

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-  
east1.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-  
east1.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

```
[hadoop@ip-172-31-8-198 ~]$ wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz  
--2025-09-20 10:18:19-- https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz  
Resolving de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)... 52.216.220.9, 52.217.228.7  
3, 3.5.16.61, ...  
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|52.216.220.9|:443... conn  
ected.  
HTTP request sent, awaiting response... 200 OK  
Length: 4079310 (3.9M) [application/x-gzip]  
Saving to: 'mysql-connector-java-8.0.25.tar.gz'  
  
100%[=====>] 4,079,310 --.-K/s in 0.02s  
  
2025-09-20 10:18:19 (161 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]
```

```
[hadoop@ip-172-31-8-198 ~]$ tar -xvf mysql-connector-java-8.0.25.tar.gz
mysql-connector-java-8.0.25/
mysql-connector-java-8.0.25/src/
mysql-connector-java-8.0.25/src/build/
mysql-connector-java-8.0.25/src/build/java/
mysql-connector-java-8.0.25/src/build/java/documentation/
mysql-connector-java-8.0.25/src/build/java/instrumentation/
mysql-connector-java-8.0.25/src/build/misc/
mysql-connector-java-8.0.25/src/build/misc/debian.in/
mysql-connector-java-8.0.25/src/build/misc/debian.in/source/
mysql-connector-java-8.0.25/src/demo/
mysql-connector-java-8.0.25/src/demo/java/
mysql-connector-java-8.0.25/src/demo/java/demo/
mysql-connector-java-8.0.25/src/demo/java/demo/x/
```

```
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableSelectTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableUpdateTest.java
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-172-31-8-198 ~]$
```

```
[hadoop@ip-172-31-8-198 ~]$ cd mysql-connector-java-8.0.25/
[hadoop@ip-172-31-8-198 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-172-31-8-198 mysql-connector-java-8.0.25]$ ls /usr/lib/sqoop/lib/
ant-contrib-1.0b3.jar          jackson-core-2.6.7.jar      parquet-column-1.6.0.jar
ant-eclipse-1.0-jvm1.2.jar    jackson-core-asl-1.9.13.jar  parquet-common-1.6.0.jar
avro-1.8.2.jar                jackson-databind-2.6.7.4.jar  parquet-encoding-1.6.0.jar
avro-mapred-1.8.2-hadoop2.jar jackson-mapper-asl-1.9.13.jar  parquet-format-2.2.0-rc1.jar
aws-glue-datacatalog-hive2-client.jar kite-data-core-1.1.0.jar      parquet-generator-1.6.0.jar
commons-codec-1.4.jar         kite-data-hive-1.1.0.jar      parquet-hadoop-1.6.0.jar
commons-compress-1.8.1.jar     kite-data-mapreduce-1.1.0.jar  parquet-jackson-1.6.0.jar
commons-io-1.4.jar            kite-hadoop-compatibility-1.1.0.jar postgresql-jdbc.jar
commons-jexl-2.1.1.jar        mariadb-connector-java.jar    RedshiftJDBC.jar
commons-lang3-3.4.jar         mysql-connector-java-8.0.25.jar slf4j-api-1.6.1.jar
commons-logging-1.1.1.jar     opencsv-2.3.jar              snappy-java-1.1.7.3.jar
hsqldb-1.8.0.10.jar           paranamer-2.7.jar            xz-1.5.jar
jackson-annotations-2.6.0.jar  parquet-avro-1.6.0.jar
[hadoop@ip-172-31-8-198 mysql-connector-java-8.0.25]$
```

1 : Importing AWS RDS Card Member into Hadoop

```
[hadoop@ip-172-31-8-198 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]
25/09/20 10:22:14 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
```

```

FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=85081
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)=9547776
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=3108
  Total vcore-milliseconds taken by all map tasks=3108
  Total megabyte-milliseconds taken by all map tasks=9547776
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)=72
  CPU time spent (ms)=1490
  Physical memory (bytes) snapshot=343183360
  Virtual memory (bytes) snapshot=4644302848
  Total committed heap usage (bytes)=317194240
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
25/09/20 10:22:35 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 17.7235 seconds (4.688 KB/sec)
25/09/20 10:22:35 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-8-198 mysql-connector-java-8.0.25]$
```

2 : 999 Records Received


```
[hadoop@ip-172-31-8-198 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.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
25/09/20 10:24:38 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
25/09/20 10:24:38 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
25/09/20 10:24:38 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

3 : Importing AWS RDS Member Score into Hadoop

```
FILE: Number of write operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=19980
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)=9424896
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=3068
  Total vcore-milliseconds taken by all map tasks=3068
  Total megabyte-milliseconds taken by all map tasks=9424896
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)=73
  CPU time spent (ms)=1440
  Physical memory (bytes) snapshot=336470016
  Virtual memory (bytes) snapshot=4638531584
  Total committed heap usage (bytes)=317718528
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=19980
25/09/20 10:24:56 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 15.1455 seconds (1.2883 KB/sec)
25/09/20 10:24:56 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-8-198 mysql-connector-java-8.0.25]$
```

4 : 999 Records Received