

1. Login to our EMR instance, switch to the root user via **sudo -i** and after that connect the mysql connector using the commands which are given below:

- **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**

```
root@ip-10-0-14-59:~/mysql-connector-java-8.0.25
[hadop@ip-10-0-14-59 ~]$ sudo -i
EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRRRRRRRRR
E:|||||EEEEEEEEEEEEEEEE M:|||||M M:|||||M R:|||||EEEEEEEEEEEE
RR:|||||EEEEEEEEEEEEEEEE M:|||||M M:|||||M R:|||||RRRRRRRRRRRR
E:|||||E EEEEE M:|||||M M:|||||M RR:|||||R R:|||||R
E:|||||E M:|||||M:M:|||||M M:|||||M R:|||||R R:|||||R
E:|||||EEEEEEEEEEEE M:|||||M M:|||||M R:|||||RRRRRRRRRRRR
E:|||||EEEEEEEEEEEE M:|||||M M:|||||M R:|||||RRRRRRRRRRRR
E:|||||EEEEEEEEEEEE M:|||||M M:|||||M R:|||||RRRRRRRRRRRR
E:|||||E M:|||||M M:|||||M M:|||||M R:|||||R R:|||||R
E:|||||E EEEEE M:|||||M M:|||||M R:|||||R R:|||||R
RR:|||||EEEEEEEEEEEEEEEE M:|||||M M:|||||M R:|||||R R:|||||R
R:|||||EEEEEEEEEEEEEEEE M:|||||M M:|||||M R:|||||R R:|||||R
EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR

[root@ip-10-0-14-59 ~]# wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
--2024-02-05 18:24:10-- 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)... 16.182.72.161, 52.216.42.153, 52.216.43.161, ...
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|16.182.72.161|:443... connected.
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 --R/s in 0.06s

2024-02-05 18:24:10 (61.6 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]

[root@ip-10-0-14-59 ~]# 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/demo/java/demo/x/devapi/
mysql-connector-java-8.0.25/src/generated/
mysql-connector-java-8.0.25/src/generated/java/
mysql-connector-java-8.0.25/src/generated/java/com/
mysql-connector-java-8.0.25/src/generated/java/com/mysql/
mysql-connector-java-8.0.25/src/generated/java/com/mysql/cj/
mysql-connector-java-8.0.25/src/generated/java/com/mysql/cj/x/
mysql-connector-java-8.0.25/src/generated/java/com/mysql/cj/x/protobuf/
```

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

```
root@ip-10-0-14-59:~/mysql-connector-java-8.0.25
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/SaslPrepTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/SequentialIdCaseTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/ServerControllerTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/SimpleTransformer.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/SplitDBDotNameTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/StatementTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/StringUtilTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/TestLifecycleInterceptor.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/TransactionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/TraversalTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/UpdateabilityTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/UtilTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/simple/XATest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/AsyncTests.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/BaseDevAPITestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/TestXDevAPIRequirements.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/AsyncQueryTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/BaseCollectionTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/BaseTableTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/BindTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionAddTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionFindTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionModifyTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionRemoveTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CompressionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/DevAPIBaseTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/IPv6SupportTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/MetadadataTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/ResultTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SchemaTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SecureSessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SessionFailoverTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableDeleteTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableInsertTest.java
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/MySQLSessionTest.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
[roo@ip-10-0-14-59 ~]# cd mysql-connector-java-8.0.25/
[roo@ip-10-0-14-59 ~]# sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib
[roo@ip-10-0-14-59 ~]#
```

