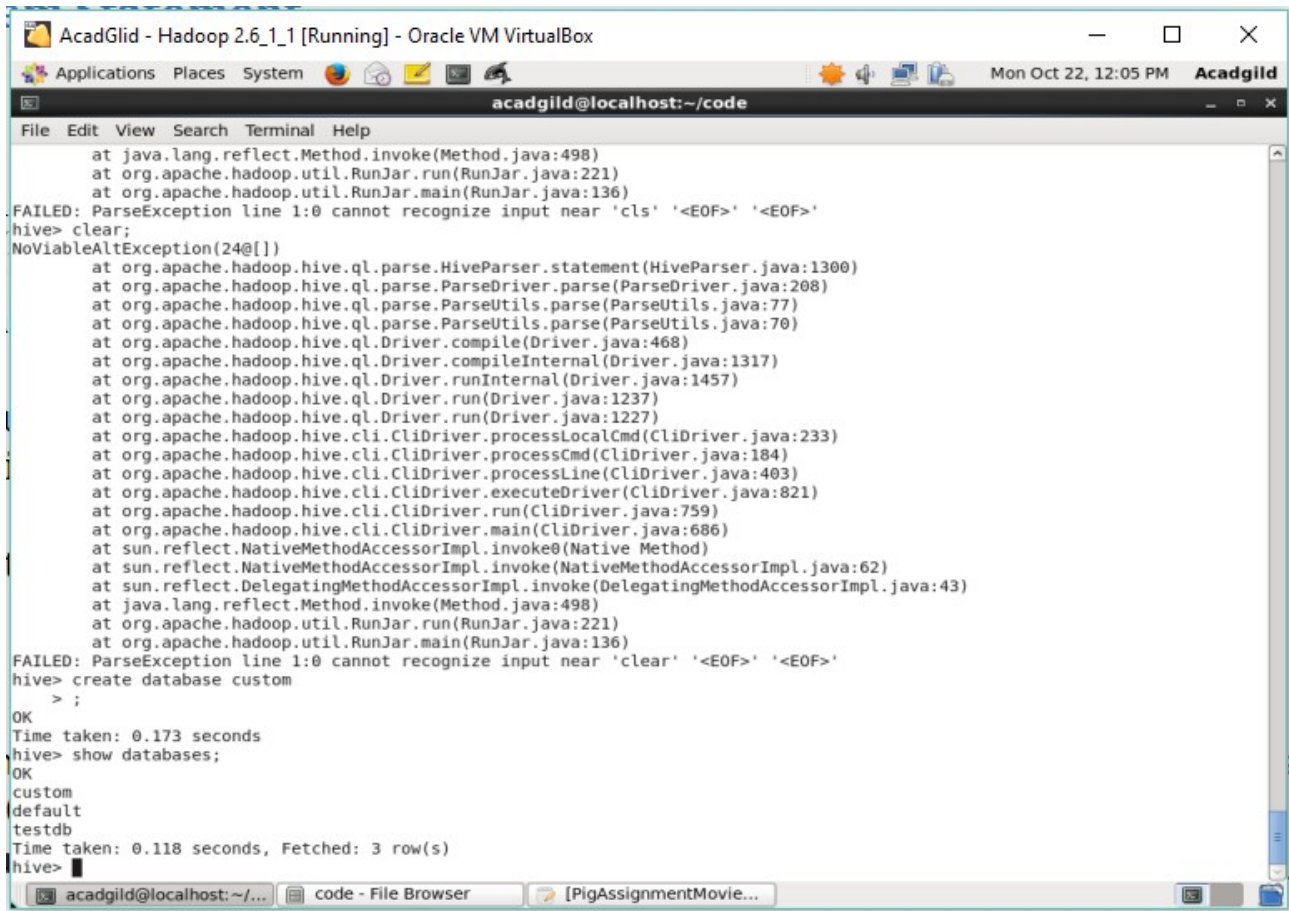


Task 1

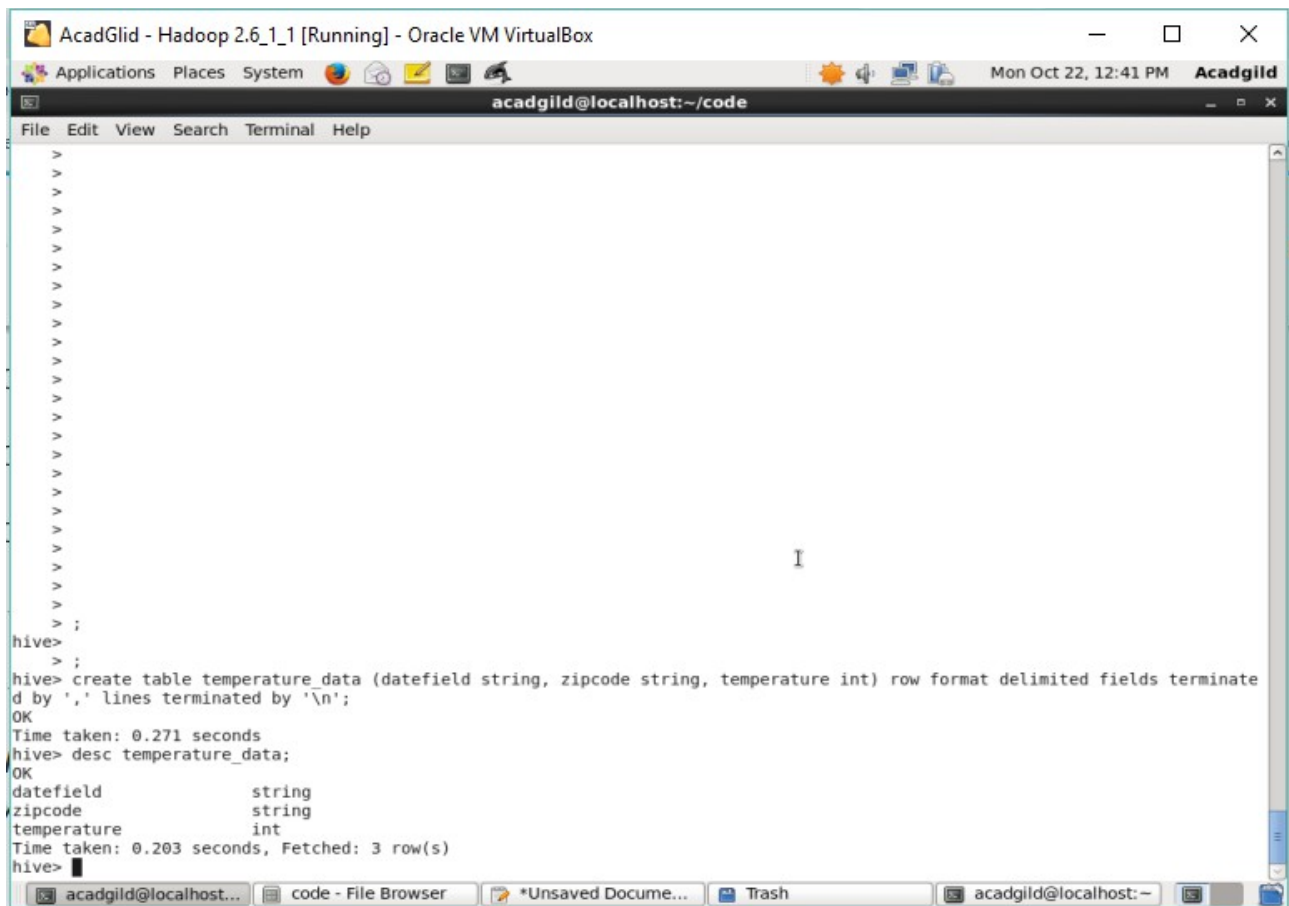
1. Create database named 'custom'



The screenshot shows a terminal window titled 'AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox'. The terminal is running Hive commands. The first command is 'clear;', which results in a 'FAILED: ParseException line 1:0 cannot recognize input near 'cls' '<EOF>' '<EOF>'' error. The second command is 'create database custom', which is successful. The third command is 'show databases;', which lists 'custom', 'default', and 'testdb'. The terminal also shows the time taken for each command and the number of rows fetched for the 'show databases;' command.

```
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 1:0 cannot recognize input near 'cls' '<EOF>' '<EOF>'
hive> clear;
NoViableAltException(24@[])
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1300)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
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:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:0 cannot recognize input near 'clear' '<EOF>' '<EOF>'
hive> create database custom
> ;
OK
Time taken: 0.173 seconds
hive> show databases;
OK
custom
default
testdb
Time taken: 0.118 seconds, Fetched: 3 row(s)
hive>
```

