//Creating database custom

CREATE DATABASE IF NOT EXISTS custom;

//Creating table

CREATE TABLE IF NOT EXISTS custom.temperature\_details

(

start\_date string,

zip\_code int,

temperature int

)

row format delimited

fields terminated by ',';

//Loading data

LOAD DATA LOCAL

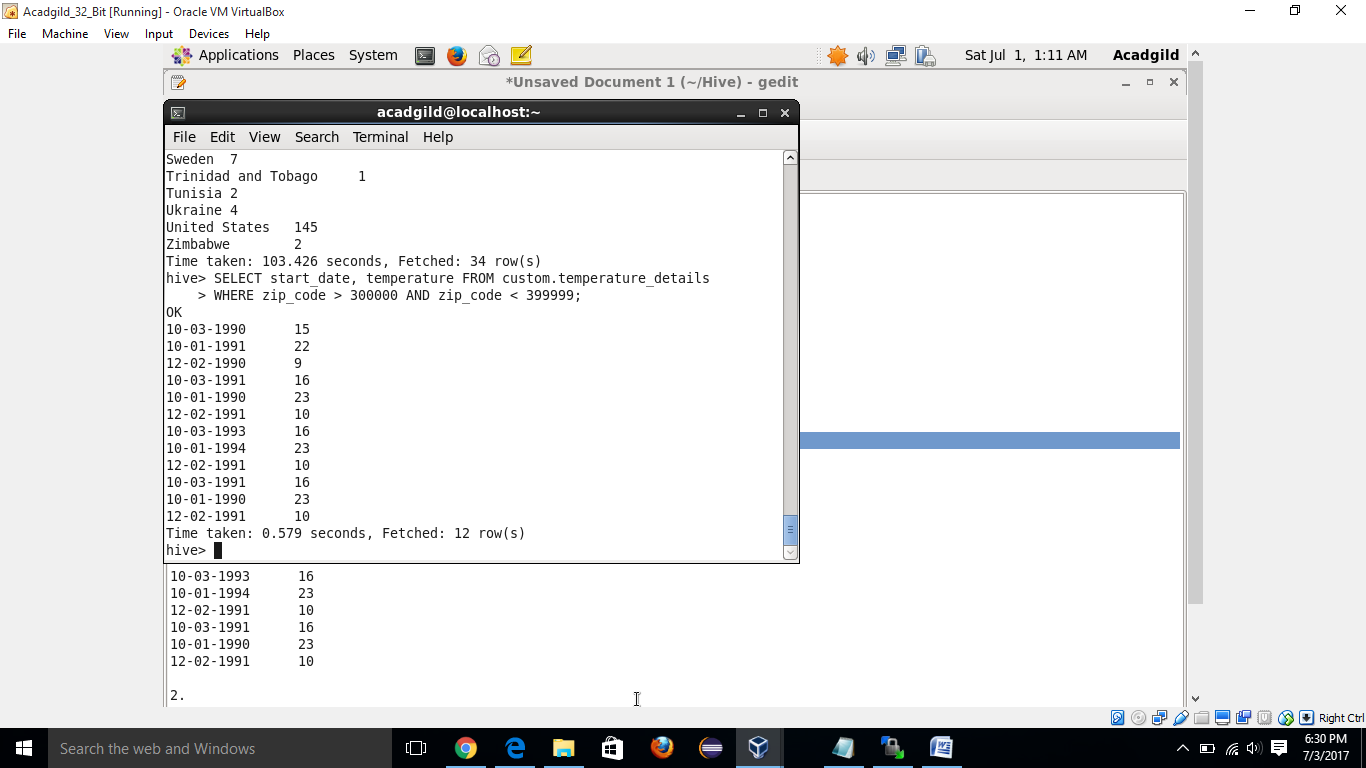
INPATH '/home/acadgild/Hive/dataset.txt'

INTO TABLE temperature\_details;

