

Assignment-8.1

Task 1

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.

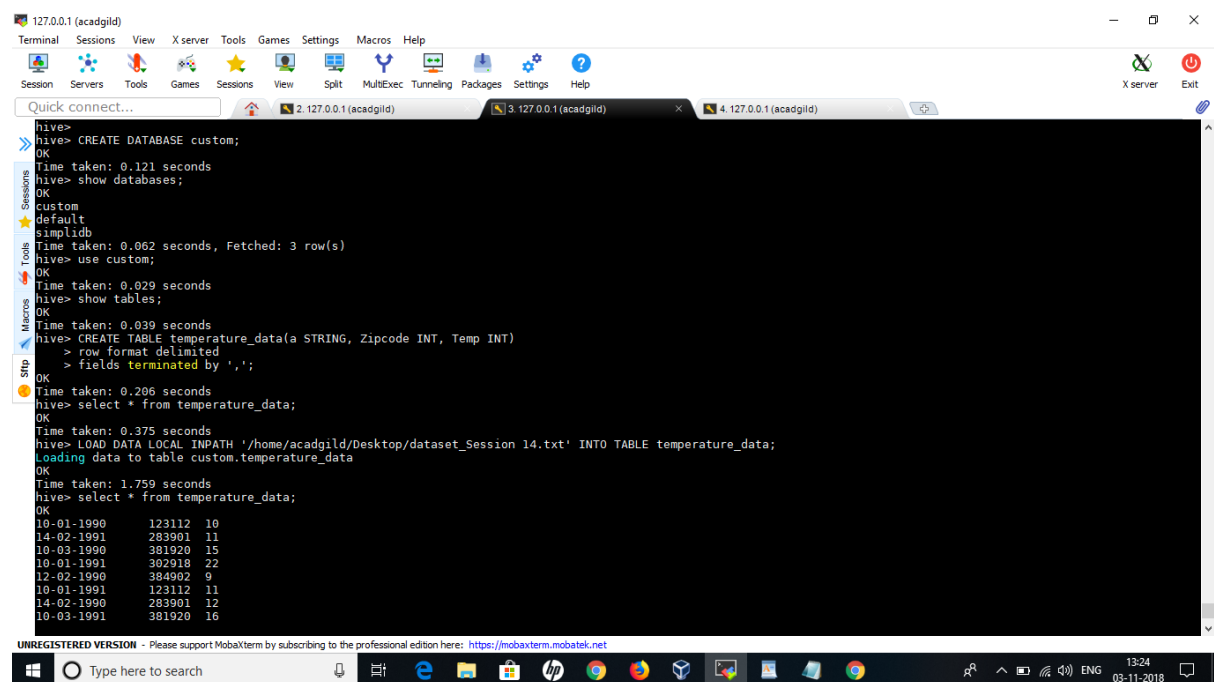
Solution: *CREATE DATABASE custom;*

CREATE TABLE temperature_data(a STRING, Zipcode INT, Temp INT)

row format delimited

fields terminated by ',';

LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/dataset_Session 14.txt' INTO TABLE temperature_data;



```
h1ve> CREATE DATABASE custom;
OK
Time taken: 0.121 seconds
h1ve> show databases;
OK
custom
default
simp1db
Time taken: 0.062 seconds, Fetched: 3 row(s)
h1ve> use custom;
OK
Time taken: 0.029 seconds
h1ve> show tables;
OK
Time taken: 0.039 seconds
h1ve> CREATE TABLE temperature_data(a STRING, Zipcode INT, Temp INT)
> row format delimited
> fields terminated by ',';
OK
Time taken: 0.206 seconds
h1ve> select * from temperature_data;
OK
Time taken: 0.375 seconds
h1ve> LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/dataset_Session 14.txt' INTO TABLE temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 1.759 seconds
h1ve> 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
```

The screenshot shows a MobaXterm terminal window with a Hive query executed. The query is `select * from temperature_data;`. The output displays 20 rows of data, each containing a date, a temperature value, and a zip code. The data is as follows:

Date	Temperature	Zip Code
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

The terminal also shows the execution time: `Time taken: 0.372 seconds, Fetched: 20 row(s)`. The MobaXterm interface includes a menu bar, a toolbar, and a sidebar with session and tool icons. The Windows taskbar at the bottom shows the search bar and various application icons.

