

Assignment for 8th Sesion

Starting Hive After Starting mysql to make sure the will not be any error

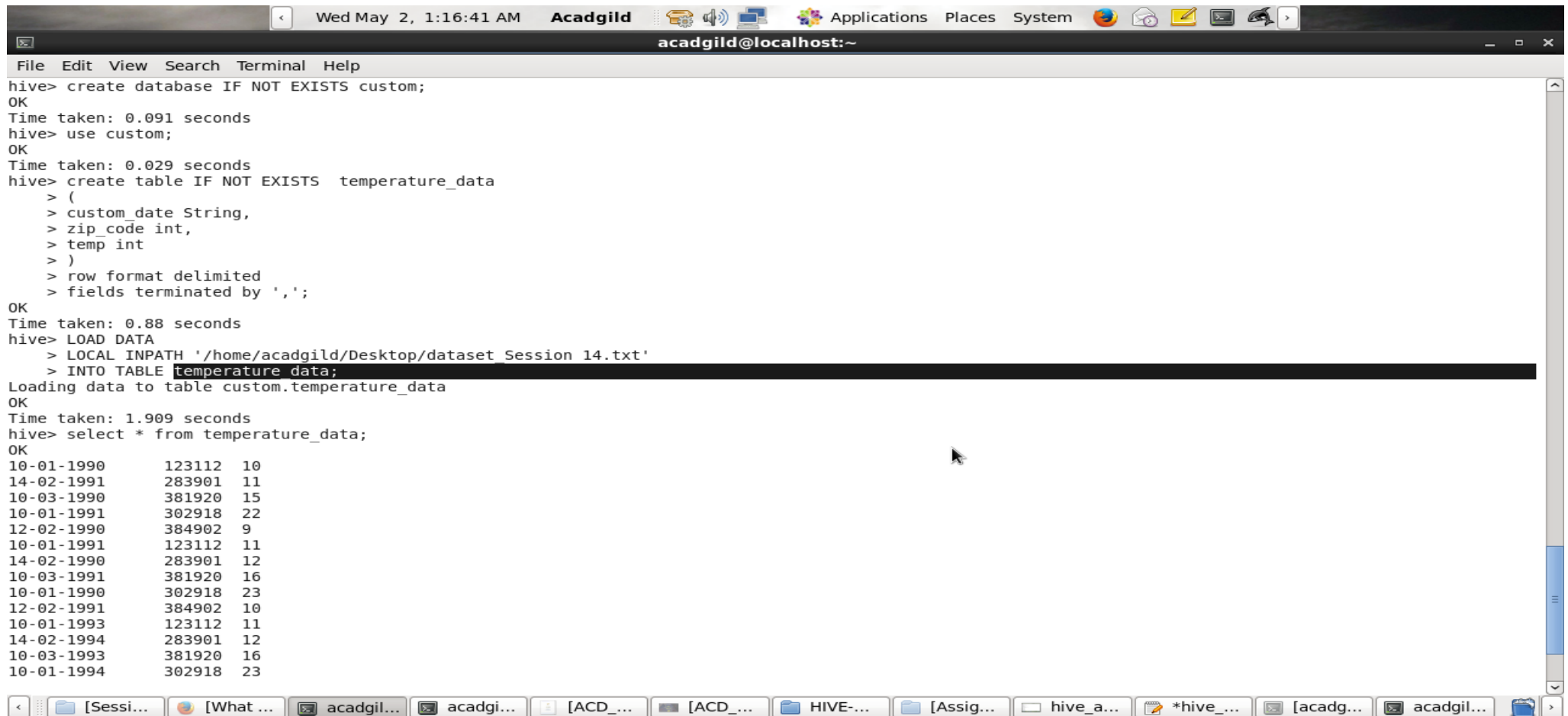


The screenshot shows a terminal window titled 'acadgild@localhost:~'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal output shows the execution of 'start-all.sh', which starts Hadoop services (namenode, datanode, secondary namenode, yarn daemons, resource manager, and node manager). It also shows the execution of 'sudo service mysqld start', which starts MySQL. The output indicates that MySQL is started successfully and that there is new mail in the spool directory. The prompt returns to 'acacgild@localhost ~]' after the 'hive' command is entered.

```
[acacgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/05/01 23:36:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acacgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acacgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acacgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acacgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acacgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acacgild-secondarynamenode-localhost.localdomain.out
18/05/01 23:36:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acacgild/install/hadoop/hadoop-2.6.5/logs/yarn-acacgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acacgild/install/hadoop/hadoop-2.6.5/logs/yarn-acacgild-nodemanager-localhost.localdomain.out
[acacgild@localhost ~]$ sudo service mysqld start
[sudo] password for acadgild:
Starting mysqld:
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost ~]$ hive
```

