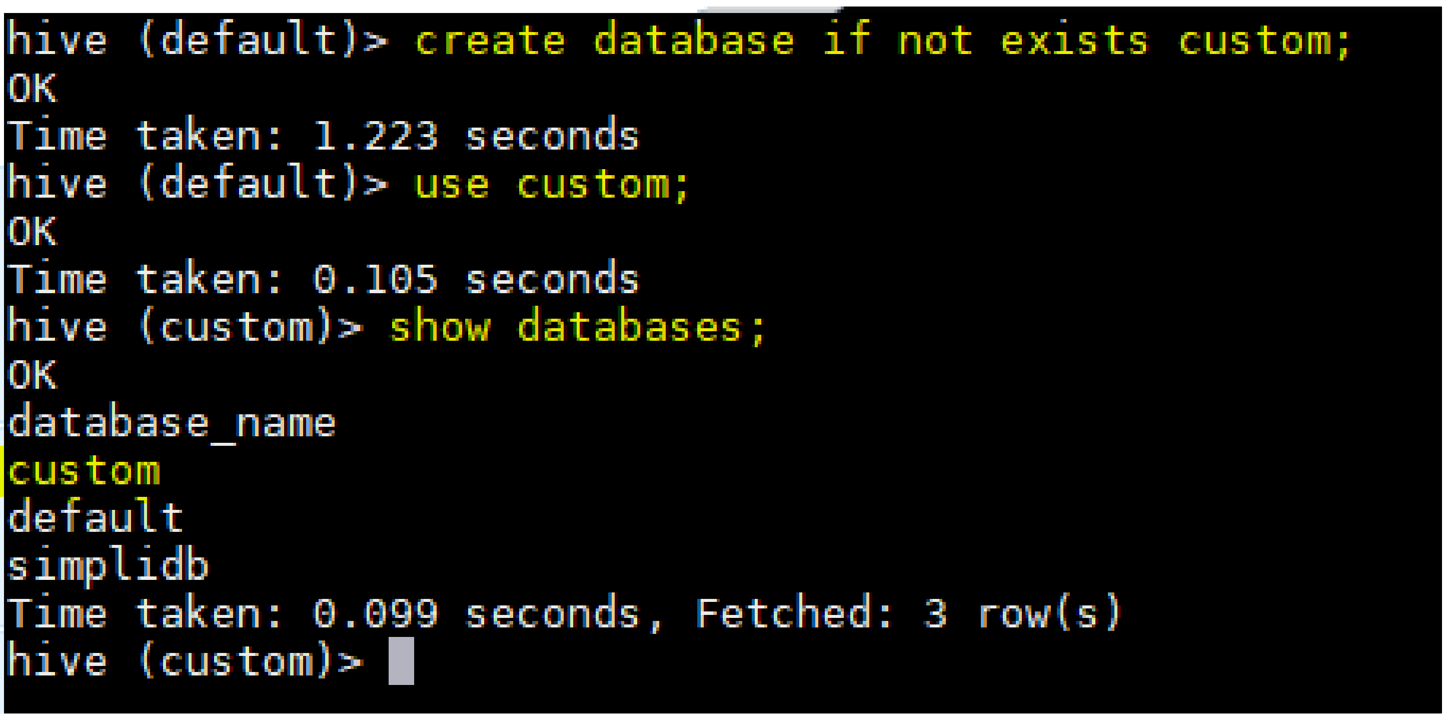
Task 1-

Create a database named 'custom'.

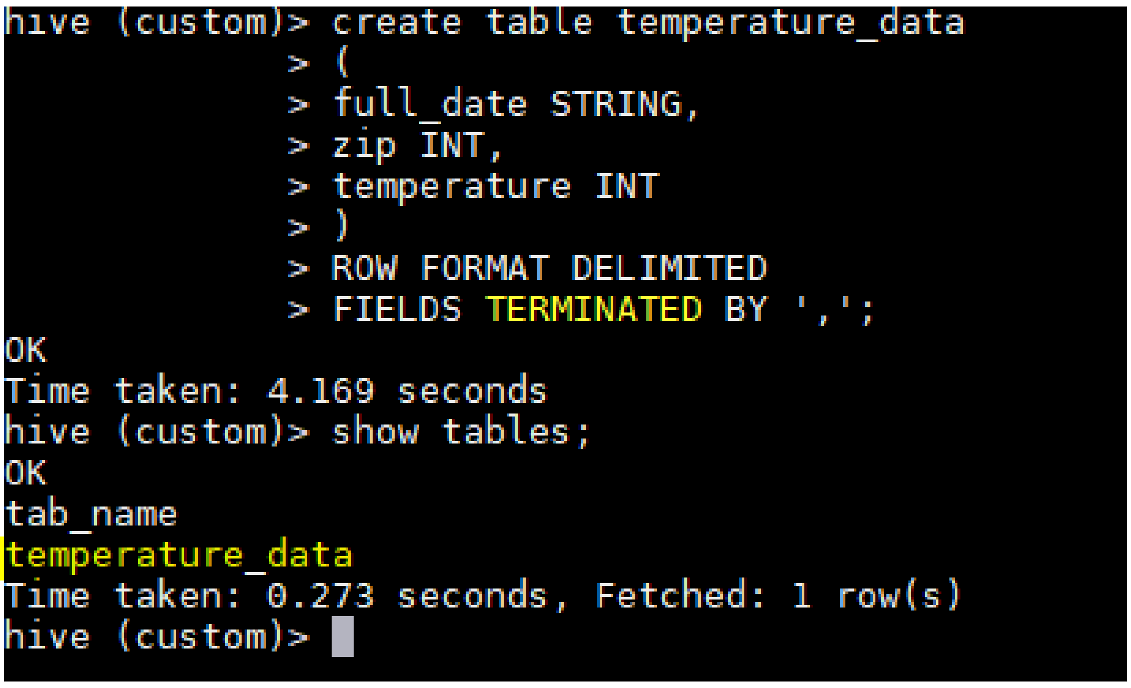
Below screenshot shows the creation of database ‘custom’



```
hive (default)> create database if not exists custom;
OK
Time taken: 1.223 seconds
hive (default)> use custom;
OK
Time taken: 0.105 seconds
hive (custom)> show databases;
OK
database_name
custom
default
simplidb
Time taken: 0.099 seconds, Fetched: 3 row(s)
hive (custom)> 
```