2. create a table named temperature_data with columns
date (mm-dd-yyyy) format
zip code
temperature
Table to be loaded from comma -delimited file



The screenshot shows a VirtualBox window titled "AcadGlid - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox". The window contains a terminal window titled "acadgild@localhost:~/code". The terminal shows a Hive session with the following commands and output:

```
hive> ;
hive> ;
hive> create table temperature_data (datefield string, zipcode string, temperature int) row format delimited fields terminated by ',' lines terminated by '\n';
OK
Time taken: 0.271 seconds
hive> desc temperature_data;
OK
datefield          string
zipcode            string
temperature        int
Time taken: 0.203 seconds, Fetched: 3 row(s)
hive>
```

The terminal window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The bottom of the window shows a taskbar with icons for "acadgild@localhost...", "code - File Browser", "*Unsaved Docume...", "Trash", and "acadgild@localhost:~".

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

AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox

Applications Places System Mon Oct 22, 12:23 PM Acadgild

acadgild@localhost:~/code

```
File Edit View Search Terminal Help
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
at sun.reflect.NativeMethodAccessorImpl.invoke(Native Method)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
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' 'date' ',' in column name or primary key or foreign key
hive> create table temperature_data (datefield date, zipcode string, temperature int) row format delimited fields terminated
by ',' lines terminated by '\n';
OK
Time taken: 2.047 seconds
hive> show tables;
OK
temperature_data
Time taken: 0.117 seconds, Fetched: 1 row(s)
hive> describe temprature_data;
FAILED: SemanticException [Error 10001]: Table not found temprature_data
hive> describe temperature_data;
OK
datefield          date
zipcode            string
temperature         int
Time taken: 0.456 seconds, Fetched: 3 row(s)
hive> load data local inpath '/home/acadgild/input/dataset_Session_14.txt' into table temperature_data;
Loading data to table testdb.temperature_data
OK
Time taken: 5.098 seconds
hive>
```

AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox

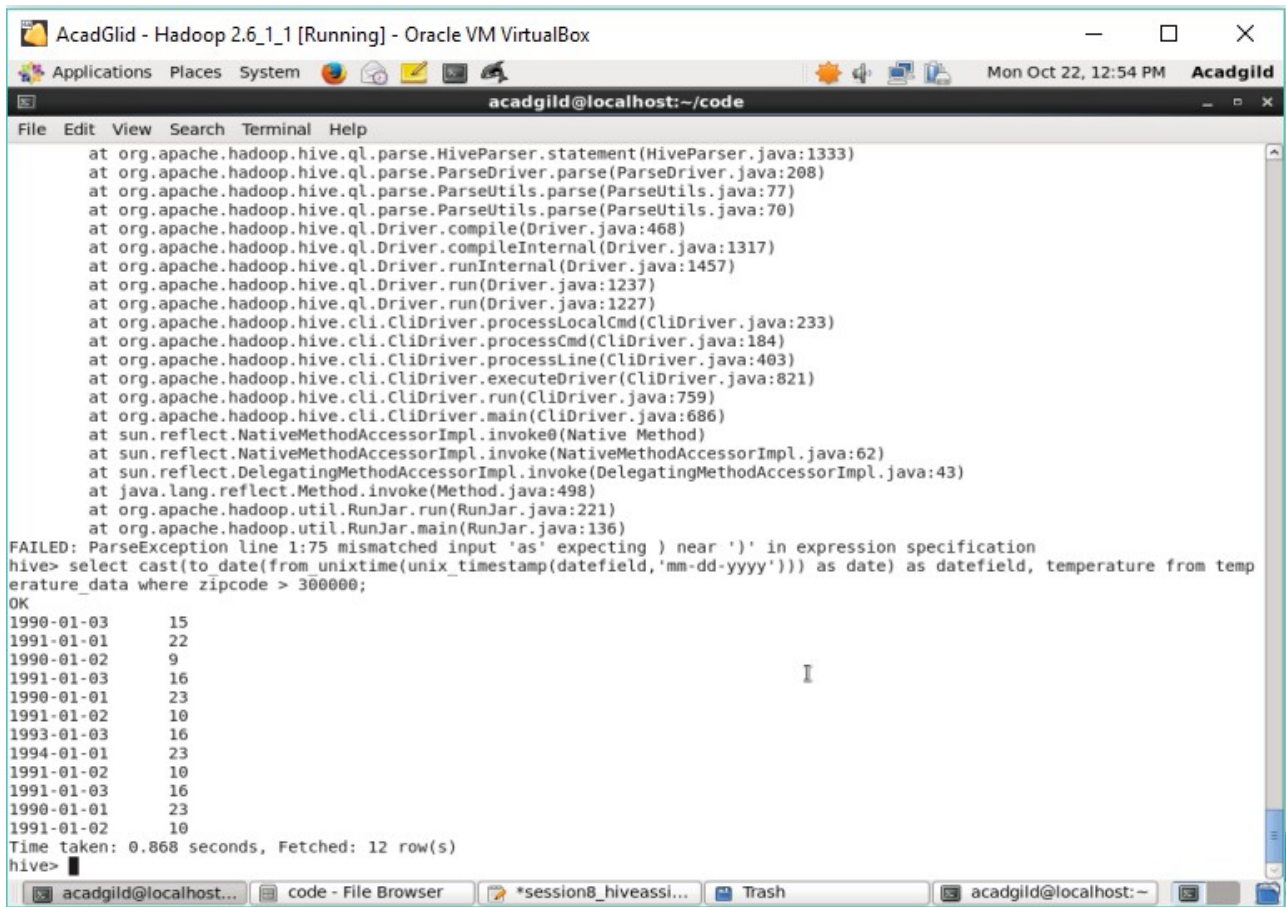
Applications Places System Mon Oct 22, 12:43 PM Acadgild