Assignment for 8th Sesion

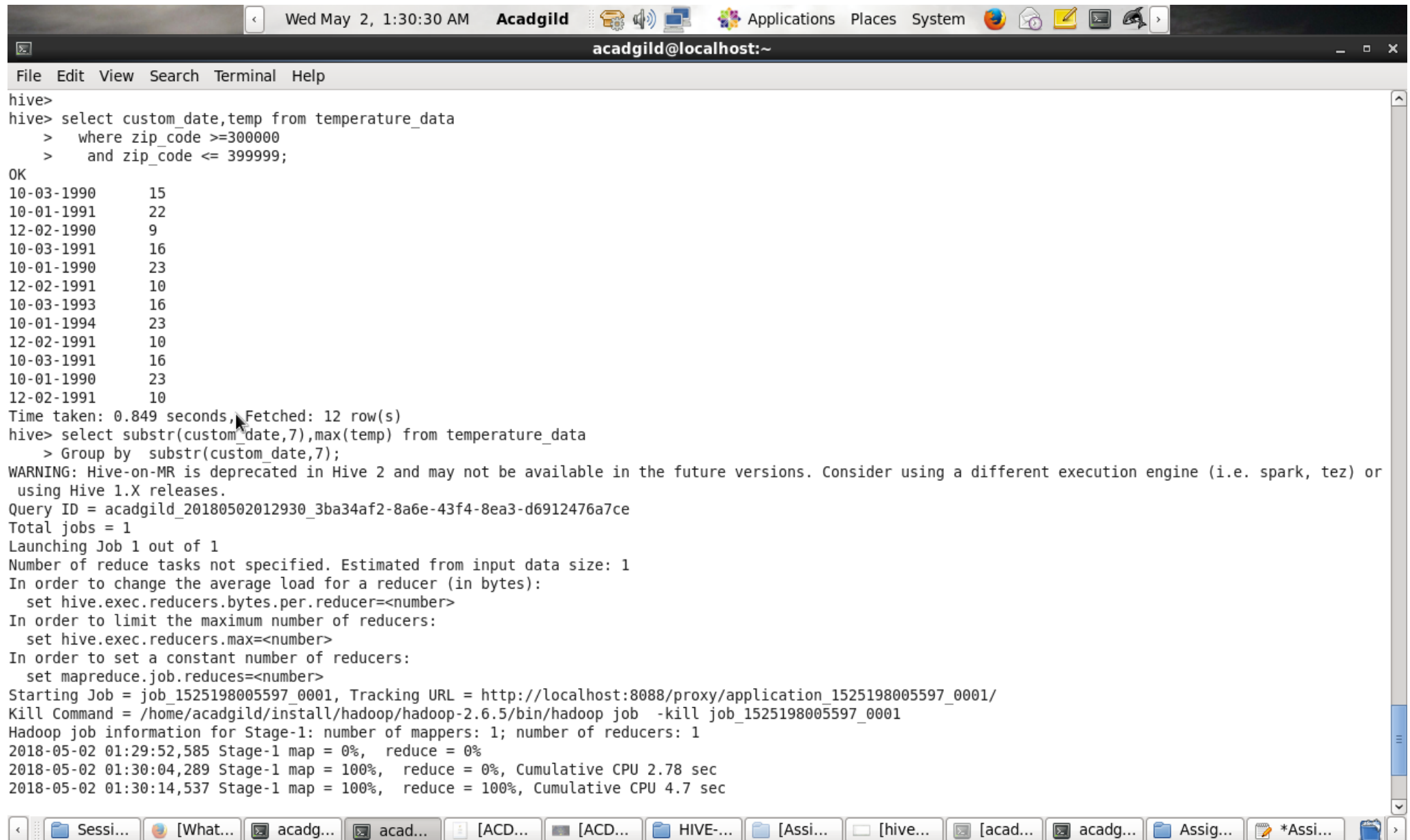
Creating database if not exists already ,
using the same database to create tables,
creating table as described in assignment default date format in hive is YYYY-MM-DD
So creating date as DD-MM-YYYY for custom format,
loading data from local file, selected all records to make sure they are loaded.



```
Wed May 2, 1:16:41 AM Acadgild Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
hive> create database IF NOT EXISTS custom;
OK
Time taken: 0.091 seconds
hive> use custom;
OK
Time taken: 0.029 seconds
hive> create table IF NOT EXISTS temperature_data
> (
>   custom_date String,
>   zip_code int,
>   temp int
> )
> row format delimited
> fields terminated by ',';
OK
Time taken: 0.88 seconds
hive> 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: 1.909 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
```

