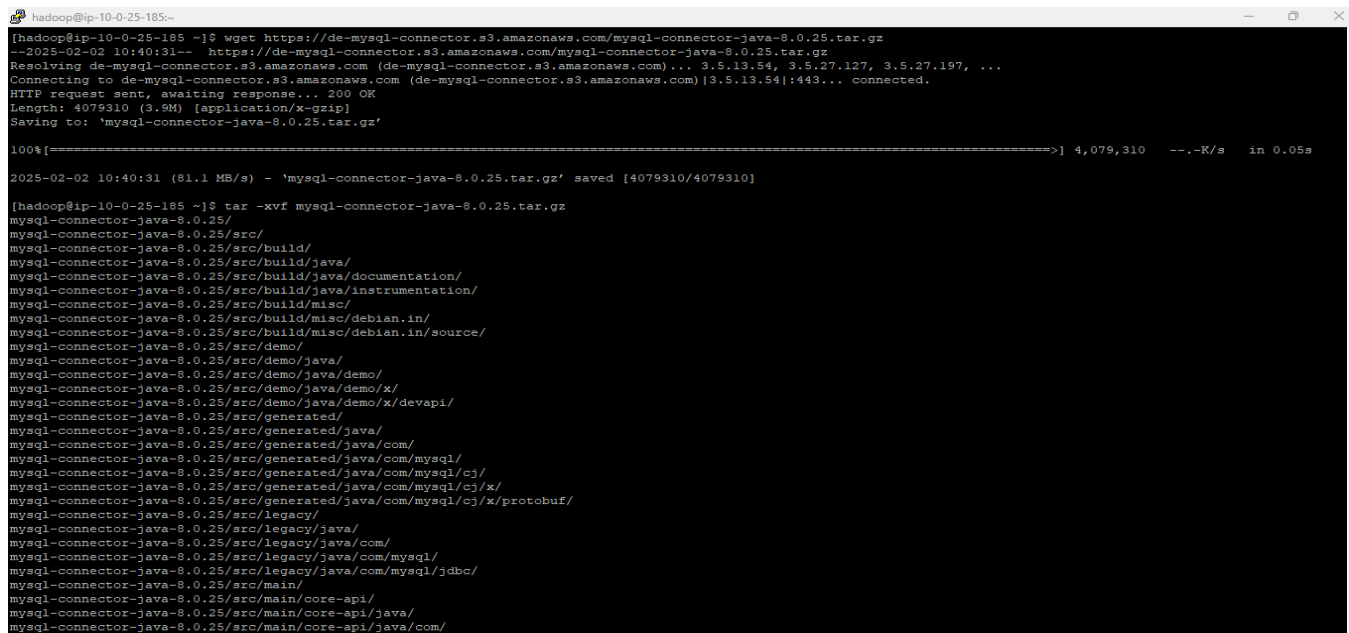


Task 02

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

A. Get the SQL connector and unzip it:

```
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
```

A terminal window titled 'hadoop@ip-10-0-25-185--' shows the execution of two commands. First, 'wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz' is run, which successfully downloads a 3.9MB file. Second, 'tar -xvf mysql-connector-java-8.0.25.tar.gz' is run, which extracts the contents of the tarball. The extraction process lists various files and directories, including 'src/build/java', 'src/build/java/documentation', 'src/build/misc', 'src/build/misc/debian.in', 'src/demo', 'src/demo/java', 'src/demo/java/demo', 'src/demo/java/demo/x', 'src/demo/java/demo/x/devapi', 'src/generated', 'src/generated/java', 'src/generated/java/com', 'src/generated/java/com/mysql', 'src/generated/java/com/mysql/cj', 'src/generated/java/com/mysql/cj/x', 'src/generated/java/com/mysql/cj/x/protobuf', 'src/legacy', 'src/legacy/java', 'src/legacy/java/com', 'src/legacy/java/com/mysql', 'src/legacy/java/com/mysql/jdbc', 'src/main', 'src/main/core-api', 'src/main/core-api/java', and 'src/main/core-api/java/com'.

```
[hadoop@ip-10-0-25-185 ~]$ wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
--2025-02-02 10:40:31-- 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)... 3.5.13.54, 3.5.27.127, 3.5.27.197, ...
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|3.5.13.54|: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 --K/s in 0.05s

2025-02-02 10:40:31 (81.1 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]

[hadoop@ip-10-0-25-185 ~]$ 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/
mysql-connector-java-8.0.25/src/legacy/
mysql-connector-java-8.0.25/src/legacy/java/
mysql-connector-java-8.0.25/src/legacy/java/com/
mysql-connector-java-8.0.25/src/legacy/java/com/mysql/
mysql-connector-java-8.0.25/src/legacy/java/com/mysql/jdbc/
mysql-connector-java-8.0.25/src/main/
mysql-connector-java-8.0.25/src/main/core-api/
mysql-connector-java-8.0.25/src/main/core-api/java/
mysql-connector-java-8.0.25/src/main/core-api/java/com/
```

