Assignment 6.2 ( Hive Introduction)

Task1: Fetch date and temperature from temperature\_data where zip code is greater an 300000 and less than 399999

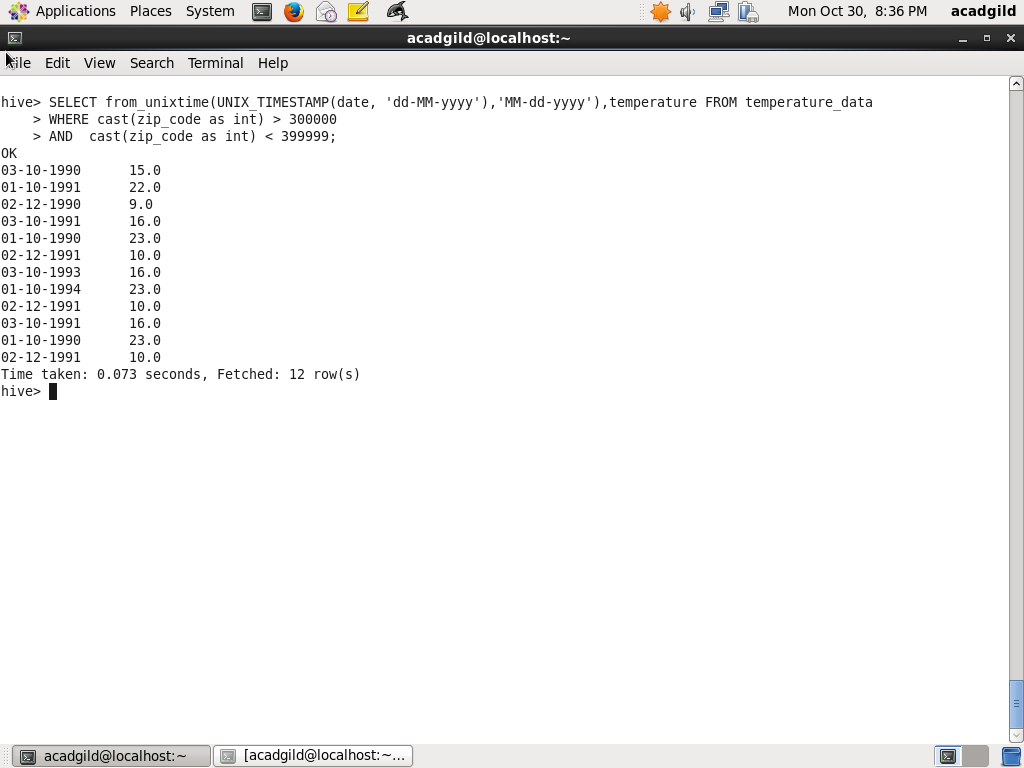
The following query will perform this task zip code is checked if it is between

300000 and 399999

SELECT from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy'), 'MM-dd-yyyy'),temperature FROM temperature\_data

WHERE cast(zip\_code as int) > 300000 AND cast(zip\_code as int) < 399999;

Screenshot of query with output is as below:



Task2: Calculate maximum temperature of each year

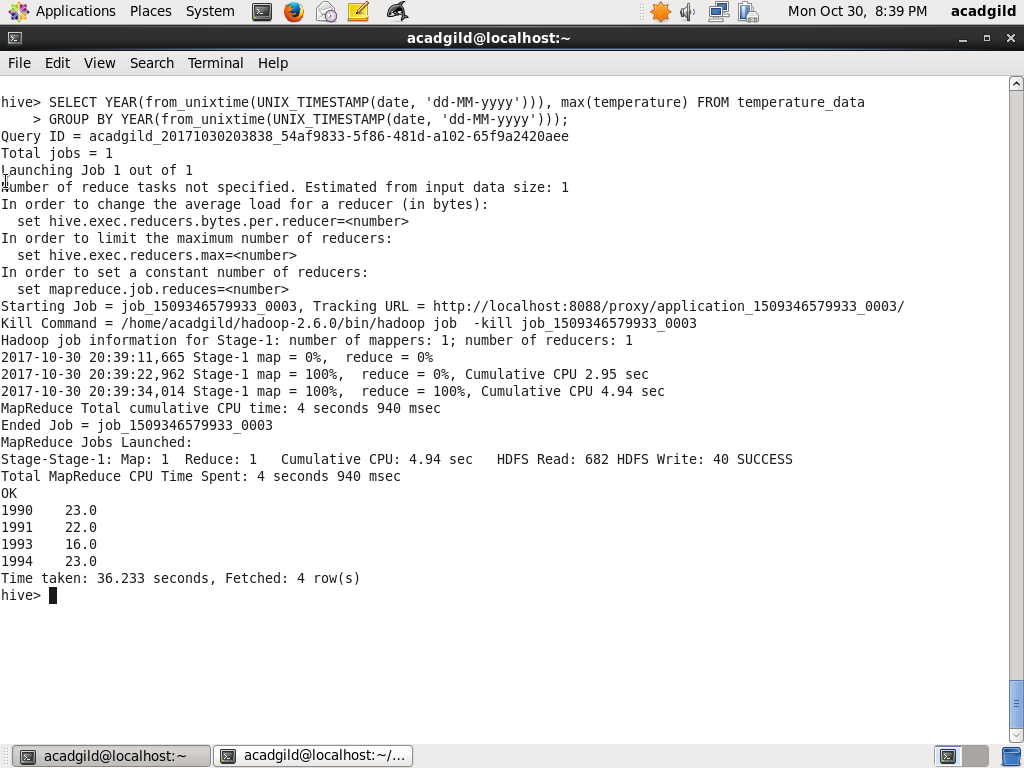
The following query will do the task, YEAR function is to get Year and max function is used

To get the maximum temperature group by year

SELECT YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy'))), max(temperature) FROM temperature\_data

GROUP BY YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy')));

Screenshot of Hive query and output is as below:



Task3:

Calculate maximum temperature corresponding to each year having at least 2 entries

The following query will do the task, YEAR function is to get Year and max function is used

To get the maximum temperature group by year and count function is used to check number

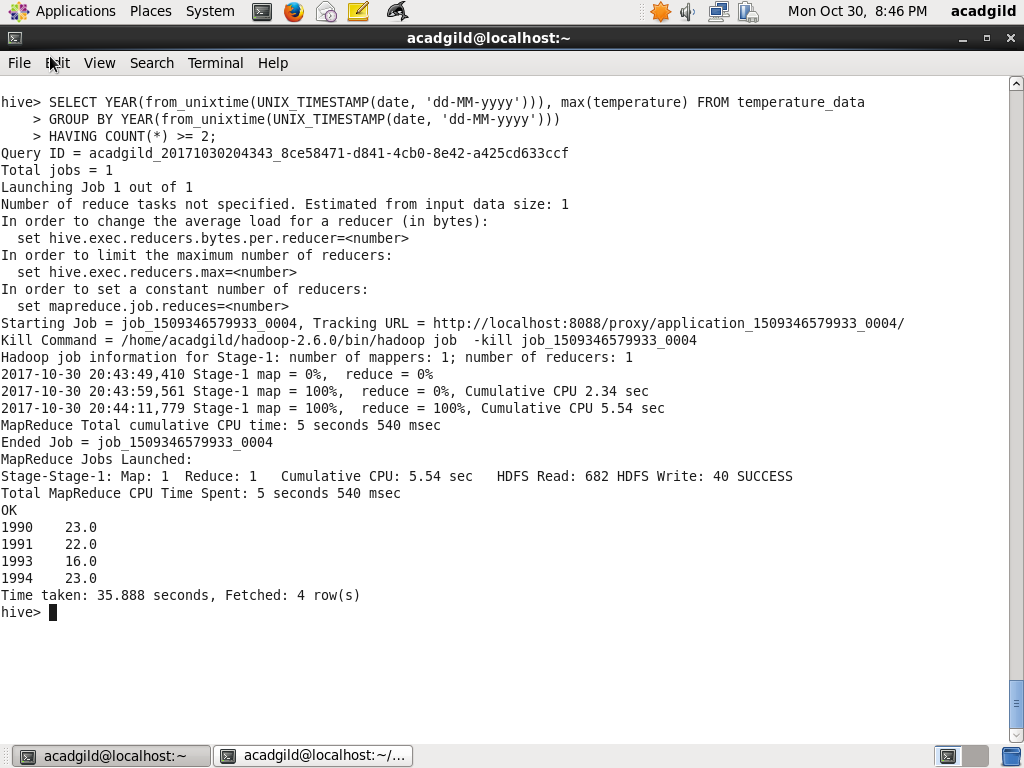
Of records of year is equal to or greater than 2

