

```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
HiveShim: hive --help
Applications Places System
Thu Mar 16, 10:14 AM cloudera
cloudera@quickstart-~
File Edit View Search Terminal Help
[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> create database hive_1;
OK
Time taken: 4.82 seconds
hive> use hive_1;
OK
Time taken: 0.178 seconds
hive> create table department_data
> (
>   dept_id int,
>   dept_name string,
>   manager_id int,
>   salary int
> )
> row format delimited
> fields terminated by ','
> ;
OK
Time taken: 2.489 seconds
hive> show tables;
OK
department_data
Time taken: 0.511 seconds, Fetched: 1 row(s)
hive> describe department_data;
OK
dept_id          int
dept_name        string
manager_id       int
salary           int
Time taken: 0.684 seconds, Fetched: 4 row(s)
hive> describe formatted department_data;
OK
# col_name      data_type      comment
dept_id         int
dept_name       string
manager_id      int
salary          int

# Detailed Table Information
Database:      hive_1
Owner:         cloudera
CreateTime:    Thu Mar 16 10:03:32 PDT 2023
LastAccessTime: UNKNOWN
Protect Mode:   None
Retention:     0

```

```
cloudera-quickstart-vm-5.13.0-0-vmtoolsd [Running] - Oracle VM VirtualBox
Applications Places System
Thu Mar 16, 11:18 AM cloudera
cloudera@quickstart:/tmp/hive_prac

File Edit View Search Terminal Help
Location: hdfs://quickstart.cloudera:8020/user/hive/warehouse/hive1.db/department_data
Table Type: MANAGED_TABLE
Table Parameters:
  transient_lastDedTime 1678986212

# Storage Information
SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat: org.apache.hadoop.mapred.TextInputFormat
OutputFormat: org.apache.hadoop.hive.q.ln.HiveIgnoreKeyTextOutputFormat
orcat
Compressed: No
Num Buckets: -1
Bucket Columns: []
Sort Columns: []
Storage Desc Params:
  field.delim serialization.format
Time taken: 0.42 seconds, Fetched: 30 row(s)
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory.release() is invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ cd data
[cloudera@quickstart data]$ ls
depart_data.csv
[cloudera@quickstart data]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart data]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart data]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart data]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a98b4c82b4c_resources hiveserde2_data_spark2 hiveserde2_data_spark2
3379d74f-6e25-4c5a-93d6-7172af8f717_resources hiveserde2_data_spark2 hiveserde2_data_spark2
8028b05a-f684-4d48-3144-a4e8a8d4bc_resources hiveserde2_data_spark2 hiveserde2_data_spark2
cloudera hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar1613581116795580364 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar201880632308460 hiveserde2_data_spark2 hiveserde2_data_spark2
hadoop-unjar253723280854304934 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
hiveserde2_data_spark2 hiveserde2_data_spark2 hiveserde2_data_spark2
[cloudera@quickstart ~]$ cd /tmp/hive1/
bash: cd: /tmp/hive1/: No such file or directory
[cloudera@quickstart ~]$ ls /tmp/hive*
ls: cannot open directory /tmp/hive: Permission denied
[cloudera@quickstart ~]$ cd
[cloudera@quickstart ~]$ mkdir /tmp/hive_prac
[cloudera@quickstart ~]$ ls /tmp
0816d664-7653-4272-b074-1a8ed8ed94b_resources hiveserde2_data_spark hiveserde2_data_spark2
1ab39fb5-a3e5-47f9-8e43-a
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:tmp/hive_prac

File Edit View Search Terminal Help
hadoop-unjar1615581316759580384 hspferdata_hive impala-scratch Jetty_0_0_0_9995_thrift___.vgz9l orbit-cloudera
hadoop-unjar20018086323300460 hspferdata_hdfs Jetty_0_0_0_18002_hiveserver2___.any- Jetty_localhost_52587_datanode___.sl974y pulse-0r1kzWTF8568
hadoop-unjar2537320868954389434 hspferdata_impala Jetty_0_0_0_19088_jobhistory___.djqltw Jetty_localhost_58160_datanode___.drsi7 tmpRSHBqX
