```
Last login: Wed Jan  8 19:12:58 on ttys000
rupeshande@Rupeshs-MacBook-Air ~ % ssh -ohostkeyalgorithms=ssh-rsa cloudera@192.168.64.5
cloudera@192.168.64.5's password:
Last login: Wed Jan  8 05:43:16 2025 from 192.168.64.1
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
[cloudera@quickstart ~]$ hadoop dfsadmin -safemode leave
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Safe mode is OFF
[cloudera@quickstart ~]$ #sqoop-hive imports begins....
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #Let's load some data nin mysql tables then we can load that tables to hive and we can observe those in hdfs(relation)
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ mysql -uroot -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 22
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

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

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| cm                 |
| firehose           |
| hue                |
| metastore          |
| mysql              |
| nav                |
| navms              |
| oozie              |
| prodb              |
| retail_db          |
| rman               |
| sentry             |
+--------------------+
13 rows in set (0.03 sec)

mysql> create database masterrates;
Query OK, 1 row affected (0.00 sec)

mysql> use masterrates;
Database changed
mysql> show tables;
Empty set (0.01 sec)

mysql> create table user_data(id INT, name VARCHAR(25), amount INT);
Query OK, 0 rows affected (0.07 sec)

mysql> INSERT INTO user_data(id, name, amount)
    -> VALUES(1, 'Govinda', 1000),
    -> (2, 'Kesava', 2000),
    -> (3, 'Vasudeva', 5000),
    -> (4, 'Hare ram', 7000),
    -> (5, 'Achutha', 4000);
Query OK, 5 rows affected (0.06 sec)
Records: 5  Duplicates: 0  Warnings: 0

mysql> select * from user_data;
+------+----------+--------+
| id   | name     | amount |
+------+----------+--------+
|    1 | Govinda  |   1000 |
|    2 | Kesava   |   2000 |
|    3 | Vasudeva |   5000 |
|    4 | Hare ram |   7000 |
|    5 | Achutha  |   4000 |
+------+----------+--------+
5 rows in set (0.01 sec)

mysql> quit;
Bye
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #let's do sqoop import to hive tables...
[cloudera@quickstart ~]$ #before that lets check hive tables...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
default
test
Time taken: 4.081 seconds, Fetched: 2 row(s)
hive> use test;
OK
Time taken: 0.363 seconds
hive> show tables;
OK
Time taken: 0.369 seconds
hive> create database metadata;
OK
Time taken: 11.144 seconds
hive> use metadata;
OK
Time taken: 0.523 seconds
hive> show tables;
OK
Time taken: 0.245 seconds
hive> create table sqoop_import_data(id int, name string, amount int);
OK
Time taken: 1.859 seconds
hive> select * from sqoop_import_data;
OK
Time taken: 2.598 seconds
hive> quit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ # lets import sqoop mysql table data in to hive sqoop_import_data table...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #here goes sqoop command...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/masterrates --username root --password cloudera --table user_data --m 1 --hive-import --hive-database metadata --hive-table sqoop_i
mport_data
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
25/01/08 06:20:23 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
25/01/08 06:20:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
25/01/08 06:20:23 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
25/01/08 06:20:23 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
25/01/08 06:20:24 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
25/01/08 06:20:24 INFO tool.CodeGenTool: Beginning code generation
25/01/08 06:20:27 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:20:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:20:28 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/fc9a69cab7e10a0c6eb9499a0785c0d3/user_data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
25/01/08 06:20:46 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/fc9a69cab7e10a0c6eb9499a0785c0d3/user_data.jar
25/01/08 06:20:46 WARN manager.MySQLManager: It looks like you are importing from mysql.
