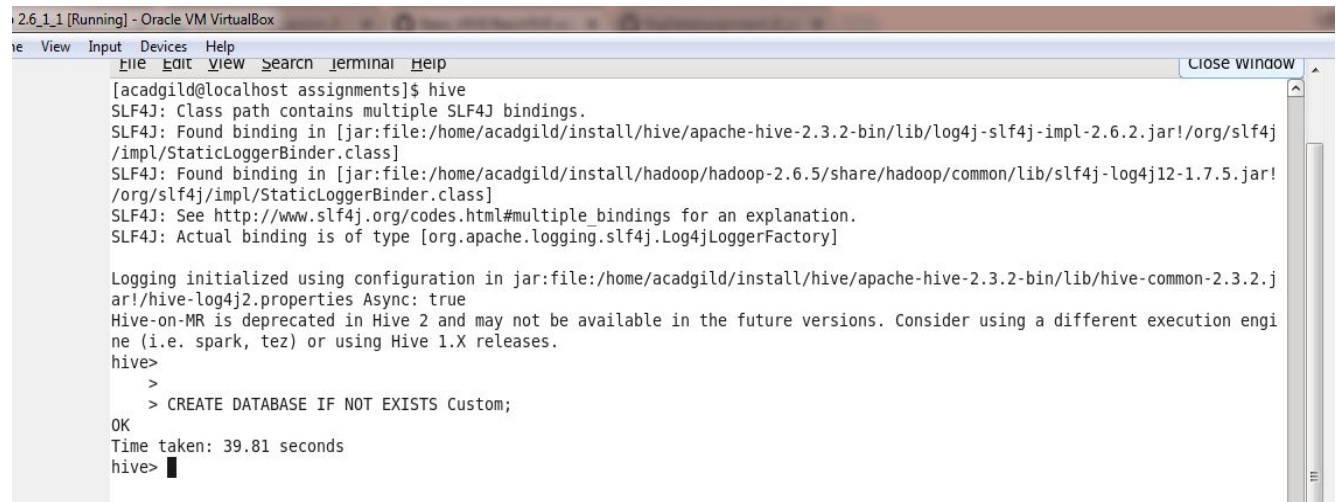


Task 1

Create a database named 'custom'.



```
2.6_1.1 [Running] - Oracle VM VirtualBox
File Edit View Search Terminal Help
[acadgild@localhost assignments]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/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]

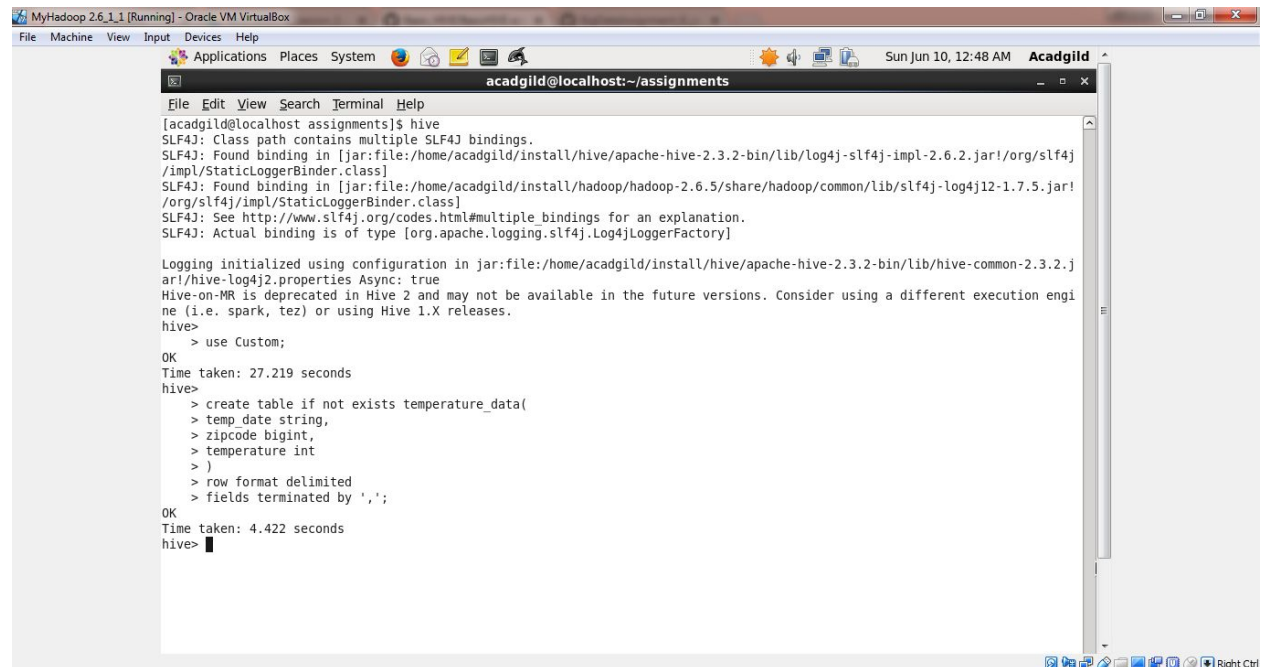
Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
>
> CREATE DATABASE IF NOT EXISTS Custom;
OK
Time taken: 39.81 seconds
hive>
```

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.



```
MyHadoop 2.6_1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
acadgild@localhost:~/assignments
File Edit View Search Terminal Help
[acadgild@localhost assignments]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/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/acadgild/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>
> use Custom;
OK
Time taken: 27.219 seconds
hive>
> create table if not exists temperature_data(
> temp_date string,
> zipcode bigint,
> temperature int
> )
> row format delimited
> fields terminated by ',';
OK
Time taken: 4.422 seconds
hive>
```

