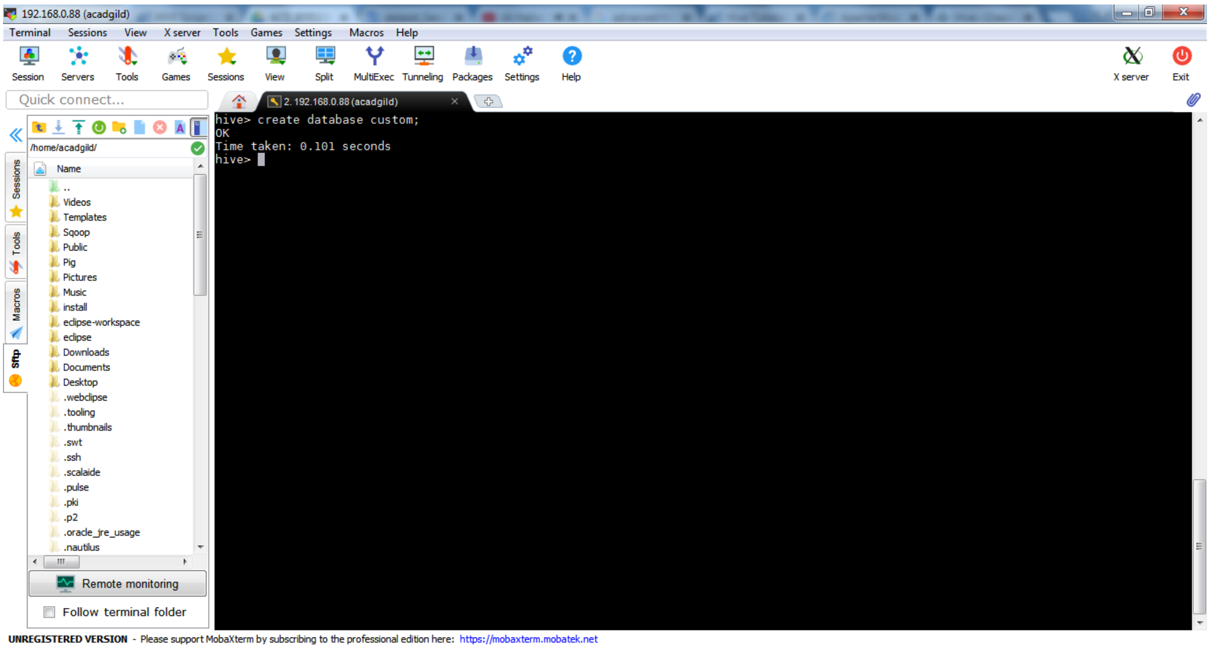
5. Problem Statement

Task 1

Create a database named 'custom'.

Script-:

create database custom;



```
hive> create database custom;
OK
Time taken: 0.101 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.

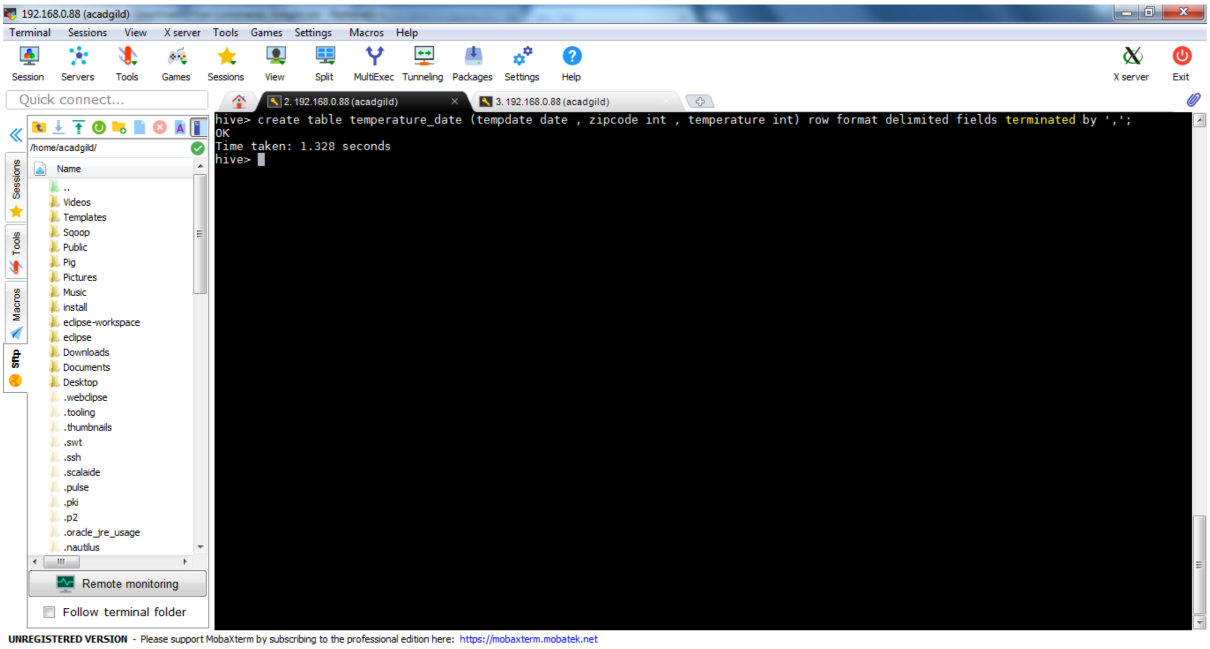
Script-:

Use custom;

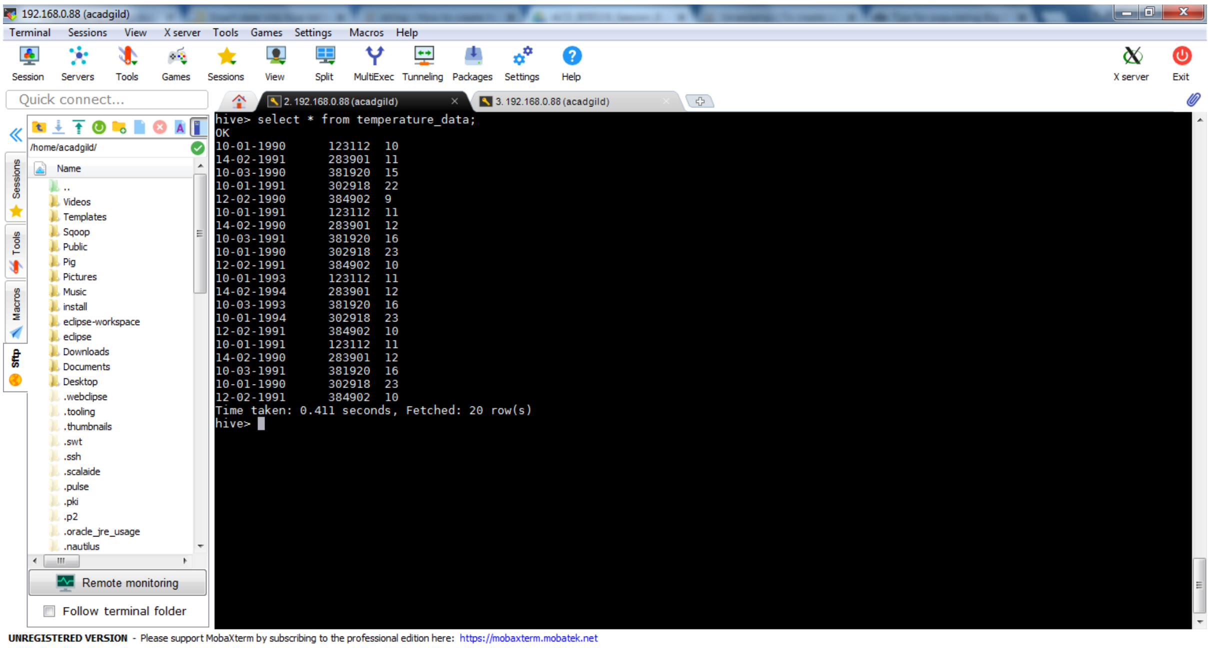
create table temperature_data (tempdate date , zipcode int , temperature int) row format delimited fields terminated by ',';

LOAD DATA LOCAL INPATH '/home/acadgild/dataset_Session14.txt' into table temperature_data;

Output-:



```
hive> create table temperature_date (tempdate date , zipcode int , temperature int) row format delimited fields terminated by ',';
OK
Time taken: 1.328 seconds
hive> 
```


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
..
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
.oracle_jre_usage
.nautilus
Remote monitoring
Follow terminal folder
Sessions
Tools
Macros
Sftp
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

```
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: 0.411 seconds, Fetched: 20 row(s)
hive> 
```

Task 2

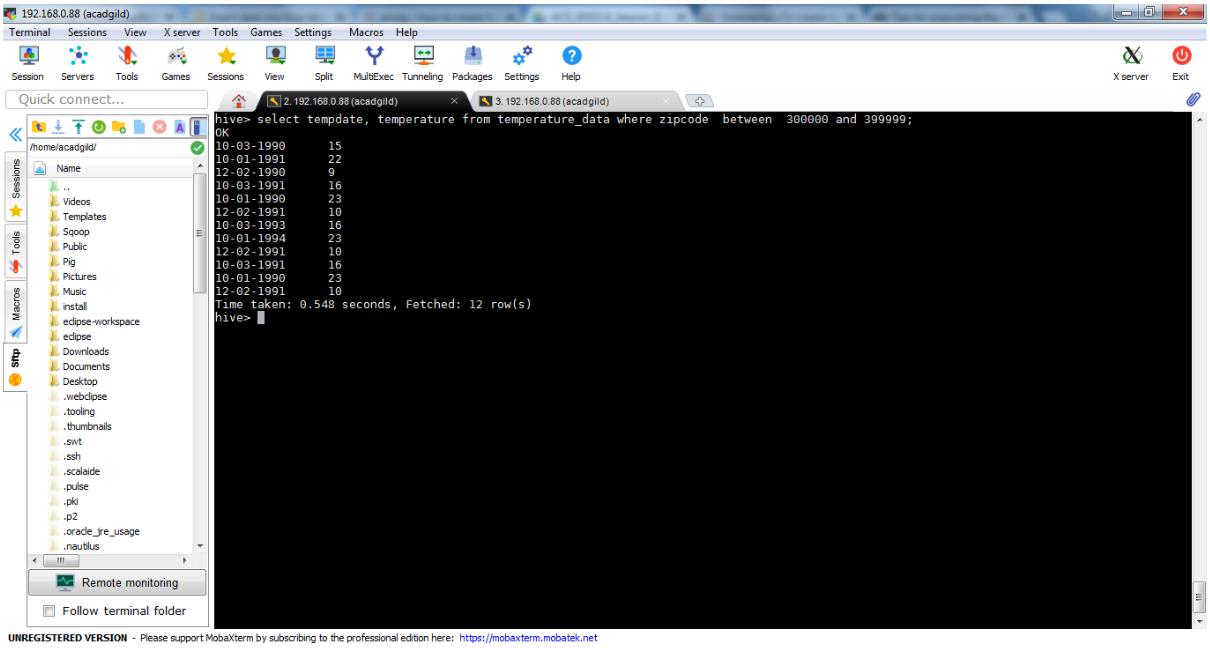
- Fetch date and temperature from temperature_data where zip code is greater than

300000 and less than 399999.

Script-:

select tempdate, temperature from temperature_data where zipcode between 300000 and 399999;

Output-:


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
.oracle_jre_usage
.nautilus
Remote monitoring
Follow terminal folder
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

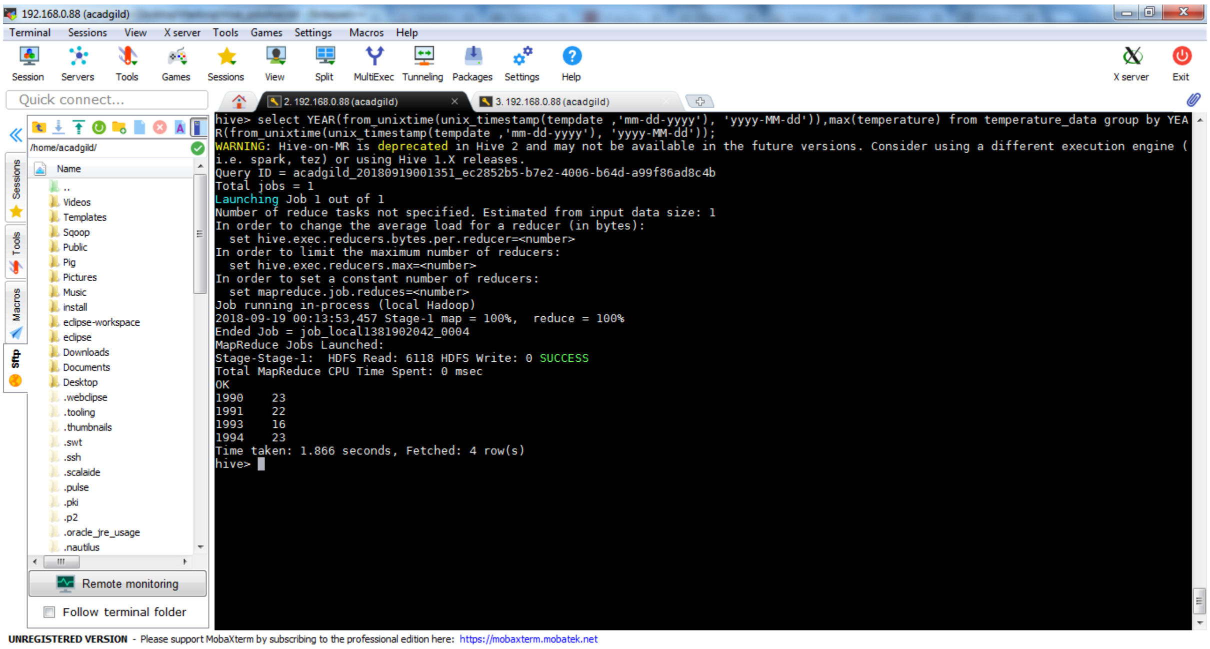
```
hive> select tempdate, temperature from temperature_data where zipcode  between  300000 and 399999;
OK
10-03-1990      15
10-01-1991      22
12-02-1990      9
10-03-1991      16
10-01-1990      23
12-02-1991      10
10-03-1993      16
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 0.548 seconds, Fetched: 12 row(s)
hive>
```

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

Script-:

select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd'));

Output-:


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
.oracle_jre_usage
.nautilus
Remote monitoring
Follow terminal folder
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

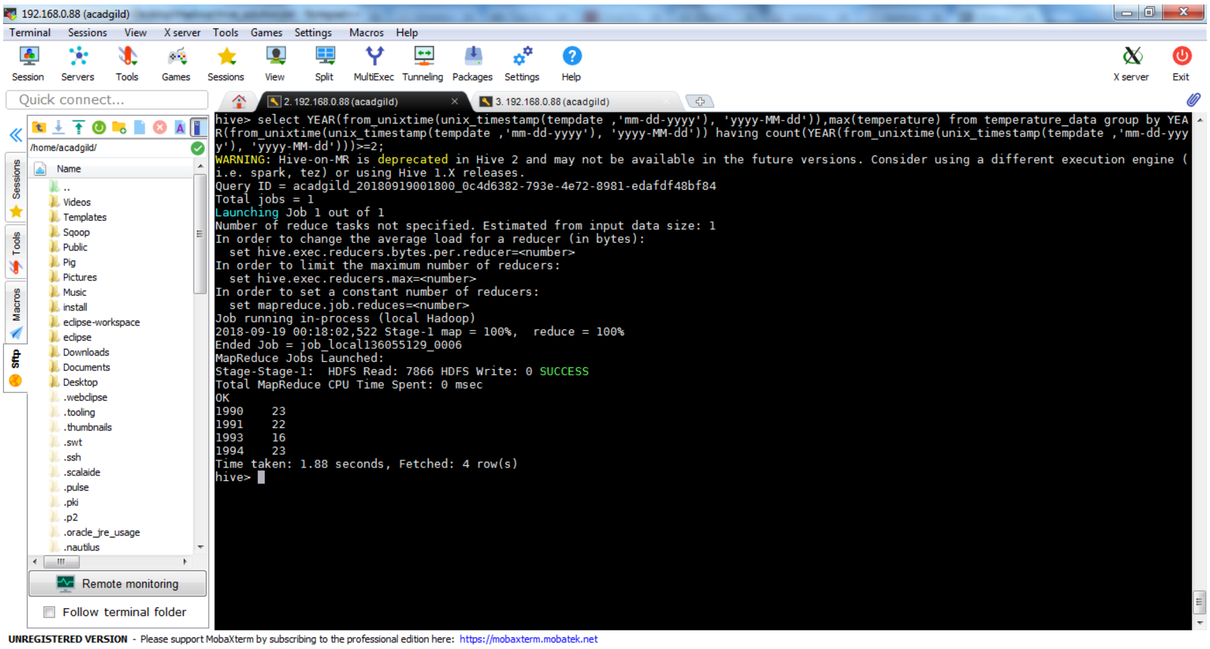
```
hive> select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEA
R(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd'));
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (
i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180919001351_ec2852b5-b7e2-4006-b64d-a99f86ad8c4b
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>
Job running in-process (local Hadoop)
2018-09-19 00:13:53,457 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1381902042_0004
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 6118 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 1.866 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.

Script-:

select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')) having count(YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')))>=2;

Output-:


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
..
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
.oracle_jre_usage
.nautilus
Remote monitoring
Follow terminal folder
Sessions
Tools
Macros
Sftp
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

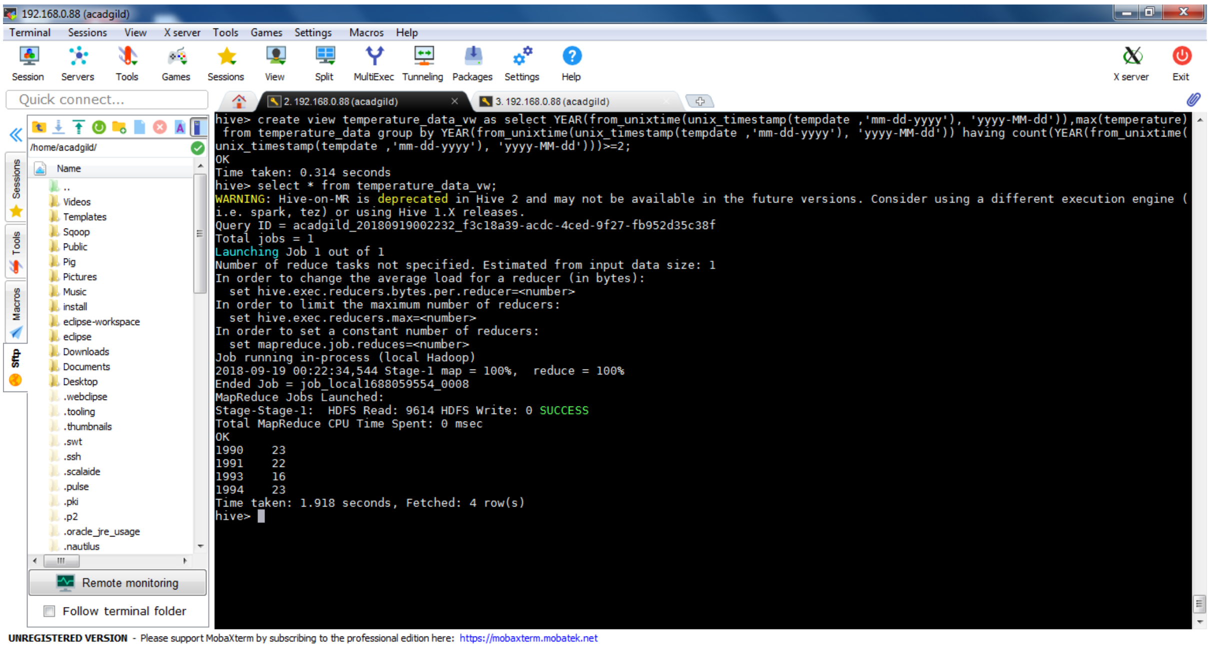
```
hive> select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEA
R(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')) having count(YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyy
y'), 'yyyy-MM-dd')))>=2;
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 = acadgild_20180919001800_0c4d6382-793e-4e72-8981-edafdf48bf84
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>
Job running in-process (local Hadoop)
2018-09-19 00:18:02,522 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local136055129_0006
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 7866 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 1.88 seconds, Fetched: 4 row(s)
hive>
```

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

Script-:

create view temperature_data_vw as select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')) having count(YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')))>=2;

Output-:


192.168.0.88 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
X server Exit
Quick connect...
2. 192.168.0.88 (acadgild)
3. 192.168.0.88 (acadgild)
/home/acadgild/
Name
..
Videos
Templates
Sqoop
Public
Pig
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
.webclipse
.tooling
.thumbnails
.swt
.ssh
.scalaide
.pulse
.pki
.p2
.oracle_jre_usage
.nautilus
Remote monitoring
Follow terminal folder
Sessions
Tools
Macros
Sftp
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

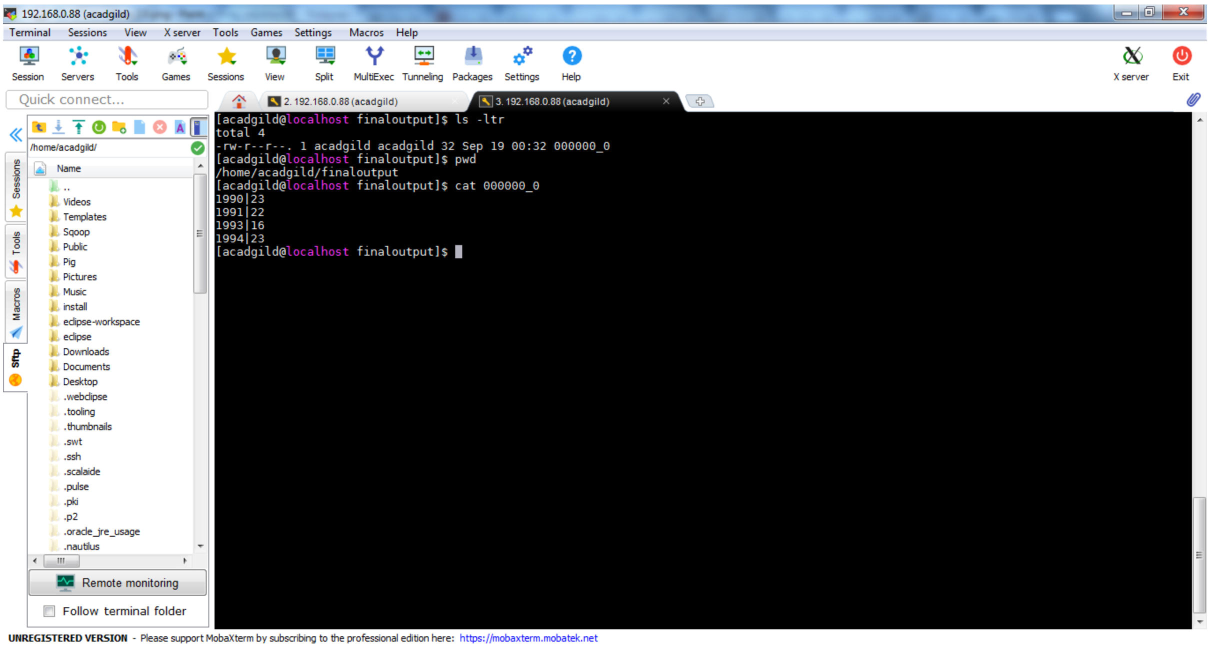
```
hive> create view temperature_data_vw as select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature)
 from temperature_data group by YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')) having count(YEAR(from_unixtime(
unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')))>=2;
OK
Time taken: 0.314 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 execution engine (
i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180919002232_f3c18a39-acdc-4ced-9f27-fb952d35c38f
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>
Job running in-process (local Hadoop)
2018-09-19 00:22:34,544 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1688059554_0008
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 9614 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 1.918 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.

Script-:

insert overwrite local directory '/home/acadgild/finaloutput.txt' row format delimited fields terminated by '|' select YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')),max(temperature) from temperature_data group by YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')) having count(YEAR(from_unixtime(unix_timestamp(tempdate ,'mm-dd-yyyy'), 'yyyy-MM-dd')))>=2;

Output-:



```
[acadgild@localhost finaloutput]$ ls -ltr
total 4
-rw-r--r--. 1 acadgild acadgild 32 Sep 19 00:32 000000_0
[acadgild@localhost finaloutput]$ pwd
/home/acadgild/finaloutput
[acadgild@localhost finaloutput]$ cat 000000_0
1990|23
1991|22
1993|16
1994|23
[acadgild@localhost finaloutput]$ 
```

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net