

## Data Ingestion from the RDS to HDFS using Sqoop

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

#### 1. Command to check if the directory is not present already:

```
hadoop fs -rm -r /user/root/SRC_ATM_TRANS
```

```
hadoop@ip-172-31-94-95:~$
[getfattr [-R] [-n name | -d] [-e en] <path>]
[getmerge [-nl] [-skip-empty-file] <src> <localdst>]
[help [cmd ...]]
[ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]]
[mkdir [-p] <path> ...]
[moveFromLocal <localsrc> ... <dst>]
[moveToLocal <src> <localdst>]
[mv <src> ... <dst>]
[put [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
[renameSnapshot <snapshotDir> <oldName> <newName>]
[rm [-f] [-r] [-R] [-skipTrash] [-safely] <src> ...]
[rmdir [--ignore-fail-on-non-empty] <dir> ...]
[setfacl [-R] [(t-b)k] {-m|-x <acl_spec>} <path>][--set <acl_spec> <pa
th>]]
[setfattr [-n name [-v value] | -x name] <path>]
[setrep [-R] [-w] <rep> <path> ...]
[stat [format] <path> ...]
[tail [-f] <file>]
[test -[defsz] <path>]
[test [-ignoreCrc] <src> ...]
[touchz <path> ...]
[truncate [-w] <length> <path> ...]
[usage [cmd ...]]

Generic options supported are
-conf <configuration file>      specify an application configuration file
-D <property=value>             use value for given property
-fs <file:///hdfs://namenode:port> specify default filesystem URL to use, overr
ides 'fs.defaultFS' property from configurations.
-jt <local|resourceManager:port> specify a ResourceManager
-files <comma separated list of files> specify comma separated files to be co
pied to the map reduce cluster
-libjars <comma separated list of jars> specify comma separated jar files to
include in the classpath.
-archives <comma separated list of archives> specify comma separated archives
to be unarchived on the compute machines.

The general command line syntax is
command [genericOptions] [commandOptions]

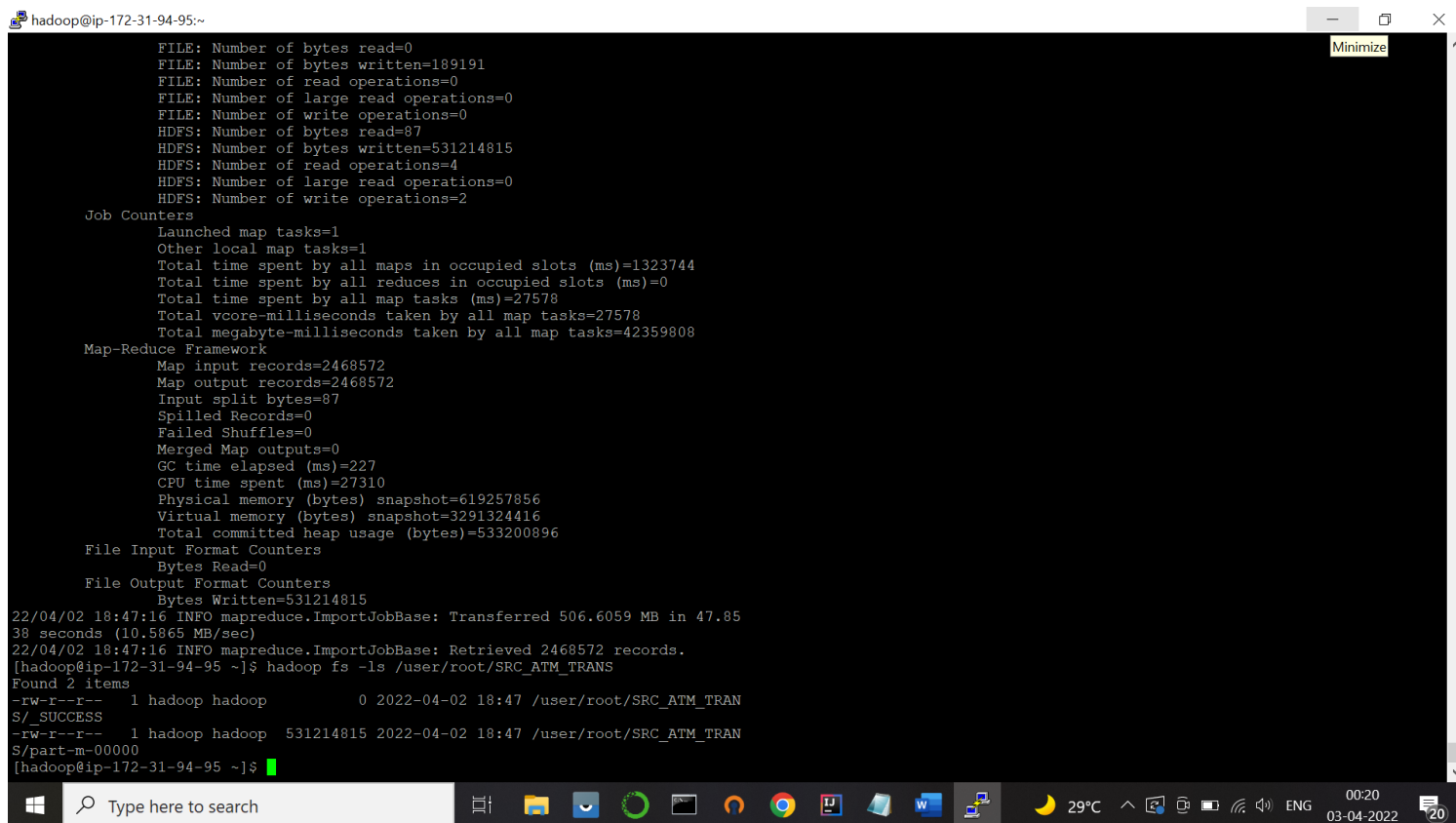
[hadoop@ip-172-31-94-95 ~]$ hadoop fs -rm -r /user/root/SRC_ATM_TRANS
rm: '/user/root/SRC_ATM_TRANS': No such file or directory
[hadoop@ip-172-31-94-95 ~]$ ^C
[hadoop@ip-172-31-94-95 ~]$
```