hadoop-unjar3403436857533133159 hspferdata_mapred Jetty_0_0_0_50070_hdfs___.w2cu88 keyring-20kcnx
hbase-hbase hspferdata_oozie Jetty_0_0_0_50090_secondary___.y8aanv keyring-40z5n3
hspferdata_sentry Jetty_0_0_0_60010_master___.q3mwon libnetty-transport-native-epoll2007954657341198839.so

[cloudera@quickstart ~]$ cat data
cat: data: Is a directory
[cloudera@quickstart ~]$ cd data
[cloudera@quickstart data]$ ls
depart_data.csv
[cloudera@quickstart data]$ mv ~/depart_data /tmp/hive_prac
mv: cannot stat '/home/cloudera/depart_data': No such file or directory
[cloudera@quickstart data]$ mv ~/data/depart_data /tmp/hive_prac
mv: cannot stat '/home/cloudera/data/depart_data': No such file or directory
[cloudera@quickstart data]$ cd
[cloudera@quickstart ~]$ mv ~/data/depart_data /tmp/hive_prac
mv: cannot stat '/home/cloudera/data/depart_data': No such file or directory
[cloudera@quickstart ~]$ mv ~/data/depart_data.csv /tmp/hive_prac
[cloudera@quickstart ~]$ cd /tmp/hive_prac
[cloudera@quickstart tmp/hive_prac]$ ls
depart_data.csv
[cloudera@quickstart tmp/hive_prac]$ 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 hive_1;
OK
Time taken: 0.876 seconds
hive> show tables;
OK
department_data
Time taken: 0.629 seconds, Fetched: 1 row(s)
hive> exit;
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 tmp/hive_prac]$ 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> load data local inpath 'file:///tmp/hive_prac/depart_data.csv' into tab
le department_data;
FAILED: SemanticException [Error 18001]: Line 1:74 Table not found 'departmen
t_data'
hive> use hive_1;
OK
Time taken: 0.297 seconds
hive> load data local inpath 'file:///tmp/hive_prac/depart_data.csv' into tab
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:tmp/hive_prac

File Edit View Search Terminal Help
le department_data;
Loading data to table hive_1.department_data
Table hive_1.department_data stats: [numFiles=1, totalSize=6551]
OK
Time taken: 2.462 seconds
hive> select * from department_data;
OK
10 Administration 200 1700
20 Marketing 201 1800
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
150 Shareholder Services NULL 1700
160 Benefits NULL 1700
170 Manufacturing NULL 1700
180 Construction NULL 1700
190 Contracting NULL 1700
200 Operations NULL 1700
210 IT Support NULL 1700
220 NOC NULL 1700
230 IT Helpdesk NULL 1700
240 Government Sales NULL 1700
250 Retail Sales NULL 1700
260 Recruiting NULL 1700
270 Payroll NULL 1700
Time taken: 1.475 seconds, Fetched: 27 row(s)
hive> select count(*) from department table;
FAILED: SemanticException [Error 18001]: Line 1:21 Table not found 'departmen
table'
hive> select count(*) from department data;
FAILED: SemanticException [Error 18001]: Line 1:21 Table not found 'departmen
table'
hive> select count(*) from department_data;
Query ID = cloudera_20230316113333_8c36967-7668-430c-9fcd-8175b1d91e88
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 4:41 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
le department data;
Loading data to table hive_1.department data
Table hive_1.department data stats: [numFiles=1, totalSize=655]
OK
Time taken: 2.462 seconds
hive> select * from department data;
OK
10 Administration 200 1700
20 Marketing 201 1800
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
150 Shareholder Services NULL 1700
160 Benefits NULL 1700
170 Manufacturing NULL 1700
180 Construction NULL 1700
190 Contracting NULL 1700
200 Operations NULL 1700
210 IT Support NULL 1700
220 NOC NULL 1700
230 IT Helpdesk NULL 1700
240 Government Sales NULL 1700
250 Retail Sales NULL 1700
260 Recruiting NULL 1700
270 Payroll NULL 1700
Time taken: 1.475 seconds, Fetched: 27 row(s)
hive> select count(*) from department table;
FAILED: SemanticException [Error 10001]: Line 1:21 Table not found 'department