25/01/08 06:20:46 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
25/01/08 06:20:46 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
```

```
25/01/08 06:20:46 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
25/01/08 06:20:46 INFO mapreduce.ImportJobBase: Beginning import of user_data
25/01/08 06:20:46 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
25/01/08 06:20:48 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
25/01/08 06:20:56 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
25/01/08 06:20:57 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/01/08 06:21:21 INFO db.DBInputFormat: Using read commited transaction isolation
25/01/08 06:21:22 INFO mapreduce.JobSubmitter: number of splits:1
25/01/08 06:21:23 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1736338927837_0001
25/01/08 06:21:29 INFO impl.YarnClientImpl: Submitted application application_1736338927837_0001
25/01/08 06:21:30 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1736338927837_0001/
25/01/08 06:21:30 INFO mapreduce.Job: Running job: job_1736338927837_0001
25/01/08 06:23:04 INFO mapreduce.Job: Job job_1736338927837_0001 running in uber mode : false
25/01/08 06:23:04 INFO mapreduce.Job:  map 0% reduce 0%
25/01/08 06:24:04 INFO mapreduce.Job:  map 100% reduce 0%
25/01/08 06:24:07 INFO mapreduce.Job: Job job_1736338927837_0001 completed successfully
25/01/08 06:24:09 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=151554
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=76
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=51237
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=51237
                Total vcore-milliseconds taken by all map tasks=51237
                Total megabyte-milliseconds taken by all map tasks=52466688
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=350
                CPU time spent (ms)=4900
                Physical memory (bytes) snapshot=102141952
                Virtual memory (bytes) snapshot=2729472000
                Total committed heap usage (bytes)=52822016
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=76
25/01/08 06:24:09 INFO mapreduce.ImportJobBase: Transferred 76 bytes in 192.7937 seconds (0.3942 bytes/sec)
25/01/08 06:24:09 INFO mapreduce.ImportJobBase: Retrieved 5 records.
25/01/08 06:24:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:24:10 INFO hive.HiveImport: Loading uploaded data into Hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.12.0.jar!/hive-log4j.properties
OK
Time taken: 14.666 seconds
Loading data to table metadata.sqoop_import_data
Table metadata.sqoop_import_data stats: [numFiles=1, totalSize=76]
OK
Time taken: 6.335 seconds
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets login in to hive and check the data..
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use metadata;
OK
Time taken: 2.488 seconds
hive> select * from sqoop_import_data;
OK
1       Govinda 1000
2       Kesava  2000
3       Vasudeva        5000
4       Hare ram        7000
5       Achutha 4000
Time taken: 4.738 seconds, Fetched: 5 row(s)
hive> quit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #what if I didn't mention hive database --- then data will go to hive default location instead of mentioned hive database location...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/masterrates --username root --password cloudera --m 1 --table user_data --hive-import --hive-table userdata
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
25/01/08 06:44:18 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
25/01/08 06:44:18 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
25/01/08 06:44:18 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
25/01/08 06:44:18 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
25/01/08 06:44:19 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
25/01/08 06:44:19 INFO tool.CodeGenTool: Beginning code generation
25/01/08 06:44:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:44:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:44:23 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/06009a0bb46a9b97ad92b1efcc468c90/user_data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
25/01/08 06:44:42 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/06009a0bb46a9b97ad92b1efcc468c90/user_data.jar
25/01/08 06:44:42 WARN manager.MySQLManager: It looks like you are importing from mysql.
25/01/08 06:44:42 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
25/01/08 06:44:42 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
25/01/08 06:44:42 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
25/01/08 06:44:42 INFO mapreduce.ImportJobBase: Beginning import of user_data
25/01/08 06:44:42 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
25/01/08 06:44:45 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
25/01/08 06:44:53 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
25/01/08 06:44:54 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/01/08 06:45:07 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 06:45:09 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 06:45:13 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 06:45:15 INFO db.DBInputFormat: Using read commited transaction isolation
25/01/08 06:45:15 INFO mapreduce.JobSubmitter: number of splits:1
25/01/08 06:45:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1736338927837_0002
25/01/08 06:45:19 INFO impl.YarnClientImpl: Submitted application application_1736338927837_0002
25/01/08 06:45:20 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1736338927837_0002/
25/01/08 06:45:20 INFO mapreduce.Job: Running job: job_1736338927837_0002
25/01/08 06:46:26 INFO mapreduce.Job: Job job_1736338927837_0002 running in uber mode : false
25/01/08 06:46:26 INFO mapreduce.Job:  map 0% reduce 0%
```

```
25/01/08 06:47:09 INFO mapreduce.Job:  map 100% reduce 0%
25/01/08 06:47:12 INFO mapreduce.Job: Job job_1736338927837_0002 completed successfully
25/01/08 06:47:13 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=151403
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=76
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=39016
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=39016
                Total vcore-milliseconds taken by all map tasks=39016
                Total megabyte-milliseconds taken by all map tasks=39952384
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=339
                CPU time spent (ms)=4830
                Physical memory (bytes) snapshot=101806080
                Virtual memory (bytes) snapshot=2729467904
                Total committed heap usage (bytes)=52822016
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=76
25/01/08 06:47:13 INFO mapreduce.ImportJobBase: Transferred 76 bytes in 140.0924 seconds (0.5425 bytes/sec)
25/01/08 06:47:13 INFO mapreduce.ImportJobBase: Retrieved 5 records.