## 2. Sqoop import command:

```
sqoop import \
--connect jdbc:mysql://upgraddetest.cyaiecl9bmnf.useast1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS \
--m 1
```

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

```
hadoop fs -ls /user/root/SRC_ATM_TRANS
```



```
hadoop@ip-172-31-94-95:~$ sqoop import \
--connect jdbc:mysql://upgraddetest.cyaiecl9bmnf.useast1.rds.amazonaws.com/testdatabase \
--table SRC_ATM_TRANS \
--username student password STUDENT123 \
--target-dir /user/root/SRC_ATM_TRANS \
--m 1

FILE: Number of bytes read=0
FILE: Number of bytes written=189191
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)=1323744
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=27578
  Total vcore-milliseconds taken by all map tasks=27578
  Total megabyte-milliseconds taken by all map tasks=42359808

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)=227
  CPU time spent (ms)=27310
  Physical memory (bytes) snapshot=619257856
  Virtual memory (bytes) snapshot=3291324416
  Total committed heap usage (bytes)=533200896

File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=531214815
22/04/02 18:47:16 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 47.85
38 seconds (10.5865 MB/sec)
22/04/02 18:47:16 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[hadoop@ip-172-31-94-95 ~]$ hadoop fs -ls /user/root/SRC_ATM_TRANS
Found 2 items
-rw-r--r-- 1 hadoop hadoop 0 2022-04-02 18:47 /user/root/SRC_ATM_TRAN
S/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 531214815 2022-04-02 18:47 /user/root/SRC_ATM_TRAN
S/part-m-00000
[hadoop@ip-172-31-94-95 ~]$
```

