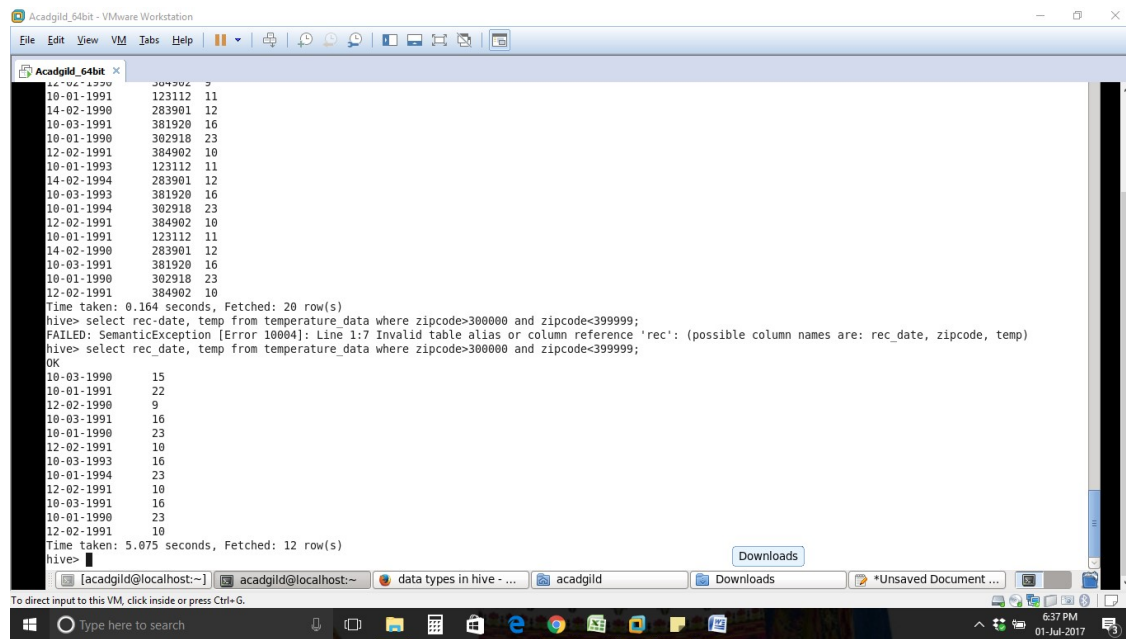


## Assignment 14.2

### Problem Statement

Fetch date and temperature from temperature\_data where zip code is greater than 300000 and less than 399999.

Select rec\_date,temp from temperature\_data where zipcode>300000 and zipcode<399999;

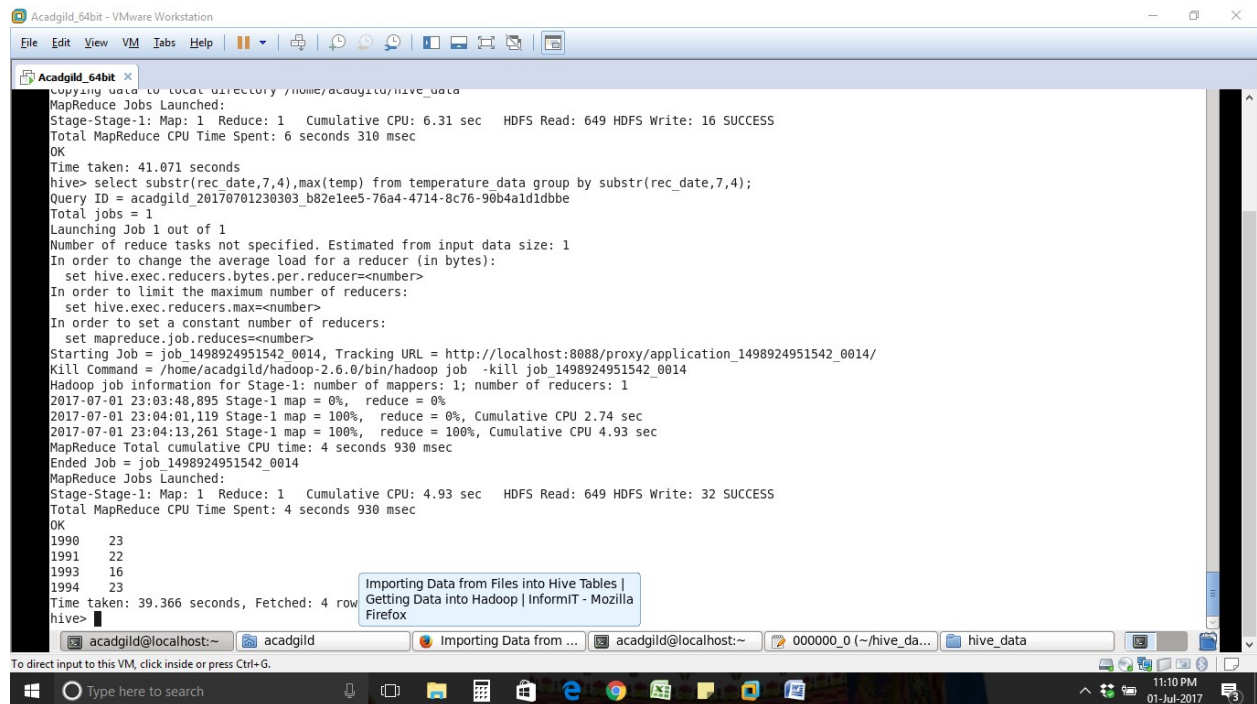


```
Acadgild_64bit x
14-02-1990 304902 9
10-01-1991 123112 11
14-02-1990 283901 12
10-03-1991 381920 16
10-01-1990 302918 23
12-02-1991 384902 10
10-01-1993 123112 11
14-02-1994 283901 12
10-03-1993 381920 16
10-01-1994 302918 23
12-02-1991 384902 10
10-01-1991 123112 11
14-02-1990 283901 12
10-03-1991 381920 16
10-01-1990 302918 23
12-02-1991 384902 10
Time taken: 0.164 seconds, Fetched: 20 row(s)
hive> select rec_date, temp from temperature_data where zipcode>300000 and zipcode<399999;
FAILED: SemanticException [Error 10004]: Line 1:7 Invalid table alias or column reference 'rec': (possible column names are: rec_date, zipcode, temp)
hive> select rec_date, temp 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: 5.075 seconds, Fetched: 12 row(s)
hive>
```

## Problem Statement

Calculate maximum temperature corresponding to every year from temperature\_data table.

Select substr(rec\_date,7,4) , max(temp) from temperature\_data group by substr(rec\_date,7,4);



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Copying data to local directory /home/acadgild/hive_data
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 649 HDFS Write: 16 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
Time taken: 41.071 seconds
hive> select substr(rec_date,7,4),max(temp) from temperature_data group by substr(rec_date,7,4);
Query ID = acadgild_20170701230303_b82e1ee5-76a4-4714-8c76-90b4a1d1dbbe
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_1498924951542_0014, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0014/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 23:03:48,895 Stage-1 map = 0%, reduce = 0%
2017-07-01 23:04:01,119 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.74 sec
2017-07-01 23:04:13,261 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.93 sec
MapReduce Total cumulative CPU time: 4 seconds 930 msec
Ended Job = job_1498924951542_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.93 sec HDFS Read: 649 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 930 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 39.366 seconds, Fetched: 4 rows
hive>
```

Importing Data from Files into Hive Tables | Getting Data into Hadoop | InformIT - Mozilla Firefox

acadgild@localhost:~\$ acadgild

000000\_0 (~hive\_da...)

hive\_data

To direct input to this VM, click inside or press Ctrl+G.

