**Session 06: Hive Introduction**

**Assignment 1**

Create a database named 'custom'.

Create a table named temperature\_data inside custom having below fields:

1. date (mm-dd-yyyy) format

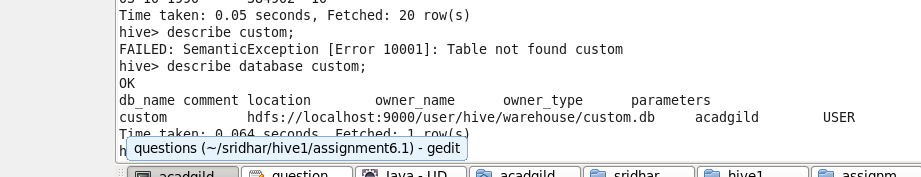
2. zip code

3. temperature

The table will be loaded from comma-delimited file.

Load the dataset.txt (which is ',' delimited) in the table.

CREATE DATABASE IF NOT EXISTS custom;



1. **First we will create a staging table with data type String so that we can store all the values.**

CREATE TABLE stg\_temperature\_data(

date string,

Zip\_code string,

temperature string

)

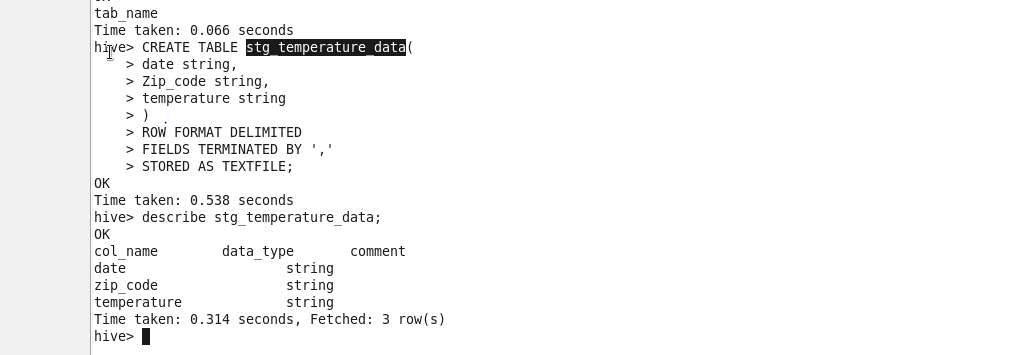
ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

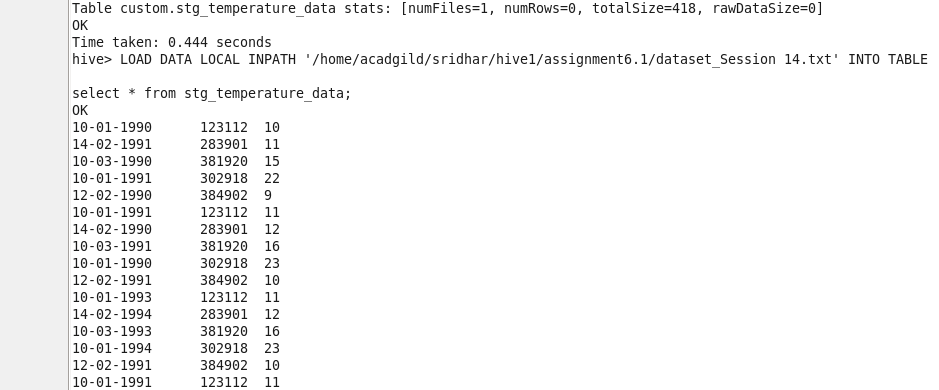
STORED AS TEXTFILE;

1. **Load the staging table:**

LOAD DATA LOCAL INPATH '/home/acadgild/sridhar/hive1/assignment6.1/dataset\_Session 14.txt' INTO TABLE custom.stg\_temperature\_data;



Output records:



1. **Next we will create final table with proper data types , which could be used for data processing**

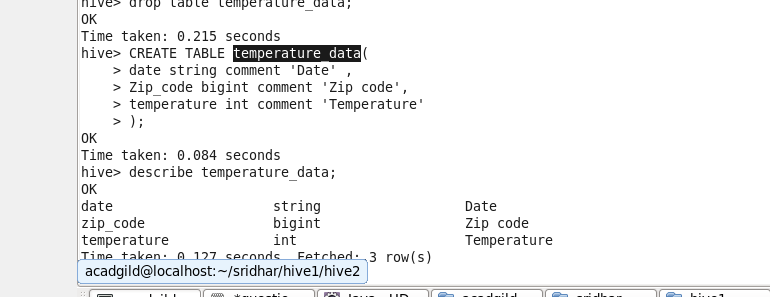
CREATE TABLE temperature\_data(

date string comment 'Date' ,

Zip\_code int comment 'Zip code',

temperature string comment 'Temperature'

