

TASK 2:

For creating a table in HBASE

Step 1: login to root user using command `sudo -l`

Step 2: then give command `hbase shell`

Step 3: for creating table give command `create "trip_log_hbase", "cf1"` where `trip_log_hbase` is table name and `cf1` is table column family name.

Step 4: for ingesting the civ file data in hbase database exit hbase cluster using `exit` command the in root user give

`sqoop import --connect jdbc:mysql://demodb.c0tcmelmutf0.us-`

`east-1.rds.amazonaws.com:3306/yellowtrip_data --username admin --password user1234 --table`

`data --hbase-table trip_log_hbase --column-family cf1 --hbase-create-table --hbase-row-key`

`tpep_pickup_datetime,tpep_dropoff_datetime --hbase-bulkload --split-by payment_type`

```
-- jupyter_mac.command -- python - jupyter_mac.command
~/.Downloads -- python p.py
~/MyEC2.pem hadoop@ec2-44-203-179-2.compute-1.amazonaws.com +

E:::::E      EEEEE M:::~::~:M M:::~::~:M RR:::~::R R:::~::R
E:::::E      M:::~::~:M M:::~::~:M M:::~::~:M R:::~::R R:::~::R
E:::::E      M:::~::~:M M:::~::~:M M:::~::~:M R:::~::R R:::~::R
E:::::E      M:::~::~:M M:::~::~:M M:::~::~:M R:::~::R R:::~::R
E:::::E      M:::~::~:M M:::~::~:M M:::~::~:M R:::~::R R:::~::R
E:::::E      M:::~::~:M M:::~::~:M M:::~::~:M R:::~::R R:::~::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:::::E      EEEEE~::~:E M:::~::~:M M:::~::~:M R:::~::R R:::~::R
E:::::E      EEEEE~::~:E M:::~::~:M M:::~::~:M RR:::~::R R:::~::R
EEEEEEEEEEEEEEEE M:::~::~:M M:::~::~:M M:::~::~:M RR:::~::R RR:::~::R

([root@ip-172-31-90-250 ~]# sqoop import --connect jdbc:mysql://demodb.c0tcmelmutf0.us-east-1.rds.amazonaws.com:3306/yellowtrip_data --username admin --password user1234 --table data --hbase-table trip_log_hbase --column-family cf1 --hbase-create-table --hbase-row-key tpep_pickup_datetime,tpep_dropoff_datetime --hbase-bulkload --split-by payment_type
Warning: /usr/lib/sqoop/./hcatalog does not exist! Hcatalog jobs will fail.
Please set $HCAT_HOME to the root of your Hcatalog installation.
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/07/10 16:39:20 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: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
23/07/10 16:39:20 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/07/10 16:39:20 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/07/10 16:39:20 INFO tool.CodeGenTool: Beginning code generation
23/07/10 16:39:21 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `data` AS t LIMIT 1
23/07/10 16:39:21 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `data` AS t LIMIT 1
23/07/10 16:39:21 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/0e684572bae9d2efde4e5ec31508ee63/data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/07/10 16:39:22 INFO orm.CompilationManager: Writing jar files: /tmp/sqoop-root/compile/0e684572bae9d2efde4e5ec31508ee63/data.jar
23/07/10 16:39:22 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/07/10 16:39:22 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/07/10 16:39:23 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/07/10 16:39:23 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/07/10 16:39:23 INFO mapreduce.ImportJobBase: Beginning import of data
23/07/10 16:39:23 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
23/07/10 16:39:23 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
23/07/10 16:39:25 INFO mapreduce.HBaseImportJob: Creating missing HBase table trip_log_hbase
23/07/10 16:39:27 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class?>...) method has been deprecated since it is easy to use incorrectly. Most users should rely on addDependen
cyJars(Job) instead. See HBASE-8386 for more details.