table'
hive> select count(*) from department data;
FAILED: SemanticException [Error 10001]: Line 1:21 Table not found 'department
data'
hive> select count(*) from department data;
Query ID = cloudera_20230316113333_0ca369d7-7668-430c-9fcd-9775b1d91e88
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
  set mapreduce.job.reduces=<number>
  set mapreduce.job.reduces.max=<number>
Starting Job = job_1678959727197_0001, Tracking URL = http://quickstart.cloud
era:0008/pxy/application/1678959727197_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678959727197_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-16 11:34:06,475 Stage-1 map = 0%, reduce = 0%
2023-03-16 11:34:33,191 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.83 sec
2023-03-16 11:34:51,168 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.28 sec
MapReduce Total cumulative CPU time: 17 seconds 280 msec
Ended Job = job_1678959727197_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 17.28 sec HDFS Read: 8222 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 280 msec
OK
27
Time taken: 99.746 seconds, Fetched: 1 row(s)
hive>
> exit;
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 hive_prac]$ 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> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart hive_prac]$ hadoop fs -ls /user/hive/warehouse/hive_prac
db
ls: /user/hive/warehouse/hive_prac.db: No such file or directory
[cloudera@quickstart hive_prac]$ hadoop fs -ls /user/hive/warehouse/hive_prac
db/department_data
ls: /user/hive/warehouse/hive_prac.db/department_data: No such file or dire
ctory
[cloudera@quickstart hive_prac]$ cd
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/hive_prac.db/depa
rtment_data
ls: /user/hive/warehouse/hive_prac.db/department_data: No such file or dire
ctory
[cloudera@quickstart ~]$ hadoop fs -mkdir /tmp/hive_prac1
[cloudera@quickstart ~]$ cd /tmp/hive_prac1
bash: cd: /tmp/hive_prac1: No such file or directory
[cloudera@quickstart ~]$ hadoop fs -ls
[cloudera@quickstart ~]$ hadoop fs -ls /tmp/hive_prac1
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 4:43 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678959727197_0001, Tracking URL = http://quickstart.cloud
era:0008/pxy/application/1678959727197_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678959727197_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-16 11:34:06,475 Stage-1 map = 0%, reduce = 0%
2023-03-16 11:34:33,191 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.83 sec
2023-03-16 11:34:51,168 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 17.28 sec
MapReduce Total cumulative CPU time: 17 seconds 280 msec
Ended Job = job_1678959727197_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 17.28 sec HDFS Read: 8222 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 280 msec
OK
27
Time taken: 99.746 seconds, Fetched: 1 row(s)
hive>
> exit;
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 hive_prac]$ 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> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart hive_prac]$ hadoop fs -ls /user/hive/warehouse/hive_prac
db
ls: /user/hive/warehouse/hive_prac.db: No such file or directory
[cloudera@quickstart hive_prac]$ hadoop fs -ls /user/hive/warehouse/hive_prac
db/department_data
ls: /user/hive/warehouse/hive_prac.db/department_data: No such file or dire
ctory
[cloudera@quickstart hive_prac]$ cd
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/hive_prac.db/depa
rtment_data
ls: /user/hive/warehouse/hive_prac.db/department_data: No such file or dire
ctory
[cloudera@quickstart ~]$ hadoop fs -mkdir /tmp/hive_prac1
[cloudera@quickstart ~]$ cd /tmp/hive_prac1
bash: cd: /tmp/hive_prac1: No such file or directory
[cloudera@quickstart ~]$ hadoop fs -ls
[cloudera@quickstart ~]$ hadoop fs -ls /tmp/hive_prac1