2. Switch to root user (**sudo -i**) and Install HBase using: **pip install happybase**

```
root@ip-10-0-5-33:~#
Installing : python3-rpm-macros-3-60.amzn2.0.1.noarch 1/4
Installing : python3-rpm-macros-3-60.amzn2.0.1.noarch 2/4
Installing : python3-rpm-macros-3-60.amzn2.0.1.noarch 3/4
Installing : python3-devel-3.7.16-1.amzn2.0.4.x86_64 4/4
Verifying : python3-rpm-macros-3-60.amzn2.0.1.noarch 1/4
Verifying : python3-rpm-macros-3-60.amzn2.0.1.noarch 2/4
Verifying : python3-rpm-macros-3-60.amzn2.0.1.noarch 3/4
Verifying : python3-devel-3.7.16-1.amzn2.0.4.x86_64 4/4

Installed:
python3-devel.x86_64 0:3.7.16-1.amzn2.0.4

Dependency Installed:
python3-rpm-macros.noarch 0:3-60.amzn2.0.1 python3-rpm-macros.noarch 0:3-60.amzn2.0.1 python3-rpm-macros.noarch 0:3-60.amzn2.0.1

Complete!
[root@ip-10-0-5-33 ~]# pip install happybase
WARNING: Running pip install with root privileges is generally not a good idea. Try 'pip3 install --user' instead.
Collecting happybase
  Downloading happybase-1.2.0.tar.gz (40 kB)
    |#####| 40 kB 9.4 MB/s
Requirement already satisfied: six in /usr/local/lib/python3.7/site-packages (from happybase) (1.13.0)
Collecting thriftypy2>=0.4
  Downloading thriftypy2-0.4.17.tar.gz (519 kB)
    |#####| 519 kB 36.2 MB/s
Installing build dependencies ... done
WARNING: Missing build requirements in pyproject.toml for thriftypy2>=0.4 from https://files.pythonhosted.org/packages/1d/5c/852a627317a75e0ec19f42b955ef115b0906c43ee4c7595c112a652f0b20/thriftypy2-0.4.17.tar.gz#sha256=190f35c32da9146d1fdd822f46b6a0ad543572ea405ca6853b4ec7b128efbc0d (from happybase).
WARNING: the project does not specify a build backend, and pip cannot fall back to setuptools without 'wheel'.
Getting requirements to build wheel ... done
Installing backend dependencies ... done
Preparing wheel metadata ... done
Collecting ply<4.0,>=3.4
  Downloading ply-3.11-py2.py3-none-any.whl (49 kB)
    |#####| 49 kB 10.2 MB/s
Using legacy 'setup.py install' for happybase, since package 'wheel' is not installed.
Building wheels for collected packages: thriftypy2
  Building wheel for thriftypy2 (PEP 517) ... done
  Created wheel for thriftypy2: filename=thriftypy2-0.4.17-cp37m-linux_x86_64.whl size=1308606 sha256=9c8b2de9ce8c8c57d8d82a4f69ef6987047704ceb032497cf94f2a6ec8a12ffb
  Stored in directory: /root/.cache/pip/wheels/6a/f3/2b/9a4b02cc3cdf27f9a9c9101d3df6de13e975caef66c7dcb77
Successfully built thriftypy2
Installing collected packages: ply, thriftypy2, happybase
  Running setup.py install for happybase ... done
ERROR: After October 2020 you may experience errors when installing or updating packages. This is because pip will change the way that it resolves dependency conflicts.
We recommend you use --use-feature=2020-resolver to test your packages with the new resolver before it becomes the default.
thriftypy2 0.4.17 requires six<=1.15, but you'll have six 1.13.0 which is incompatible.
Successfully installed happybase-1.2.0 ply-3.11 thriftypy2-0.4.17
[root@ip-10-0-5-33 ~]#
```

3. Go to HBase Shell using: **hbase shell** command.

4. Create a table in HBase using: **create 'trip_2017_hbase', 'cf1'**

5. Use **list** command to check whether the table is created or not.

6. Use **describe** command to describe the table.

7. **exit** from hbase shell

8. Switch to root user (**sudo -i**)

9. Using **sqoop import** command to ingest data from RDS into the HBase table.

```
sqoop import \
--connect jdbc:mysql://mapreducedb.c5ekieu6m7o3.us-east-1.rds.amazonaws.com:3306/trip \
--username harshuser \
--password harsh123 \
--table trip_2017 \
--hbase-table trip_2017_hbase \
--column-family cf1 \
--hbase-create-table \
--hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime \
```

--hbase-bulkload \
--split-by payment_type

Explanation of sqoop import command:

It is a “sqoop” command that imports data from a MySQL database table “trip_2017” into an HBase table “trip_2017_hbase”.

