

1. Creating hbase table.

Command Used:

create 'trip_details', 'td'

Explanation: trip_details is the hbase table name and td is the column family name

```
[root@ip-172-31-59-44 ~]# hbase shell
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Fri Apr 17 15:18:24 UTC 2020

hbase(main):001:0> create 'trip_details', 'td'
0 row(s) in 2.5780 seconds

=> Hbase::Table - trip_details
hbase(main):002:0>
```

2. Sqoop Import

Command Used:

```
sqoop import --connect jdbc:mysql://nyc-tlc-mr-db.cv4nmpnzjxab.us-east-1.rds.amazonaws.com:3306/taximrdb --username admin -P --table yellow_taxi_trips --hbase-table trip_details --column-family td --hbase-row-key trip_id
```

Explanation: --connect is used to connect to mysql database using the --username and -P (password).

--table is the mysql database table name which is the source table

--hbase-table is the hbase table name which is the target table

--column-family is the column family name of the hbase table

--hbase-row-key is the primary key of the mysql database table that will be used as rowkey for the hbase table

```
[root@ip-172-31-58-150 hadoop]# sqoop import --connect jdbc:mysql://nyc-tlc-mr-rds.cv4nmpnzjxab.us-east-1.rds.amazonaws.com:3306/taximrdb --username admin -P --table yellow_taxi_trips --hbase-table trip_details --column-family td --hbase-row-key trip_id
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/04/26 07:00:26 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]
Enter password:
23/04/26 07:00:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/04/26 07:00:30 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 the driver class is generally unnecessary.
23/04/26 07:00:30 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow_taxi_trips' AS t LIMIT 1
23/04/26 07:00:31 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'yellow_taxi_trips' AS t LIMIT 1
23/04/26 07:00:31 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/325816c8054d55edf2a187242b0d5aac/yellow_taxi_trips.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/04/26 07:00:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/325816c8054d55edf2a187242b0d5aac/yellow_taxi_trips.jar
23/04/26 07:00:35 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/04/26 07:00:35 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/04/26 07:00:35 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/04/26 07:00:35 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/04/26 07:00:35 INFO mapreduce.ImportJobBase: Beginning import of yellow taxi trips
23/04/26 07:00:36 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/04/26 07:00:36 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/04/26 07:00:39 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. Most users should rely on addDependencyJars(Job) instead. See HBASE-9386 for more details.
23/04/26 07:00:39 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-58-150.ec2.internal/172.31.58.150:8032
23/04/26 07:00:41 INFO db.DBInputFormat: Using read committed transaction isolation
23/04/26 07:00:41 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN('trip_id'), MAX('trip_id') FROM 'yellow_taxi_trips'
23/04/26 07:00:41 INFO db.IntegerSplitter: Split size: 4733622; Num splits: 4 from: 1 to: 18934490
23/04/26 07:00:41 INFO mapreduce.JobSubmitter: number of splits:4
23/04/26 07:00:42 INFO impl.YarnClientImpl: Submitted application application_1682483230379_0002
23/04/26 07:00:42 INFO mapreduce.Job: The url to track the job: http://ip-172-31-58-150.ec2.internal:20888/proxy/application_1682483230379_0002/
23/04/26 07:00:42 INFO mapreduce.Job: Running job: job_1682483230379_0002
23/04/26 07:00:50 INFO mapreduce.Job: Job job_1682483230379_0002 running in uber mode : false
23/04/26 07:00:50 INFO mapreduce.Job: map 0% reduce 0%
```

```
hbase(main):001:0> get 'trip_details', '1'
COLUMN                                CELL
td:DOLocationID                      timestamp=1681792570962, value=236
td:PULocationID                      timestamp=1681792570962, value=140
td:RatecodeID                        timestamp=1681792570962, value=1
td:VendorID                          timestamp=1681792570962, value=1
td:airport_fee                       timestamp=1681792570962, value=0.0
td:congestion_surcharge              timestamp=1681792570962, value=0.0
td:extra                             timestamp=1681792570962, value=0.5
td:fare_amount                       timestamp=1681792570962, value=6.5
td:improvement_surcharge             timestamp=1681792570962, value=0.3
td:mta_tax                           timestamp=1681792570962, value=0.5
td:passenger_count                   timestamp=1681792570962, value=1
td:payment_type                      timestamp=1681792570962, value=2
td:store_and_fwd_flag                timestamp=1681792570962, value=N
td:tip_amount                        timestamp=1681792570962, value=0.0
td:tolls_amount                      timestamp=1681792570962, value=0.0
td:total_amount                      timestamp=1681792570962, value=7.8
td:tpep_dropoff_date                 timestamp=1681792570962, value=2017-01-01 00:37:48.0
td:tpep_pickup_datetime              timestamp=1681792570962, value=2017-01-01 00:32:05.0
td:trip_distance                     timestamp=1681792570962, value=1.2
1 row(s) in 0.9100 seconds

hbase(main):002:0> █
```