Assignment for 8th Sesion

Selecting Date & temp on condition as shown below;



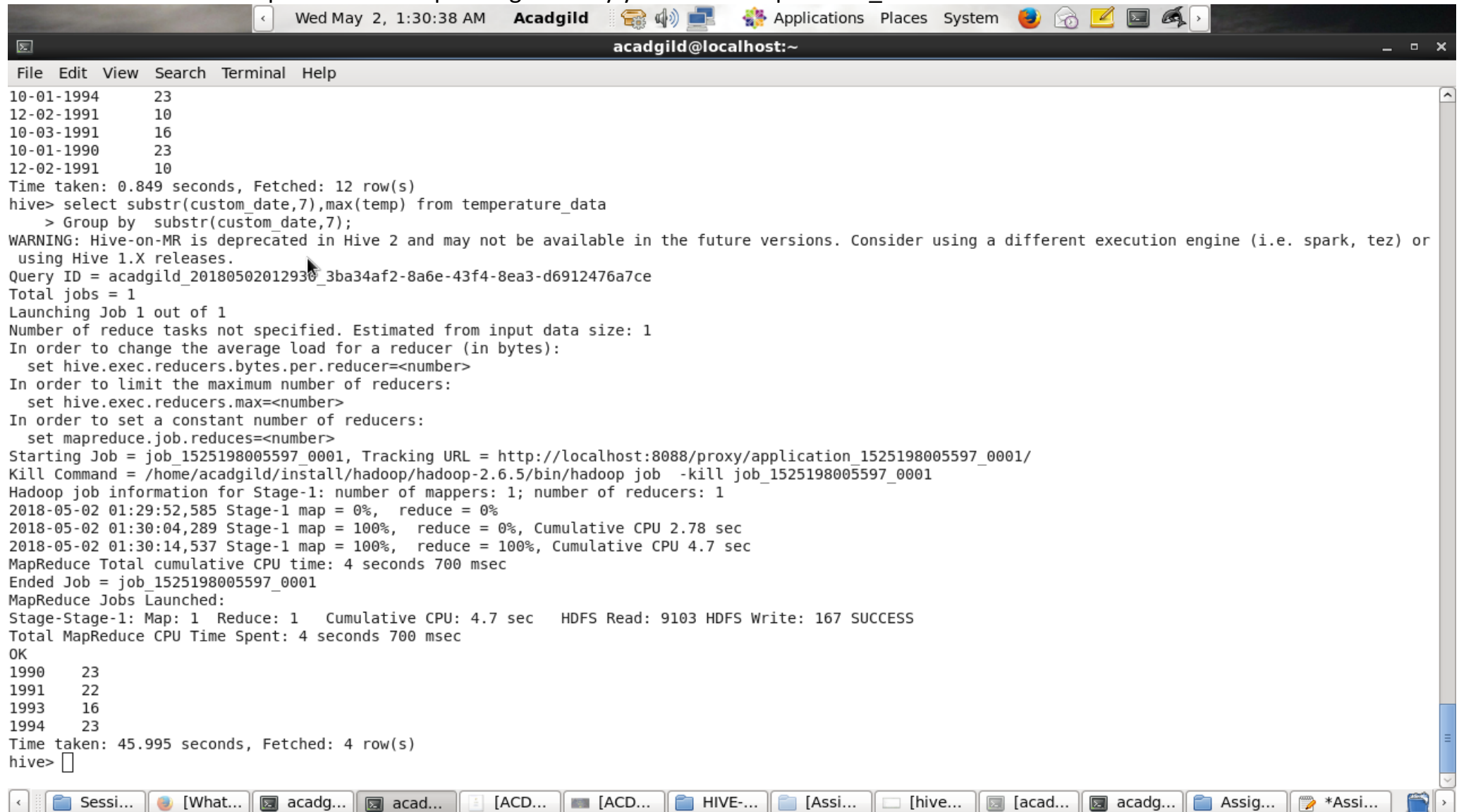
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help) and a system bar at the top (Wed May 2, 1:30:30 AM, Acadgild, and various application icons). The terminal displays the following content:

```
hive>
hive> select custom_date,temp from temperature_data
  >   where zip_code >=300000
  >   and zip_code <= 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: 0.849 seconds, Fetched: 12 row(s)
hive> select substr(custom_date,7),max(temp) from temperature_data
  > Group by  substr(custom_date,7);
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_20180502012930_3ba34af2-8a6e-43f4-8ea3-d6912476a7ce
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_1525198005597_0001, Tracking URL = http://localhost:8088/proxy/application_1525198005597_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525198005597_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-02 01:29:52,585 Stage-1 map = 0%,  reduce = 0%
2018-05-02 01:30:04,289 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.78 sec
2018-05-02 01:30:14,537 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.7 sec
```

