1.Run SQL connector

wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz

tar -xvf mysql-connector-java-8.0.25.tar.gz

cd mysql-connector-java-8.0.25/

sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/

2. creation of HBase table and Sqoop command to ingest data from RDS into the HBase table: -

sqoop import \

--connect "jdbc:mysql://mysqlassignment1.cuyf7nwfvt5s.us-east-1.rds.amazonaws.com:3306/demo" \

--username srbhrungta --password srbhrungta \

--table nyc_yellow_taxi \

--columns "VendorID,tpep_pickup_datetime,tpep_dropoff_datetime,Passenger_count,Trip_distance,RateCodeID,Store_and_fwd_flag,PULocationID,DOLocationID,Payment_type,\

Fare_amount,Extra,MTA_tax,Tip_amount,Tolls_amount,Improvement_surcharge,Total_amount" \

--hbase-create-table \

--hbase-table nyc_yellow_taxi \

--column-family Trip_details \

--hbase-row-key VendorID,tpep_pickup_datetime,tpep_dropoff_datetime \

-m 1

```
EEEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M::::::M         M::::::M R::::::::::::::R
EE:::::EEEEEEEEE::::E M:::::::M       M:::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M::::::::M     M::::::::M RR::::R      R::::R
  E::::E             M:::::::::M   M:::::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE    M::::::M:::M M:::M::::::M   R:::RRRRRR:::::R
  E::::::::::::::E    M:::::M M:::M:::M M:::::M   R:::::::::::RR
  E:::::EEEEEEEEEE    M:::::M  M:::::M  M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M   M:::M   M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M    MMM    M:::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M:::::M           M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M           M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEE MMMMMMM           MMMMMMM RRRRRRR      RRRRRR

[root@ip-172-31-35-158 ~]# sqoop import \
> --connect "jdbc:mysql://mysqlassignment1.cuyf7nwfvt5s.us-east-1.rds.amazonaws.com:3306/demo" \
> --username srbhrungta --password srbhrungta \
> --table nyc_yellow_taxi \
> --columns "VendorID,tpep_pickup_datetime,tpep_dropoff_datetime,Passenger_count,Trip_distance,RateCodeID,Store_and_fwd_flag,PULocationID,DOLocationID,Payment_type,\
> Fare_amount,Extra,MTA_tax,Tip_amount,Tolls_amount,Improvement_surcharge,Total_amount" \
> --hbase-create-table \
> --hbase-table nyc_yellow_taxi \
> --column-family Trip_details \
> --hbase-row-key VendorID,tpep_pickup_datetime,tpep_dropoff_datetime \
> -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.
23/01/22 14:07:54 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: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
23/01/22 14:07:55 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/01/22 14:07:55 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/01/22 14:07:55 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 SPI and ma
nual loading of the driver class is generally unnecessary.
23/01/22 14:07:56 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/22 14:07:56 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/22 14:07:56 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/5183b6371ccf7edd17ef532320c69314/nyc_yellow_taxi.java uses or overrides a deprecated API.
```

```
> --hbase-table nyc_yellow_taxi \
> --column-family Trip_details \
> --hbase-row-key VendorID,tpep_pickup_datetime,tpep_dropoff_datetime \
> -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.
23/01/22 14:07:54 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: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
23/01/22 14:07:55 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/01/22 14:07:55 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/01/22 14:07:55 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 SPI and ma
nual loading of the driver class is generally unnecessary.
23/01/22 14:07:56 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/22 14:07:56 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/22 14:07:56 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/5183b6371ccf7edd17ef532320c69314/nyc_yellow_taxi.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/01/22 14:07:59 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/5183b6371ccf7edd17ef532320c69314/nyc_yellow_taxi.jar
23/01/22 14:07:59 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/01/22 14:07:59 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/01/22 14:07:59 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/01/22 14:07:59 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/01/22 14:07:59 INFO mapreduce.ImportJobBase: Beginning import of nyc_yellow_taxi
23/01/22 14:08:00 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/01/22 14:08:00 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/01/22 14:08:02 INFO mapreduce.HBaseImportJob: Creating missing HBase table nyc_yellow_taxi
23/01/22 14:08:05 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. M
ost users should rely on addDependencyJars(Job) instead. See HBASE-8386 for more details.
23/01/22 14:08:05 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-35-158.ec2.internal/172.31.35.158:8032
