

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

Solution:

Steps are as follows:

1. Connect to EMR cluster

Public dns for connection: ec2-3-92-192-184.compute-1.amazonaws.com

```
Using username "hadoop".
Authenticating with public key "imported-openssh-key"

      _I_ _I_ )
      _I_ ( /  Amazon Linux 2 AMI
      __I\__I__I

https://aws.amazon.com/amazon-linux-2/
8 package(s) needed for security, out of 9 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::::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 M:::M M:::M R:::R R:::R
E::::EEEEEEEEEE M::::M M:::M M:::M M:::M R:::RRRRRR::::R
E::::::::::E 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
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-30-2-71 ~]$
```

2. Download mysql connector and copy it to sqoop lib location.

Commands:

```
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
```

```

[hadoop@ip-172-30-2-71 ~]$ sudo -i
EEEEEEEEEEEEEEEEEEEE MAAAAAAAA MAAAAAAAA RRRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::M M:::::M R:::::R R:::::R
EE:::::EEEEEEEEEE::E M:::::M M:::::M R:::::R RRRRRR:::::R
E::::E EEEEE M:::::M M:::::M RR:::::R R:::::R
E::::E M:::::M:M M:::::M R:::::R R:::::R
E:::::EEEEEEEEEE M:::::M M:::M R:::R RRRRRR:::::R
E:::::EEEEEE M:::::M M:::M M:::::M R:::::R R:::::R
E:::::EEEEEEEEEE M:::::M M:::M M:::::M R:::R RRRRRR:::::R
E::::E M:::::M M:::M M:::::M R:::::R R:::::R
E::::E EEEEE M:::::M M:::M M:::::M R:::::R R:::::R
EE:::::EEEEEEEEEE::E M:::::M M:::::M R:::::R R:::::R
E::::::::::::::::::::E M:::::M M:::::M RR:::::R R:::::R
EEEEEEEEEEEEEEEEEEEE MAAAAAAAA MAAAAAAAA RRRRRR
[root@ip-172-30-2-71 ~]# wget https://de-mysql-connector.s3.amazonaws.com/mysql-connector-java-8.0.25.tar.gz
--2023-08-30 08:53:22-- 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) : 52.217.141.193, 3.5.25.67, 52.217.78.212, ...
Connecting to de-mysql-connector.s3.amazonaws.com (de-mysql-connector.s3.amazonaws.com)|52.217.141.193|: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 25.2MB/s in

2023-08-30 08:53:23 (25.2 MB/s) - 'mysql-connector-java-8.0.25.tar.gz' saved [4079310/4079310]

[root@ip-172-30-2-71 ~]# tar -xvf mysql-connector-java-8.0.25.tar.gz

[root@ip-172-30-2-71 mysql-connector-java-8.0.25]# sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib
[root@ip-172-30-2-71 mysql-connector-java-8.0.25]#

```

3. Check the connection with RDS db

```
mysql -h demodb1.cr0zubg47bor.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
```

```

root@ip-172-30-2-71 mysql-connector-java-8.0.25]# mysql -h demodbl.cr0zubg47bor.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 278
Server version: 8.0.33 Source distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MySQL [(none)]> show databases;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| performance_schema |
| sys |
| taxidb |
+-----+
6 rows in set (0.04 sec)

MySQL [(none)]> exit
bye

```

4. Run sqoop import command to take data from RDS system to hdfs for **testing**.

```
sqoop import --connect jdbc:mysql://demodb1.cr0zubg47bor.us-east-1.rds.amazonaws.com:3306/taxidb \
--table Vendor \
--username admin --password admin123 \
--split-by VendorID \
--target-dir /user/root/Vendor_data \
-m 1
```