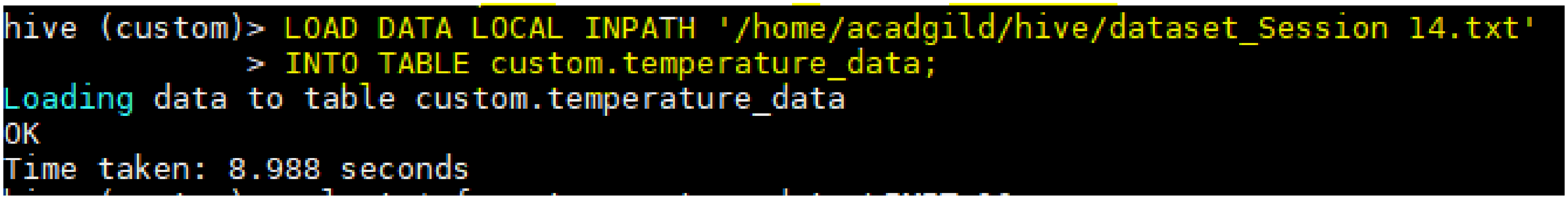
Create a table named temperature_data inside custom having below fields:
1. date (mm-dd-yyyy) format
2. zip code
3. temperature

Below screenshot shows the table being created temperature_data-



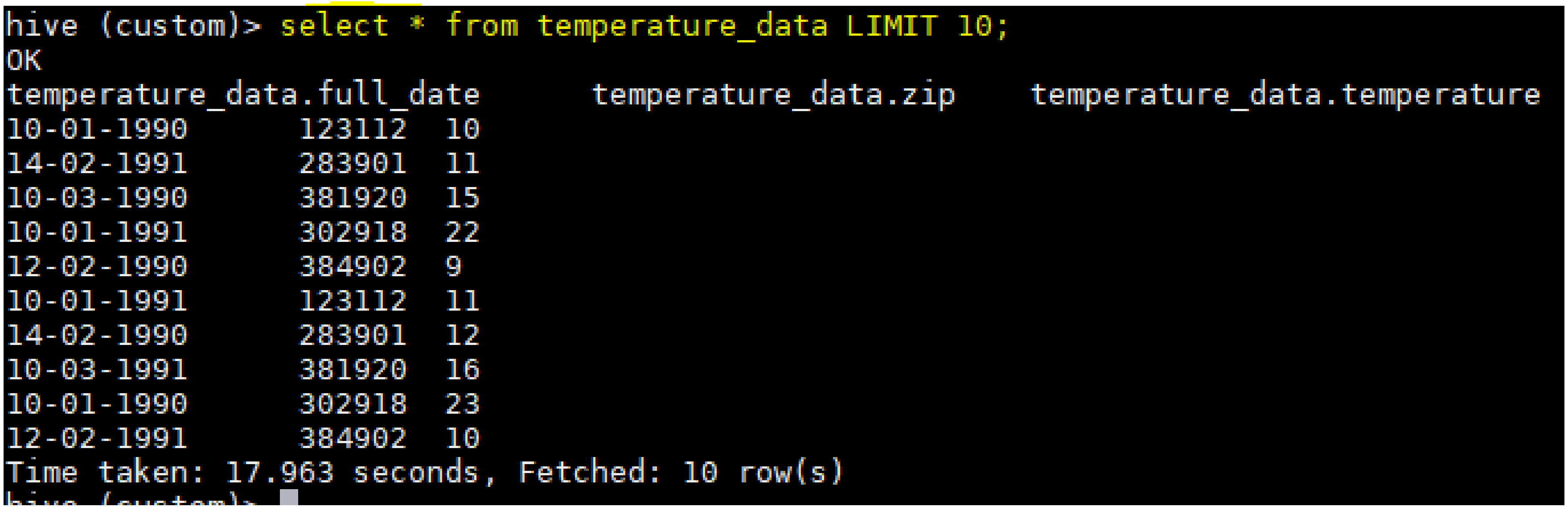
```
hive (custom)> create table temperature_data
             > (
             > full_date STRING,
             > zip INT,
             > temperature INT
             > )
             > ROW FORMAT DELIMITED
             > FIELDS TERMINATED BY ',';
OK
Time taken: 4.169 seconds
hive (custom)> show tables;
OK
tab_name
temperature_data
Time taken: 0.273 seconds, Fetched: 1 row(s)
hive (custom)> 
```