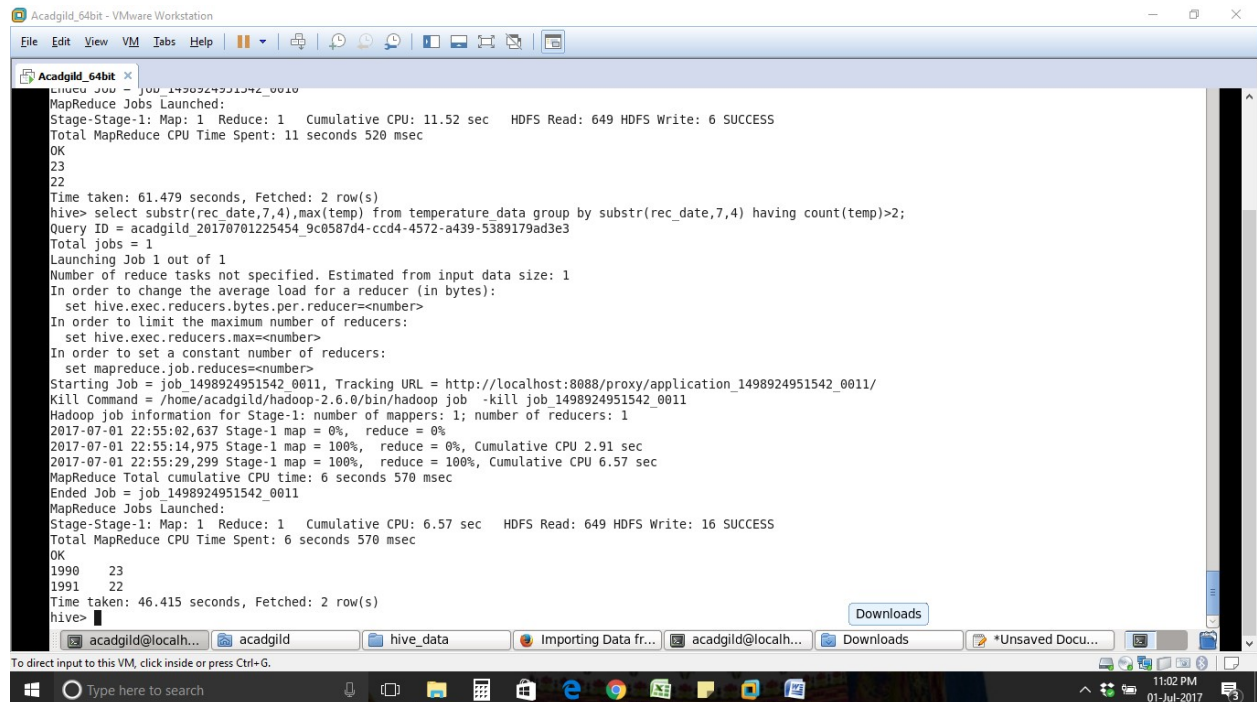
Type here to search

11:10 PM 01-Jul-2017

## Problem Statement

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

Select substr(rec\_date,7,4) , max(temp) from temperature\_data group by substr(rec\_date,7,4) having count(temp)>2;



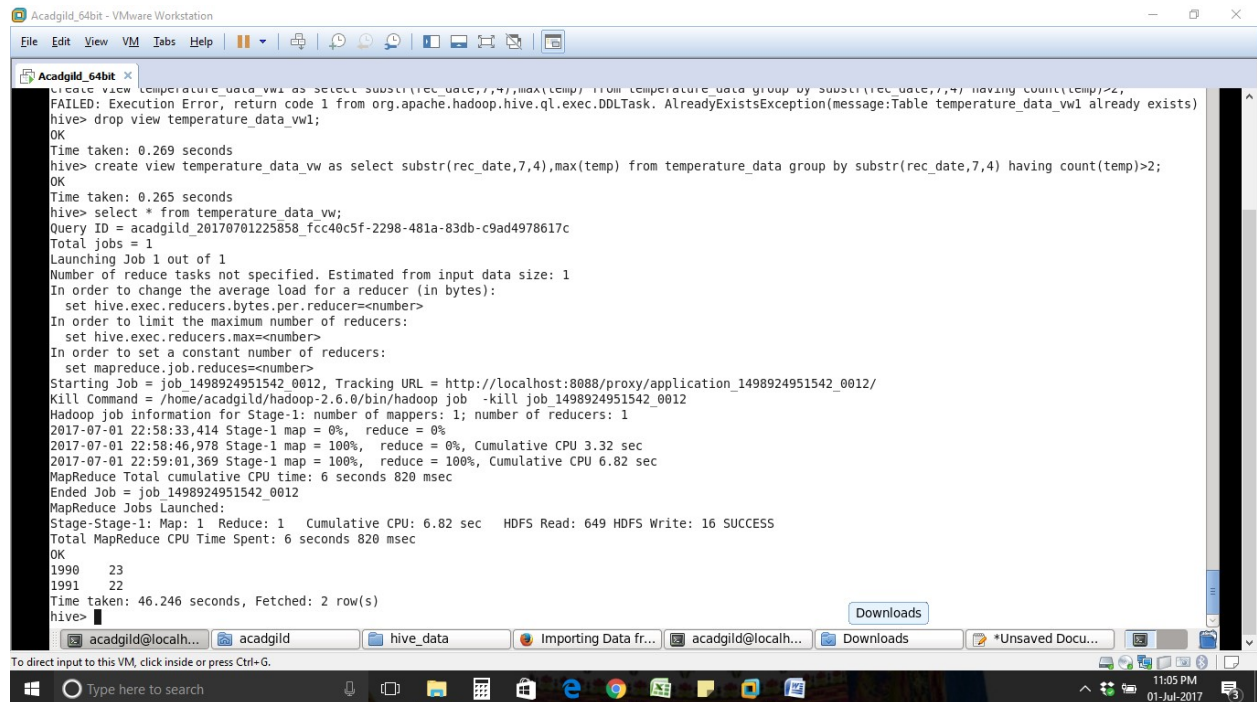
```
Acadgild_64bit - VMware Workstation
File Edit View VM Jobs Help
Acadgild_64bit
Linux Job - job_1498924951542_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.52 sec HDFS Read: 649 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 520 msec
OK
23
22
Time taken: 61.479 seconds, Fetched: 2 row(s)
hive> select substr(rec_date,7,4),max(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;
Query ID = acadgild_20170701225454_9c0587d4-ccd4-4572-a439-5389179ad3e3
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_1498924951542_0011, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0011/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:55:02,637 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:55:14,975 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.91 sec
2017-07-01 22:55:29,299 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.57 sec
MapReduce Total cumulative CPU time: 6 seconds 570 msec