Map input records=18880595

```

MySQL [(none)]> exit
bye
[root@ip-172-30-2-71 mysql-connector-java-8.0.25]# sqoop import --connect jdbc:mysql://demo-dbl-cr0zubg47bor.us-east-1.rds.amazonaws.com:3306/taxidb \
> --table Vendor \
> --username admin --password admin123 \
> --split-by VendorID \
> --target-dir /user/root/Vendor_data \
> -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-reloadj-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-reloadj-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.ReloadjLoggerFactory]
2023-08-30 09:00:17,803 INFO manager.Sqoop: Running Sqoop version: 1.4.7
2023-08-30 09:00:17,832 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2023-08-30 09:00:17,938 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2023-08-30 09:00:17,938 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 load
ing of the driver class is generally unnecessary.
2023-08-30 09:00:18,417 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 09:00:18,547 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 09:00:18,567 INFO ocm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2023-08-30 09:00:23,401 INFO ocm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/dff21a4f9e50d21bbfd173d745f9b34c/Vendor.jar
2023-08-30 09:00:23,416 WARN manager.MySQLManager: It looks like you are importing from mysql.
2023-08-30 09:00:23,416 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2023-08-30 09:00:23,416 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2023-08-30 09:00:23,416 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2023-08-30 09:00:23,422 INFO mapreduce.ImportJobBase: Beginning import of Vendor
2023-08-30 09:00:23,553 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2023-08-30 09:00:24,117 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2023-08-30 09:00:24,402 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at ip-172-30-2-71.ec2.internal/172.30.2.71:8032
2023-08-30 09:00:24,827 INFO client.AHSProxy: Connecting to Application History server at ip-172-30-2-71.ec2.internal/172.30.2.71:10200
2023-08-30 09:00:25,083 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1693384615004_0001
2023-08-30 09:00:26,264 INFO job.InputFormat: Using read committed transaction isolation
2023-08-30 09:00:26,301 INFO mapreduce.JobSubmitter: number of splits=1
2023-08-30 09:00:26,933 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1693384615004_0001
2023-08-30 09:00:26,933 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-08-30 09:00:27,117 INFO conf.Configuration: resource-types.xml not found

```

```

2023-08-30 09:00:34,770 INFO mapreduce.Job: Job job_1693384615004_0001 running in uber mode : false
2023-08-30 09:00:34,771 INFO mapreduce.Job: map 0% reduce 0%
2023-08-30 09:02:30,225 INFO mapreduce.Job: map 100% reduce 0%
2023-08-30 09:02:30,229 INFO mapreduce.Job: Job job_1693384615004_0001 completed successfully
2023-08-30 09:02:30,329 INFO mapreduce.Job: Counters: 33

  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=298385
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=85
    HDFS: Number of bytes written=1966321996
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0

  Job Counters
    Launched map tasks=1
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=10785216
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=112346
    Total vcore-milliseconds taken by all map tasks=112346
    Total megabyte-milliseconds taken by all map tasks=345126912

  Map-Reduce Framework
    Map input records=18880595
    Map output records=18880595
    Input split bytes=85
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=567
    CPU time spent (ms)=118240
    Physical memory (bytes) snapshot=1081307136
    Virtual memory (bytes) snapshot=4419592192
    Total committed heap usage (bytes)=526385152
    Peak Map Physical memory (bytes)=1081307136
    Peak Map Virtual memory (bytes)=4419592192

  File Input Format Counters
    Bytes Read=0

  File Output Format Counters
    Bytes Written=1966321996
2023-08-30 09:02:30,333 INFO mapreduce.ImportJobBase: Transferred 1.8313 GB in 126.2061 seconds (14.8585 MB/sec)
2023-08-30 09:02:30,335 INFO mapreduce.ImportJobBase: Retrieved 18880595 records

```

5. Run sqoop import command to import data from RDS to Hbase

Testing: loading 10 records from CSV to check import to hbase is running fine or not.

Command:

- `head -n 10 yellow_tripdata_2017-01.csv > yellow_tripdata_2017-01_test.csv`
- `create database taxi_test;`
- `use taxi_test;`
- `LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-01_test.csv'`
`INTO TABLE Vendor_test`
`FIELDS TERMINATED BY ','`
`LINES TERMINATED BY '\n'`
`IGNORE 1 LINES;`

```

[kroot@isp-172-30-2-64 ~]# head -n 10 yellow_tripdata_2017-01.csv > yellow_tripdata_2017-01_test.csv
[kroot@isp-172-30-2-64 ~]# cat yellow_tripdata_2017-01_test.csv
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,
trip_amount,tolls_amount,improvement_surchage,total_amount,congestion_surchage,airport_fee
1,2017-01-01,00:32:05,2017-01-01,00:37:48,1,1.2,1,N,140,236,2,6,5,0,0,5,0,5,0,0,0,0,0,3,7,8,,
1,2017-01-01,00:49:25,2017-01-01,00:47:42,2,0,7,1,N,237,140,2,5,0,0,0,5,0,5,0,0,0,0,0,3,6,3,,
1,2017-01-01,00:53:59,2017-01-01,00:53:59,2,0,8,1,N,140,237,2,5,5,0,0,0,5,0,5,0,0,0,0,0,3,6,8,,
1,2017-01-01,00:36:42,2017-01-01,00:41:09,1,1,1,N,41,42,2,6,0,0,0,5,0,5,0,0,0,0,0,0,3,7,3,,
1,2017-01-01,00:07:41,2017-01-01,00:18:16,1,3,0,1,N,48,263,2,1,1,0,0,0,5,0,5,0,0,0,0,0,3,12,3,,
1,2017-01-01,00:20:52,2017-01-01,00:24:59,2,0,7,1,N,236,262,2,5,0,0,0,5,0,5,0,0,0,0,0,3,6,3,,
1,2017-01-01,00:33:49,2017-01-01,00:42:38,2,1,6,1,N,236,238,1,8,0,0,0,5,0,5,1,85,0,0,0,3,11,15,,
1,2017-01-01,00:48:22,2017-01-01,00:52:15,2,0,6,1,N,238,239,1,5,0,0,0,5,0,5,1,25,0,0,0,3,7,55,,
1,2017-01-01,00:57:12,2017-01-01,01:06:28,2,1,0,1,N,239,46,1,7,5,0,0,0,5,1,75,0,0,0,3,10,55,,

```

[illegible]

- ```
➤ sqoop import --connect jdbc:mysql://demodb1.cr0zubg47bor.us-east-1.rds.amazonaws.com:3306/taxitest \
--table Vendor_test \
--username admin --password admin123 \
--hbase-table vendor_t1 \
--column-family cf1 \
--hbase-create-table \
--hbase-row-key "tpep_pickup_datetime,tpep_dropoff_datetime" \
--split-by VendorID
```

Explanation of above command:

Sqoop import will load the data from RDS db to Hbase through MYSQL jdbc connector.