acadgild@localhost:~/code

```
File Edit View Search Terminal Help
> ;
hive> create table temperature_data (datefield string, zipcode string, temperature int) row format delimited fields terminate
d by ',' lines terminated by '\n';
OK
Time taken: 0.271 seconds
hive> desc temperature_data;
OK
datefield          string
zipcode            string
temperature         int
Time taken: 0.203 seconds, Fetched: 3 row(s)
hive> load data local inpath '/home/acadgild/input/dataset_Session_14.txt' into table temperature_data;
Loading data to table testdb.temperature_data
OK
Time taken: 1.845 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
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
Time taken: 0.512 seconds, Fetched: 20 row(s)
hive>
```

Task 2

1. Select date and temperature where zipcode > 300000



```
AcadGild - Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System Mon Oct 22, 12:54 PM Acadgild
acadgild@localhost:~/code
File Edit View Search Terminal Help
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1333)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:208)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:77)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:70)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:468)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1317)
at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1457)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1237)
at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1227)
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:233)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:184)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:403)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:686)
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:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:75 mismatched input 'as' expecting ')' near ')' in expression specification
hive> select cast(to date(from unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as date) as datefield, temperature from temp
erature_data where zipcode > 300000;
OK
1990-01-03      15
1991-01-01      22
1990-01-02       9
1991-01-03      16
1990-01-01      23
1991-01-02      10
1993-01-03      16
1994-01-01      23
1991-01-02      10
1991-01-03      16
1990-01-01      23
1991-01-02      10
Time taken: 0.868 seconds, Fetched: 12 row(s)
hive>
```

2. getting grouped on year max temperature

AcadGild - Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox

Applications Places System Mon Oct 22, 1:06 PM Acadgild

acadgild@localhost:~/code

File Edit View Search Terminal Help

```
OK
1990 10
1991 11
1990 15
1991 22
1990 9
1991 11
1990 12
1991 16
1990 23
1991 10
1993 11
1994 12
1993 16
1994 23
1991 10
1991 11
1990 12
1991 16
1990 23
1991 10
Time taken: 0.401 seconds, Fetched: 20 row(s)
hive> select year(cast(to date(from unixtime(unix timestamp(datefield,'mm-dd-yyyy')) as date)), max(temperature) from temper
ature_data group by year(cast(to date(from unixtime(unix timestamp(datefield,'mm-dd-yyyy')) as date)));
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20181022130635_139aa09d-b62d-4068-a411-7ed4b069dbbf
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0002, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0002
code - File Browser
```

acadgild@localhost... [code - File Browser] *session8_hiveassi... Trash [acadgild@localhos...]