B. Go to the directory and run the following commands:

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

```

build.xml  CHANGES  INFO  BIN  INFO  SRC  LICENSE  mysql-connector-java-8.0.25.jar  README  src
[hadoop@ip-10-0-22-234 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-10-0-22-234 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://mydb2.cske577dghlr.us-east-1.rds.amazonaws.com:3306/ride_data \
> --username admin \
> --password Sarshad070121 \
> --table ride_log \
> --hbase-table ride_log_hbase \
> --column-family colfam1 \
> --hbase-create-table \
> --hbase-row-key pickup_datetime,dropoff_datetime \
> --hbase-bulkload \
> --split-by payment_type
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/slf4j-reload4j-1.7.33.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]
2025-02-02 15:24:01,294 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2025-02-02 15:24:01,332 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2025-02-02 15:24:01,489 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.

```

- C. One we run the above commands we are ready to import the sqoop import from MySQL to Hbase:

```

sqoop import \

--connect jdbc:mysql://RDS endpoint:3306/ride_data \

--username admin \

--password Password\

--table ride_log \

--hbase-table ride_log_hbase \

--column-family colfam1 \

--hbase-create-table \

--hbase-row-key pickup_datetime,dropoff_datetime \

--hbase-bulkload \

--split-by payment_type

```

Explanation:

--connect: Indicates the JDBC URL for the MySQL database connection.

jdbc:mysql://...: The hostname, port (336), and database name (ride_data) are all included in the connection string.

--username & --password: Enter login information for the MySQL database.

--table: Specifies the import of the ride_log table from MySQL.

--hbase-table: The name of the HBase table (ride_log_hbase) into which the data will be imported.

The HBase column family (colfam1) is defined by the --column-family option. Data is arranged into column families in HBase.

If the ride_log_hbase table in HBase does not already exist, --hbase-create-table tells Sqoop to build it.

--hbase-row-key: Indicates the pickup_datetime and dropoff_datetime column combination to be used as HBase's unique row key. Every row in HBase is guaranteed to be uniquely recognized by this combination.

Bulk loading, which is more effective for huge datasets, is enabled with --hbase-bulkload. It prepares HFiles and loads them straight into HBase rather than adding data row by row.