It will read table Vendor\_test from data base and Create vendor\_t1 into hbase system by using -hbase-create-table parameter with column family as cf1 and hbase-row-key "tpep\_pickup\_datetime,tpep\_dropoff\_datetime" from RDS table with split on VendorID

```
[root@ip-172-30-2-64 ~]# sqoop import --connect jdbc:mysql://demodbl.cr0rubg47bor.us-east-1.rds.amazonaws.com:3306/taxitest \
--hbase-table vendor_t1 \
> --table Vendor_test \
> --username admin --password admin123 \
> --hbase-table vendor_t1 \
> --column-family cf1 \
> --hbase-create-table \
> --hbase-row-key "tpep_pickup_datetime,tpep_dropoff_datetime" \
> --split-by VendorID
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]
2023-08-30 14:29:59,434 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2023-08-30 14:29:59,465 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2023-08-30 14:29:59,562 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2023-08-30 14:29:59,563 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.
2023-08-30 14:29:59,902 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor_test' AS t LIMIT 1
2023-08-30 14:29:59,948 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor_test' AS t LIMIT 1
2023-08-30 14:29:59,971 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /usr/lib/hadoop-mapreduce
2023-08-30 14:30:02,417 ERROR orm.CompilationManager: Could not rename /tmp/sqoop-root/compile/f78a3203841cabdc51b7820eddba7e2c/Vendor_test.java to /root/.Vendor_test.java.
or: Destination '/root/.Vendor_test.java' already exists
2023-08-30 14:30:02,417 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/f78a3203841cabdc51b7820eddba7e2c/Vendor_test.jar
2023-08-30 14:30:02,433 WARN manager.MySQLManager: It looks like you are importing from mysql.
2023-08-30 14:30:02,433 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2023-08-30 14:30:02,433 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2023-08-30 14:30:02,434 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2023-08-30 14:30:02,456 INFO mapreduce.ImportJobBase: Beginning import of Vendor_test
2023-08-30 14:30:02,570 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2023-08-30 14:30:02,596 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2023-08-30 14:30:04,733 INFO mapreduce.HBaseImportJob: Creating missing HBase table vendor_t1
2023-08-30 14:30:07,001 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly.
Users should rely on addDependencyJars(Job) instead. See HBASE-8386 for more details.
2023-08-30 14:30:07,230 INFO client.DefaultNoHARMFallowerProxyProvider: Connecting to ResourceManager at ip-172-30-2-64.ec2.internal/172.30.2.64:8032
2023-08-30 14:30:07,414 INFO client.AHSProxy: Connecting to Application History server at ip-172-30-2-64.ec2.internal/172.30.2.64:10200
2023-08-30 14:30:08,083 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1693404136954_0002
2023-08-30 14:30:12,723 INFO db.DBInputFormat: Using read committed transaction isolation
2023-08-30 14:30:12,724 INFO db.DataDrivenDBInputFormat: RoundingValsQuery: SELECT MIN('VendorID'), MAX('VendorID') FROM 'Vendor_test'
```

```
2023-08-30 14:30:20,151 INFO mapreduce.Job: Job job_1693404136954_0002 running in uber mode : false
2023-08-30 14:30:20,152 INFO mapreduce.Job: map 0% reduce 0%
2023-08-30 14:30:27,283 INFO mapreduce.Job: map 100% reduce 0%
2023-08-30 14:30:27,288 INFO mapreduce.Job: Job job_1693404136954_0002 completed successfully
2023-08-30 14:30:27,377 INFO mapreduce.Job: Counters: 33
File System Counters
 FILE: Number of bytes read=0
 FILE: Number of bytes written=337102
 FILE: Number of read operations=0
 FILE: Number of large read operations=0
 FILE: Number of write operations=0
 HDFS: Number of bytes read=109
 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
 HDFS: Number of bytes read erasure-coded=0
Job Counters
 Launched map tasks=1
 Other local map tasks=1
 Total time spent by all maps in occupied slots (ms)=430080
 Total time spent by all reduces in occupied slots (ms)=0
 Total time spent by all map tasks (ms)=4480
 Total vcore-milliseconds taken by all map tasks=4480
 Total megabyte-milliseconds taken by all map tasks=13762560
Map-Reduce Framework
 Map input records=9
 Map output records=9
 Input split bytes=109
 Spilled Records=0
 Failed Shuffles=0
 Merged Map outputs=0
 GC time elapsed (ms)=118
 CPU time spent (ms)=3900
 Physical memory (bytes) snapshot=414576640
 Virtual memory (bytes) snapshot=4464443392
 Total committed heap usage (bytes)=380108800
 Peak Map Physical memory (bytes)=414576640
 Peak Map Virtual memory (bytes)=4464443392
File Input Format Counters
 Bytes Read=0
File Output Format Counters
 Bytes Written=0
2023-08-30 14:30:27,381 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 20.3759 seconds (0 bytes/sec)
```

- Scan hbase table to check table is created with data or not.

```

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-0, rUnknown, Fri Jun 30 17:23:19 UTC 2023
Took 0.0017 seconds
hbase:001:0> scan 'vendor_tl'
ROW
COLUMN+CELL
column=cfl:DOLocationID, timestamp=2023-08-30T14:30:25.740, value=263
column=cfl:PULocationID, timestamp=2023-08-30T14:30:25.740, value=48
column=cfl:RatecodeID, timestamp=2023-08-30T14:30:25.740, value=1
column=cfl:VendorID, timestamp=2023-08-30T14:30:25.740, value=1
column=cfl:airport_fee, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:congestion_surcharge, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:extra, timestamp=2023-08-30T14:30:25.740, value=0.5
column=cfl:fare_amount, timestamp=2023-08-30T14:30:25.740, value=11.0
column=cfl:improvement_surcharge, timestamp=2023-08-30T14:30:25.740, value=0.3
column=cfl:mta_tax, timestamp=2023-08-30T14:30:25.740, value=0.5
column=cfl:passenger_count, timestamp=2023-08-30T14:30:25.740, value=1
column=cfl:payment_type, timestamp=2023-08-30T14:30:25.740, value=2
column=cfl:store_and_fwd_flag, timestamp=2023-08-30T14:30:25.740, value=N
column=cfl:tip_amount, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:tolls_amount, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:total_amount, timestamp=2023-08-30T14:30:25.740, value=12.3
column=cfl:trip_distance, timestamp=2023-08-30T14:30:25.740, value=3.0
column=cfl:DOLocationID, timestamp=2023-08-30T14:30:25.740, value=262
column=cfl:PULocationID, timestamp=2023-08-30T14:30:25.740, value=236
column=cfl:RatecodeID, timestamp=2023-08-30T14:30:25.740, value=1
column=cfl:VendorID, timestamp=2023-08-30T14:30:25.740, value=1
column=cfl:airport_fee, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:congestion_surcharge, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:extra, timestamp=2023-08-30T14:30:25.740, value=0.5
column=cfl:fare_amount, timestamp=2023-08-30T14:30:25.740, value=5.0
column=cfl:improvement_surcharge, timestamp=2023-08-30T14:30:25.740, value=0.3
column=cfl:mta_tax, timestamp=2023-08-30T14:30:25.740, value=0.5
column=cfl:passenger_count, timestamp=2023-08-30T14:30:25.740, value=2
column=cfl:payment_type, timestamp=2023-08-30T14:30:25.740, value=2
column=cfl:store_and_fwd_flag, timestamp=2023-08-30T14:30:25.740, value=N
column=cfl:tip_amount, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:tolls_amount, timestamp=2023-08-30T14:30:25.740, value=0.0
column=cfl:total_amount, timestamp=2023-08-30T14:30:25.740, value=6.3
column=cfl:trip_distance, timestamp=2023-08-30T14:30:25.740, value=0.7