The bottom of the image shows a taskbar with several open application windows, including 'Sessi...', '[What...', 'acadg...', 'ACD...', 'HIVE-...', '[Assi...', '[hive...', '[acad...', 'acadg...', 'Assig...', and '*Assi...'.

Assignment for 8th Sesion

Calculate maximum temperature corresponding to every year from temperature_data table.



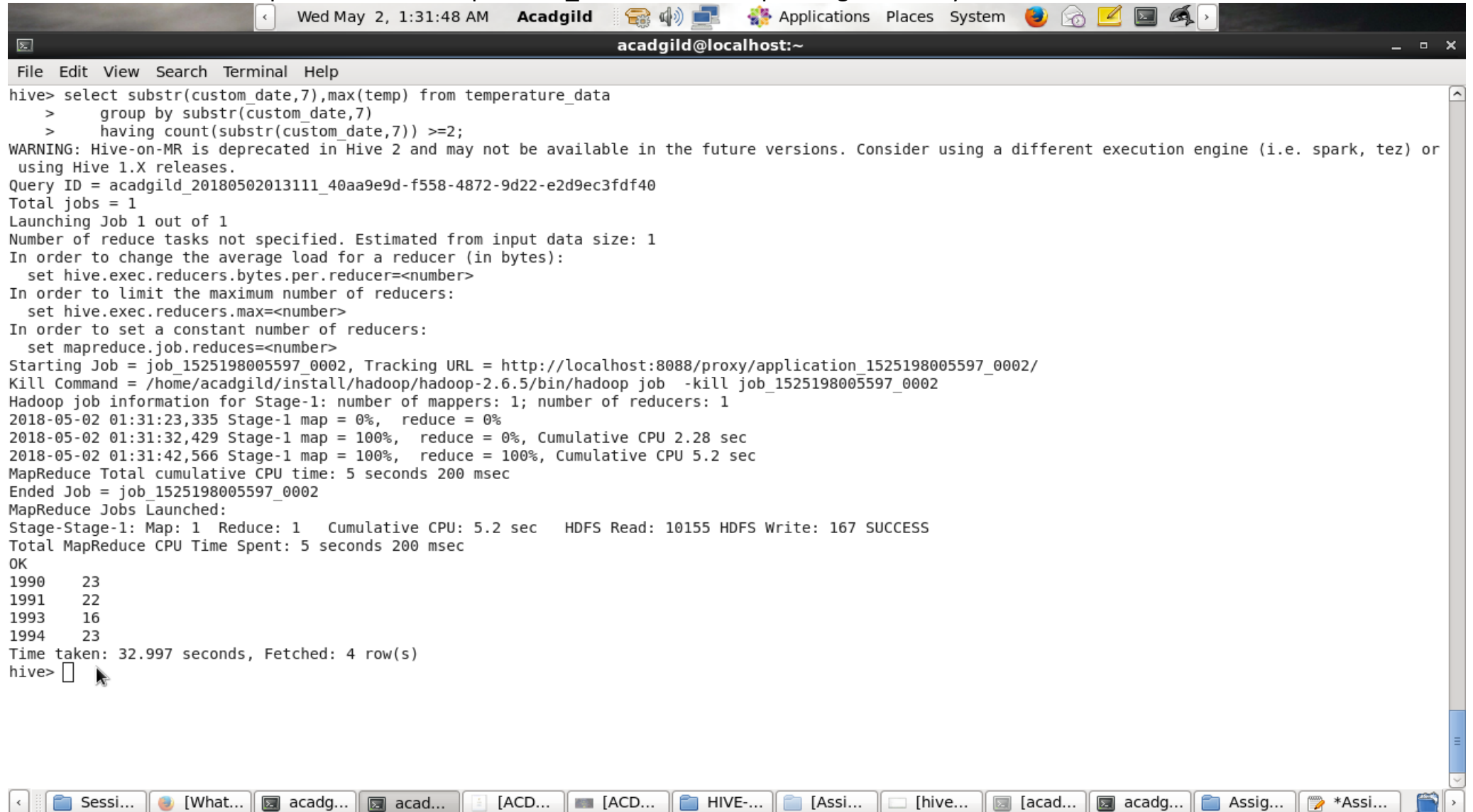
The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal displays the following content:

```
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 0.849 seconds, Fetched: 12 row(s)
hive> select substr(custom_date,7),max(temp) from temperature_data
> Group by substr(custom_date,7);
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_20180502012930_3ba34af2-8a6e-43f4-8ea3-d6912476a7ce
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_1525198005597_0001, Tracking URL = http://localhost:8088/proxy/application_1525198005597_0001/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525198005597_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-02 01:29:52,585 Stage-1 map = 0%, reduce = 0%
2018-05-02 01:30:04,289 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.78 sec
2018-05-02 01:30:14,537 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.7 sec
MapReduce Total cumulative CPU time: 4 seconds 700 msec
Ended Job = job_1525198005597_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.7 sec HDFS Read: 9103 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 700 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 45.995 seconds, Fetched: 4 row(s)
hive>
```