Ended Job = job_1498924951542_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.57 sec HDFS Read: 649 HDFS Write: 16 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 570 msec
OK
1990 23
1991 22
Time taken: 46.415 seconds, Fetched: 2 row(s)
hive>
```

## Problem Statement

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

Create view `temperature_data_vw` as `Select substr(rec_date,7,4) , max(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;`

Select `*` from `temperature_data_vw`;

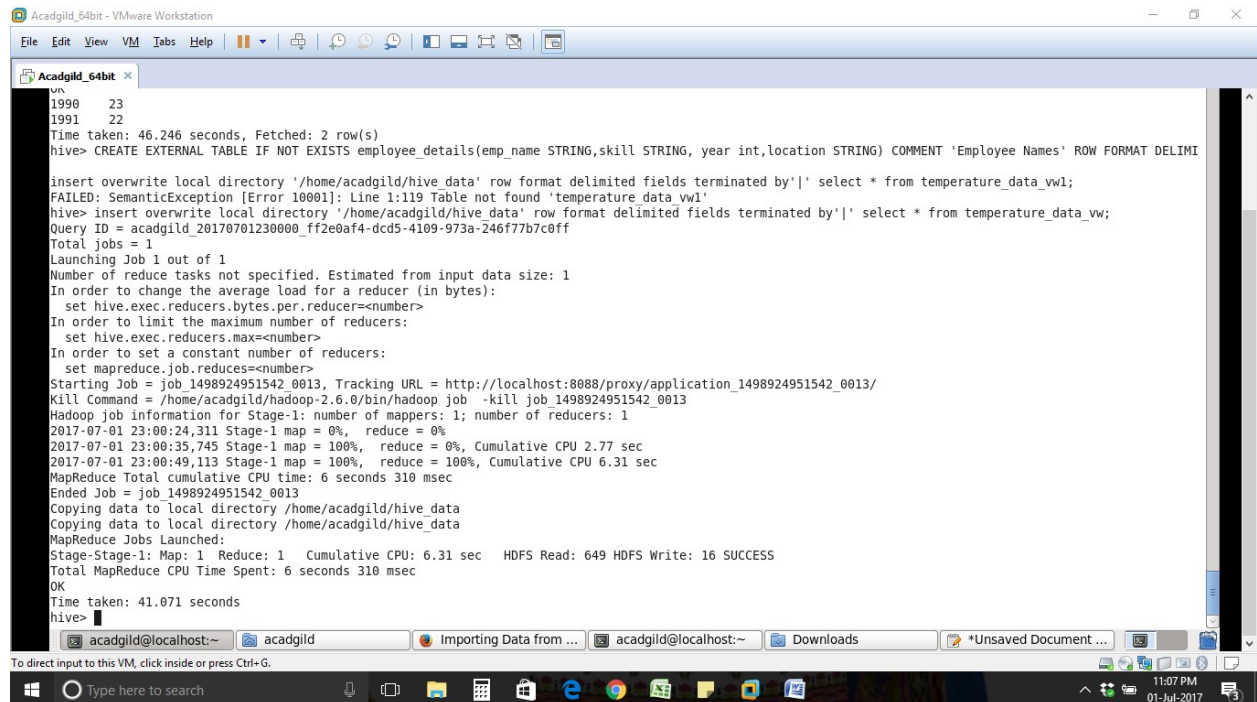


```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
hive> create view temperature_data_vw1 as select substr(rec_date,7,4),max(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table temperature_data_vw1 already exists)
OK
hive> drop view temperature_data_vw1;
OK
Time taken: 0.269 seconds
hive> create view temperature_data_vw as select substr(rec_date,7,4),max(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;
OK
Time taken: 0.265 seconds
hive> select * from temperature_data vw;
Query ID = acadgild_20170701225858_ffc40c5f-2298-481a-83db-c9ad4978617c
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_1498924951542_0012, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0012/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:58:33,414 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:58:46,978 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.32 sec