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

Hive> SELECT start\_date, temperature FROM custom.temperature\_details

Hive> WHERE zip\_code > 300000 AND zip\_code < 399999;

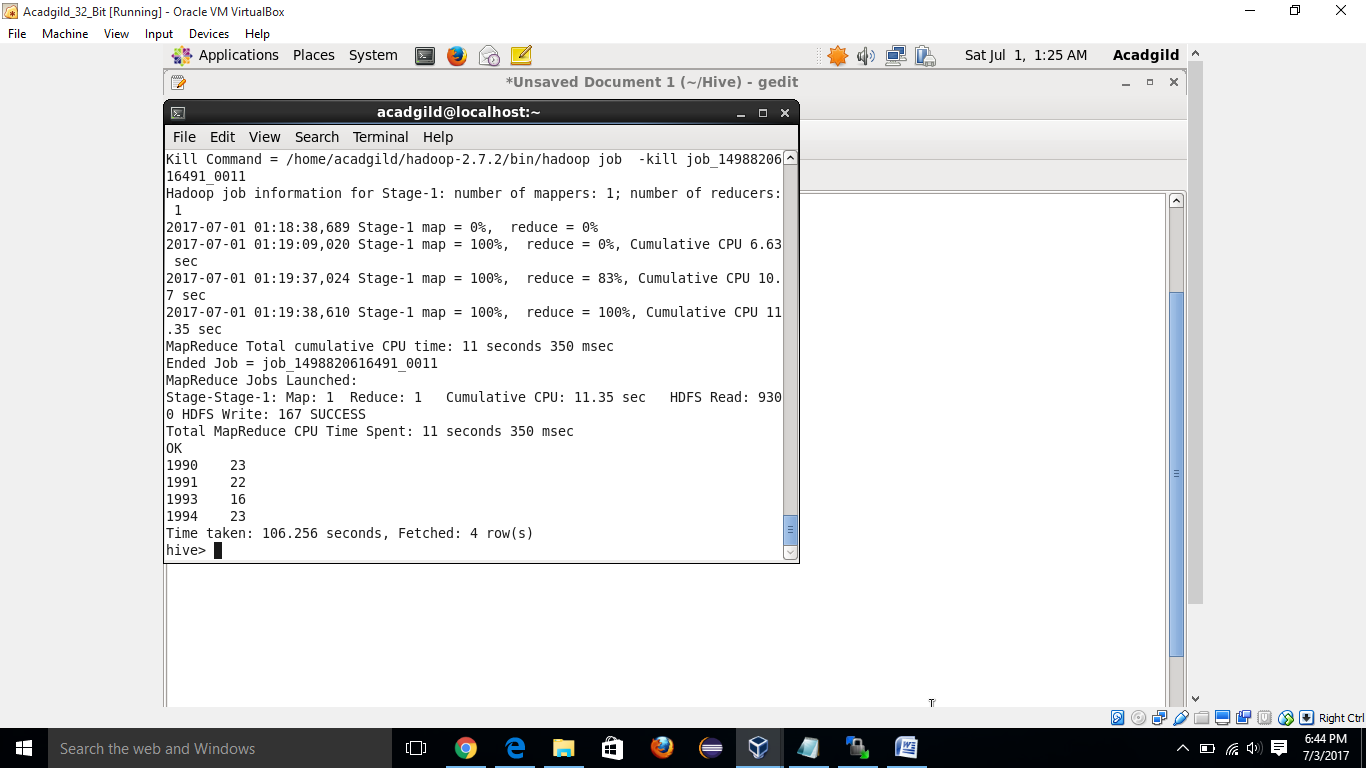


1. Calculate maximum temperature corresponding to every year from temperature\_data table.

Hive> SELECT SUBSTR(start\_date,7,4) AS year, MAX(temperature)

Hive>FROM custom.temperature\_details

Hive>GROUP BY SUBSTR(start\_date,7,4);