## Screenshot of the imported data:

```
hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
```

## Screenshot of the imported data:

```
hadoop@ip-172-31-94-95:~$ cat /user/root/SRC_ATM_TRANS/part-m-00000
1,2621449,Gistrup,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,22,NCR,Roskilde,KÅfÅ,benhavnsvej,65,4000,55.642,12.106,DKK,Mastercard - on-us,328,Withdrawal,,,55.642,12.080,2614481,Roskil
de,276.764,1038,94,4,273,0.000,88,804,Clouds,overcast clouds
2017,January,24,Tuesday,18,Inactive,52,NCR,FarsÅfÅ,,Torvet,8,9640,56.771,9.340,DKK,MasterCard,1350,Withdrawal,,,56.773,9.339,2622310,Farso,276.114,1037,92,4,
250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Inactive,81,NCR,Spar KÅfÅ,bmand TornhÅfÅ,j,TornhÅfÅ,jvej,4,9220,57.026,10.002,DKK,Visa Dankort,5764,Withdrawal,,,56.994,9.991,2621
449,Gistrup,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,11,NCR,Sauersvej,Fridtjof Nansens Vej,2,9210,57.023,9.940,DKK,MasterCard,4772,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresun
db,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,22,NCR,Roskilde,KÅfÅ,benhavnsvej,65,4000,55.642,12.106,DKK,Visa Dankort,6568,Withdrawal,,,55.642,12.080,2614481,Roskilde,27
6.764,1038,94,4,273,0.000,88,804,Clouds,overcast clouds
2017,January,24,Tuesday,18,Inactive,73,NCR,HÅfÅ,jbjerg,Rosenvangsalle,194,8270,56.119,10.192,DKK,Visa Dankort - on-us,8525,Withdrawal,,,56.157,10.211,2624652
,Arhus,275.089,1036,92,4,252,0.000,12,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,Mastercard - on-us,2919,Withdrawal,,,56.715,10.117,2620952,Ha
dsund,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,6,NCR,Fredericia,SjÅfÅ,llandsgade,33,7000,55.564,9.757,DKK,VISA,9019,Withdrawal,,,55.566,9.753,2621951,Fredericia,275.764,1
034,96,3,250,0.000,36,802,Clouds,scattered clouds
2017,January,24,Tuesday,18,Active,41,Diebold Nixdorf,Skagen,Sct. Laurentiivej,36,9990,57.723,10.590,DKK,Visa Dankort - on-us,6340,Withdrawal,,,57.721,10.584,
2613939,Skagen,277.439,1031,95,3,259,0.000,92,804,Clouds,overcast clouds
2017,January,24,Tuesday,18,Active,1,NCR,NÅfÅ,stved,Farimagsvej,8,4700,55.233,11.763,DKK,Visa Dankort,4642,Withdrawal,,,55.230,11.761,2616038,Naestved,275.839
,1039,96,4,278,0.125,68,500,Rain,light rain
2017,January,24,Tuesday,18,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.716,10.114,DKK,Mastercard - on-us,4189,Withdrawal,,,56.715,10.117,2620952,Ha
dsund,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Inactive,2,NCR,Vejgaard,Hadsundvej,20,9000,57.043,9.950,DKK,Visa Dankort,7019,Withdrawal,,,57.048,9.935,2616235,NÅfÅ,rresundby,276
.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,10,NCR,NÅfÅ,rresundby,Torvet,6,9400,57.059,9.922,DKK,Visa Dankort,2527,Withdrawal,,,57.048,9.919,2624886,Aalborg,276.114,10
37,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,62,Diebold Nixdorf,Tørndrup,Bymidt,2,9575,56.815,10.057,DKK,Visa Dankort - on-us,1578,Withdrawal,,,56.715,10.117,2620952,
Hadsund,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,64,NCR,Horsens,GrÅfÅ,nlandsvej,5,8700,55.859,9.854,EUR,Mastercard - on-us,309,Withdrawal,,,55.861,9.850,2619771,Horsens,275
.764,1034,96,3,250,0.000,36,802,Clouds,scattered clouds
2017,January,24,Tuesday,18,Active,31,NCR,Slagelse,Mariendals Alle,29,4200,55.398,11.342,DKK,Visa Dankort,6813,Withdrawal,,,55.403,11.355,2613460,Slagelse,276
.114,1039,92,4,275,0.000,88,804,Clouds,overcast clouds
2017,January,24,Tuesday,18,Active,57,NCR,HillerÅfÅ,d,KÅfÅ,benhavnsvej,31,3400,55.933,12.314,DKK,MasterCard,3137,Withdrawal,,,55.767,12.433,2621394,Gladsakse
Kommune,276.114,1034,96,3,283,0.195,80,500,Rain,light rain
2017,January,24,Tuesday,18,Inactive,30,NCR,NykÅfÅ,bing Mors,Kirketorvet,1,7900,56.795,8.860,DKK,VISA,5245,Withdrawal,,,56.793,8.853,2615964,Nykobing Mors,277
.139,1037,88,4,248,0.000,68,803,Clouds,broken clouds
2017,January,24,Tuesday,18,Active,39,NCR,Svenstrup,GodthÅfÅ,bsvej,14,9230,56.973,9.851,DKK,Visa Dankort - on-us,6816,Withdrawal,,,56.972,9.848,2612021,Svenst
rup,276.114,1037,92,4,250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,10,NCR,NÅfÅ,rresundby,Torvet,6,9400,57.059,9.922,DKK,VISA,8345,Withdrawal,,,57.048,9.919,2624886,Aalborg,276.114,1037,92,4,
250,0.000,20,801,Clouds,few clouds
2017,January,24,Tuesday,18,Active,13,NCR,SÅfÅ,by,Vestergade,3,9300,57.334,10.515,DKK,Mastercard - on-us,1277,Withdrawal,,,57.441,10.537,2621927,Frederikshavn
,279.089,1037,98,7,263,0.000,44,802,Clouds,scattered clouds
2017,January,24,Tuesday,18,Active,35,NCR,Aabybro,ÅfÅ,estergade,6,
```