--connect: specifies the JDBC connection for the MySQL database.
--username: specifies the username to use when connecting to the MySQL database.
--password: specifies the password to use when connecting to the MySQL database.
--table: specifies the name of the MySQL table to import data from.
--hbase-table: specifies the name of the HBase table to import data into.
--column-family: specifies the name of the column family in HBase where the imported data will be stored.
--hbase-create-table: creates an HBase table if it does not exist.
--hbase-row-key: specifies one or more columns from the MySQL table that will be used as the row key in HBase.
--hbase-bulkload: uses HBase bulk load feature for faster data loading.
--split-by: specifies a column from the MySQL table that will be used to split data into multiple HBase regions. In this command, data is imported from the MySQL table “trip_2017” into an HBase table named “trip_2017_hbase”. The column family in HBase where the imported data will be stored is named “cf1”. The row key in HBase is composed of two columns from the MySQL table: “tpep_pickup_datetime” and “tpep_dropoff_datetime”. The data is loaded into HBase using a bulk load feature for faster loading. The data is split into multiple regions based on the column “payment_type”.

```
root@ip-10-0-5-33:~#
24/02/05 22:11:10 INFO mapreduce.Job: map 100% reduce 72%
24/02/05 22:11:46 INFO mapreduce.Job: map 100% reduce 73%
24/02/05 22:12:22 INFO mapreduce.Job: map 100% reduce 74%
24/02/05 22:12:50 INFO mapreduce.Job: map 100% reduce 75%
24/02/05 22:13:34 INFO mapreduce.Job: map 100% reduce 76%
24/02/05 22:14:11 INFO mapreduce.Job: map 100% reduce 77%
24/02/05 22:14:47 INFO mapreduce.Job: map 100% reduce 78%
24/02/05 22:15:29 INFO mapreduce.Job: map 100% reduce 79%
24/02/05 22:16:05 INFO mapreduce.Job: map 100% reduce 80%
24/02/05 22:16:41 INFO mapreduce.Job: map 100% reduce 81%
24/02/05 22:17:17 INFO mapreduce.Job: map 100% reduce 82%
24/02/05 22:17:59 INFO mapreduce.Job: map 100% reduce 83%
24/02/05 22:18:35 INFO mapreduce.Job: map 100% reduce 84%
24/02/05 22:19:11 INFO mapreduce.Job: map 100% reduce 85%
24/02/05 22:19:47 INFO mapreduce.Job: map 100% reduce 86%
24/02/05 22:20:23 INFO mapreduce.Job: map 100% reduce 87%
24/02/05 22:21:06 INFO mapreduce.Job: map 100% reduce 88%
24/02/05 22:21:48 INFO mapreduce.Job: map 100% reduce 89%
24/02/05 22:22:24 INFO mapreduce.Job: map 100% reduce 90%
24/02/05 22:23:06 INFO mapreduce.Job: map 100% reduce 91%
24/02/05 22:23:36 INFO mapreduce.Job: map 100% reduce 92%
24/02/05 22:24:12 INFO mapreduce.Job: map 100% reduce 93%
24/02/05 22:24:48 INFO mapreduce.Job: map 100% reduce 94%
24/02/05 22:25:31 INFO mapreduce.Job: map 100% reduce 95%
24/02/05 22:26:07 INFO mapreduce.Job: map 100% reduce 96%
24/02/05 22:26:43 INFO mapreduce.Job: map 100% reduce 97%
24/02/05 22:27:19 INFO mapreduce.Job: map 100% reduce 98%
24/02/05 22:27:55 INFO mapreduce.Job: map 100% reduce 99%
24/02/05 22:28:31 INFO mapreduce.Job: map 100% reduce 100%
24/02/05 22:28:47 INFO mapreduce.Job: Job job_1707167464418_0001 completed successfully
24/02/05 22:28:47 INFO mapreduce.Job: Counters: 50
File System Counters
  FILE: Number of bytes read=12992489560
  FILE: Number of bytes written=17880486760
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=591
  HDFS: Number of bytes written=24000059912
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=5
Job Counters
  Killed map tasks=1
  Launched map tasks=5
  Launched reduce tasks=1
  Other local map tasks=5
  Total time spent by all maps in occupied slots (ms)=73106160
  Total time spent by all reducers in occupied slots (ms)=125335776
  Total time spent by all map tasks (ms)=1523045
```

```
root@ip-10-0-5-33:~#
Total time spent by all reduce tasks (ms)=1305581
Total vcore-milliseconds taken by all map tasks=1523045
Total vcore-milliseconds taken by all reduce tasks=1305581
Total megabyte-milliseconds taken by all map tasks=2339397120
Total megabyte-milliseconds taken by all reduce tasks=4010744832
Map-Reduce Framework
  Map input records=18880595
  Map output records=302089520
  Map output bytes=41200198406
  Map output materialized bytes=4886618403
  Input split bytes=591
  Combine input records=0
  Combine output records=0
  Reduce input groups=18842048
  Reduce shuffle bytes=4886618403
  Reduce input records=302089520
  Reduce output records=301472768
  Spilled Records=110473907
  Shuffled Maps =5
  Failed Shuffles=0
  Merged Map outputs=5
  GC time elapsed (ms)=13560
  CPU time spent (ms)=2838820
  Physical memory (bytes) snapshot=4761481216
  Virtual memory (bytes) snapshot=21303877632
  Total committed heap usage (bytes)=3735027712
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=24000059912
24/02/05 22:28:47 INFO mapreduce.ImportJobBase: Transferred 22.3518 GB in 2,283.5919 seconds (10.0229 MB/sec)
24/02/05 22:28:47 INFO mapreduce.ImportJobBase: Retrieved 302089520 records.
24/02/05 22:28:47 WARN mapreduce.LoadIncrementalHFiles: managed connection cannot be used for bulkload. Creating unmanaged connection.
24/02/05 22:28:47 WARN mapreduce.LoadIncrementalHFiles: Skipping non-directory hdfs://ip-10-0-5-33.ec2.internal:8020/user/root/trip_2017/_SUCCESS
24/02/05 22:28:47 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2-hbase.properties
24/02/05 22:28:47 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
24/02/05 22:28:47 INFO impl.MetricsSystemImpl: HBase metrics system started
24/02/05 22:28:47 WARN mapreduce.LoadIncrementalHFiles: Trying to bulk load hfile hdfs://ip-10-0-5-33.ec2.internal:8020/user/root/trip_2017/cf1/3d94e6b0d0b54f91bbbe1e128eedae9c with size: 1190767393 bytes can be problematic as it may lead to oversplitting.
24/02/05 22:28:48 WARN mapreduce.LoadIncrementalHFiles: Trying to bulk load hfile hdfs://ip-10-0-5-33.ec2.internal:8020/user/root/trip_2017/cf1/e2c1a91cd32c485b9a18cd4053bca9e4 with size: 1190767397 bytes can be problematic as it may lead to oversplitting.
24/02/05 22:28:48 INFO Configuration.deprecation: hbase.offheapcache.minblocksize is deprecated. Instead, use hbase.blockcache.minblocksize
root@ip-10-0-5-33 ~#
```