Task 2

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

Solution: *SELECT a, Temp from temperature_data WHERE Zipcode BETWEEN 300000 AND 399999;*

The screenshot shows a MobaXterm terminal window with a filtered Hive query executed. The query is `SELECT a, Temp from temperature_data WHERE Zipcode BETWEEN 300000 AND 399999;`. The output displays 12 rows of data, each containing a date and a temperature value. The data is as follows:

Date	Temperature
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

The terminal also shows the execution time: `Time taken: 0.574 seconds, Fetched: 12 row(s)`. The MobaXterm interface is consistent with the first screenshot, showing the same menu bar, toolbar, and sidebar. The Windows taskbar at the bottom shows the search bar and various application icons.

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

Solution: *SELECT year,MAX(t1.Temp) as temperature*

FROM (select SUBSTRING(a,7,4) year,

Temp from temperature_data) t1

GROUP BY year;

The screenshot shows a MobaXterm terminal window with the following content:

```

hive> SELECT year,MAX(t1.Temp) as temperature
> FROM (select SUBSTRING(a,7,4) year,
> Temp from temperature_data) t1
> GROUP BY year;
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_20181016173011_aae9a006-b740-4f48-8000-623f2e8fc6ba
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)
2018-10-16 17:30:13,342 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1975067203_0006
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 8642 HDFS Write: 978 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 2.281 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.

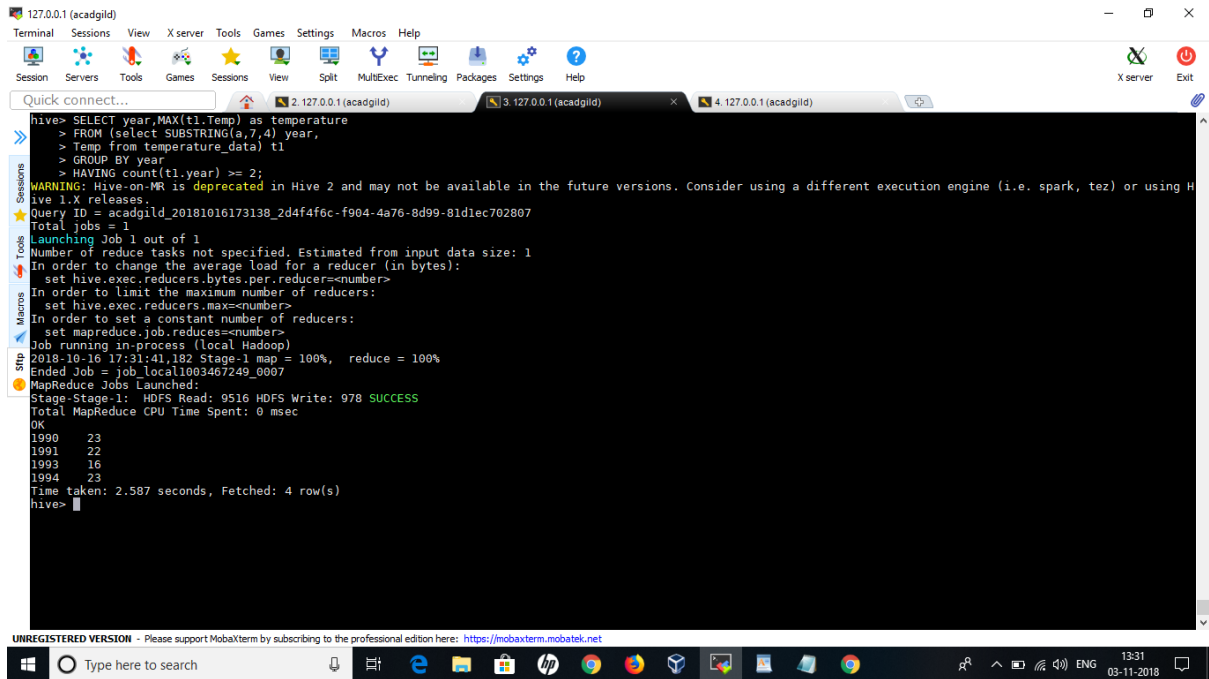
Solution: *SELECT year,MAX(t1.Temp) as temperature*

FROM (select SUBSTRING(a,7,4) year,

Temp from temperature_data) t1

GROUP BY year

HAVING count(t1.year) >= 2;