25/01/08 06:47:14 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 06:47:14 INFO hive.HiveImport: Loading uploaded data into Hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.12.0.jar!/hive-log4j.properties
OK
Time taken: 19.454 seconds
Loading data to table default.userdata
Table default.userdata stats: [numFiles=1, totalSize=76]
OK
Time taken: 4.784 seconds
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets check in hive..
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
default
metadata
test
Time taken: 4.824 seconds, Fetched: 3 row(s)
hive> use default;
OK
Time taken: 0.245 seconds
hive> show tables;
OK
userdata
Time taken: 0.285 seconds, Fetched: 1 row(s)
hive> select * from userdata;
OK
1       Govinda 1000
2       Kesava  2000
3       Vasudeva        5000
4       Hare ram        7000
5       Achutha 4000
Time taken: 2.923 seconds, Fetched: 5 row(s)
hive> quit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #see this is how data will be moved...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #this data can also be accessed using hdfs as well... lets check those..
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/
Found 3 items
drwxrwxrwx   - cloudera supergroup          0 2025-01-08 06:17 /user/hive/warehouse/metadata.db
drwxrwxrwx   - cloudera supergroup          0 2021-02-04 16:19 /user/hive/warehouse/test.db
drwxrwxrwx   - cloudera supergroup          0 2025-01-08 06:47 /user/hive/warehouse/userdata
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #here we have the same data that we have in hive databases...
[cloudera@quickstart ~]$ #lets check the data in those files...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/userdata/
Found 1 items
-rw-r--r--   1 cloudera cloudera         76 2025-01-08 06:47 /user/hive/warehouse/userdata/part-m-00000
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/userdata/part-m-00000
1Govinda1000
2Kesava2000
3Vasudeva5000
4Hare ram7000
5Achutha4000
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/metadata.db/
Found 1 items
drwxrwxrwx   - cloudera supergroup          0 2025-01-08 06:24 /user/hive/warehouse/metadata.db/sqoop_import_data
[cloudera@quickstart ~]$ hadoop fs -cat /user/hive/warehouse/metadata.db/sqoop_import_data/part-m-00000
1Govinda1000
2Kesava2000
3Vasudeva5000
4Hare ram7000
5Achutha4000
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets create two more tables in mysql and lets import all those tables once at a time to hive using sqoop...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ mysql -uroot -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 31
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

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

mysql> use masterrates;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+------------------------+
| Tables_in_masterrates |
+------------------------+
| user_data              |
+------------------------+
1 row in set (0.02 sec)

mysql> create table students(id int, name VARCHAR(100), fee INT, course_data VARCHAR(50));
Query OK, 0 rows affected (0.09 sec)
```

```
mysql> INSERT INTO students(id, name, fee, course_data)
    -> VALUES (1, 'Rupesh', 3000, 'Computers'),
    -> (2, 'Krishna', 2000, 'Vedas');
Query OK, 2 rows affected (0.04 sec)
Records: 2  Duplicates: 0  Warnings: 0

mysql> select * from students;
+------+---------+------+-------------+
| id   | name    | fee  | course_data |
+------+---------+------+-------------+
|    1 | Rupesh  | 3000 | Computers   |
|    2 | Krishna | 2000 | Vedas       |
+------+---------+------+-------------+
2 rows in set (0.03 sec)

mysql> CREATE TABLE orders(id INT AUTO_INCREMENT, books VARCHAR(50), other_data VARCHAR(30));
ERROR 1075 (42000): Incorrect table definition; there can be only one auto column and it must be defined as a key
mysql> CREATE TABLE orders(id INT, books VARCHAR(50), other_data VARCHAR(30));
Query OK, 0 rows affected (0.03 sec)

mysql> insert into orders values(1, 'Gita', 'Nothing');
Query OK, 1 row affected (0.01 sec)

mysql> select * from orders;
+------+-------+------------+
| id   | books | other_data |
+------+-------+------------+
|    1 | Gita  | Nothing    |
+------+-------+------------+
1 row in set (0.02 sec)

mysql> show tables;
+----------------------+
| Tables_in_masterrates |
+----------------------+
| orders               |
| students             |
| user_data            |
+----------------------+
3 rows in set (0.01 sec)

mysql> quit;
Bye
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets import all this tables directly to hive using sqoop...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ sqoop import-all-tables --connect jdbc:mysql://localhost/masterrates --username root --password cloudera --m 1 --warehouse-dir /user/cloudera/sqoop10
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
25/01/08 07:26:14 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.12.0
25/01/08 07:26:14 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
25/01/08 07:26:16 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
25/01/08 07:26:19 INFO tool.CodeGenTool: Beginning code generation
25/01/08 07:26:19 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders` AS t LIMIT 1
25/01/08 07:26:20 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `orders` AS t LIMIT 1
25/01/08 07:26:20 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/orders.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
25/01/08 07:26:37 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/orders.jar
25/01/08 07:26:37 WARN manager.MySQLManager: It looks like you are importing from mysql.