23/07/10 16:39:27 WARN mapreduce.TableMapReduceUtil: The addDependencyJars(Configuration, Class?>...) method has been deprecated since it is easy to use incorrectly. Most users should rely on addDependen
cyJars(Job) instead. See HBASE-8386 for more details.
23/07/10 16:39:27 INFO zlib.ZlibFactory: Successfully loaded & initialized native-zlib library
23/07/10 16:39:27 INFO compress.CodecPool: Got brand-new compressor: [deflate]
23/07/10 16:39:27 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-90-250.ec2.internal/172.31.90.250:8032
23/07/10 16:39:33 INFO db.DBInputFormat: Using read committed transaction isolation
23/07/10 16:39:33 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`payment_type`), MAX(`payment_type`) FROM `data`
23/07/10 16:39:33 INFO db.IntegerSplitter: Split size: 0. Num splits: 4 from: 2 to: 2
23/07/10 16:39:33 INFO mapreduce.JobSubmitter: number of splits:1
23/07/10 16:39:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1689083667402_0001
23/07/10 16:39:34 INFO impl.YarnClientImpl: Submitted application application_1689083667402_0001
23/07/10 16:39:34 INFO mapreduce.Job: The url to track the job: http://ip-172-31-90-250.ec2.internal:20888/proxy/application_1689083667402_0001/
23/07/10 16:39:34 INFO mapreduce.Job: Running job: job_1689083667402_0001
23/07/10 16:39:42 INFO mapreduce.Job: Job job_1689083667402_0001 running in uber mode : false
23/07/10 16:39:42 INFO mapreduce.Job: map 0% reduce 0%
23/07/10 16:39:49 INFO mapreduce.Job: map 100% reduce 0%
23/07/10 16:39:50 INFO mapreduce.Job: map 100% reduce 100%
23/07/10 16:39:56 INFO mapreduce.Job: Job job_1689083667402_0001 completed successfully
23/07/10 16:39:56 INFO mapreduce.Job: Counters: 49
```

```

FILE: Number of bytes read=882
FILE: Number of bytes written=456925
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=119
HDFS: Number of bytes written=8065
HDFS: Number of read operations=7
HDFS: Number of large read operations=0
HDFS: Number of write operations=3
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=419232
  Total time spent by all reduces in occupied slots (ms)=738944
  Total time spent by all map tasks (ms)=4367
  Total time spent by all reduce tasks (ms)=3807
  Total vcore-milliseconds taken by all map tasks=4367
  Total vcore-milliseconds taken by all reduce tasks=3807
  Total megabyte-milliseconds taken by all map tasks=13415424
  Total megabyte-milliseconds taken by all reduce tasks=23390288
Map-Reduce Framework
  Map input records=2
  Map output records=34
  Map output bytes=4918
  Map output materialized bytes=878
  Input split bytes=119
  Combine input records=0
  Combine output records=0
  Reduce input groups=2
  Reduce shuffle bytes=878
  Reduce input records=34
  Reduce output records=34
  Spilled Records=68
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=218
  CPU time spent (ms)=6730
  Physical memory (bytes) snapshot=981692416
  Virtual memory (bytes) snapshot=11945779968
  Total committed heap usage (bytes)=1802962944
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=8065
23/07/09 15:07:47 INFO mapreduce.ImportJobBase: Transferred 7.8662 KB in 23.7119 seconds (339.7029 bytes/sec)
23/07/09 15:07:47 INFO mapreduce.ImportJobBase: Retrieved 34 records.
23/07/09 15:07:47 WARN mapreduce.LoadIncrementalHFiles: managed connection cannot be used for bulkload. Creating unmanaged connection.
23/07/09 15:07:47 INFO impl.MetricsConfig: loaded properties from hadoop-metrics2-hbase.properties
23/07/09 15:07:48 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
23/07/09 15:07:48 INFO impl.MetricsSystemImpl: HBase metrics system started
23/07/09 15:07:48 INFO Configuration.deprecation: hbase.offheapcache.minblocksize is deprecated. Instead, use hbase.blockcache.minblocksize
[root@ip-172-31-92-203 ~]#

```