```

7 Now let us try with without `-hbase-create-table` parameter.

We can create multiple column families such as payment details, trip\_details and vendor\_details but currently creating one single family as 'Vendor\_details' to keeping it simple.

Create Hbase table and column family with below commands:

- List
- create 'Vendor','Vendor\_details'
- describe 'Vendor'

```

root@ip-172-30-2-64 ~]# hbase shell
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/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]
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-0, rUnknown, Fri Jun 30 17:23:19 UTC 2023
Took 0.0018 seconds
hbase:001:0> list
TABLE
1 row(s)
Took 0.4555 seconds
> {}
hbase:002:0> create 'Vendor','Vendor_details'
Created table Vendor
Took 1.1545 seconds
> Hbase::Table - Vendor
hbase:003:0> list
TABLE
Vendor
1 row(s)
Took 0.0101 seconds
> ["Vendor"]
hbase:004:0> describe 'Vendor'
Table Vendor is ENABLED
Vendor
COLUMN FAMILIES DESCRIPTION
NAME => 'Vendor_details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL
=> 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'
1 row(s)
Quota is disabled
Took 0.1195 seconds
hbase:005:0>

```



Or If you want to create multiple columns families

- create create 'Vendor','Vendor\_details','Trip\_details','Payment\_details'

```
hbase:001:0> create 'Vendor','Vendor details','Trip details','Payment details'
Created table Vendor
Took 1.0564 seconds
=> hbase:Table - Vendor
hbase:002:0> list
TABLE
Vendor
1 row(s)
Took 0.0205 seconds
=> [[Vendor]]
hbase:003:0> describe 'Vendor'
Table Vendor is ENABLED
Vendor
COLUMN FAMILIES DESCRIPTION
(NAME => 'Payment details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
(NAME => 'Trip details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
(NAME => 'Vendor details', BLOOMFILTER => 'ROW', IN_MEMORY => 'false', VERSIONS => '1', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
3 row(s)
Quota is disabled
Took 0.1200 seconds
hbase:004:0> []
```