);



1. **Loading data into target table:**

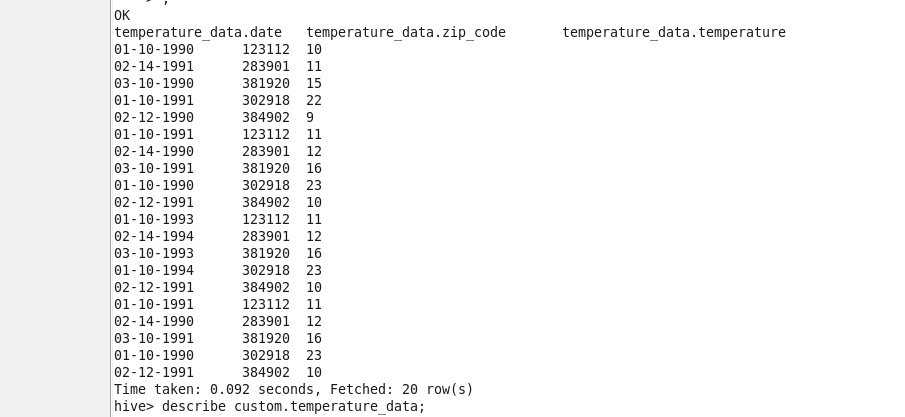
INSERT OVERWRITE TABLE custom.temperature\_data

select

from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy'),'MM-dd-yyyy') as date,

Zip\_code,

temperature from custom.stg\_temperature\_data;



1. **Storing a date column as string just for formatting does not make, yields result which we can do using DATE functions**

So we will alter the type of date column from string to date using to\_date function which converts a string to date type

With Date as our data type we can make much use of processing like getting day, week, month, year etc using date functions on data type

ALTER TABLE temperature\_data CHANGE column date date date;

DESCRIBE temperature\_data;

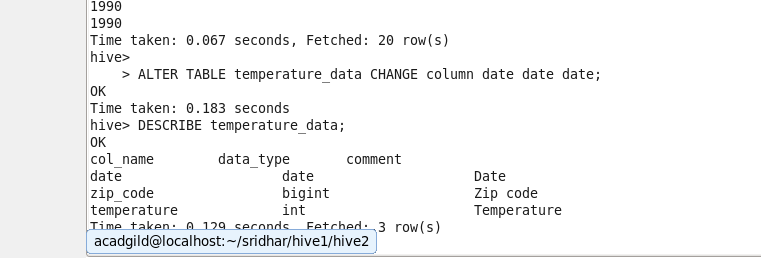
INSERT OVERWRITE TABLE custom.temperature\_data

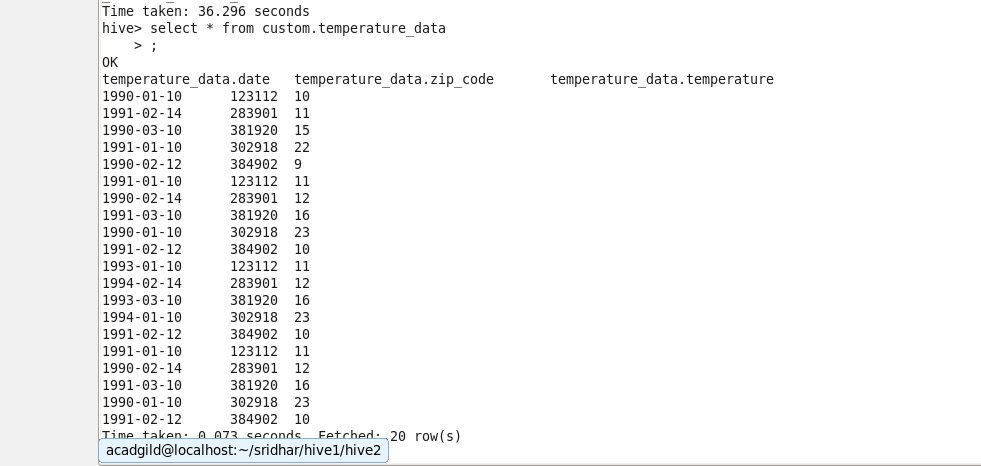
select

**to\_date(from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy'))) as date ,**

Zip\_code,

temperature from custom.stg\_temperature\_data;





**If we still want to view the date in format of (mm-DD-YYYY) we can use query like this:**

select

concat(

lpad(cast(MONTH(to\_date(from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy')))) as string),2,0),"-",

lpad(cast(DAY(to\_date(from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy')))) as string),2,0),"-",

lpad(cast(YEAR(to\_date(from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy')))) as string),2,0)

)as date,

Zip\_code,

temperature

from temperature\_data;

