

## 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.cyaie1c9bmnf.us-east-1.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:

```
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS |head
cat: `/user/root/SRC_ATM_TRANS': Is a directory
```

```
[root@ip-172-31-83-37 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS |head
Found 2 items
-rw-r--r--  1 root hadoop      0 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/_SUCCESS
-rw-r--r--  1 root hadoop 531214815 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/part-m-00000
```

```
hadoop fs -ls /user/root/SRC_ATM_TRANS/part-m-00000 |head
```

```
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000 |head
```

### Screenshot of the imported data:

```
[root@ip-172-31-83-37 ~]# sqoop import --connect jdbc:mysql://upgradtest.cyaiei
c9bmnf.us-east-1.rds.amazonaws.com/testdatabase --table SRC_ATM_TRANS --username s
tudent --password STUDENT123 --target-dir /user/root/SRC_ATM_TRANS --m 1
Warning: /usr/lib/sqoop/./hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/lib/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/10/04 07:30:25 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/
org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1
.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
22/10/04 07:30:25 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
22/10/04 07:30:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
22/10/04 07:30:25 INFO tool.CodeGenTool: Beginning code generation
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class
is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SP
I and manual loading of the driver class is generally unnecessary.
22/10/04 07:30:26 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/10/04 07:30:26 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM 'SRC_ATM_TRANS' AS t LIMIT 1
22/10/04 07:30:26 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/ha
doop-mapreduce
Note: /tmp/sqoop-root/compile/d151a00715ab75e3e98c2bca58a7eb88/SRC_ATM_TRANS.jav
a uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/10/04 07:30:28 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root
/compile/d151a00715ab75e3e98c2bca58a7eb88/SRC_ATM_TRANS.jar
22/10/04 07:30:28 WARN manager.MySQLManager: It looks like you are importing fro
m mysql.
22/10/04 07:30:28 WARN manager.MySQLManager: This transfer can be faster! Use th
```

```
e --direct
22/10/04 07:30:28 WARN manager.MySQLManager: option to exercise a MySQL-specific
fast path.
22/10/04 07:30:28 INFO manager.MySQLManager: Setting zero DATETIME behavior to c
onvertToNull (mysql)
22/10/04 07:30:28 INFO mapreduce.ImportJobBase: Beginning import of SRC_ATM_TRAN
S
22/10/04 07:30:29 INFO Configuration.deprecation: mapred.jar is deprecated. Inst
ead, use mapreduce.job.jar
22/10/04 07:30:30 INFO Configuration.deprecation: mapred.map.tasks is deprecated
. Instead, use mapreduce.job.maps
22/10/04 07:30:30 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-3
1-83-37.ec2.internal/172.31.83.37:8032
22/10/04 07:30:36 INFO db.DBInputFormat: Using read committed transaction isolati
on
22/10/04 07:30:36 INFO mapreduce.JobSubmitter: number of splits:1
22/10/04 07:30:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
64866349974_0001
22/10/04 07:30:37 INFO impl.YarnClientImpl: Submitted application application_16
64866349974_0001
22/10/04 07:30:37 INFO mapreduce.Job: The url to track the job: http://ip-172-31
-83-37.ec2.internal:20888/proxy/application_1664866349974_0001/
22/10/04 07:30:37 INFO mapreduce.Job: Running job: job_1664866349974_0001
22/10/04 07:30:47 INFO mapreduce.Job: Job job_1664866349974_0001 running in uber
mode : false
22/10/04 07:30:47 INFO mapreduce.Job: map 0% reduce 0%
22/10/04 07:31:14 INFO mapreduce.Job: map 100% reduce 0%
22/10/04 07:31:15 INFO mapreduce.Job: Job job_1664866349974_0001 completed succe
ssfully
22/10/04 07:31:16 INFO mapreduce.Job: Counters: 30
    File System Counters
        FILE: Number of bytes read=0
        FILE: Number of bytes written=189352
        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)=1222320
        Total time spent by all reduces in occupied slots (ms)=0
        Total time spent by all map tasks (ms)=25465
        Total vcore-milliseconds taken by all map tasks=25465
        Total megabyte-milliseconds taken by all map tasks=39114240
    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)=218
CPU time spent (ms)=28660
Physical memory (bytes) snapshot=626712576
Virtual memory (bytes) snapshot=3349901312
Total committed heap usage (bytes)=534773760

File Input Format Counters
  Bytes Read=0

File Output Format Counters
  Bytes Written=531214815
22/10/04 07:31:16 INFO mapreduce.ImportJobBase: Transferred 506.6059 MB in 45.96
33 seconds (11.022 MB/sec)
22/10/04 07:31:16 INFO mapreduce.ImportJobBase: Retrieved 2468572 records.
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS |head cat: '/u
ser/root/SRC_TRANS':Is a directory'/user/root/SRC_TRANS':Is a directory
head: cannot open 'cat:' for reading: No such file or directory
head: cannot open '/user/root/SRC_TRANS:Is' for reading: No such file or directo
ry
head: cannot open 'a' for reading: No such file or directory
head: cannot open 'directory/user/root/SRC_TRANS:Is' for reading: No such file o
r directory
head: cannot open 'a' for reading: No such file or directory
head: cannot open 'directory' for reading: No such file or directory
cat: '/user/root/SRC_ATM_TRANS': Is a directory
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS |head
cat: '/user/root/SRC_ATM_TRANS': Is a directory
[root@ip-172-31-83-37 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS |head
Found 2 items
-rw-r--r--  1 root hadoop          0 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/
_SUCCESS
-rw-r--r--  1 root hadoop  531214815 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/
part-m-00000
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
|head
2017,January,1,Sunday,0,Active,1,NCR,NÃfÃ{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,NÃfÃ,rresundby,280.640,102
0,93,9,250,0.590,92,500,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,NÃfÃ,rresundby,280.640,1020,93,9
,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃfÃ¥dhusstrÃfÃ{det,12,7430,56.139,
9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,6,2