AcadGild - Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox

Applications Places System Mon Oct 22, 1:07 PM Acadgild

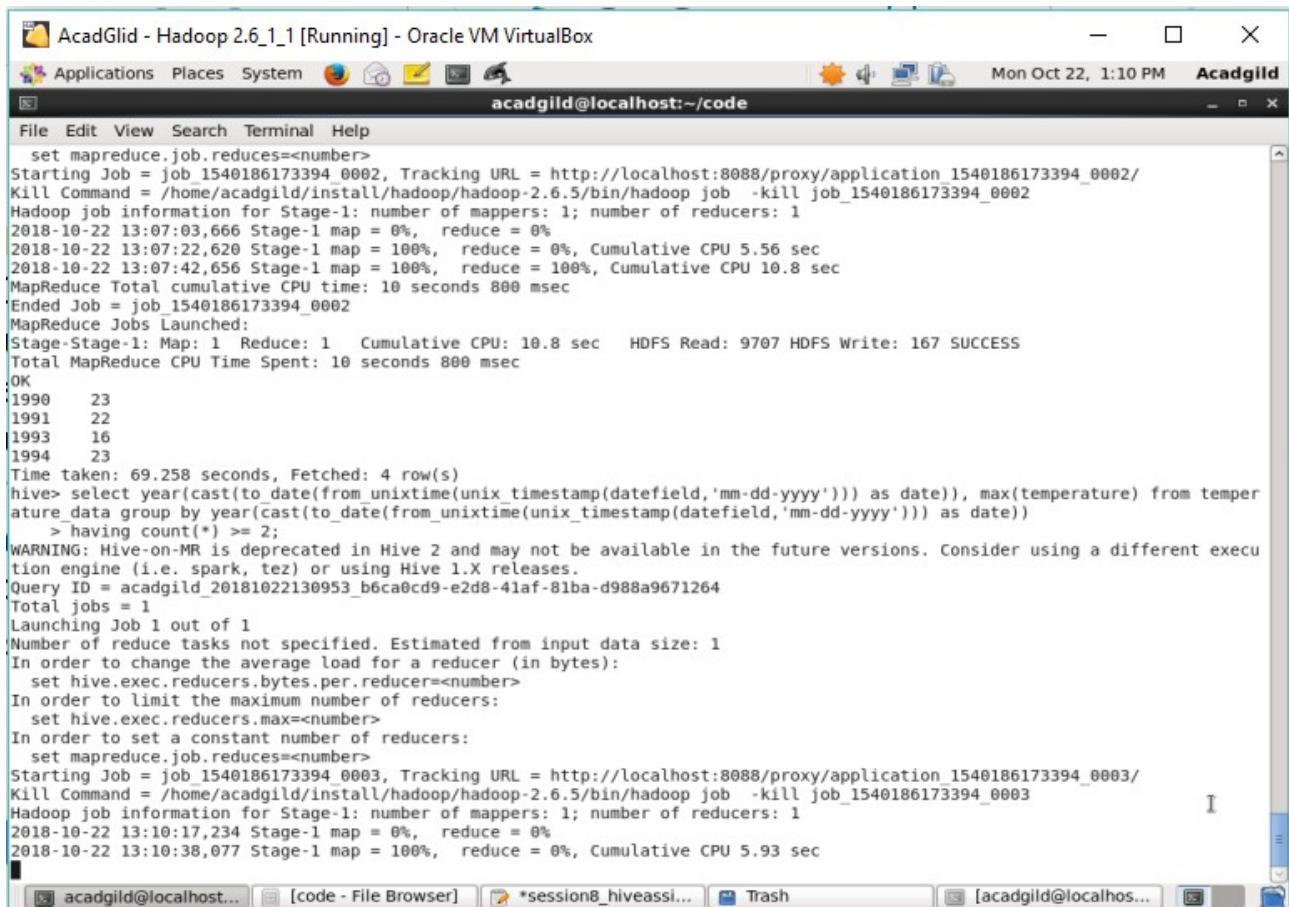
acadgild@localhost:~/code

File Edit View Search Terminal Help

```
1991 10
1991 11
1990 12
1991 16
1990 23
1991 10
Time taken: 0.401 seconds, Fetched: 20 row(s)
hive> select year(cast(to date(from unixtime(unix timestamp(datefield,'mm-dd-yyyy')) as date)), max(temperature) from temper
ature_data group by year(cast(to date(from unixtime(unix timestamp(datefield,'mm-dd-yyyy')) as date)));
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20181022130635_139aa09d-b62d-4068-a411-7ed4b069dbbf
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0002, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-10-22 13:07:03,666 Stage-1 map = 0%, reduce = 0%
2018-10-22 13:07:22,620 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.56 sec
2018-10-22 13:07:42,656 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.8 sec
MapReduce Total cumulative CPU time: 10 seconds 800 msec
Ended Job = job_1540186173394_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.8 sec HDFS Read: 9707 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 800 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 69.258 seconds, Fetched: 4 row(s)
hive>
```

acadgild@localhost... [code - File Browser] *session8_hiveassi... Trash [acadgild@localhos...]

3. select temperature from those years which have at least 2 entries



```
AcadGild - Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Mon Oct 22, 1:10 PM Acadgild
acadgild@localhost:~/code
File Edit View Search Terminal Help
set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0002, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-10-22 13:07:03,666 Stage-1 map = 0%, reduce = 0%
2018-10-22 13:07:22,620 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.56 sec
2018-10-22 13:07:42,656 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.8 sec
MapReduce Total cumulative CPU time: 10 seconds 800 msec
Ended Job = job_1540186173394_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.8 sec HDFS Read: 9707 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 800 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 69.258 seconds, Fetched: 4 row(s)
hive> select year(cast(to_date(from_unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as date)), max(temperature) from temper
ature data group by year(cast(to_date(from_unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as date))
> having count(*) >= 2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20181022130953_b6ca0cd9-e2d8-41af-81ba-d988a9671264
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
In order to set a constant number of reducers:
set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0003, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-10-22 13:10:17,234 Stage-1 map = 0%, reduce = 0%
