

Assignment 14.1

Introduction to Hive

Problem Statement:

Create a database named 'custom'.

Create a table named temperature_data inside custom having below fields:

1. date (mm-dd-yyyy) format
2. zip code
3. temperature

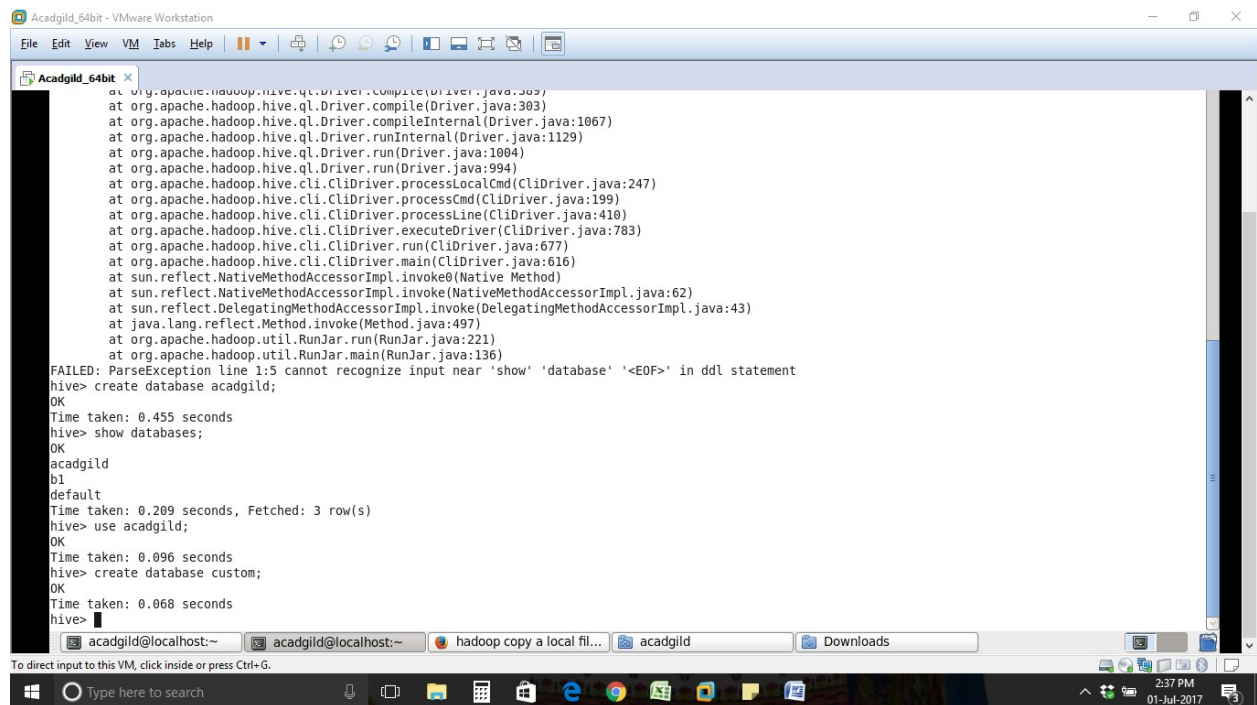
The table will be loaded from comma-delimited file.

Load the dataset.txt (which is ',' delimited) in the table.