23/01/22 14:08:10 INFO db.DBInputFormat: Using read commited transaction isolation
23/01/22 14:08:10 INFO mapreduce.JobSubmitter: number of splits:1
23/01/22 14:08:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1674396306124_0001
23/01/22 14:08:11 INFO impl.YarnClientImpl: Submitted application application_1674396306124_0001
23/01/22 14:08:11 INFO mapreduce.Job: The url to track the job: http://ip-172-31-35-158.ec2.internal:20888/proxy/application_1674396306124_0001/
23/01/22 14:08:11 INFO mapreduce.Job: Running job: job_1674396306124_0001
23/01/22 14:08:23 INFO mapreduce.Job: Job job_1674396306124_0001 running in uber mode : false
23/01/22 14:08:23 INFO mapreduce.Job:  map 0% reduce 0%
```

```
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.5041.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.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]
23/01/19 16:46:22 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/01/19 16:46:22 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/01/19 16:46:22 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 SPI and ma
nual loading of the driver class is generally unnecessary.
23/01/19 16:46:24 INFO manager.SqlManager: Executing SQL statements SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/19 16:46:24 INFO manager.SqlManager: Executing SQL statements SELECT t.* FROM `nyc_yellow_taxi` AS t LIMIT 1
23/01/19 16:46:24 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/1d4e93136935e7ad7b9932ad73a513e5/nyc_yellow_taxi.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/01/19 16:46:30 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/1d4e93136935e7ad7b9932ad73a513e5/nyc_yellow_taxi.jar
23/01/19 16:46:30 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/01/19 16:46:30 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/01/19 16:46:30 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/01/19 16:46:30 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/01/19 16:46:30 INFO mapreduce.ImportJobBase: Beginning import of nyc_yellow_taxi
23/01/19 16:46:30 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/01/19 16:46:30 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/01/19 16:46:30 INFO mapreduce.HBaseImportJob: Creating missing HBase table nyc_yellow_taxi
23/01/19 16:46:31 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. M
ost users should rely on addDependencyJars(Job) instead. See HBASE-8386 for more details.
23/01/19 16:46:36 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-45-200.ec2.internal/172.31.45.200:8032
23/01/19 16:46:44 INFO db.DBInputFormat: Using read committed transaction isolation
23/01/19 16:46:44 INFO mapreduce.JobSubmitter: number of splits:1
23/01/19 16:46:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1674146772261_0001
23/01/19 16:46:45 INFO impl.YarnClientImpl: Submitted application application_1674146772261_0001
23/01/19 16:46:45 INFO mapreduce.Job: The url to track the job: http://ip-172-31-45-200.ec2.internal:20888/proxy/application_1674146772261_0001/
23/01/19 16:46:45 INFO mapreduce.Job: Running job: job_1674146772261_0001
23/01/19 16:46:59 INFO mapreduce.Job: Job job_1674146772261_0001 running in uber mode : false
23/01/19 16:46:59 INFO mapreduce.Job:  map 0% reduce 0%
23/01/19 17:22:04 INFO mapreduce.Job:  map 100% reduce 0%
23/01/19 17:22:05 INFO mapreduce.Job: Job job_1674146772261_0001 completed successfully
23/01/19 17:22:06 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=224909
                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
```

```
23/01/19 16:46:45 INFO mapreduce.Job: Running job: job_1674146772261_0001
23/01/19 16:46:59 INFO mapreduce.Job: Job job_1674146772261_0001 running in uber mode : false
23/01/19 16:46:59 INFO mapreduce.Job:  map 0% reduce 0%
23/01/19 17:22:04 INFO mapreduce.Job:  map 100% reduce 0%
23/01/19 17:22:05 INFO mapreduce.Job: Job job_1674146772261_0001 completed successfully
23/01/19 17:22:06 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=224909
                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=0
                HDFS: Number of read operations=1
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=100939488
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=2102954
                Total vcore-milliseconds taken by all map tasks=2102954
                Total megabyte-milliseconds taken by all map tasks=3230043616
        Map-Reduce Framework
                Map input records=18880595
                Map output records=18880595
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=13125
                CPU time spent (ms)=942260
                Physical memory (bytes) snapshot=816644096
                Virtual memory (bytes) snapshot=3371978752
                Total committed heap usage (bytes)=674234368
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
23/01/19 17:22:06 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 2,130.3122 seconds (0 bytes/sec)
23/01/19 17:22:06 INFO mapreduce.ImportJobBase: Retrieved 18880595 records.
[root@ip-172-31-45-200 ~]#
```