```

```
[root@ip-172-31-83-37 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS |head
Found 2 items
-rw-r--r-- 1 root hadoop          0 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/
_SUCCESS
-rw-r--r-- 1 root hadoop 531214815 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/
part-m-00000
[root@ip-172-31-83-37 ~]# hadoop fs -cat /user/root/SRC_ATM_TRANS/part-m-00000
|head
2017,January,1,Sunday,0,Active,1,NCR,NÃfÃ|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,NÃfÃ,rresundby,280.640,102
0,93,9,250,0.590,92,500,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,NÃfÃ,rresundby,280.640,1020,93,9
,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Inactive,3,NCR,Ikast,RÃfÃ¥dhusstrÃfÃ|det,12,7430,56.139,
9.154,DKK,VISA,4166,Withdrawal,,,56.139,9.158,2619426,Ikast,281.150,1011,100,6,2
40,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,4,NCR,Svogerslev,BrÃfÃ,nsager,1,4000,55.634,12.01
8,DKK,MasterCard,5153,Withdrawal,,,55.642,12.080,2614481,Roskilde,280.610,1014,8
7,7,260,0.000,88,701,Mist,mist
2017,January,1,Sunday,0,Active,5,NCR,Nibe,Torvet,1,9240,56.983,9.639,DKK,MasterC
ard,3269,Withdrawal,,,56.981,9.639,2616483,Nibe,280.640,1020,93,9,250,0.590,92,5
00,Rain,light rain
2017,January,1,Sunday,0,Active,6,NCR,Fredericia,SjÃfÃ|llandsgade,33,7000,55.564,
9.757,DKK,MasterCard,887,Withdrawal,,,55.566,9.753,2621951,Fredericia,281.150,10
14,93,7,230,0.290,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,7,Diebold Nixdorf,Hjallerup,Hjallerup Centret,18,
9320,57.168,10.148,DKK,Mastercard - on-us,4626,Withdrawal,,,57.165,10.146,262027
5,Hjallerup,280.640,1020,93,9,250,0.590,92,500,Rain,light rain
2017,January,1,Sunday,0,Active,8,NCR,GlyngÃfÃ,re,FÃfÃ|rgevej,1,7870,56.762,8.867
,DKK,MasterCard,470,Withdrawal,,,56.793,8.853,2615964,Nykobing Mors,281.150,1011
,100,6,240,0.000,75,300,Drizzle,light intensity drizzle
2017,January,1,Sunday,0,Active,9,Diebold Nixdorf,Hadsund,Storegade,12,9560,56.71
6,10.114,DKK,VISA,8473,Withdrawal,,,56.715,10.117,2620952,Hadsund,280.640,1020,9
3,9,250,0.590,92,500,Rain,light rain
cat: Unable to write to output stream.
[root@ip-172-31-83-37 ~]# hadoop fs -ls /user/root/SRC_ATM_TRANS/part-m-00000 |
head
-rw-r--r-- 1 root hadoop 531214815 2022-10-04 07:31 /user/root/SRC_ATM_TRANS/
part-m-00000
[root@ip-172-31-83-37 ~]#
```