1. To create database 'custom'

Command is

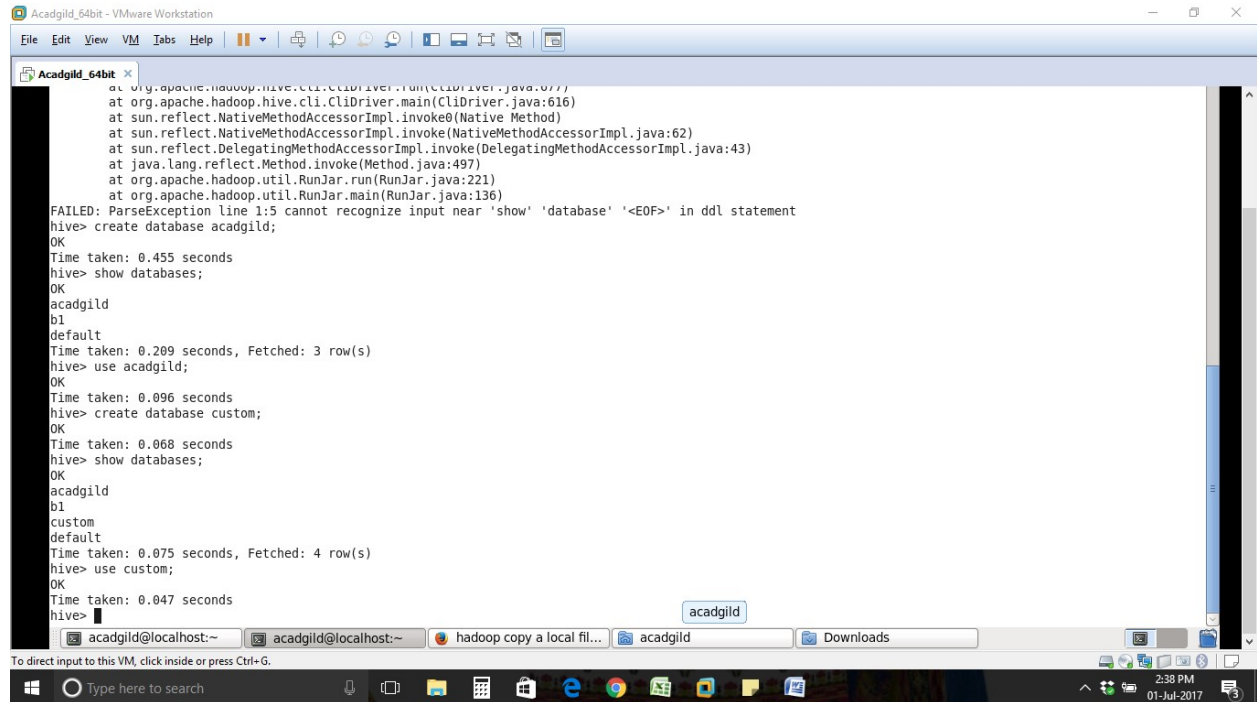
create database custom;



```
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:303)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:303)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1067)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1129)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1004)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:994)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:247)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:199)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:410)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:783)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:677)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:616)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:497)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:5 cannot recognize input near 'show' 'database' '<EOF>' in ddl statement
hive> create database acadgild;
OK
Time taken: 0.455 seconds
hive> show databases;
OK
acadgild
default
Time taken: 0.209 seconds, Fetched: 3 row(s)
hive> use acadgild;
OK
Time taken: 0.096 seconds
hive> create database custom;
OK
Time taken: 0.068 seconds
hive>
```