SELECT YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy'))), max(temperature) FROM temperature\_data

GROUP BY YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy')))

HAVING COUNT(\*) >= 2;

Screenshot of Hive query and output is as below:



Task 4: Create a view of last query and store it as temperature\_data\_dw

Using CREATE VIEW temperature\_data\_vw with last query the view is created. Query is as below:

CREATE VIEW temperature\_data\_vw AS

SELECT YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy'))), max(temperature) FROM temperature\_data

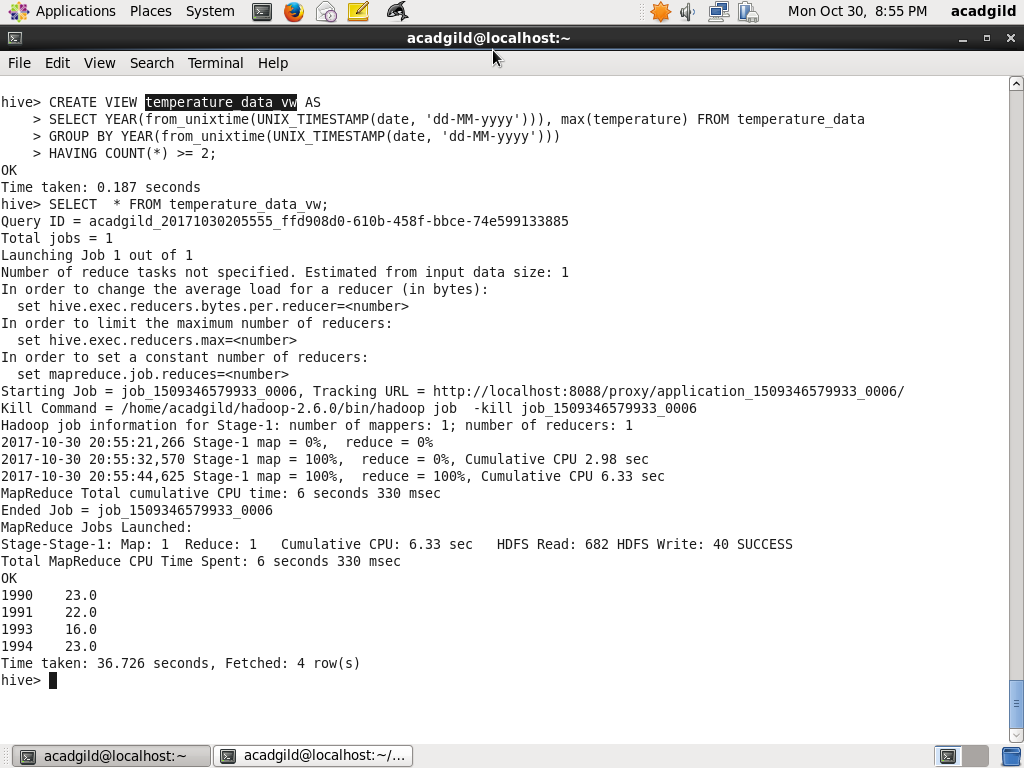
GROUP BY YEAR(from\_unixtime(UNIX\_TIMESTAMP(date, 'dd-MM-yyyy')))

HAVING COUNT(\*) >= 2;

The records of temperature\_data\_vw are selected using SELECT query as below

SELECT \* FROM temperature\_data\_vw

The screenshot of query and output is as below:



Task5:

Export contents from view temperature\_data\_vw to local file system with | delimited

Using the query below contents of temperature\_data\_vw is saved to local directory

/home/acadgild/assignment\_6.2/output with | delimited

INSERT OVERWRITE LOCAL DIRECTORY ‘/home/acadgild/assignment\_6.2/output' ROW FORMAT DELIMITED

FIELDS TERMINATED BY '|'

SELECT \* FROM temperature\_data\_vw;

The list of files under /home/acadgild/assignment\_6.2/output is viewed

!ls –l /home/acadgild/assignment\_6.2/output

The content is displayed using:

!cat /home/acadgild/assignment\_6.2/output/000000\_0

Screenshot for the query and content is as below:

