

## Data Ingestion from the RDS to HDFS using Sqoop

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

```
sqoop import \  
--connect jdbc:mysql://upgraddetest.cyaiele9bmnf.us-east-1.rds.amazonaws.com/testdatabase \  
--table SRC_ATM_TRANS \  
--username student \  
--password STUDENT123 \  
--target-dir /user/root/spar_node_bank_data\  
--delete-target-dir \  
--m 1
```

### Sqoop Import status

```
[hadoop@ip-172-31-17-138 ~]$ sqoop import \  
> --connect jdbc:mysql://upgraddetest.cyaiele9bmnf.us-east-1.rds.amazonaws.com/testdatabase \  
> --table SRC_ATM_TRANS \  
> --username student \  
> --password STUDENT123 \  
> --target-dir /user/root/spar_node_bank_data\  
> --delete-target-dir \  
> --m 1
```

```
File Output Format Counters  
  Bytes Written=531214815  
23/10/06 19:15:09 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 42.6975 seconds (11.865 MB/sec)  
23/10/06 19:15:09 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
```

### Command used to see the files in directory in HDFS:

hadoop fs -ls /user/root/spar\_node\_bank\_data--delete-target-dir

```
[hadoop@ip-172-31-17-138 ~]$ hadoop fs -ls /user/root/spar_node_bank_data--delete-target-dir  
Found 2 items  
-rw-r--r-- 1 hadoop hadoop 0 2023-10-06 19:15 /user/root/spar_node_bank_data--delete-target-dir/_SUCCESS  
-rw-r--r-- 1 hadoop hadoop 531214815 2023-10-06 19:15 /user/root/spar_node_bank_data--delete-target-dir/part-m-00000
```

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

hadoop fs -cat /user/root/spar\_node\_bank\_data--delete-target-dir/part-m-00000 | head -n 5

### Sample data – first 5 rows

```
[hadoop@ip-172-31-17-138 ~]$ hadoop fs -cat /user/root/spar_node_bank_data--delete-target-dir/part-m-00000 | head -n 5  
2017,January,1,Sunday,0,Active,1,NCR,NÅfÅ:stved,Farimagvej,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,NÅfÅ,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,NÅfÅ,rresundby,280.640,1020,9  
3,9,250,0.590,92,500,Rain,light rain  
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÅfÅVdhustrÅfÅ:det,12,7430,56.139,9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,  
6,240,0.000,75,300,Drizzle,light intensity drizzle  
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÅfÅ,nsager,1,4000,55.634,12.018,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,2614481,Roskilde,280.610,101  
4,87,7,260,0.000,88,701,Mist,mist
```

### Command to see count of total records loaded in HDFS

```
hadoop fs -cat /user/root/spar_node_bank_data--delete-target-dir/part-m-00000 | wc -l
```

```
[hadoop@ip-172-31-17-138 ~]$ hadoop fs -cat /user/root/spar_node_bank_data--delete-target-dir/part-m-00000 | wc -l  
2468572  
[hadoop@ip-172-31-17-138 ~]$
```