Now run sqoop import command with above created table

- sqoop import --connect jdbc:mysql://demodb1.cr0zubg47bor.us-east-1.rds.amazonaws.com:3306/taxidb \
- table Vendor \
- username admin --password admin123 \
- hbase-table Vendor \
- column-family Vendor\_details \
- hbase-row-key "tpep\_pickup\_datetime,tpep\_dropoff\_datetime" \
- split-by VendorID

```
[root@ip-172-30-2-64 ~]# sqoop import --connect jdbc:mysql://demodb1.cr0zubg47bor.us-east-1.rds.amazonaws.com:3306/taxidb \
> --table Vendor \
> --username admin --password admin123 \
> --hbase-table Vendor \
> --column-family Vendor_details \
> --hbase-row-key "tpep_pickup_datetime,tpep_dropoff_datetime" \
> --split-by VendorID
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/aiwe/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]
2023-08-30 14:58:10,492 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2023-08-30 14:58:10,519 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2023-08-30 14:58:10,617 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2023-08-30 14:58:10,618 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 load
ing of the driver class is generally unnecessary.
2023-08-30 14:58:10,960 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 14:58:11,007 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 14:58:11,025 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2023-08-30 14:58:13,538 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/d3f1e44ddc5c262c771382a83f6eeb17/Vendor.jar
2023-08-30 14:58:13,557 WARN manager.MySQLManager: It looks like you are importing from mysql.
2023-08-30 14:58:13,557 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2023-08-30 14:58:13,557 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2023-08-30 14:58:13,557 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2023-08-30 14:58:13,590 INFO mapreduce.ImportJobBase: Beginning import of Vendor
2023-08-30 14:58:13,690 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2023-08-30 14:58:13,708 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2023-08-30 14:58:15,448 WARN mapreduce.TableMapReduceUtil: The addDependencyVars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. Most
users should rely on addDependencyVars(Job) instead. See HBASE-8386 for more details.
2023-08-30 14:58:15,670 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at ip-172-30-2-64.ec2.internal/172.30.2.64:8032
2023-08-30 14:58:15,834 INFO client.AHSProxy: Connecting to Application History server at ip-172-30-2-64.ec2.internal/172.30.2.64:10200
2023-08-30 14:58:16,486 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1693404136954_0003
2023-08-30 14:58:19,458 INFO db.DBInputFormat: Using read committed transaction isolation
2023-08-30 14:58:19,458 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN('VendorID'), MAX('VendorID') FROM 'Vendor'
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 lo
2023-08-30 14:58:10,960 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 14:58:11,007 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Vendor' AS t LIMIT 1
2023-08-30 14:58:11,025 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
2023-08-30 14:58:13,538 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/d3f1e44ddc5c262c771382a83f6eeb17/Vendor.jar
2023-08-30 14:58:13,557 WARN manager.MySQLManager: It looks like you are importing from mysql.
2023-08-30 14:58:13,557 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
2023-08-30 14:58:13,557 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
2023-08-30 14:58:13,557 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2023-08-30 14:58:13,590 INFO mapreduce.ImportJobBase: Beginning import of Vendor
2023-08-30 14:58:13,690 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
2023-08-30 14:58:13,708 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
2023-08-30 14:58:15,448 WARN mapreduce.TableMapReduceUtil: The addDependencyVars(Configuration, Class<?>...) method has been deprecated since it is easy to use incorrectly. Mo
users should rely on addDependencyVars(Job) instead. See HBASE-8386 for more details.
2023-08-30 14:58:15,670 INFO client.DefaultNoHARMFaloverProxyProvider: Connecting to ResourceManager at ip-172-30-2-64.ec2.internal/172.30.2.64:8032
2023-08-30 14:58:16,486 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1693404136954_0003
2023-08-30 14:58:19,458 INFO db.DBInputFormat: Using read committed transaction isolation
2023-08-30 14:58:19,458 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN('VendorID'), MAX('VendorID') FROM 'Vendor'
2023-08-30 14:59:01,508 INFO db.IntegerSplitter: Split size: 0; Num splits: 4 from: 1 to: 2
2023-08-30 14:59:01,567 INFO mapreduce.JobSubmitter: number of splits:2
2023-08-30 14:59:02,003 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2023-08-30 14:59:02,219 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1693404136954_0003
2023-08-30 14:59:02,219 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-08-30 14:59:02,391 INFO conf.Configuration: resource-types.xml not found
2023-08-30 14:59:02,391 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-08-30 14:59:02,449 INFO impl.YarnClientImpl: Submitted application application_1693404136954_0003
2023-08-30 14:59:02,492 INFO mapreduce.Job: The url to track the job: http://ip-172-30-2-64.ec2.internal:20888/proxy/application_1693404136954_0003/
2023-08-30 14:59:02,492 INFO mapreduce.Job: Running job: job_1693404136954_0003
2023-08-30 14:59:09,558 INFO mapreduce.Job: Job job_1693404136954_0003 running in uber mode : false
2023-08-30 14:59:09,559 INFO mapreduce.Job: map 0% reduce 0%
```