--split-by: Specifies the payment_type column that will be used to divide the data for processing in parallel. By enabling Sqoop to execute many mappers concurrently, each handling a data chunk based on payment_type, this enhances performance.

```
2025-02-03 19:45:42,490 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2025-02-03 19:45:42,541 INFO mapreduce.ImportJobBase: Beginning import of ride_log
2025-02-03 19:45:42,659 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2025-02-03 19:45:42,700 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2025-02-03 19:45:44,846 INFO mapreduce.HBaseImportJob: Creating missing HBase table ride_log_hbase
2025-02-03 19:45:45,717 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use
tly. Most users should rely on addDependencyJars(Job) instead. See HBASE-8386 for more details.
2025-02-03 19:45:45,718 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use
tly. Most users should rely on addDependencyJars(Job) instead. See HBASE-8386 for more details.
2025-02-03 19:45:46,581 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
2025-02-03 19:45:46,582 INFO compress.CodecPool: Got brand-new compressor [.deflate]
2025-02-03 19:45:47,071 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at ip-10-0-16-118.ec2.internal/10.0.16.118:8032
2025-02-03 19:45:47,250 INFO client.AHSPProxy: Connecting to Application History server at ip-10-0-16-118.ec2.internal/10.0.16.118:10200
2025-02-03 19:45:47,440 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1738610451542_0001
2025-02-03 19:45:49,939 INFO db.DBInputFormat: Using read committed transaction isolation
2025-02-03 19:45:49,940 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN('payment_type'), MAX('payment_type') FROM 'ride_log'
2025-02-03 19:46:43,880 INFO db.IntegerSplitter: Split size: 1; Num splits: 4 from: 1 to: 5
2025-02-03 19:46:44,022 INFO mapreduce.JobSubmitter: number of splits:5
2025-02-03 19:46:44,200 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-pu
enabled
2025-02-03 19:46:44,366 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1738610451542_0001
2025-02-03 19:46:44,366 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-02-03 19:46:44,591 INFO conf.Configuration: resource-types.xml not found
2025-02-03 19:46:44,591 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-02-03 19:46:44,935 INFO impl.YarnClientImpl: Submitted application application_1738610451542_0001
2025-02-03 19:46:44,987 INFO mapreduce.Job: The url to track the job: http://ip-10-0-16-118.ec2.internal:20888/proxy/application_1738610451542_0001/
2025-02-03 19:46:44,987 INFO mapreduce.Job: Running job: job_1738610451542_0001
2025-02-03 19:46:54,135 INFO mapreduce.Job: Job job_1738610451542_0001 running in uber mode : false
2025-02-03 19:46:54,136 INFO mapreduce.Job: map 0% reduce 0%
2025-02-03 19:49:01,933 INFO mapreduce.Job: map 20% reduce 0%
2025-02-03 19:49:02,938 INFO mapreduce.Job: map 40% reduce 0%
2025-02-03 19:49:03,941 INFO mapreduce.Job: map 60% reduce 0%
2025-02-03 19:49:20,994 INFO mapreduce.Job: map 60% reduce 20%
```

```
2025-02-03 20:10:01,259 INFO mapreduce.Job: map 96% reduce 27%
2025-02-03 20:10:07,272 INFO mapreduce.Job: map 100% reduce 27%
2025-02-03 20:10:19,296 INFO mapreduce.Job: map 93% reduce 27%
2025-02-03 20:13:02,638 INFO mapreduce.Job: map 100% reduce 27%
2025-02-03 20:14:12,779 INFO mapreduce.Job: map 100% reduce 67%
2025-02-03 20:16:12,010 INFO mapreduce.Job: map 100% reduce 68%
2025-02-03 20:18:18,262 INFO mapreduce.Job: map 100% reduce 69%
2025-02-03 20:20:30,505 INFO mapreduce.Job: map 100% reduce 70%
2025-02-03 20:22:42,762 INFO mapreduce.Job: map 100% reduce 71%