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
cloudera@quickstart:~$ ls
cloudera-manager Documents kerberos Public workspace
cs_api.py Downloads lib Templates
create eclipse Music test2.txt
data enterprise-deployment.json parcels Videos
Desktop express-deployment.json Pictures vigya
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> exit;
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:~$ hadoop fs -ls /user/hive/hive_1.db/department_data
ls: /user/hive/hive_1.db/department_data: No such file or directory
cloudera@quickstart:~$ hadoop fs -ls /user/hive/hive_1.db/department_data
ls: /user/hive/hive_1.db/department_data: No such file or directory
cloudera@quickstart:~$ hadoop fs -ls /user/hive/warehouse/hive_1.db/department_data
Found 1 items
-rwxrwxrwx 1 cloudera supergroup 655 2023-03-16 11:30 /user/hive/warehouse/hive_1.db/department_data/depart_data.csv
cloudera@quickstart:~$ hadoop fs -ls /user/hive/warehouse/hive_1.db/department_data/depart_data.csv
-rwxrwxrwx 1 cloudera supergroup 655 2023-03-16 11:30 /user/hive/warehouse/hive_1.db/department_data/depart_data.csv
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> select * from department_data;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'department_data'
hive> select * from depart_data;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'depart_data'
hive> select * from depart_data.csv;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'csv'
hive> use hive 1;
OK
Time taken: 0.192 seconds
hive> select * from department_data;
OK
10 Administration 200 1700
20 Marketing 201 1800
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
150 Shareholder Services NULL 1700
160 Benefits NULL 1700
170 Manufacturing NULL 1700
180 Construction NULL 1700
190 Contracting NULL 1700
200 Operations NULL 1700
210 IT Support NULL 1700
220 NOC NULL 1700
230 IT Helpdesk NULL 1700
240 Government Sales NULL 1700
250 Retail Sales NULL 1700
260 Recruiting NULL 1700
270 Payroll NULL 1700
Time taken: 1.701 seconds, Fetched: 27 row(s)
hive> set hive.cli.print.header = true;
hive> select * from department_data;
OK
department_data.dept_id department_data.dept_name department_data.manager_id department_data.salary
10 Administration 200 1700
20 Marketing 201 1800
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
150 Shareholder Services NULL 1700
160 Benefits NULL 1700
170 Manufacturing NULL 1700
180 Construction NULL 1700
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 5:37 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
208 Operations NULL 1788
218 IT Support NULL 1788
228 NOC NULL 1788
238 IT Helpdesk NULL 1788
248 Government Sales NULL 1788
258 Retail Sales NULL 1788
268 Recruiting NULL 1788
278 Payroll NULL 1788
Time taken: 0.253 seconds, Fetched: 27 row(s)
hive> exit;
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 ~]$ hadoop fs -ls /tmp
Found 4 items
drwxr-xr-x - mapred mapred 0 2017-10-23 09:15 /tmp/hadoop-yarn
drwxr-xr-x - hive supergroup 0 2023-03-16 06:35 /tmp/hive
drwxr-xr-x - cloudera supergroup 0 2023-03-16 16:25 /tmp/hive_prac1
drwxr-xr-x - mapred hadoop 0 2017-10-23 09:17 /tmp/logs
[cloudera@quickstart ~]$ hadoop fs -mkdir /tmp/hive_prac1
mkdir: '/tmp/hive_prac1': File exists
[cloudera@quickstart ~]$ hadoop fs -put /tmp/hive_prac/depart_data.csv /tmp/hive_prac1
[cloudera@quickstart ~]$ hdfs dfs -ls /tmp/hive_prac1
Found 1 items
-rw-r--r-- 1 cloudera supergroup 655 2023-03-16 16:51 /tmp/hive_prac1/depart_data.csv
[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 hive1;
OK
Time taken: 0.911 seconds
hive> create table department_data_from_hdfs
> (
> dept_id int,
> dept_name string,
> manager_id int,
> salary int
> )
> row format delimited
> fields terminated by ',';
OK
Time taken: 1.165 seconds
hive> show tables
> exit;