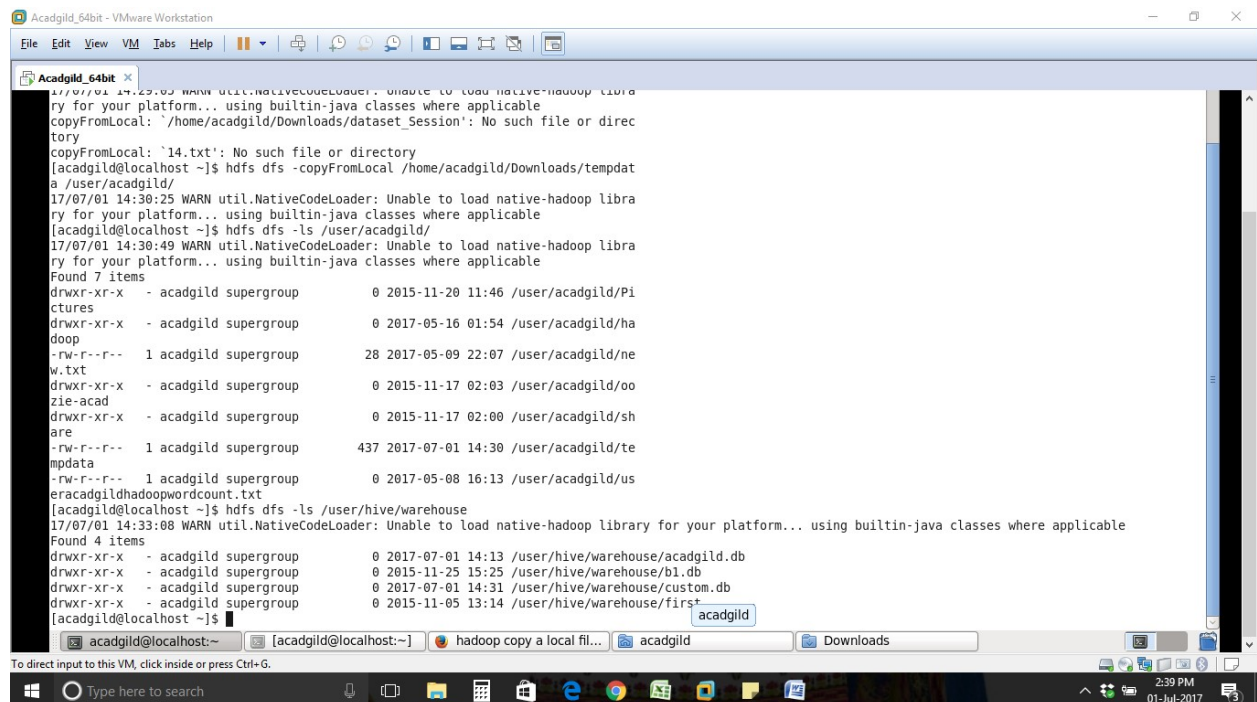
Command : show databases;

Shows databases in hive



```
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:877)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:616)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:497)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:5 cannot recognize input near 'show' 'database' '<EOF>' in ddl statement
hive> create database acadgild;
OK
Time taken: 0.455 seconds
hive> show databases;
OK
acadgild
b1
default
Time taken: 0.209 seconds, Fetched: 3 row(s)
hive> use acadgild;
OK
Time taken: 0.096 seconds
hive> create database custom;
OK
Time taken: 0.068 seconds
hive> show databases;
OK
acadgild
b1
custom
default
Time taken: 0.075 seconds, Fetched: 4 row(s)
hive> use custom;
OK
Time taken: 0.047 seconds
hive>
```

Checking whether a new folder in same name as the database 'custom' is created in hdfs by hive