The terminal window has a taskbar at the bottom with several open applications: Sessi..., [What..., acadg..., acad..., [ACD..., [ACD..., HIVE..., [Assi..., [hive..., [acad..., acadg..., Assig..., *Assi...

Assignment for 8th Sesion

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

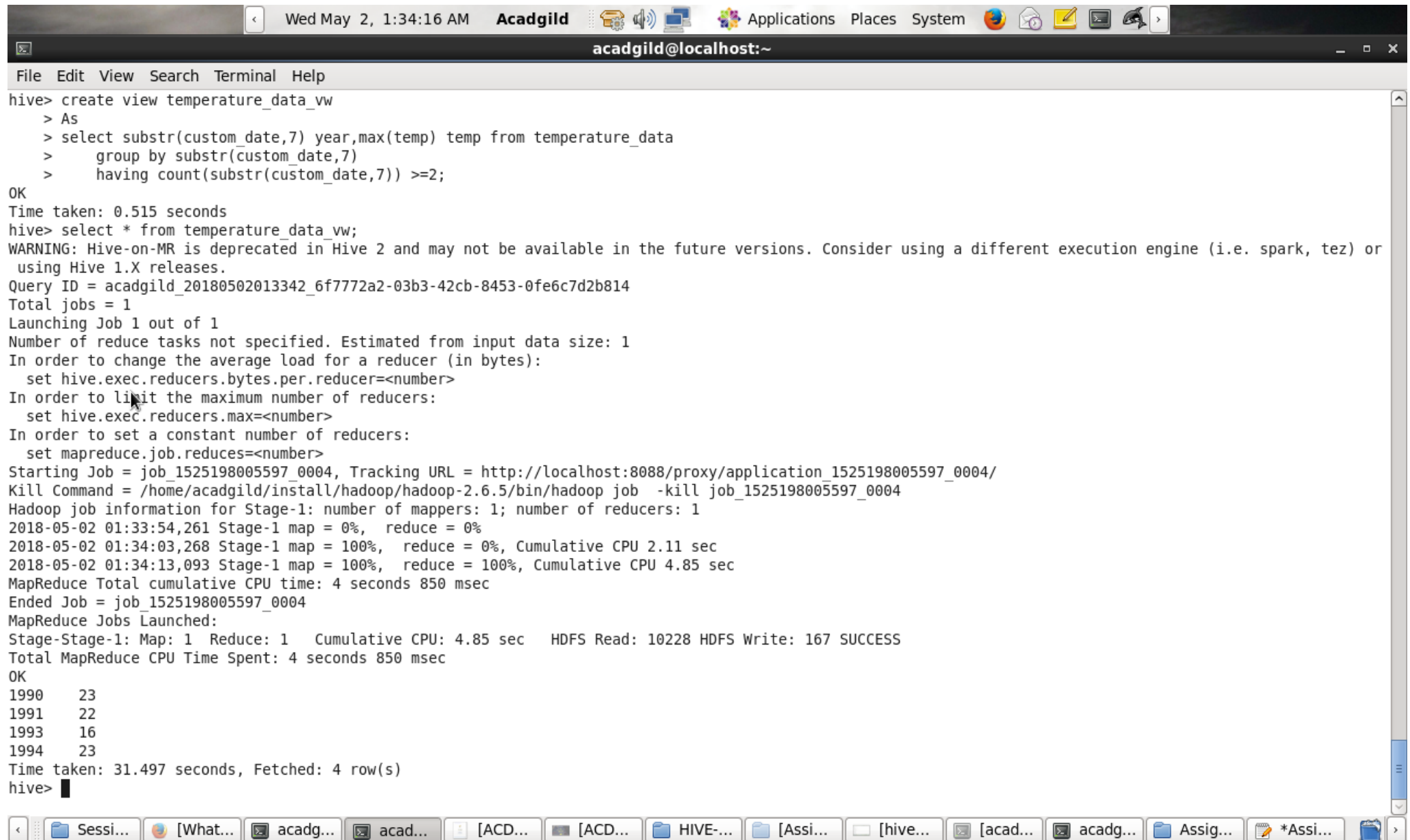


```
Wed May 2, 1:31:48 AM Acadgild Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
hive> select substr(custom_date,7),max(temp) from temperature_data
>   group by substr(custom_date,7)
>   having count(substr(custom_date,7)) >=2;
WARNING: Hive-on-MR is deprecated in 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_20180502013111_40aa9e9d-f558-4872-9d22-e2d9ec3fdf40
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_1525198005597_0002, Tracking URL = http://localhost:8088/proxy/application_1525198005597_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525198005597_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-02 01:31:23,335 Stage-1 map = 0%, reduce = 0%
2018-05-02 01:31:32,429 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.28 sec
2018-05-02 01:31:42,566 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.2 sec
MapReduce Total cumulative CPU time: 5 seconds 200 msec
Ended Job = job_1525198005597_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.2 sec HDFS Read: 10155 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 200 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 32.997 seconds, Fetched: 4 row(s)
hive>
```

Sessi... [What... acadg... acad... [ACD... [ACD... HIVE... [Assi... [hive... [acad... acadg... Assig... *Assi...

Assignment for 8th Sesion

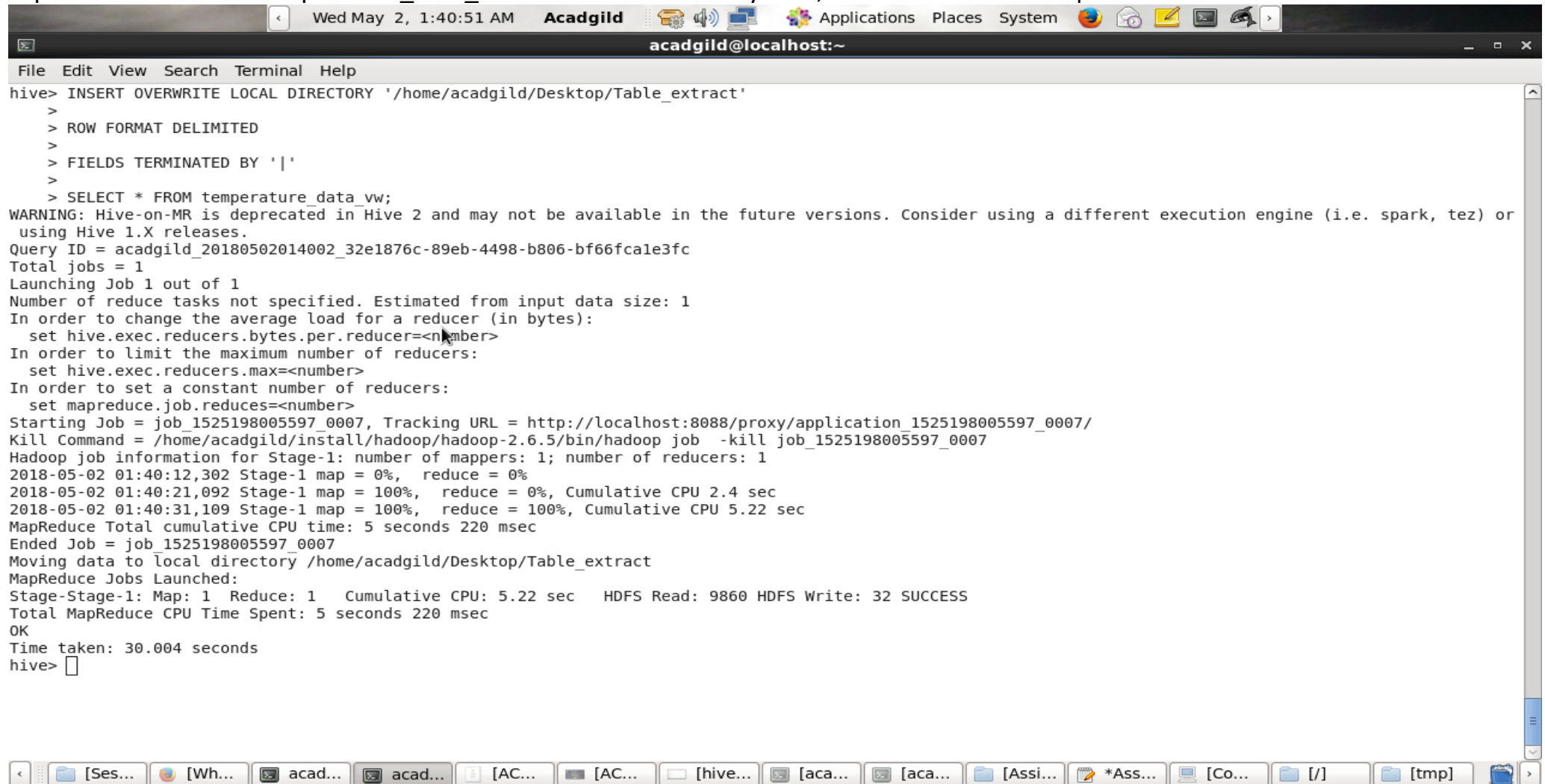
Create a view on the top of last query, name it temperature_data_vw & selecting all recods from view.



```
Wed May 2, 1:34:16 AM  Acadgild  Applications  Places  System  acadgild@localhost:~  
File Edit View Search Terminal Help  
hive> create view temperature_data_vw  
> As  
> select substr(custom_date,7) year,max(temp) temp from temperature_data  
> group by substr(custom_date,7)  
> having count(substr(custom_date,7)) >=2;  
OK  
Time taken: 0.515 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_20180502013342_6f7772a2-03b3-42cb-8453-0fe6c7d2b814  
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_1525198005597_0004, Tracking URL = http://localhost:8088/proxy/application_1525198005597_0004/  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525198005597_0004  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2018-05-02 01:33:54,261 Stage-1 map = 0%, reduce = 0%  
2018-05-02 01:34:03,268 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.11 sec  
2018-05-02 01:34:13,093 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.85 sec  
MapReduce Total cumulative CPU time: 4 seconds 850 msec  
Ended Job = job_1525198005597_0004  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.85 sec HDFS Read: 10228 HDFS Write: 167 SUCCESS  
Total MapReduce CPU Time Spent: 4 seconds 850 msec  
OK  
1990 23  
1991 22  
1993 16  
1994 23  
Time taken: 31.497 seconds, Fetched: 4 row(s)  
hive>
```


Assignment for 8th Sesion

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



```
Wed May 2, 1:40:51 AM Acadgild Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/Desktop/Table_extract'
>
> 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_20180502014002_32e1876c-89eb-4498-b806-bf66fca1e3fc
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_1525198005597_0007, Tracking URL = http://localhost:8088/proxy/application_1525198005597_0007/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1525198005597_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-02 01:40:12,302 Stage-1 map = 0%, reduce = 0%
2018-05-02 01:40:21,092 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.4 sec
2018-05-02 01:40:31,109 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.22 sec
MapReduce Total cumulative CPU time: 5 seconds 220 msec
Ended Job = job_1525198005597_0007
Moving data to local directory /home/acadgild/Desktop/Table_extract
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.22 sec HDFS Read: 9860 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 220 msec
OK
Time taken: 30.004 seconds
hive>
```

Assignment for 8th Sesion

Making sure they have exported as required;

