

## Data Ingestion from the RDS to HDFS using Sqoop

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

```
sqoop import \  
--connect jdbc:mysql://upgradawsrds1.cyaieic9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \  
--username upgraduser \  
--password upgraduser \  
--table card_member \  
--target-dir /user/CCFD_project/card_member \  
-m 1
```

```
sqoop import \  
--connect jdbc:mysql://upgradawsrds1.cyaieic9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \  
--username upgraduser \  
--password upgraduser \  
--table member_score \  
--target-dir /user/CCFD_project/member_score \  
-m 1
```

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

```
hadoop fs -ls /user/CCFD_project/card_member
```

```
hadoop fs -ls /user/CCFD_project/member_score
```

## Screenshot of the imported data

```
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TransactionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/package-info.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/InternalXBaseTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/MysqlxSessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAsyncTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAuthTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/package-info.java
[hadoop@ip-172-31-26-104 ~]$
[hadoop@ip-172-31-26-104 ~]$ cd mysql-connector-java-8.0.25/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgradawsrds1.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table card_member \
> --target-dir /user/CCFD_project/card_member \
> -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.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]
23/02/13 21:42:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
23/02/13 21:42:00 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
```

Console Screenshot 1 : Importing AWS RDS Card Member into Hadoop

```
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)=198240
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4130
  Total vcore-milliseconds taken by all map tasks=4130
  Total megabyte-milliseconds taken by all map tasks=6343680
Map-Reduce Framework
  Map input records=999
  Map output records=999
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=91
  CPU time spent (ms)=2190
  Physical memory (bytes) snapshot=326868992
  Virtual memory (bytes) snapshot=3313852416
  Total committed heap usage (bytes)=233832448
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 18.2504 seconds (4.5526 KB/sec)
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 2 : 999 Records Received

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
Input split bytes=87
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=91
CPU time spent (ms)=2190
Physical memory (bytes) snapshot=326868992
Virtual memory (bytes) snapshot=3313852416
Total committed heap usage (bytes)=233832448

File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 18.2504 seconds (4.5526 KB/sec)
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgradawsrds1.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table member_score \
> --target-dir /user/CCFD_project/member_score \
> -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
```

Console Screenshot 3 : Importing AWS RDS Member Score into Hadoop

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
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)=197328
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4111
  Total vcore-milliseconds taken by all map tasks=4111
  Total megabyte-milliseconds taken by all map tasks=6314496
Map-Reduce Framework
  Map input records=999
  Map output records=999
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=91
  CPU time spent (ms)=1590
  Physical memory (bytes) snapshot=335671296
  Virtual memory (bytes) snapshot=3305467904
  Total committed heap usage (bytes)=326631424
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=19980
23/02/13 21:44:06 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 20.2668 seconds (985.8498 bytes/sec)
23/02/13 21:44:06 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 4 : 999 Records Received