcore-milliseconds taken by all map tasks=23854
                Total megabyte-milliseconds taken by all map tasks=36639744
        Map-Reduce Framework
                Map input records=2468572
                Map output records=2468572
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=315
                CPU time spent (ms)=26480
                Physical memory (bytes) snapshot=625647616
                Virtual memory (bytes) snapshot=3286245376
                Total committed heap usage (bytes)=535298048
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=531214815
22/11/19 03:10:40 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 45.46
22 seconds (11.1435 MB/sec)
22/11/19 03:10:40 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-88-50 ~]#
```

2. Command used to see the list of imported data in HDFS:

hadoop fs -ls /user/root/bank_atm_project

```
[root@ip-172-31-88-50 ~]# hadoop fs -ls /user/root/bank_atm_project
Found 2 items
-rw-r--r--   1 root hadoop          0 2022-11-19 03:10 /user/root/bank_atm_project/_SUCCESS
-rw-r--r--   1 root hadoop  531214815 2022-11-19 03:10 /user/root/bank_atm_project/part-m-00000
```

3. Command used to view the imported data in HDFS:

hadoop fs -cat /user/root/bank_atm_project/part-m-00000

Screenshot of imported data

```
989,100,4,200,0.138,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,24,NCR,Hobro,Adelgade,31,9500,56.638,9.794,DKK,Mastercard - on-us,7780,Withdrawal,,,56.643,9.790,2620167,Hobro,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,20,NCR,Bispensgade,Bispensgade,35,9800,57.453,9.996,DKK,Mastercard - on-us,2830,Withdrawal,,,57.464,9.982,2620214,Hjorring,
278.589,998,99,8,210,0.000,44,802,Clouds,scattered clouds
2017,December,31,Sunday,23,Active,28,NCR,LÃfÃ¸gstÃfÃ¸r,ÃfËœsterbrogade,8,9670,56.967,9.253,DKK,VISA,812,Withdrawal,,,56.962,9.258,2617467,Logstor,277.589,999
,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,28,NCR,LÃfÃ¸gstÃfÃ¸r,ÃfËœsterbrogade,8,9670,56.967,9.253,DKK,VISA,5568,Withdrawal,,,56.962,9.258,2617467,Logstor,99
9,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃfÃ¸rrebro,18,9881,57.541,10.200,DKK,MasterCard,5403,Withdrawal,,,57.471,10.203,2614010,Sindal,277.589,999
,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,49,NCR,Bindslev,NÃfÃ¸rrebro,18,9881,57.541,10.200,DKK,MasterCard,312,Withdrawal,,,57.471,10.203,2614010,Sindal,277.589,999,
87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,56,Diebold Nixdorf,KÃfÃ¸ge,SÃfÃ¸ndre Alle,1,4600,55.454,12.181,DKK,Visa Dankort - on-us,5351,Withdrawal,,,55.588,12.251,262
1218,Greve Kommune,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,103,Diebold Nixdorf,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Mastercard - on-us,5821,Withdrawal,,,57.048,9.935,2616235,
NÃfÃ¸rresundby,277.589,999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,1,NCR,NÃfÃ¦stved,Farimagsvej,8,4700,55.233,11.763,DKK,Visa Dankort,3049,Withdrawal,,,55.230,11.761,2616038,Naestved,280.150
,991,100,5,210,1.175,92,500,Rain,light rain
2017,December,31,Sunday,23,Active,85,Diebold Nixdorf,KÃfÃ¸benhavn,Regnbuepladsen,5,1550,55.676,12.571,DKK,MasterCard,7785,Withdrawal,,,55.676,12.566,2618425,
Copenhagen,280.150,992,100,3,190,0.000,90,804,Clouds,overcast clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,Mastercard - on-us,9777,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,
999,87,6,208,0.000,76,803,Clouds,broken clouds
2017,December,31,Sunday,23,Active,15,NCR,Vestre,Kastetvej,36,9000,57.053,9.905,DKK,MasterCard,5531,Withdrawal,,,57.048,9.919,2624886,Aalborg,277.589,999,87,6
,208,0.000,76,803,Clouds,broken clouds
```