FAILED: ParseException line 3:8 extraneous input 'exit' expecting EOF near '<
EOF>'
hive> load data inpath '/tmp/hive_prac1/'
> load data inpath '/tmp/hive_prac1/';

Cloudera Live: Welco... Cloudera Live: Welco... Cloudera Live: Welco... cloudera@quickstart:~$
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 5:37 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
MismatchedTokenException[154]=141
    at org.apache.hadoop.hive.ql.parse.HiveParser.loadStatement(HiveParser.java:1769)
    at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1544)
    at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1065)
    at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:281)
    at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:166)
    at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1356)
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1473)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1285)
    at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:226)
    at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:175)
    at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:38)
    at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:781)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:699)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:634)
    at sun.reflect.NativeMethodAccessorImpl.invoke(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 2:8 mismatched input 'load' expecting INTO near '
/tmp/hive_prac1/' in load statement
hive> load data inpath '/tmp/hive_prac1/' into table department_data_from_hdf
s;
Loading data to table hive1.department_data_from_hdfs
Table hive1.department_data_from_hdfs Stats: [numFiles=1, totalSize=655]
OK
Time taken: 1.834 seconds
hive> set hive.cli.print.header = true;
hive> select * from department_data_from_hdfs;
OK
department_data_from_hdfs.dept_id    department_data_from_hdfs.dept_name d
```



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 5:38 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
department_data_from_hdfs.salary
10 Administration 200 1700
20 Marketing 201 1800
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
150 Shareholder Services NULL 1700
160 Benefits NULL 1700
170 Manufacturing NULL 1700
180 Construction NULL 1700
190 Contracting NULL 1700
200 Operations NULL 1700
210 IT Support NULL 1700
220 HOC NULL 1700
230 IT Helpdesk NULL 1700
240 Government Sales NULL 1700
250 Retail Sales NULL 1700
260 Recruiting NULL 1700
270 Payroll NULL 1700
Time taken: 1.082 seconds, Fetched: 27 row(s)
hive> create external table department_data_external
> (
> dept_id int,
> dept_name string,
> manager_id int,
> salary int
> )
> row format delimited
> fields terminated by ','
> location '/tmp/hive_prac/';
OK
Time taken: 0.349 seconds
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ hadoop fs -put /tmp/hive_prac/department_data.csv /tmp/h
ive_prac/
[cloudera@quickstart ~]$ hadoop fs -ls /tmp/hive_prac
Found 1 items
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Mar 16, 5:39 PM cloudera
cloudera@quickstart:~$
File Edit View Search Terminal Help
~$ cd /tmp/hive_prac
1/department_data.csv
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4
j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use hive_prac;
FAILED: SemanticException [Error 10072]: Database does not exist: hive_prac
hive>
> set hive.cli.print.header = true;
hive> select * from department_data_external;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'departmen
t_data_external'
hive> select * from department_data_external;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'departmen
t_data_external'
hive> exit;
WARN: The method class org.apache.commons.logging.impl.SLF4JLogFactory#releas
e() was invoked.
WARN: Please see http://www.slf4j.org/codes.html#release for an explanation.