2017-07-01 22:59:01,369 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.82 sec
MapReduce Total cumulative CPU time: 6 seconds 820 msec
Ended Job = job_1498924951542_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.82 sec HDFS Read: 649 HDFS Write: 16 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 820 msec
OK
1990 23
1991 22
Time taken: 46.246 seconds, Fetched: 2 row(s)
hive>
```

## Problem Statement

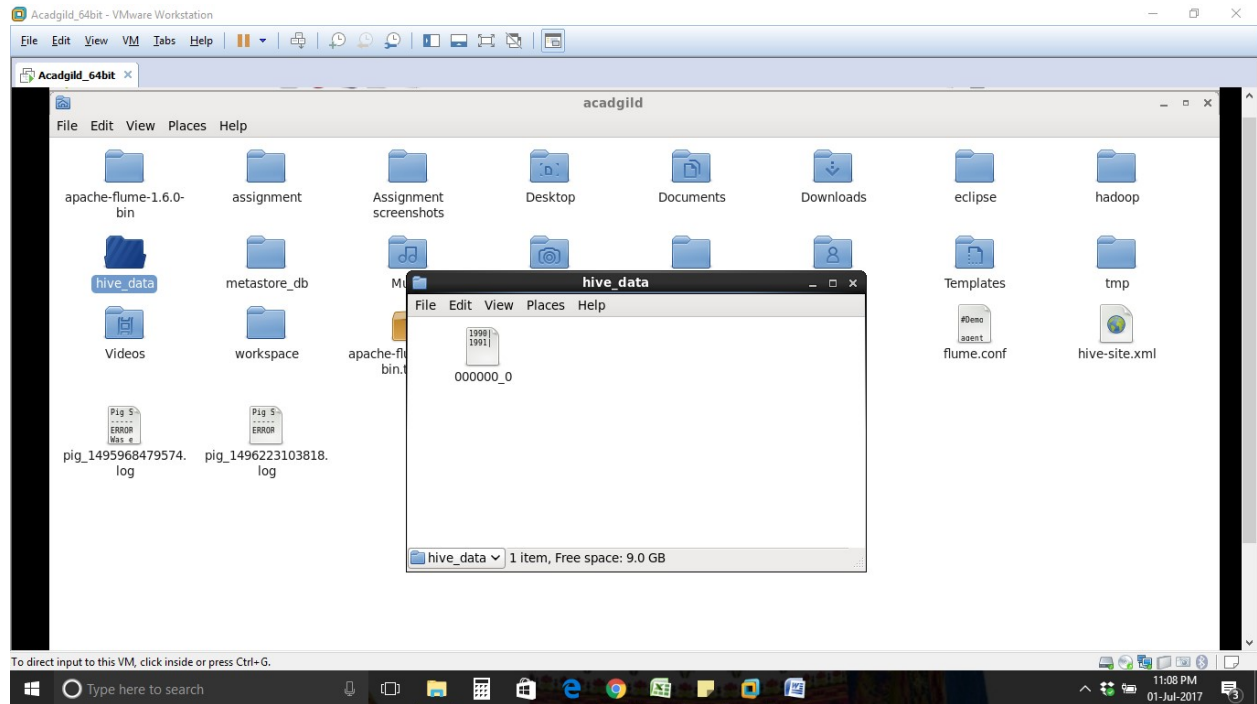
Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

Insert overwrite local directory '/home/acadgild/hive\_data' row format delimited fields terminated by '|' select \* from temperature\_data\_vw;

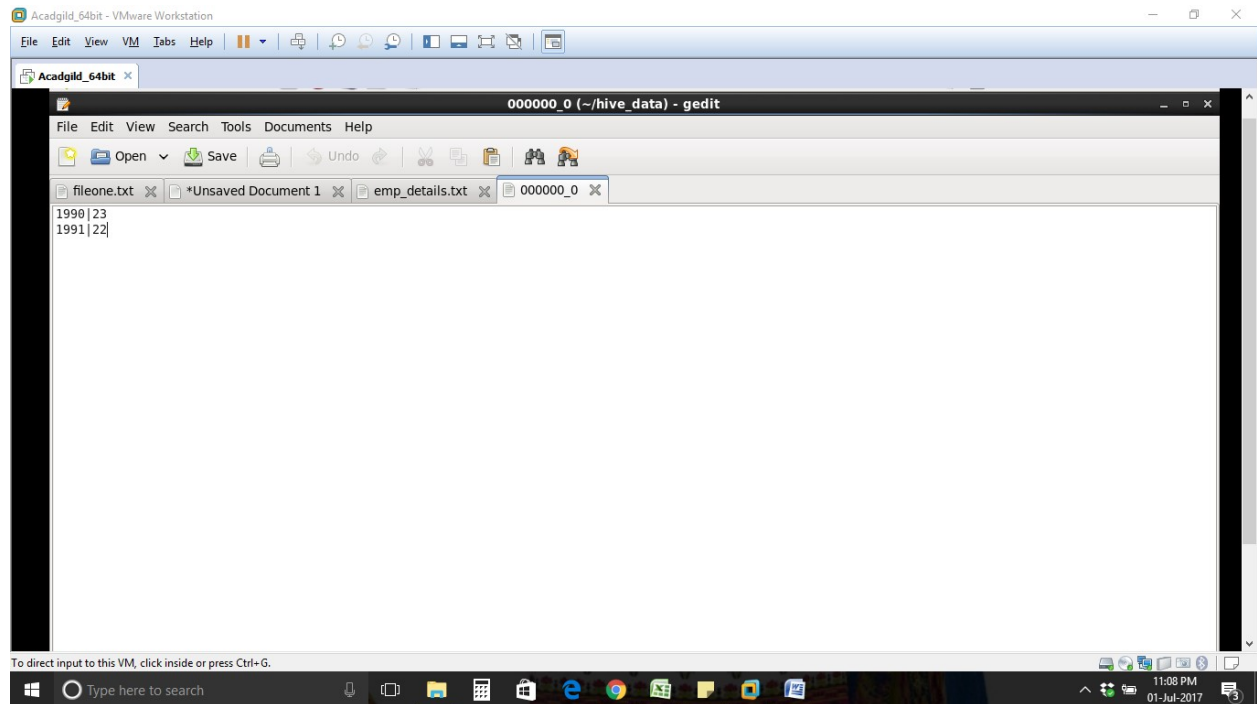


```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
UN
1990 23
1991 22
Time taken: 46.246 seconds, Fetched: 2 row(s)
hive> CREATE EXTERNAL TABLE IF NOT EXISTS employee_details(emp_name STRING,skill STRING, year int,location STRING) COMMENT 'Employee Names' ROW FORMAT DELIMI
insert overwrite local directory '/home/acadgild/hive_data' row format delimited fields terminated by '|' select * from temperature_data_vw1;
FAILED: SemanticException [Error 10001]: Line 1:119 Table not found 'temperature_data_vw1'
hive> insert overwrite local directory '/home/acadgild/hive_data' row format delimited fields terminated by '|' select * from temperature_data_vw;
Query ID = acadgild_20170701230000_ff2e0af4-dcd5-4109-973a-246f77b7c0ff
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_1498924951542_0013, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0013/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 23:00:24,311 Stage-1 map = 0%, reduce = 0%
2017-07-01 23:00:35,745 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.77 sec
2017-07-01 23:00:49,113 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.31 sec
MapReduce Total cumulative CPU time: 6 seconds 310 msec
Ended Job = job_1498924951542_0013
Copying data to local directory /home/acadgild/hive_data
Copying data to local directory /home/acadgild/hive_data
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.31 sec HDFS Read: 649 HDFS Write: 16 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 310 msec
OK
Time taken: 41.071 seconds
hive>
```

The contents of view is written in local directory :



Contents of the file in which data is exported from hive, the fields are '|' separated.



