Task 2. Use Sqoop command to ingest the data from RDS into the HBase Table.

Solution: Here's the step by step procedure for achieving the above task:

Step1: We will first download mysql connector and unzip the file.

tar -xzvf mysql-connector-j-9.1.0.tar.gz

```
[hadoop@ip-172-31-32-108 ~]$ sudo wget https://cdn.mysql.com//Downloads/Connector-J/mysql-connector-j-9.1.0.tar.gz
--2025-01-07 05:28:51--  https://cdn.mysql.com//Downloads/Connector-J/mysql-connector-j-9.1.0.tar.gz
Resolving cdn.mysql.com (cdn.mysql.com)... 184.25.42.21, 2600:1408:c400:188d::1d68, 2600:1408:c400:1884::1d68
Connecting to cdn.mysql.com (cdn.mysql.com)|184.25.42.21|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4485702 (4.3M) [application/x-tar-gz]
Saving to: 'mysql-connector-j-9.1.0.tar.gz'

mysql-connector-j-9.1.0.tar.gz                100%[=========================================================================================>]   4.28M  --.-KB/s    in 0.1s

2025-01-07 05:28:51 (34.3 MB/s) - 'mysql-connector-j-9.1.0.tar.gz' saved [4485702/4485702]

[hadoop@ip-172-31-32-108 ~]$ tar -xzvf mysql-connector-j-9.1.0.tar.gz
mysql-connector-j-9.1.0/
mysql-connector-j-9.1.0/src/
mysql-connector-j-9.1.0/src/build/
mysql-connector-j-9.1.0/src/build/java/
mysql-connector-j-9.1.0/src/build/java/documentation/
mysql-connector-j-9.1.0/src/build/java/instrumentation/
mysql-connector-j-9.1.0/src/build/misc/
mysql-connector-j-9.1.0/src/build/misc/debian.in/
mysql-connector-j-9.1.0/src/build/misc/debian.in/source/
mysql-connector-j-9.1.0/src/demo/
mysql-connector-j-9.1.0/src/demo/java/
mysql-connector-j-9.1.0/src/demo/java/demo/
mysql-connector-j-9.1.0/src/demo/java/demo/x/
mysql-connector-j-9.1.0/src/demo/java/demo/x/devapi/
mysql-connector-j-9.1.0/src/generated/
mysql-connector-j-9.1.0/src/generated/java/
mysql-connector-j-9.1.0/src/generated/java/com/
mysql-connector-j-9.1.0/src/generated/java/com/mysql/
mysql-connector-j-9.1.0/src/generated/java/com/mysql/cj/
```

Step 2: sqoop list-databases --connect jdbc:mysql://database-1.crkohtdqsjvo.us-east-1.rds.amazonaws.com --username admin --password *******

```
[hadoop@ip-172-31-32-108 ~]$ sqoop list-databases --connect jdbc:mysql://database-1.crkohtdqsjvo.us-east-1.rds.amazonaws.com --username admin --password [redacted]
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/log4j-slf4j-impl-2.17.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
2025-01-07 05:53:09,366 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2025-01-07 05:53:09,404 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2025-01-07 05:53:09,534 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the d
river class is generally unnecessary.
mysql
information_schema
performance_schema
sys
assignment
[hadoop@ip-172-31-32-108 ~]$
```

Step 3: sqoop import --connect jdbc:mysql://database-1.crkohtdqsjvo.us-east-1.rds.amazonaws.com/assignment --username admin --password ******* --table TaxiTripData --target-dir /TaxiTripDataHbase --m 1

```
Try --help for usage instructions.
[hadoop@ip-172-31-32-108 ~]$ sqoop import --connect jdbc:mysql://database-1.crkohtdqsjvo.us-east-1.rds.amazonaws.com/assignment --username admin --password [redacted] --table TaxiTripDa
-target-dir /TaxiTripDataHbase --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-reload4j-1.7.36.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: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/log4j-slf4j-impl-2.17.2.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.Reload4jLoggerFactory]
2025-01-07 06:02:05,986 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2025-01-07 06:02:06,028 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2025-01-07 06:02:06,189 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2025-01-07 06:02:06,189 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 manual loading of t
river class is generally unnecessary.
2025-01-07 06:02:07,118 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TaxiTripData` AS t LIMIT 1
2025-01-07 06:02:07,197 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `TaxiTripData` AS t LIMIT 1
2025-01-07 06:02:07,218 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2025-01-07 06:02:13,861 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/d5899caf4a53970835b47c2ee400f855/TaxiTripData.jar
2025-01-07 06:02:13,880 WARN manager.MySQLManager: It looks like you are importing from mysql.
2025-01-07 06:02:13,880 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2025-01-07 06:02:13,880 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2025-01-07 06:02:13,881 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
```

Step 4: Now we can see that sqoop has ingested the data from RDS into Hbase table.