[cloudera@quickstart ~]$ hive
[cloudera@quickstart ~]$
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4
j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use hive_prac;
FAILED: SemanticException [Error 10072]: Database does not exist: hive_prac
hive> use hive_prac;
FAILED: SemanticException [Error 10072]: Database does not exist: hive_prac
hive> use hive_i;
OK
Time taken: 0.15 seconds
hive> select * from department_data_external;
OK
10 Administration 200 1700
20 Marketing 201 1800
30 Purchasing 114 1700
40 Human Resources 203 2400
50 Shipping 121 1500
60 IT 103 1400
70 Public Relations 204 2700
80 Sales 145 2500
90 Executive 100 1700
100 Finance 108 1700
110 Accounting 205 1700
120 Treasury NULL 1700
130 Corporate Tax NULL 1700
140 Control And Credit NULL 1700
```



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
hive> load data local inpath 'file:///tmp/hive_1/array_data.csv' into table employee;
FAILED: SemanticException Line 1:23 Invalid path 'file:///tmp/hive_1/array_data.csv': No files matching path file:///tmp/hive_1/array_data.csv
hive> load data local inpath 'file:///tmp/hive_prac/array_data.csv' into table employee;
Loading data to table default.employee
Table default.employee stats: [numFiles=1, totalSize=215]
OK
Time taken: 1.871 seconds
hive> use hive_prac;
FAILED: SemanticException [Error 10072]: Database does not exist: hive_prac
hive> use hive_1;
OK
Time taken: 0.896 seconds
hive> select * from employee;
OK
Time taken: 1.891 seconds
hive> select * from employee;
OK
Time taken: 0.244 seconds
hive> exit;
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 ~]$ 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> select * from employee;
OK
101 Amit [HADOOP,HIVE,SPARK,BIG-DATA]
102 Sumit [HIVE,HOZZIE,HADOOP,SPARK,STORM]
103 Rohit [KAFKA,CASSANDRA,HBASE]
Time taken: 2.391 seconds, Fetched: 3 row(s)
hive> set hive.cli.print.header = true
> set hive.cli.print.header = true;
Warning: Value had a \n character in it.
hive> select * from employee;
OK
101 Amit [HADOOP,HIVE,SPARK,BIG-DATA]
102 Sumit [HIVE,HOZZIE,HADOOP,SPARK,STORM]
103 Rohit [KAFKA,CASSANDRA,HBASE]
Time taken: 0.358 seconds, Fetched: 3 row(s)
hive> select id,name,skills[0] as prime_skill from employee;
OK
101 Amit HADOOP
102 Sumit HIVE
103 Rohit KAFKA
Time taken: 0.411 seconds, Fetched: 3 row(s)
hive> select
> id,
> name,
```

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
cloudera@quickstart:~$
File Edit View Search Terminal Help
> size(skills) as size of each array,
> array contains(skills,HADOOP) as knows_hadoop,
> sort_array(skills) as sorted_array
> from employee;
FAILED: ParseException line 7:8 extraneous input 'from employee' expecting EOF near '<EOF>'
hive> from employee;
org.apache.hadoop.hive.ql.parse.HiveParser.singleFromStatement(HiveParser.java:4136)
at org.apache.hadoop.hive.ql.parse.HiveParser.fromStatement(HiveParser.java:41028)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatementExpression(HiveParser.java:40695)
at org.apache.hadoop.hive.ql.parse.HiveParser.queryStatementExpression(HiveParser.java:40586)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1529)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1065)
at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:281)
at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:166)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:522)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1356)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1473)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1285)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1275)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:226)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:175)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:38)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:781)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:699)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:634)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:666)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:13 required (...) loop did not match anything at input '<EOF>' in statement
```



```
OK
Time taken: 1.912 seconds
hive> select
  > id,
  > name,
  > details["gender"] as employee_gender
  > from employee_map_data;
OK
101      Amit      M
102      Sumit      M
103      Mansi      F
Time taken: 1.213 seconds, Fetched: 3 row(s)
hive> select
  > id,
  > name,
  > details,
  > size(details) as size_of_each_map,
  > map_keys(details) as distinct_map_keys,
  > map_values(details) as distinct_map_values
  > from employee_map_data;
OK
101      Amit      [{"age":21,"gender":"M"}] 2 [{"age","gender"}] [{"21","M"}]
102      Sumit      [{"age":24,"gender":"M"}] 2 [{"age","gender"}] [{"24","M"}]
103      Mansi      [{"age":23,"gender":"F"}] 2 [{"age","gender"}] [{"23","F"}]
Time taken: 0.383 seconds, Fetched: 3 row(s)
hive>
```

