**Session 6: HIVE Introduction**

Assignment 6.2

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 6.1**– Write HIVE commands to achieve the tasks provided in the problem statement.

Contents

[Introduction 2](#_Toc497840190)

[Input Data set 2](#_Toc497840191)

[Problem Statement 2](#_Toc497840192)

[Prerequisite 2](#_Toc497840193)

[Task1 4](#_Toc497840194)

[HIVE Commands 4](#_Toc497840195)

[Output 4](#_Toc497840196)

[Task2 4](#_Toc497840197)

[HIVE Commands 4](#_Toc497840198)

[Output 4](#_Toc497840199)

[TASK 3 5](#_Toc497840200)

[HIVE Commands 5](#_Toc497840201)

[Output 5](#_Toc497840202)

[TASK 4 5](#_Toc497840203)

[HIVE Commands 5](#_Toc497840204)

[Output 5](#_Toc497840205)

[TASK 5 6](#_Toc497840206)

[HIVE Commands 6](#_Toc497840207)

[Output 6](#_Toc497840208)

# Introduction

In this assignment we are going to perform the below tasks using HIVE commands.

# Input Data set

<https://drive.google.com/file/d/0Bxr27gVaXO5sa0JBamZXdkpYUFk/view?usp=sharing>

# Problem Statement

1. Fetch date and temperature from temperature\_data where zip HIVE Commands is greater than 300000 and less than 399999.
2. Calculate maximum temperature corresponding to every year from temperature\_data table.
3. Calculate maximum temperature from temperature\_data table corresponding to those years which have at least 2 entries in the table.
4. Create a view on the top of last query, name it temperature\_data\_vw.
5. Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

# Prerequisite

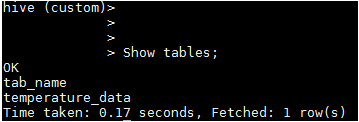
1. Create a DATABASE **“custom”** and create a TABLE “**temperature\_data”** inside **custom** having below fields:

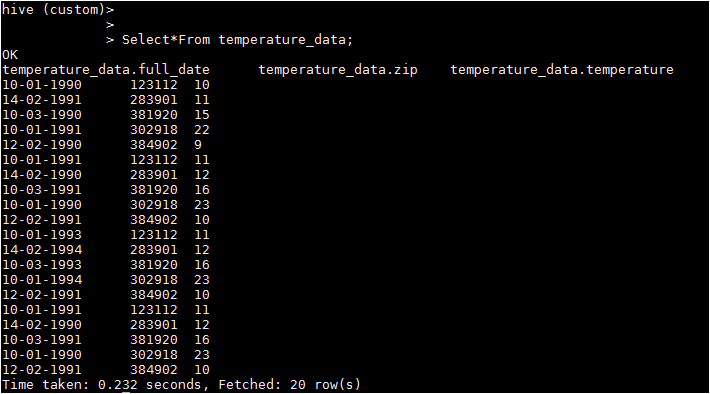
**date (mm-dd-yyyy) format**

**zip code**

**temperature**

As part of the Assignment 6.1 we had created the Database as per the above,





Hence we are going to perform the task provided for this assignment.

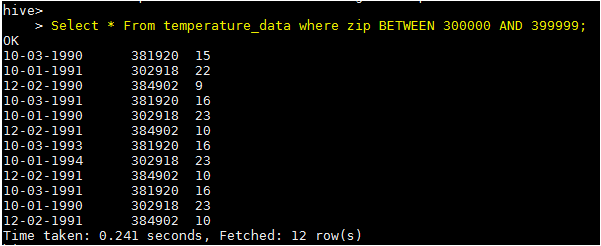
# Task1

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

### HIVE Commands

***hive (custom) >Select \* From temperature\_data where zip BETWEEN 300000 AND 399999;***

### Output

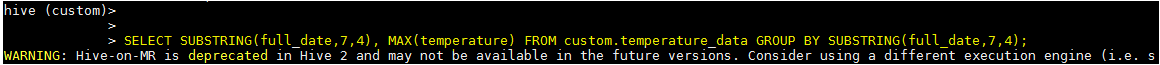


# Task2

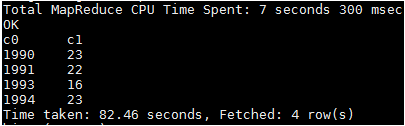
Calculate maximum temperature corresponding to every year from **temperature\_data** table.

### HIVE Commands

***hive (custom) > SELECT SUBSTRING(full\_date,7,4), MAX(temperature) FROM custom.temperature\_data GROUP BY SUBSTRING(full\_date,7,4);***



### Output



# TASK 3

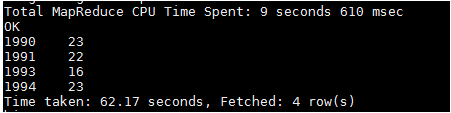
Calculate maximum temperature from **temperature\_data** table corresponding to those years which have at least **2** entries in the table.

### HIVE Commands

***hive(custom)>SELECT full\_date, MAX(t1.temperature) as temperature FROM (SELECT SUBSTRING(full\_date,7,4) full\_date, temperature FROM temperature\_data)t1 GROUP BY full\_date HAVING count(t1.full\_date)>=2;***



### Output

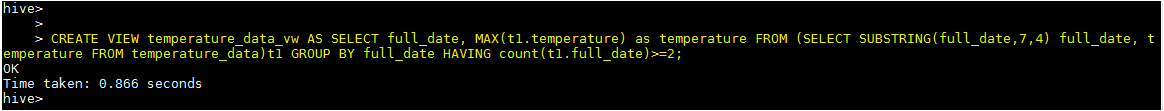


# TASK 4

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

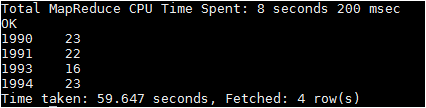
### HIVE Commands

***CREATE VIEW temperature\_data\_vw AS SELECT full\_date, MAX(t1.temperature) as temperature FROM (SELECT SUBSTRING(full\_date,7,4) full\_date, temperature FROM temperature\_data)t1 GROUP BY full\_date HAVING count(t1.full\_date)>=2;***





### Output



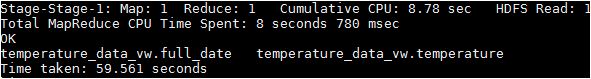
# TASK 5

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

### HIVE Commands

***INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hadoop/temperature\_data\_vw.txt' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|' SELECT \* FROM temperature\_data\_vw;***





### Output

***cat /home/acadgild/hadoop/temperature\_data\_vw.txt/\****