## HBase table verification for data insertion using count “Vendor” and scan "Vendor" command

```
[root@ip-172-30-2-64 ~]# hbase shell
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/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]
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-0, rUnknown, Fri Jun 30 17:23:19 UTC 2023
Took 0.0020 seconds
hbase:001:0> list
TABLE
Vendor
1 row(s)
Took 0.4502 seconds
=> ["Vendor"]
hbase:002:0> describe "Vendor"
Table Vendor is ENABLED
Vendor
COLUMN FAMILIES DESCRIPTION
(NAME => 'Vendor details', BLOOMFILTER => 'ROW', IN MEMORY => 'false', VERSIONS => '1', KEEP DELETED CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', COMPRESSION => 'NONE', TTL
=> 'FOREVER', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
1 row(s)
Cache is disabled
Took 0.1265 seconds
hbase:003:0> count "Vendor"
Current count: 1000, row: 2017-01-01 00:06:50.0 2017-01-01 00:14:24.0
Current count: 2000, row: 2017-01-01 00:10:49.0 2017-01-01 00:32:23.0
Current count: 3000, row: 2017-01-01 00:14:09.0 2017-01-01 00:15:47.0
Current count: 4000, row: 2017-01-01 00:17:05.0 2017-01-01 00:33:41.0
Current count: 5000, row: 2017-01-01 00:19:44.0 2017-01-01 00:29:27.0
Current count: 6000, row: 2017-01-01 00:22:20.0 2017-01-01 00:35:07.0
Current count: 7000, row: 2017-01-01 00:24:57.0 2017-01-01 00:27:01.0
Current count: 8000, row: 2017-01-01 00:27:31.0 2017-01-01 01:12:08.0
Current count: 9000, row: 2017-01-01 00:30:06.0 2017-01-01 00:56:24.0
Current count: 10000, row: 2017-01-01 00:32:41.0 2017-01-01 00:41:09.0
Current count: 11000, row: 2017-01-01 00:35:22.0 2017-01-01 01:06:12.0
Current count: 12000, row: 2017-01-01 00:37:56.0 2017-01-01 00:45:49.0
Current count: 13000, row: 2017-01-01 00:40:27.0 2017-01-01 00:52:43.0
Current count: 14000, row: 2017-01-01 00:42:58.0 2017-01-01 01:15:32.0
Current count: 15000, row: 2017-01-01 00:45:37.0 2017-01-01 00:59:36.0
Current count: 16000, row: 2017-01-01 00:48:00.0 2017-01-01 01:04:19.0
Current count: 17000, row: 2017-01-01 00:50:32.0 2017-01-01 01:04:34.0
Current count: 18000, row: 2017-01-01 00:53:06.0 2017-01-01 00:59:28.0
Current count: 19000, row: 2017-01-01 00:55:37.0 2017-01-01 01:01:34.0
Current count: 20000, row: 2017-01-01 00:58:07.0 2017-01-01 01:10:42.0
Current count: 21000, row: 2017-01-01 01:00:32.0 2017-01-01 01:20:17.0
Current count: 22000, row: 2017-01-01 01:02:57.0 2017-01-01 01:13:06.0
hbase:001:0> scan "Vendor"
ROW COLUMN+CELL
"V 2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:DOLocationID, timestamp=2023-08-30T16:33:52.488, value=234
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:FULocationID, timestamp=2023-08-30T16:33:52.488, value=249
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:RatecodeID, timestamp=2023-08-30T16:33:52.488, value=2
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:VendorID, timestamp=2023-08-30T16:33:52.488, value=2
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:airport_fee, timestamp=2023-08-30T16:33:52.488, value=0.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:congestion_surcharge, timestamp=2023-08-30T16:33:52.488, value=0.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:extra, timestamp=2023-08-30T16:33:52.488, value=0.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:fare_amount, timestamp=2023-08-30T16:33:52.488, value=52.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:improvement_surcharge, timestamp=2023-08-30T16:33:52.488, value=0.3
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:mta_tax, timestamp=2023-08-30T16:33:52.488, value=0.5
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:passenger_count, timestamp=2023-08-30T16:33:52.488, value=1
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:payment_type, timestamp=2023-08-30T16:33:52.488, value=2
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:store_and_fwd_flag, timestamp=2023-08-30T16:33:52.488, value=N
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:tip_amount, timestamp=2023-08-30T16:33:52.488, value=0.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:tolls_amount, timestamp=2023-08-30T16:33:52.488, value=0.0
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:total_amount, timestamp=2023-08-30T16:33:52.488, value=52.8
2017-01-01 00:00:00.0 2017-01-01 00:00:00.0 column=Vendor_details:trip_distance, timestamp=2023-08-30T16:33:52.488, value=0.02
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:DOLocationID, timestamp=2023-08-30T16:33:51.863, value=48
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:FULocationID, timestamp=2023-08-30T16:33:51.863, value=48
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:RatecodeID, timestamp=2023-08-30T16:33:51.863, value=1
"C 2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:VendorID, timestamp=2023-08-30T16:33:51.863, value=1
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:airport_fee, timestamp=2023-08-30T16:33:51.863, value=0.0
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:congestion_surcharge, timestamp=2023-08-30T16:33:51.863, value=0.0
2017-01-01 00:00:02.0 2017-01-01 00:03:50.0 column=Vendor_details:extra, timestamp=2023-08-30T16:33:51.863, value=0.5
```