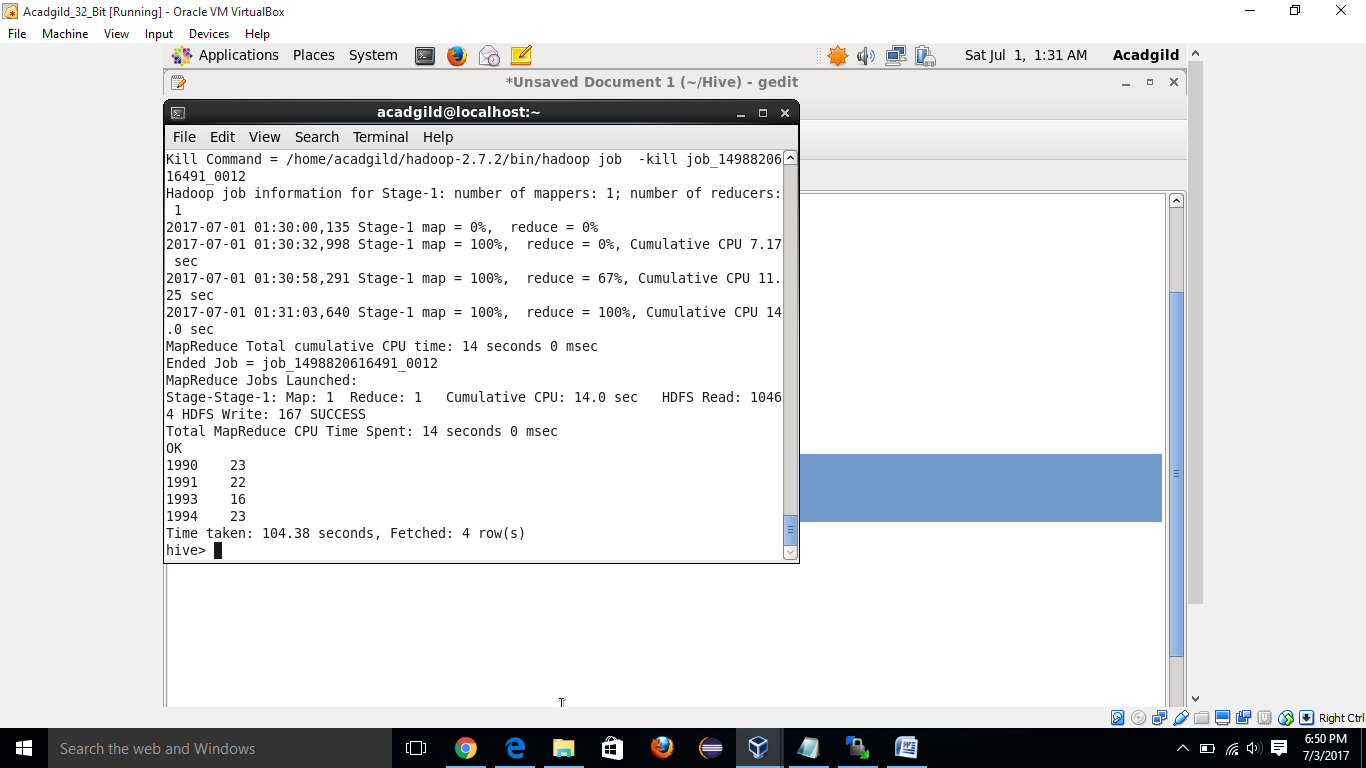
1. Calculate maximum temperature from temperature\_data table corresponding to those years which have at least 2 entries in the table.