We are then loading the table -



```
hive (custom)> LOAD DATA LOCAL INPATH '/home/acadgild/hive/dataset_Session 14.txt'
             > INTO TABLE custom.temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 8.988 seconds
```

Below screenshot shows that the table has been loaded successfully-

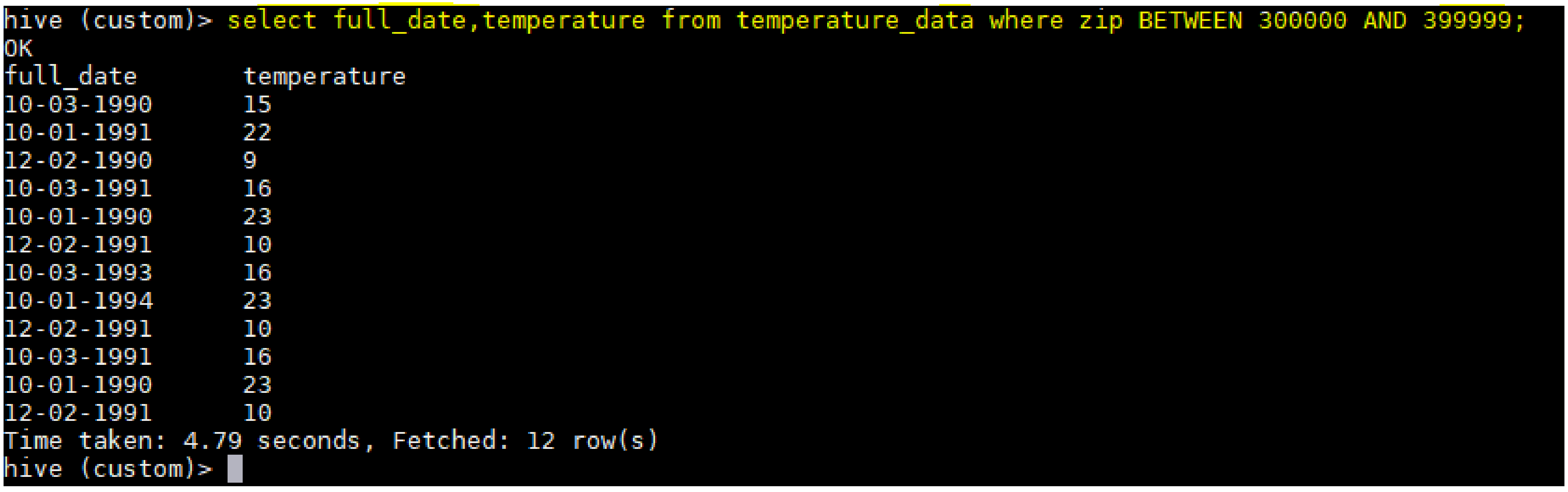


```
hive (custom)> select * from temperature_data LIMIT 10;
OK
temperature_data.full_date      temperature_data.zip    temperature_data.temperature
10-01-1990      123112  10
14-02-1991      283901  11
10-03-1990      381920  15
10-01-1991      302918  22
12-02-1990      384902  9
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
Time taken: 17.963 seconds, Fetched: 10 row(s)
hive (custom)> 
```

Task 2

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

Below query shows the same-



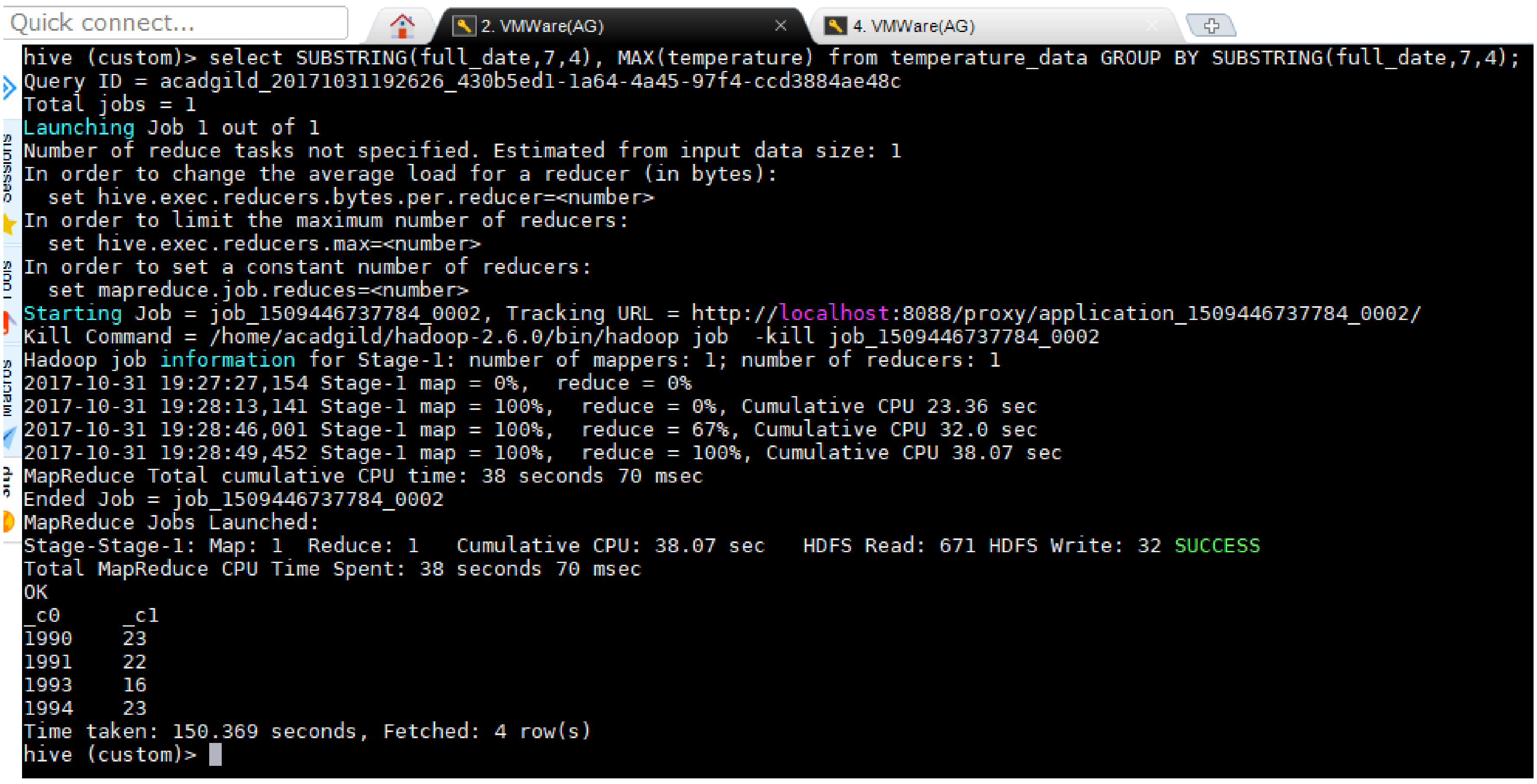
```
hive (custom)> select full_date,temperature from temperature_data where zip BETWEEN 300000 AND 399999;
OK
full_date       temperature
10-03-1990      15
10-01-1991      22
12-02-1990      9
10-03-1991      16
10-01-1990      23
12-02-1991      10
10-03-1993      16
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 4.79 seconds, Fetched: 12 row(s)
hive (custom)> 
```

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

Below query shows the same-

select substring(full_date,7,4) as year,max(temperature) as max_temp from temperature_data GROUP BY substring(full_date,7,4);

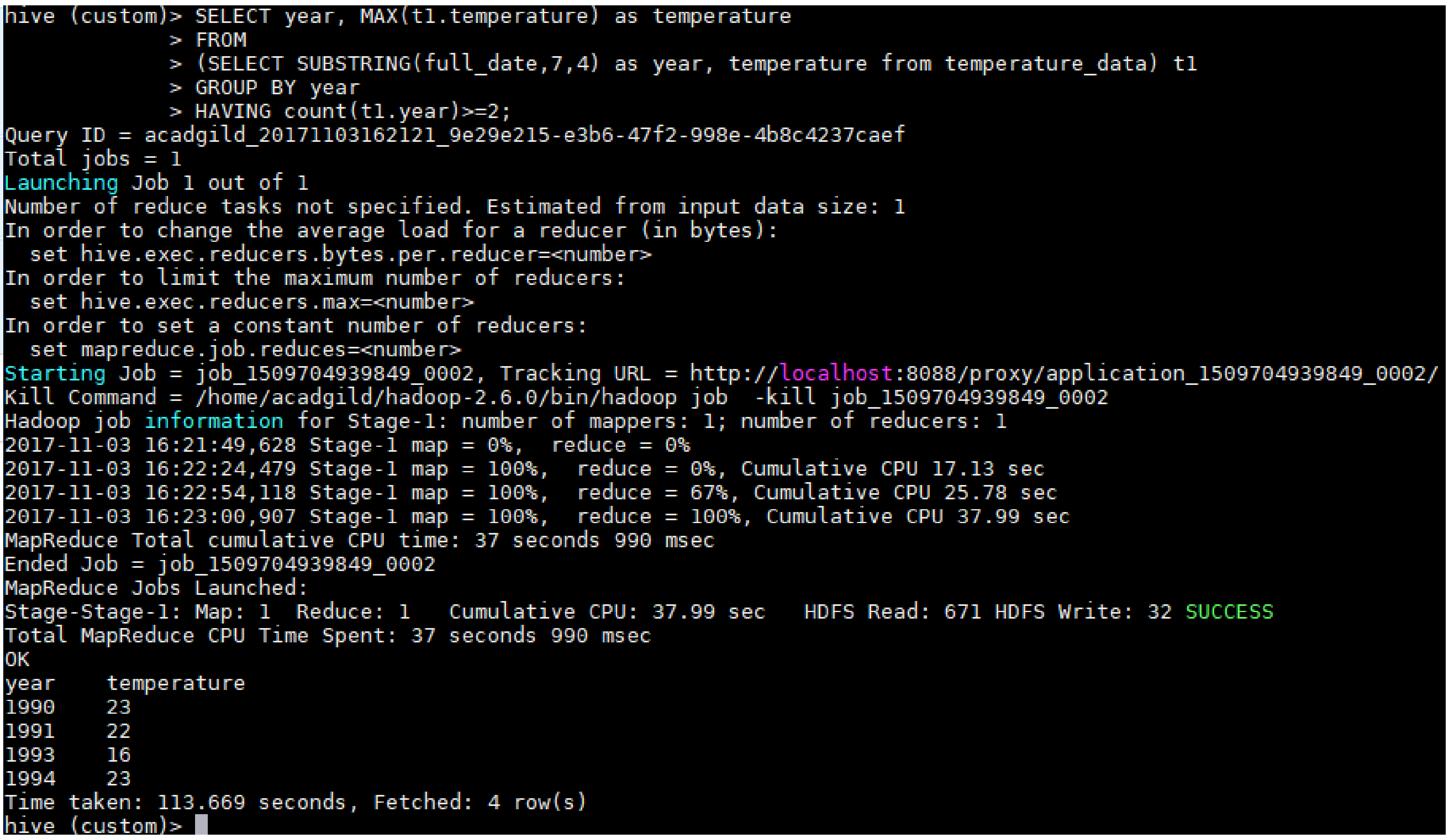
Trigger below querry to get the desired results-



```
hive (custom)> select SUBSTRING(full_date,7,4), MAX(temperature) from temperature_data GROUP BY SUBSTRING(full_date,7,4);
Query ID = acadgild_20171031192626_430b5ed1-1a64-4a45-97f4-ccd3884ae48c
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_1509446737784_0002, Tracking URL = http://localhost:8088/proxy/application_1509446737784_0002/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1509446737784_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-10-31 19:27:27,154 Stage-1 map = 0%,  reduce = 0%
2017-10-31 19:28:13,141 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 23.36 sec
2017-10-31 19:28:46,001 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 32.0 sec
2017-10-31 19:28:49,452 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 38.07 sec
MapReduce Total cumulative CPU time: 38 seconds 70 msec
Ended Job = job_1509446737784_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 38.07 sec   HDFS Read: 671 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 38 seconds 70 msec
OK
_c0     _c1
1990    23
1991    22
1993    16
1994    23
Time taken: 150.369 seconds, Fetched: 4 row(s)
hive (custom)> 
```

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

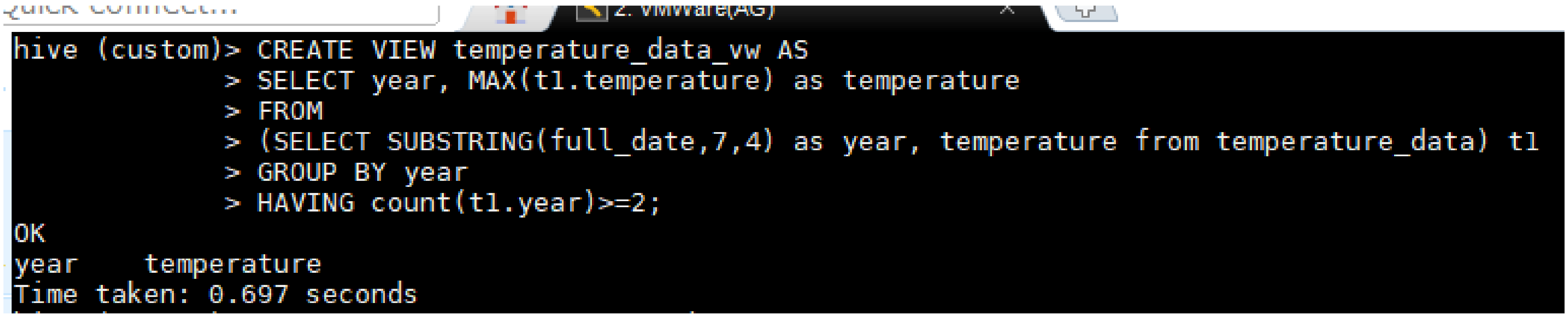
Below querry gives the desired result-



```
hive (custom)> SELECT year, MAX(t1.temperature) as temperature
             > FROM
             > (SELECT SUBSTRING(full_date,7,4) as year, temperature from temperature_data) t1
             > GROUP BY year
             > HAVING count(t1.year)>=2;
Query ID = acadgild_20171103162121_9e29e215-e3b6-47f2-998e-4b8c4237caef
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_1509704939849_0002, Tracking URL = http://localhost:8088/proxy/application_1509704939849_0002/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1509704939849_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-03 16:21:49,628 Stage-1 map = 0%,  reduce = 0%
2017-11-03 16:22:24,479 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 17.13 sec
2017-11-03 16:22:54,118 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 25.78 sec
2017-11-03 16:23:00,907 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 37.99 sec
MapReduce Total cumulative CPU time: 37 seconds 990 msec
Ended Job = job_1509704939849_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 37.99 sec   HDFS Read: 671 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 37 seconds 990 msec
OK
year    temperature
1990    23
1991    22
1993    16
1994    23
Time taken: 113.669 seconds, Fetched: 4 row(s)
hive (custom)>
```