25/01/08 07:26:37 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
25/01/08 07:26:37 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
25/01/08 07:26:37 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
25/01/08 07:26:37 INFO mapreduce.ImportJobBase: Beginning import of orders
25/01/08 07:26:37 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
25/01/08 07:26:40 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
25/01/08 07:26:45 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
25/01/08 07:26:47 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/01/08 07:26:58 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:26:59 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:00 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:02 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:02 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:02 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:04 INFO db.DBInputFormat: Using read commited transaction isolation
25/01/08 07:27:05 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:27:05 INFO mapreduce.JobSubmitter: number of splits:1
25/01/08 07:27:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1736338927837_0003
25/01/08 07:27:09 INFO impl.YarnClientImpl: Submitted application application_1736338927837_0003
25/01/08 07:27:11 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1736338927837_0003/
25/01/08 07:27:11 INFO mapreduce.Job: Running job: job_1736338927837_0003
25/01/08 07:28:11 INFO mapreduce.Job: Job job_1736338927837_0003 running in uber mode : false
25/01/08 07:28:11 INFO mapreduce.Job:  map 0% reduce 0%
25/01/08 07:29:03 INFO mapreduce.Job:  map 100% reduce 0%
25/01/08 07:29:05 INFO mapreduce.Job: Job job_1736338927837_0003 completed successfully
25/01/08 07:29:06 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=151290
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=15
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
```

```
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=43990
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=43990
                Total vcore-milliseconds taken by all map tasks=43990
                Total megabyte-milliseconds taken by all map tasks=45045760
        Map-Reduce Framework
                Map input records=1
                Map output records=1
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=380
                CPU time spent (ms)=4970
                Physical memory (bytes) snapshot=103772160
                Virtual memory (bytes) snapshot=2729771008
                Total committed heap usage (bytes)=52822016
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=15
25/01/08 07:29:06 INFO mapreduce.ImportJobBase: Transferred 15 bytes in 140.514 seconds (0.1068 bytes/sec)
25/01/08 07:29:06 INFO mapreduce.ImportJobBase: Retrieved 1 records.
25/01/08 07:29:06 INFO tool.CodeGenTool: Beginning code generation
25/01/08 07:29:07 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `students` AS t LIMIT 1
25/01/08 07:29:07 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/students.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
25/01/08 07:29:13 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/students.jar
25/01/08 07:29:13 INFO mapreduce.ImportJobBase: Beginning import of students
25/01/08 07:29:13 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
25/01/08 07:29:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/01/08 07:29:14 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:29:14 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:29:21 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:29:21 INFO db.DBInputFormat: Using read commited transaction isolation
25/01/08 07:29:22 INFO mapreduce.JobSubmitter: number of splits:1
25/01/08 07:29:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1736338927837_0004
25/01/08 07:29:23 INFO impl.YarnClientImpl: Submitted application application_1736338927837_0004
25/01/08 07:29:23 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1736338927837_0004/
25/01/08 07:29:23 INFO mapreduce.Job: Running job: job_1736338927837_0004
25/01/08 07:30:27 INFO mapreduce.Job: Job job_1736338927837_0004 running in uber mode : false
25/01/08 07:30:27 INFO mapreduce.Job:  map 0% reduce 0%
25/01/08 07:31:10 INFO mapreduce.Job:  map 100% reduce 0%
25/01/08 07:31:13 INFO mapreduce.Job: Job job_1736338927837_0004 completed successfully
25/01/08 07:31:14 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=151304
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=45
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=40204
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=40204
                Total vcore-milliseconds taken by all map tasks=40204
                Total megabyte-milliseconds taken by all map tasks=41168896
        Map-Reduce Framework
                Map input records=2
                Map output records=2
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=363
                CPU time spent (ms)=5000
                Physical memory (bytes) snapshot=102936576
                Virtual memory (bytes) snapshot=2729467904
                Total committed heap usage (bytes)=52822016
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=45
25/01/08 07:31:14 INFO mapreduce.ImportJobBase: Transferred 45 bytes in 120.4072 seconds (0.3737 bytes/sec)
25/01/08 07:31:14 INFO mapreduce.ImportJobBase: Retrieved 2 records.