Hive>SELECT SUBSTR(start\_date,7,4) AS year, MAX(temperature)

Hive>FROM custom.temperature\_details

Hive>GROUP BY SUBSTR(start\_date,7,4)

Hive>HAVING COUNT(SUBSTR(start\_date,7,4)) > 1;



1. Create a view on the top of last query, name it temperature\_data\_vw.

Hive>CREATE VIEW temperature\_data\_vw AS

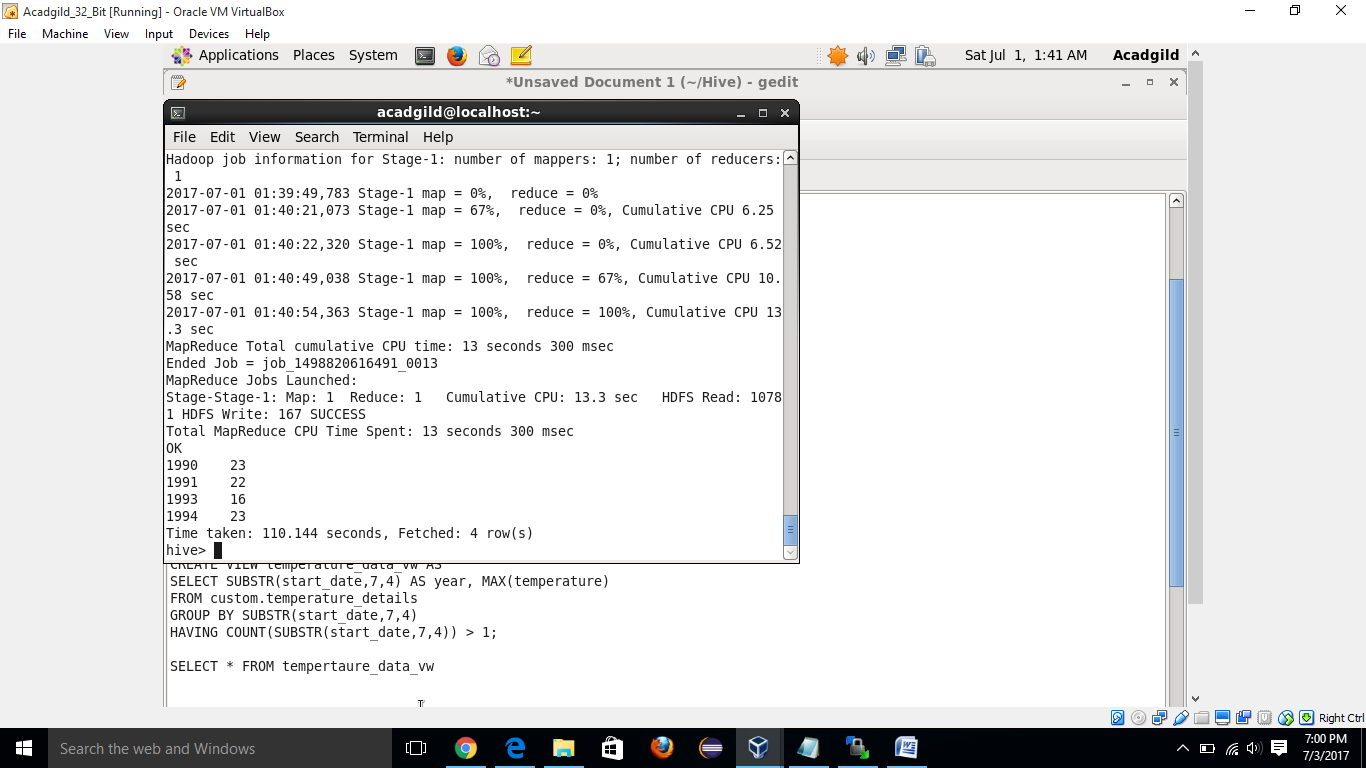
Hive>SELECT SUBSTR(start\_date,7,4) AS year, MAX(temperature)

Hive>FROM custom.temperature\_details

Hive>GROUP BY SUBSTR(start\_date,7,4)

Hive>HAVING COUNT(SUBSTR(start\_date,7,4)) > 1;

Hive>SELECT \* FROM tempertaure\_data\_vw



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

Hive>INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hive/output'

Hive>ROW FORMAT DELIMITED

Hive>FIELDS TERMINATED BY '|'

Hive>SELECT \* FROM temperature\_data\_vw