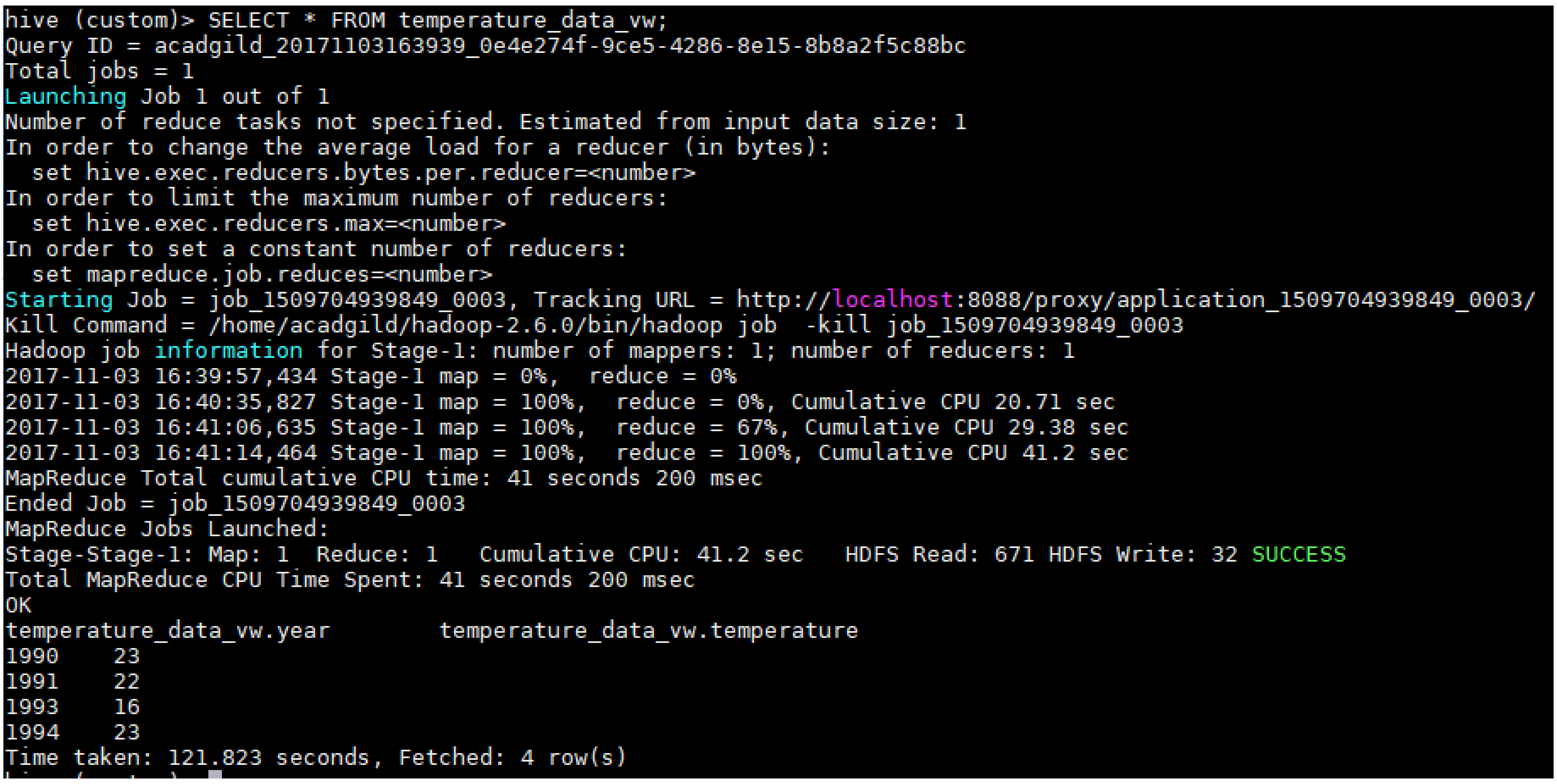
- **Create a view on the top of last query, name it temperature_data_vw.**

