**Assignment 6.2:**

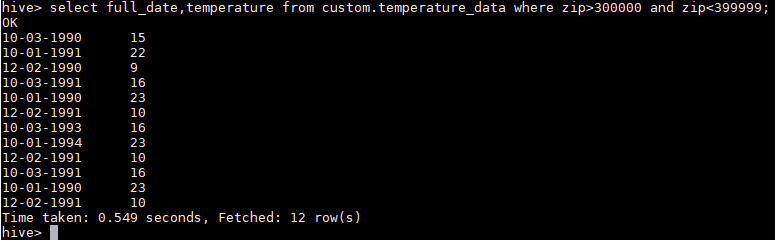
**Problem Statement:**

* **Fetch date and temperature from temperature\_data where zip code is greater than**

**300000 and less than 399999**.

**Steps**:

hive> select full\_date,temperature from custom.temperature\_data where zip>300000 and zip<399999;

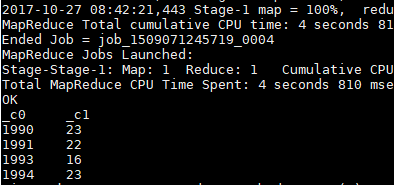


* **Calculate maximum temperature corresponding to every year from temperature\_data**

**table.**

**Steps:**

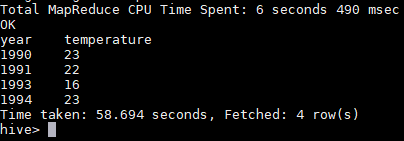
hive> SELECT SUBSTR(full\_date, 7, 10), MAX(temperature) FROM custom.temperature\_data GROUP BY SUBSTR(full\_date, 7, 10);



* **Calculate maximum temperature from temperature\_data table corresponding to those**

**years which have at least 2 entries in the table.**

**Steps:**

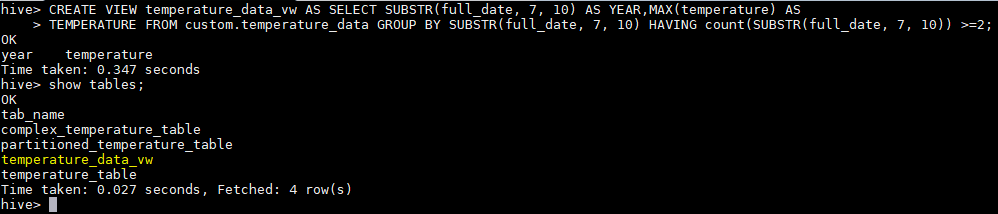
hive> SELECT SUBSTR(full\_date, 7, 10) AS YEAR,MAX(temperature) AS TEMPERATURE FROM custom.temperature\_data GROUP BY SUBSTR(full\_date, 7, 10) HAVING count(SUBSTR(full\_date, 7, 10)) >=2; 

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

**Steps:**

hive> CREATE VIEW temperature\_data\_vw AS SELECT SUBSTR(full\_date, 7, 10) AS YEAR,MAX(temperature) AS

> TEMPERATURE FROM custom.temperature\_data GROUP BY SUBSTR(full\_date, 7, 10) HAVING count(SUBSTR(full\_date, 7, 10)) >=2;



* **Export contents from temperature\_data\_vw to a file in local file system, such that each**

**file is '|' delimited.**

**Steps:**

hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hive/temperature\_data\_vw'

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY '|'

> select \* from temperature\_data\_vw;