```
hive> SELECT year,MAX(t1.Temp) as temperature
> FROM (select SUBSTRING(a,7,4) year,
> Temp from temperature_data) t1
> GROUP BY year
> HAVING count(t1.year) >= 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 H
ive 1.X releases.
Query ID = acadgild_20181016173138_2d4f4f6c-f904-4a76-8d99-81d1ec702807
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)
2018-10-16 17:31:41,182 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1003467249_0007
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 9516 HDFS Write: 978 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 2.587 seconds, Fetched: 4 row(s)
hive>
```

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

Solution: *CREATE VIEW temperature_data_vw AS*

SELECT year,MAX(t1.Temp) as temperature

FROM (select SUBSTRING(a,7,4) year,

Temp from temperature_data) t1

GROUP BY year

HAVING count(t1.year) >= 2;

The screenshot shows a MobaXterm window with a terminal session. The terminal displays the following commands and output:

```
hive> CREATE VIEW temperature_data_vw AS
> SELECT year, MAX(t1.Temp) as Temperature
> FROM (select SUBSTRING(a,7,4) year,
> Temp from temperature_data_vw) t1
> GROUP BY year
> HAVING count(t1.year) >= 2;
OK
Time taken: 0.408 seconds
hive>
```

The window title is "127.0.0.1 (acadgild)". The bottom status bar indicates "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>".

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

Solution: *INSERT OVERWRITE LOCAL DIRECTORY*

'/home/acadgild/Desktop/hive_output'

row format delimited fields terminated by '|'

*SELECT * FROM temperature_data_vw;*

The screenshot shows a MobaXterm window with a terminal session. The terminal displays the following commands and output:

```
hive> INSERT OVERWRITE LOCAL DIRECTORY
> '/home/acadgild/Desktop/hive_output'
> 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 execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20181016173504_51c04f22-3258-473b-95b7-782ffa129611
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)
2018-10-16 17:35:06,341 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1071786897_0008
Moving data to local directory /home/acadgild/Desktop/hive_output
MapReduce Jobs Launched:
Stage-Stage 1: HDFS Read: 10390 HDFS Write: 978 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 2.034 seconds
hive>
```

The window title is "127.0.0.1 (acadgild)". The bottom status bar indicates "UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: <https://mobaxterm.mobatek.net>".

127.0.0.1 (acadgild)

Terminal Sessions View X server Tools Games Settings Macros Help

Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help

Quick connect...

2 127.0.0.1 (acadgild) 3 127.0.0.1 (acadgild) 4 127.0.0.1 (acadgild)

```
[acadgild@localhost Desktop]$ ls -l
total 48
-rw-rw-r-- 1 acadgild acadgild 1287 Dec 29 2017 bashrc.txt~
-rw-rw-r-- 1 acadgild acadgild 52 Oct 16 16:05 csv_file.csv
-rw-rw-r-- 1 acadgild acadgild 92 Oct 1 00:24 Customer_Input.txt
-rw-rw-r-- 1 acadgild acadgild 366 Oct 1 05:27 custs.txt
-rw-rw-r-- 1 acadgild acadgild 437 Oct 16 03:27 dataset_session 14.txt
-rw-rw-r-- 1 acadgild acadgild 162 Oct 15 20:59 emp_details.txt
drwxrwxr-x 2 acadgild acadgild 4096 Oct 16 17:35 hive_output
-rw-rw-r-- 1 acadgild acadgild 56 Oct 15 21:49 locations.txt
drwxrwxr-x 2 acadgild acadgild 4096 Feb 2 2018 README
-rw-rw-r-- 1 acadgild acadgild 5443 Oct 1 06:09 txns.txt
-rw-rw-r-- 1 acadgild acadgild 116 Oct 15 21:44 users.txt
[acadgild@localhost Desktop]$ ls hive_output/
000000_0
[acadgild@localhost Desktop]$ cat hive_output/000000_0
1990|23
1991|22
1993|16
1994|23
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Desktop]$
```

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

Type here to search

13:37 03-11-2018