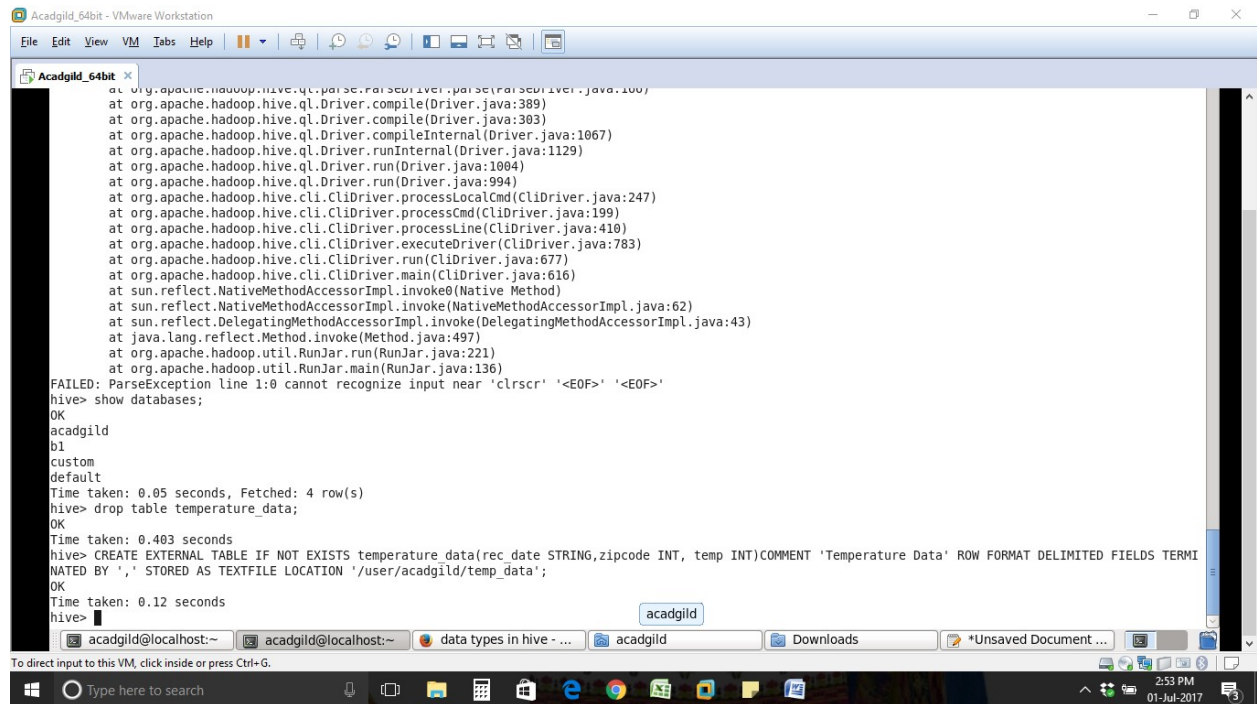


```
17/07/01 14:23:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyFromLocal: '/home/acadgild/Downloads/dataset_Session': No such file or directory
copyFromLocal: '/14.txt': No such file or directory
[acadgild@localhost ~]$ hdfs dfs -copyFromLocal /home/acadgild/Downloads/tempdata /user/acadgild/
17/07/01 14:30:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/
17/07/01 14:30:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - acadgild supergroup 0 2015-11-20 11:46 /user/acadgild/Pictures
drwxr-xr-x - acadgild supergroup 0 2017-05-16 01:54 /user/acadgild/hadoop
-rw-r--r-- 1 acadgild supergroup 28 2017-05-09 22:07 /user/acadgild/new.txt
drwxr-xr-x - acadgild supergroup 0 2015-11-17 02:03 /user/acadgild/oozie-acad
drwxr-xr-x - acadgild supergroup 0 2015-11-17 02:00 /user/acadgild/share
-rw-r--r-- 1 acadgild supergroup 437 2017-07-01 14:30 /user/acadgild/tempdata
-rw-r--r-- 1 acadgild supergroup 0 2017-05-08 16:13 /user/acadgild/useracadgildhadoopwordcount.txt
[acadgild@localhost ~]$ hdfs dfs -ls /user/hive/warehouse
17/07/01 14:33:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup 0 2017-07-01 14:13 /user/hive/warehouse/acadgild.db
drwxr-xr-x - acadgild supergroup 0 2015-11-25 15:25 /user/hive/warehouse/b1.db
drwxr-xr-x - acadgild supergroup 0 2017-07-01 14:31 /user/hive/warehouse/custom.db
drwxr-xr-x - acadgild supergroup 0 2015-11-05 13:14 /user/hive/warehouse/first
```

Creating Table temperature_data in custom database

Command

```
CREATE EXTERNAL TABLE IF NOT EXISTS temperature_data( rec_date STRING, zipcode INT, temp INT)
COMMENT 'Temperature Data' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS
TEXTFILE LOCATION '/user/acadgild/temp_data';
```

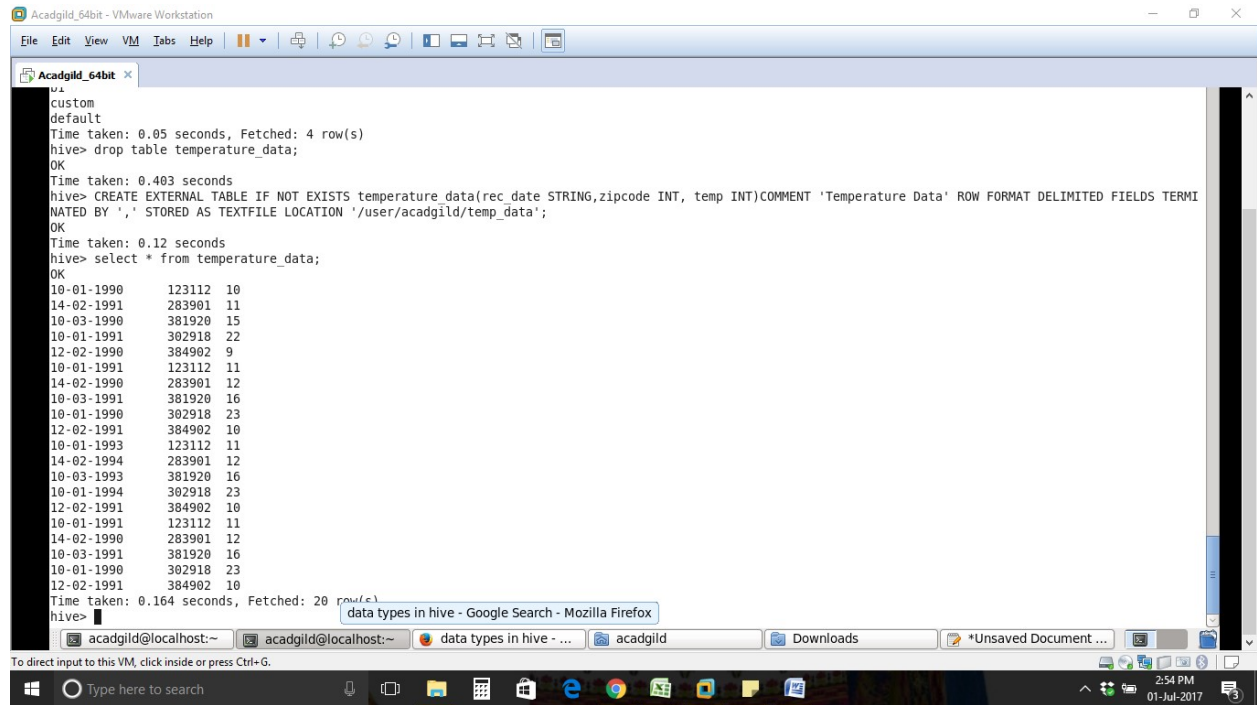


```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
at org.apache.hadoop.hive.ql.parse.ParserDriver.parse(ParserDriver.java:100)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:389)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:303)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1067)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1129)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1004)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:994)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:247)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:199)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:410)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:783)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:677)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:616)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:497)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:0 cannot recognize input near 'clrsr' '<EOF>' '<EOF>'
hive> show databases;
OK
acadgild
b1
custom
default
Time taken: 0.05 seconds, Fetched: 4 row(s)
hive> drop table temperature_data;
OK
Time taken: 0.403 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temperature_data(rec_date STRING, zipcode INT, temp INT) COMMENT 'Temperature Data' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/acadgild/temp_data';
OK
Time taken: 0.12 seconds
hive>
```

Data from table `temperature_data` is retrieved by command:

Select `*` from `temperature_data`;

Screenshot shows data from text file is loaded into the table



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Custom
default
Time taken: 0.05 seconds, Fetched: 4 row(s)
hive> drop table temperature_data;
OK
Time taken: 0.403 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temperature_data(rec_date STRING,zipcode INT, temp INT)COMMENT 'Temperature Data' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/acadgild/temp_data';
OK
Time taken: 0.12 seconds
hive> select * from temperature_data;
OK
10-01-1990      123112  10
14-02-1991      283901  11
10-03-1990      381920  15
10-01-1991      302918  22
12-02-1990      384902   9
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
10-01-1993      123112  11
14-02-1994      283901  12
10-03-1993      381920  16
10-01-1994      302918  23
12-02-1991      384902  10
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
Time taken: 0.164 seconds, Fetched: 20 row(s)
hive>
```

data types in hive - Google Search - Mozilla Firefox

acadgild@localhost:~ acadgild@localhost:~ data types in hive - ... acadgild Downloads *Unsaved Document ...

Type here to search

2:54 PM 01-Jul-2017