2025-02-03 20:24:54,999 INFO mapreduce.Job: map 100% reduce 72%
2025-02-03 20:27:01,226 INFO mapreduce.Job: map 100% reduce 73%
2025-02-03 20:29:13,458 INFO mapreduce.Job: map 100% reduce 74%
2025-02-03 20:31:19,685 INFO mapreduce.Job: map 100% reduce 75%
2025-02-03 20:33:31,923 INFO mapreduce.Job: map 100% reduce 76%
2025-02-03 20:35:44,134 INFO mapreduce.Job: map 100% reduce 77%
2025-02-03 20:37:50,347 INFO mapreduce.Job: map 100% reduce 78%
2025-02-03 20:40:02,557 INFO mapreduce.Job: map 100% reduce 79%
2025-02-03 20:42:08,774 INFO mapreduce.Job: map 100% reduce 80%
2025-02-03 20:44:20,990 INFO mapreduce.Job: map 100% reduce 81%
2025-02-03 20:46:27,194 INFO mapreduce.Job: map 100% reduce 82%
2025-02-03 20:48:39,414 INFO mapreduce.Job: map 100% reduce 83%
2025-02-03 20:50:45,612 INFO mapreduce.Job: map 100% reduce 84%
2025-02-03 20:52:57,825 INFO mapreduce.Job: map 100% reduce 85%
2025-02-03 20:55:10,036 INFO mapreduce.Job: map 100% reduce 86%
2025-02-03 20:57:16,237 INFO mapreduce.Job: map 100% reduce 87%
2025-02-03 20:59:28,452 INFO mapreduce.Job: map 100% reduce 88%
2025-02-03 21:01:34,650 INFO mapreduce.Job: map 100% reduce 89%
2025-02-03 21:03:46,869 INFO mapreduce.Job: map 100% reduce 90%
2025-02-03 21:05:59,078 INFO mapreduce.Job: map 100% reduce 91%
2025-02-03 21:08:05,279 INFO mapreduce.Job: map 100% reduce 92%
2025-02-03 21:10:17,488 INFO mapreduce.Job: map 100% reduce 93%
2025-02-03 21:12:23,689 INFO mapreduce.Job: map 100% reduce 94%
2025-02-03 21:14:35,897 INFO mapreduce.Job: map 100% reduce 95%
2025-02-03 21:16:42,096 INFO mapreduce.Job: map 100% reduce 96%
2025-02-03 21:18:54,309 INFO mapreduce.Job: map 100% reduce 97%
2025-02-03 21:21:00,508 INFO mapreduce.Job: map 100% reduce 98%
2025-02-03 21:23:12,717 INFO mapreduce.Job: map 100% reduce 99%
```

```
Bytes Read=0
File Output Format Counters
Bytes Written=25190679285
2025-02-03 21:26:28,150 INFO mapreduce.ImportJobBase: Transferred 23.4606 GB in 6,041.3923 seconds (3.9765 MB/sec)
2025-02-03 21:26:28,152 INFO mapreduce.ImportJobBase: Retrieved 320970115 records.
2025-02-03 21:26:28,202 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2025-02-03 21:26:28,246 WARN tool.LoadIncrementalHFiles: Skipping non-directory hdfs://ip-10-0-16-118.ec2.internal:8020/user/hadoop/ride_log/_SUCCESS
2025-02-03 21:26:28,368 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2-hbase.properties
2025-02-03 21:26:28,447 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-02-03 21:26:28,447 INFO impl.MetricsSystemImpl: HBase metrics system started
2025-02-03 21:26:28,560 WARN tool.LoadIncrementalHFiles: Trying to bulk load hfile hdfs://ip-10-0-16-118.ec2.internal:8020/user/hadoop/ride_log/colfam1/4fbl0b2e8f6541629cf27902010b820f with size: 11170130905 bytes can be problematic as it may lead to oversplitting.
2025-02-03 21:26:28,570 WARN tool.LoadIncrementalHFiles: Trying to bulk load hfile hdfs://ip-10-0-16-118.ec2.internal:8020/user/hadoop/ride_log/colfam1/af48561a860b43b8241403a3e02aeal with size: 11170137063 bytes can be problematic as it may lead to oversplitting.
[hadoop@ip-10-0-16-118 mysql-connector-java-8.0.25]$
```

```
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.4.17-amzn-3, rUnknown, Thu Nov  2 05:41:41 UTC 2023
Took 0.0018 seconds
hbase:001:0> list
TABLE
ride_log_hbase
1 row(s)
Took 0.5126 seconds
=> ["ride_log_hbase"]
hbase:002:0> describe 'ride_log_hbase'
Table ride_log_hbase is ENABLED
ride_log_hbase
COLUMN FAMILIES DESCRIPTION
(NAME => 'colfam1', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TT
L => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
1 row(s)
Quota is disabled
Took 0.1753 seconds
hbase:003:0> █
```