|            |            |            |            |                                                                                           |
|------------|------------|------------|------------|-------------------------------------------------------------------------------------------|
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:VendorID, timestamp=2023-08-30T16:33:48.222, value=1                |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:airport fee, timestamp=2023-08-30T16:33:48.222, value=0.0           |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:congestion surcharge, timestamp=2023-08-30T16:33:48.222, value=0.0  |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:extra, timestamp=2023-08-30T16:33:48.222, value=0.5                 |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:fare amount, timestamp=2023-08-30T16:33:48.222, value=22.0          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:improvement surcharge, timestamp=2023-08-30T16:33:48.222, value=0.3 |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:mta tax, timestamp=2023-08-30T16:33:48.222, value=0.5               |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:passenger count, timestamp=2023-08-30T16:33:48.222, value=1         |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:payment type, timestamp=2023-08-30T16:33:48.222, value=1            |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:store and fwd flg, timestamp=2023-08-30T16:33:48.222, value=N       |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:strip amount, timestamp=2023-08-30T16:33:48.222, value=4.65         |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:stolls amount, timestamp=2023-08-30T16:33:48.222, value=0.0         |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:stoll total amount, timestamp=2023-08-30T16:33:48.222, value=27.95  |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:19:19.0 | column=Vendor_details:strip distance, timestamp=2023-08-30T16:33:48.222, value=7.2        |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:DOLocationID, timestamp=2023-08-30T16:33:53.349, value=7            |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:FULocationID, timestamp=2023-08-30T16:33:53.349, value=107          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:RatecodeID, timestamp=2023-08-30T16:33:53.349, value=1              |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:VendorID, timestamp=2023-08-30T16:33:53.349, value=1                |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:airport fee, timestamp=2023-08-30T16:33:53.349, value=0.0           |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:congestion surcharge, timestamp=2023-08-30T16:33:53.349, value=0.0  |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:extra, timestamp=2023-08-30T16:33:53.349, value=0.5                 |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:fare amount, timestamp=2023-08-30T16:33:53.349, value=21.5          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:improvement surcharge, timestamp=2023-08-30T16:33:53.349, value=0.3 |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:mta tax, timestamp=2023-08-30T16:33:53.349, value=0.5               |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:passenger count, timestamp=2023-08-30T16:33:53.349, value=1         |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:payment type, timestamp=2023-08-30T16:33:53.349, value=1            |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:store and fwd flg, timestamp=2023-08-30T16:33:53.349, value=N       |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:strip amount, timestamp=2023-08-30T16:33:53.349, value=5.0          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:stolls amount, timestamp=2023-08-30T16:33:53.349, value=0.7         |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:stoll total amount, timestamp=2023-08-30T16:33:53.349, value=28.5   |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:21:34.0 | column=Vendor_details:strip distance, timestamp=2023-08-30T16:33:53.349, value=6.4        |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:DOLocationID, timestamp=2023-08-30T16:33:56.047, value=263          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:FULocationID, timestamp=2023-08-30T16:33:56.047, value=186          |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:RatecodeID, timestamp=2023-08-30T16:33:56.047, value=1              |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:VendorID, timestamp=2023-08-30T16:33:56.047, value=2                |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:airport fee, timestamp=2023-08-30T16:33:56.047, value=0.0           |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 | 00:24:24.0 | column=Vendor_details:congestion surcharge, timestamp=2023-08-30T16:33:56.047, value=0.0  |
| 2017-01-01 | 00:02:17.0 | 2017-01-01 |            |                                                                                           |