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help

Acadgild_64bit x
HIVE: EXECUTION ERROR, RETURN CODE 1 FROM org.apache.hadoop.hive.qc.ExecDDLTask: AlreadyExistsException(message: table temperature_data_vw already exists)
hive> create view temperature_data_vw1 as select max(temp),min(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;
OK
Time taken: 0.218 seconds
hive> select * from temperature_data_vw1;
Query ID = acadgild_20170701221212_510924fd-79d5-454b-8c12-35cdb7048968
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_1498924951542_0006, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0006/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:12:33,201 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:12:44,715 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.1 sec
2017-07-01 22:12:59,230 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.73 sec
MapReduce Total cumulative CPU time: 6 seconds 730 msec
Ended Job = job_1498924951542_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.73 sec HDFS Read: 649 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 730 msec
OK
23 9
22 10
Time taken: 40.352 seconds, Fetched: 2 row(s)
hive> insert overwrite local directory '/home/acadgild/hive_data' row format delimited fields terminated by '|' select * from temperature_data_vw1;
Query ID = acadgild_20170701221313_ee8eb340-d014-44e8-bcd9-ad90f8f683a9
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
Current workspace: "Workspace 1"
acadgild@localhost:~ acadgild hive_data 000000_0 (~hive_dat...
```

```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help

Acadgild_64bit x
HIVE: EXECUTION ERROR, RETURN CODE 1 FROM org.apache.hadoop.hive.qc.ExecDDLTask: AlreadyExistsException(message: table temperature_data_vw already exists)
hive> create view temperature_data_vw1 as select max(temp),min(temp) from temperature_data group by substr(rec_date,7,4) having count(temp)>2;
OK
Time taken: 0.218 seconds
hive> select * from temperature_data_vw1;
Query ID = acadgild_20170701221212_510924fd-79d5-454b-8c12-35cdb7048968
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_1498924951542_0006, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0006/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:12:33,201 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:12:44,715 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.1 sec
2017-07-01 22:12:59,230 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.73 sec
MapReduce Total cumulative CPU time: 6 seconds 730 msec
Ended Job = job_1498924951542_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.73 sec HDFS Read: 649 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 730 msec
OK
23 9
22 10
Time taken: 40.352 seconds, Fetched: 2 row(s)
hive> insert overwrite local directory '/home/acadgild/hive_data' row format delimited fields terminated by '|' select * from temperature_data_vw1;
Query ID = acadgild_20170701221313_ee8eb340-d014-44e8-bcd9-ad90f8f683a9
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_1498924951542_0007, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0007/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-01 22:13:30,228 Stage-1 map = 0%, reduce = 0%
2017-07-01 22:13:42,639 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.29 sec
2017-07-01 22:13:54,937 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.68 sec
MapReduce Total cumulative CPU time: 6 seconds 680 msec
Ended Job = job_1498924951542_0007
Copying data to local directory /home/acadgild/hive_data
Copying data to local directory /home/acadgild/hive_data
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.68 sec HDFS Read: 649 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 680 msec
OK
Time taken: 39.029 seconds
hive>
acadgild@localhost:~ acadgild hive_data 000000_0 (~hive_dat...
```