25/01/08 07:31:14 INFO tool.CodeGenTool: Beginning code generation
25/01/08 07:31:14 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `user_data` AS t LIMIT 1
25/01/08 07:31:14 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/user_data.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
25/01/08 07:31:20 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/fb5d26df9add5ea408deca4732561b93/user_data.jar
25/01/08 07:31:20 INFO mapreduce.ImportJobBase: Beginning import of user_data
25/01/08 07:31:20 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
25/01/08 07:31:21 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/01/08 07:31:28 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
25/01/08 07:31:28 INFO db.DBInputFormat: Using read commited transaction isolation
25/01/08 07:31:28 INFO mapreduce.JobSubmitter: number of splits:1
25/01/08 07:31:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1736338927837_0005
25/01/08 07:31:29 INFO impl.YarnClientImpl: Submitted application application_1736338927837_0005
25/01/08 07:31:29 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1736338927837_0005/
25/01/08 07:31:29 INFO mapreduce.Job: Running job: job_1736338927837_0005
25/01/08 07:32:29 INFO mapreduce.Job: Job job_1736338927837_0005 running in uber mode : false
25/01/08 07:32:29 INFO mapreduce.Job:  map 0% reduce 0%
25/01/08 07:33:17 INFO mapreduce.Job:  map 100% reduce 0%
25/01/08 07:33:21 INFO mapreduce.Job: Job job_1736338927837_0005 completed successfully
25/01/08 07:33:21 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=151297
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=76
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

```
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=41296
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=41296
                Total vcore-milliseconds taken by all map tasks=41296
                Total megabyte-milliseconds taken by all map tasks=42287104
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=474
                CPU time spent (ms)=4720
                Physical memory (bytes) snapshot=102891520
                Virtual memory (bytes) snapshot=2734600192
                Total committed heap usage (bytes)=52822016
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=76
25/01/08 07:33:21 INFO mapreduce.ImportJobBase: Transferred 76 bytes in 121.4293 seconds (0.6259 bytes/sec)
25/01/08 07:33:21 INFO mapreduce.ImportJobBase: Retrieved 5 records.
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets verify the data in hive and hdfs, either all tables are imported or not..
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets go first with hdfs..
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/sqoop10
Found 3 items
drwxr-xr-x   - cloudera cloudera          0 2025-01-08 07:29 /user/cloudera/sqoop10/orders
drwxr-xr-x   - cloudera cloudera          0 2025-01-08 07:31 /user/cloudera/sqoop10/students
drwxr-xr-x   - cloudera cloudera          0 2025-01-08 07:33 /user/cloudera/sqoop10/user_data
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/sqoop10/orders/part-m-00000
1,Gita,Nothing
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/sqoop10/students/part-m-00000
1,Rupesh,3000,Computers
2,Krishna,2000,Vedas
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/sqoop10/user_data/part-m-00000
1,Govinda,1000
2,Kesava,2000
3,Vasudeva,5000
4,Hare ram,7000
5,Achutha,4000
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #lets check in hive now...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use databases;
FAILED: SemanticException [Error 10072]: Database does not exist: databases
hive> use metadata;
OK
Time taken: 0.365 seconds
hive> show tables;
OK
sqoop_import_data
Time taken: 1.59 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
metadata
test
Time taken: 0.209 seconds, Fetched: 3 row(s)
hive> use default;
OK
Time taken: 0.118 seconds
hive> show tables;
OK
userdata
Time taken: 0.17 seconds, Fetched: 1 row(s)
hive> quit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#release() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ #loading data to hive form hdfs location which is /user/cloudera/sqoop10 will be taken as part of hive execution commands...
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ #done with sqoop - hive - mysql commands....
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
```