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
> create table if not exists temperature_data(
> temp_date string,
> zipcode bigint,
> temperature int
> )
> row format delimited
> fields terminated by ',';
OK
Time taken: 4.422 seconds
hive>
> load data local inpath '/home/acadgild/assignments/dataset.txt' into table temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 10.007 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: 15.773 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.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
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: 4.227 seconds, Fetched: 12 row(s)
hive>
> select temp_date, temperature from temperature_data where zipcode > 300000 and zipcode < 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: 4.227 seconds, Fetched: 12 row(s)
hive>
```

- Calculate maximum temperature corresponding to every year from temperature_data table.

MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Sun Jun 10, 1:02 AM Acadgild

acadmild@localhost:~/assignments

```
File Edit View Search Terminal Help
hive> select max(temperature), substring( temp_date, 7, 10) from temperature_data group by substring( temp_date, 7, 10 );
WARNING: 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.
Query ID = acadmild_20180610005846_a26cfb97-76e5-427e-b981-09bb4729a98e
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_1528036691222_0033, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0033/
Kill Command = /home/acadmild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0033
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-10 01:00:30,756 Stage-1 map = 0%, reduce = 0%
2018-06-10 01:01:11,815 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.5 sec
2018-06-10 01:01:42,086 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 13.15 sec
2018-06-10 01:01:44,584 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 15.19 sec
MapReduce Total cumulative CPU time: 15 seconds 190 msec
Ended Job = job_1528036691222_0033
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 15.19 sec HDFS Read: 9288 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 15 seconds 190 msec
OK
23      1990
22      1991
16      1993
23      1994
Time taken: 181.084 seconds, Fetched: 4 row(s)
hive>
```

- Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.

MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Sun Jun 10, 1:06 AM Acadgild

s8 - Google Drive - Mozilla Firefox

acadmild@localhost:~/assignments

```
File Edit View Search Terminal Help
hive> select max(temperature), substring( temp_date, 7, 10) from temperature_data group by substring( temp_date, 7, 10) having count( substring( temp_date, 7, 10 )) >=2 ;
WARNING: 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.
Query ID = acadmild_20180610010427_d3452dcl-dce4-4710-b565-497f618957ea
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_1528036691222_0034, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0034/
Kill Command = /home/acadmild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0034
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-10 01:05:10,372 Stage-1 map = 0%, reduce = 0%
2018-06-10 01:05:41,008 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.74 sec
2018-06-10 01:06:02,432 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.77 sec
MapReduce Total cumulative CPU time: 8 seconds 770 msec
Ended Job = job_1528036691222_0034
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.77 sec HDFS Read: 10181 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 770 msec
OK
23      1990
22      1991
16      1993
23      1994
Time taken: 97.396 seconds, Fetched: 4 row(s)
hive>
```

- Create a view on the top of last query, name it temperature_data_vw.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
hive> create view temperature_data_vw as select max(temperature), substring( temp_date, 7, 10) from temperature_data group by
substring( temp_date, 7, 10 ) having count( substring ( temp_date, 7 ,10 )) >=2 ;
OK
Time taken: 2.526 seconds
hive>
> select * from temperature_data_vw;
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_20180610010834_fda71574-4631-4b80-830f-2685b3f98179
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_1528036691222_0035, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0035/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0035
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-10 01:09:19,939 Stage-1 map = 0%, reduce = 0%
2018-06-10 01:09:54,200 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.13 sec
2018-06-10 01:10:55,367 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 9.13 sec
2018-06-10 01:10:56,969 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 12.97 sec
2018-06-10 01:11:09,954 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 20.61 sec
MapReduce Total cumulative CPU time: 20 seconds 610 msec
Ended Job = job_1528036691222_0035
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 20.61 sec HDFS Read: 10245 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 610 msec
OK
23 1990
22 1991
16 1993
23 1994
Time taken: 157.75 seconds, Fetched: 4 row(s)
hive>
```

- Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.

```
MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Downloads
acadgild@localhost:~/assignments
File Edit View Search Terminal Help
hive> insert overwrite local directory '/home/acadgild/assignments/'
> row format delimited
> fields terminated by '|'
> select * from temperature_data_vw ;
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_20180610012549_821b6f5f-aa14-4f94-9d74-5cadc114e69f
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_1528036691222_0037, Tracking URL = http://localhost:8088/proxy/application_1528036691222_0037/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1528036691222_0037
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-06-10 01:26:43,548 Stage-1 map = 0%, reduce = 0%
2018-06-10 01:27:05,007 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.92 sec
2018-06-10 01:27:46,600 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 11.5 sec
2018-06-10 01:28:02,098 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 20.93 sec
MapReduce Total cumulative CPU time: 20 seconds 930 msec
Ended Job = job_1528036691222_0037
Moving data to local directory /home/acadgild/assignments
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 20.93 sec HDFS Read: 9840 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 20 seconds 930 msec
OK
Time taken: 138.253 seconds
hive>
```

MyHadoop 2.6.1.1 [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System Sun Jun 10, 1:30 AM Acadgild

acadgild@localhost: ~/assignments

File Edit View Search Terminal Help

You have new mail in /var/spool/mail/acadgild

```
[acadgild@localhost assignments]$ pwd
/home/acadgild/assignments
[acadgild@localhost assignments]$ ls -ltr
total 4
-rw-r--r-- 1 acadgild acadgild 32 Jun 10 01:28 000000_0
[acadgild@localhost assignments]$ cat 000000_0
23|1990
22|1991
16|1993
23|1994
[acadgild@localhost assignments]$
```