

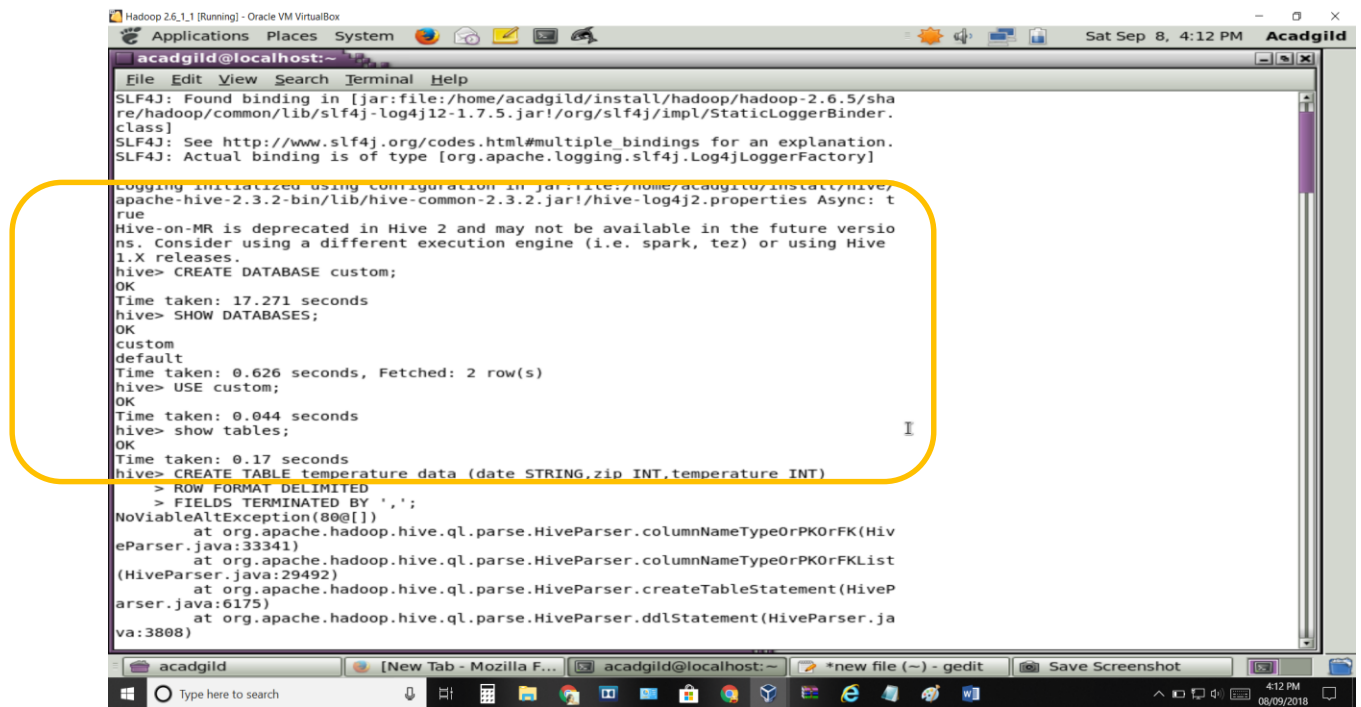
Big Data Hadoop 'Session 8 HIVE BASICS'

DATASET

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

## Task 1

- Create a database named 'custom'



```
acadmild@localhost:~$ cat /etc/hadoop/conf/hive/hive-site.xml
SLF4J: Found binding in [jar:file:/home/acadmild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.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.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/acadmild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> CREATE DATABASE custom;
OK
Time taken: 17.271 seconds
hive> SHOW DATABASES;
OK
custom
default
Time taken: 0.626 seconds, Fetched: 2 row(s)
hive> USE custom;
OK
Time taken: 0.044 seconds
hive> show tables;
OK
Time taken: 0.17 seconds
hive> CREATE TABLE temperature_data (date STRING,zip INT,temperature INT)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
Nov1ableAltException(80@[])
at org.apache.hadoop.hive.ql.parse.HiveParser.columnNameTypeOrPKOrFK(HiveParser.java:3334)
at org.apache.hadoop.hive.ql.parse.HiveParser.columnNameTypeOrPKOrFKList(HiveParser.java:29492)
at org.apache.hadoop.hive.ql.parse.HiveParser.createTableStatement(HiveParser.java:6175)
at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:3808)
```

- 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.

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:31 cannot recognize input near 'date' 'string' ',' in column name or primary key or foreign
key
hive> CREATE TABLE temperature_data(full_date string,zip int,temperature int)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.216 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hiveInput.txt' INTO TABLE custom.temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 3.4 seconds
hive> select * from custom.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: 5.709 seconds, Fetched: 20 row(s)
hive>
```

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

```
Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:31 cannot recognize input near 'date' 'string' ',' in column name or primary key or foreign
key
hive> CREATE TABLE temperature_data(full_date string,zip int,temperature int)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.216 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hiveInput.txt' INTO TABLE custom.temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 3.4 seconds
hive> select * from custom.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: 5.709 seconds, Fetched: 20 row(s)
hive>
```

## Task 2

- Fetch date and temperature from temperature\_data where zip code is greater than 300000 and less than 399999.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
10-03-1991 381920 16  
10-01-1990 302918 23  
12-02-1991 384902 10  
Time taken: 5.709 seconds, Fetched: 20 row(s)  
hive> select full_date,temperature from custom.temperature_data where zip > 300000;  
FAILED: SemanticException [Error 10004]: Line 1:7 Invalid table alias or column reference 'full_date': (possible column names are: full_date, zip, temperature)  
hive> select full_date,temperature from custom.temperature_data where zip > 300000;  
OK  
10-03-1990 15  
10-01-1991 22  
12-02-1990 9  
10-03-1991 16  
10-01-1990 23  
12-02-1991 10  
10-03-1993 16  
10-01-1994 23  
12-02-1991 10  
10-03-1991 16  
10-01-1990 23  
12-02-1991 10  
Time taken: 0.86 seconds, Fetched: 12 row(s)  
hive> select full_date,temperature from custom.temperature_data where zip > 300000 AND zip < 399999;  
OK  
10-03-1990 15  
10-01-1991 22  
12-02-1990 9  
10-03-1991 16  
10-01-1990 23  
12-02-1991 10  
10-03-1993 16  
10-01-1994 23  
12-02-1991 10  
10-03-1991 16  
10-01-1990 23  
12-02-1991 10  
Time taken: 0.622 seconds, Fetched: 12 row(s)  
hive>
```