2018-10-22 13:10:38,077 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.93 sec
```

```
AcadGlid - Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Mon Oct 22, 1:11 PM Acadgild
acadgild@localhost:~/code
File Edit View Search Terminal Help
1990 23
1991 22
1993 16
1994 23
Time taken: 69.258 seconds, Fetched: 4 row(s)
hive> select year(cast(to date(from unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as date)), max(temperature) from temper
ature_data group by year(cast(to_date(from_unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as date))
> having count(*) >= 2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20181022130953_b6ca0cd9-e2d8-41af-81ba-d988a9671264
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0003, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-10-22 13:10:17,234 Stage-1 map = 0%, reduce = 0%
2018-10-22 13:10:38,077 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.93 sec
2018-10-22 13:11:00,437 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 11.24 sec
2018-10-22 13:11:03,374 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 13.91 sec
MapReduce Total cumulative CPU time: 13 seconds 910 msec
Ended Job = job_1540186173394_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 13.91 sec HDFS Read: 10649 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 910 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 71.491 seconds, Fetched: 4 row(s)
hive>
```

4. create view on last query

```
AcadGlid - Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System Mon Oct 22, 1:13 PM Acadgild
acadgild@localhost:~/code
File Edit View Search Terminal Help
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1540186173394_0003, Tracking URL = http://localhost:8088/proxy/application_1540186173394_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1540186173394_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-10-22 13:10:17,234 Stage-1 map = 0%, reduce = 0%
2018-10-22 13:10:38,077 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.93 sec
2018-10-22 13:11:00,437 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 11.24 sec
2018-10-22 13:11:03,374 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 13.91 sec
MapReduce Total cumulative CPU time: 13 seconds 910 msec
Ended Job = job_1540186173394_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 13.91 sec HDFS Read: 10649 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 910 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 71.491 seconds, Fetched: 4 row(s)
hive> create view temperature_data_vw as select year(cast(to date(from unixtime(unix_timestamp(datefield,'mm-dd-yyyy')) as d
ate)), max(temperature) from temperature_data group by year(cast(to_date(from_unixtime(unix_timestamp(datefield,'mm-dd-yyyy')
)) as date)) having count(*) >= 2;
OK
Time taken: 0.844 seconds
hive> desc temperature_data_vw
> ;
OK
_c0 int
_c1 int
Time taken: 0.151 seconds, Fetched: 2 row(s)
hive>
```

5. export view output to a file