Creating the VIEW-



```
hive (custom)> CREATE VIEW temperature_data_vw AS
             > SELECT year, MAX(t1.temperature) as temperature
             > FROM
             > (SELECT SUBSTRING(full_date,7,4) as year, temperature from temperature_data) t1
             > GROUP BY year
             > HAVING count(t1.year)>=2;
OK
year    temperature
Time taken: 0.697 seconds
```

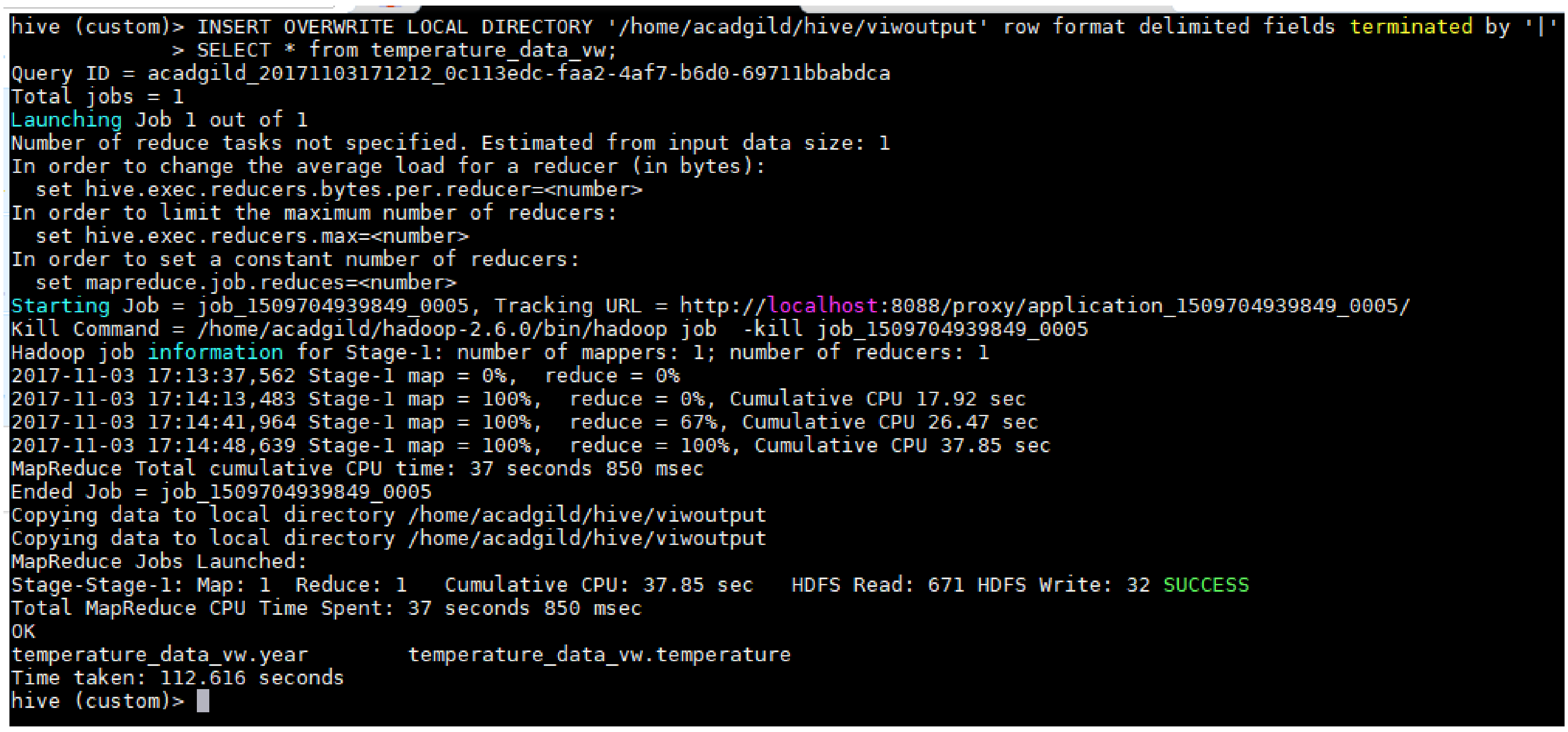
Selecting Contents of VIEW-



```
hive (custom)> SELECT * FROM temperature_data_vw;
Query ID = acadgild_20171103163939_0e4e274f-9ce5-4286-8e15-8b8a2f5c88bc
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_1509704939849_0003, Tracking URL = http://localhost:8088/proxy/application_1509704939849_0003/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1509704939849_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-03 16:39:57,434 Stage-1 map = 0%,  reduce = 0%
2017-11-03 16:40:35,827 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 20.71 sec
2017-11-03 16:41:06,635 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 29.38 sec
2017-11-03 16:41:14,464 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 41.2 sec
MapReduce Total cumulative CPU time: 41 seconds 200 msec
Ended Job = job_1509704939849_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 41.2 sec   HDFS Read: 671 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 41 seconds 200 msec
OK
temperature_data_vw.year        temperature_data_vw.temperature
1990    23
1991    22
1993    16
1994    23
Time taken: 121.823 seconds, Fetched: 4 row(s)
```

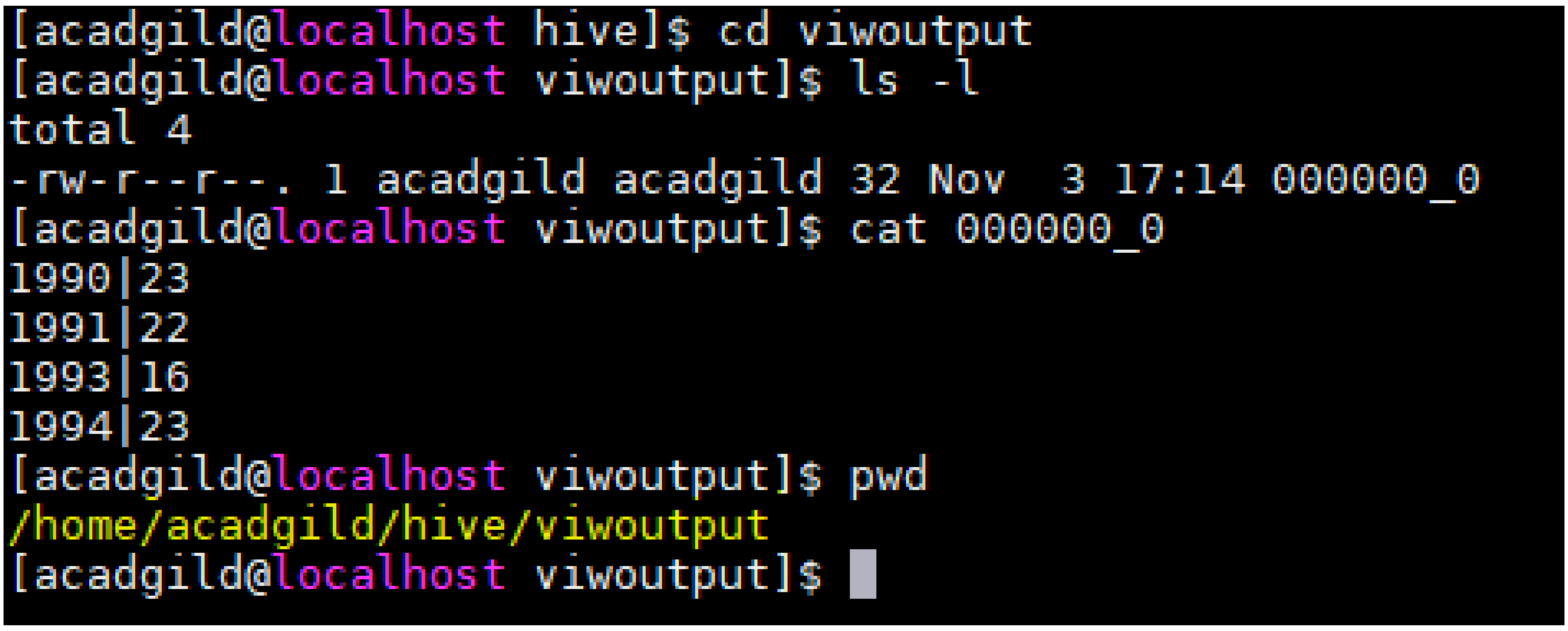
- **Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.**

Insering from VIEW to local file system-



```
hive (custom)> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hive/viwoutput' row format delimited fields terminated by '|'
             > SELECT * from temperature_data_vw;
Query ID = acadgild_20171103171212_0c113edc-faa2-4af7-b6d0-69711bbabdca
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_1509704939849_0005, Tracking URL = http://localhost:8088/proxy/application_1509704939849_0005/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job  -kill job_1509704939849_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-03 17:13:37,562 Stage-1 map = 0%,  reduce = 0%
2017-11-03 17:14:13,483 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 17.92 sec
2017-11-03 17:14:41,964 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 26.47 sec
2017-11-03 17:14:48,639 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 37.85 sec
MapReduce Total cumulative CPU time: 37 seconds 850 msec
Ended Job = job_1509704939849_0005
Copying data to local directory /home/acadgild/hive/viwoutput
Copying data to local directory /home/acadgild/hive/viwoutput
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 37.85 sec   HDFS Read: 671 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 37 seconds 850 msec
OK
temperature_data_vw.year        temperature_data_vw.temperature
Time taken: 112.616 seconds
hive (custom)>
```

Data copied in local file path-



```
[acadgild@localhost hive]$ cd viwoutput
[acadgild@localhost viwoutput]$ ls -l
total 4
-rw-r--r--. 1 acadgild acadgild 32 Nov  3 17:14 000000_0
[acadgild@localhost viwoutput]$ cat 000000_0
1990|23
1991|22
1993|16
1994|23
[acadgild@localhost viwoutput]$ pwd
/home/acadgild/hive/viwoutput
[acadgild@localhost viwoutput]$
```