**Problem Objective**

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.

Code/Script

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table temperature\_data

create table temperature\_data

(date string,zipcode int,temperature float)

row format delimited

fields terminated by ',';

--create temp table for conversion date to expected format (MM-dd-yyyy)

create table temp

(date string,zipcode int,temperature float)

row format delimited

fields terminated by ',';

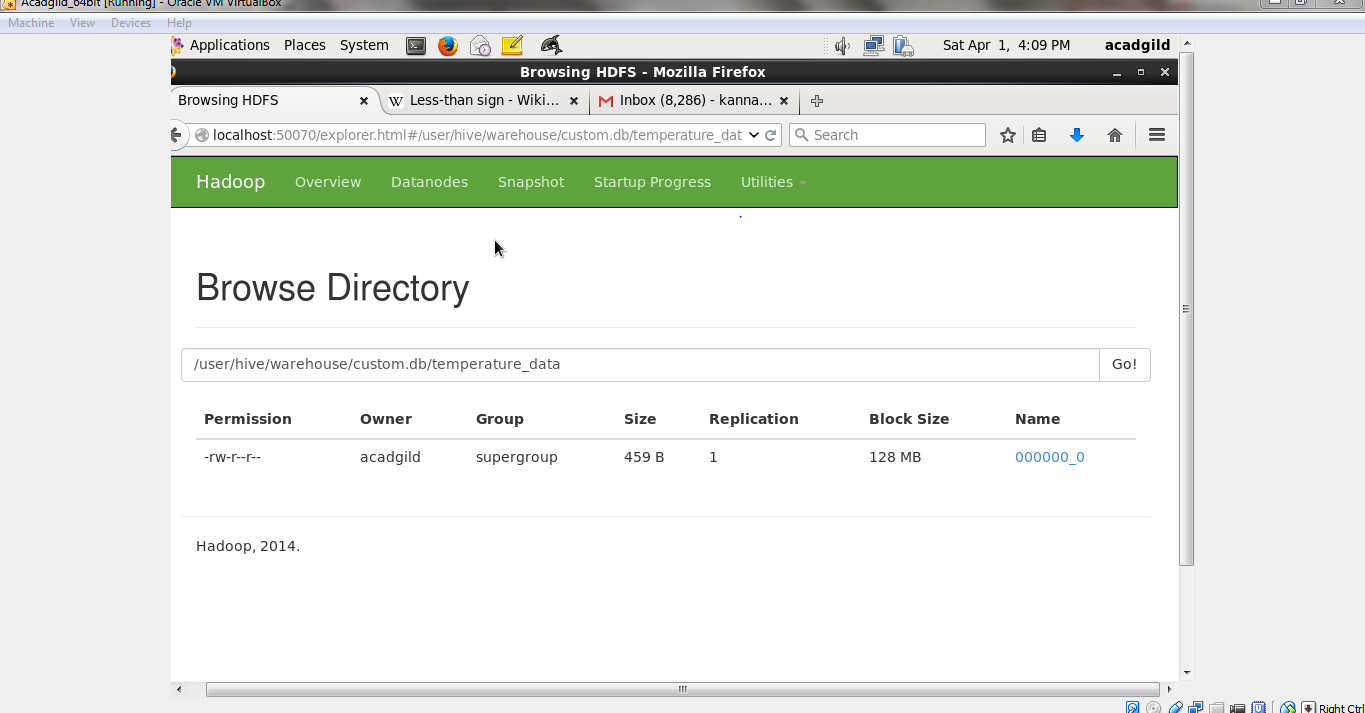
-- load data into temperature\_data

load data local inpath '/home/acadgild/Downloads/dataset\_Session\_14.txt' into table temp;

-- move data from temp to temperature\_data

insert into table temperature\_data select from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy'),'MM-dd-yyyy'),zipcode,temperature from temp;

Output



Log/Steps

[acadgild@localhost ~]$ hive

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

hive> create database custom;

OK

Time taken: 4.071 seconds

hive> use custom;

OK

Time taken: 0.165 seconds

hive> create table temperature\_data

> (date string,zipcode int,temperature float)

> row format delimited

> fields terminated by ',';

OK

Time taken: 2.081 seconds

hive> create table temp

> (date string,zipcode int,temperature float)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.494 seconds

hive> load data local inpath '/home/acadgild/Downloads/dataset\_Session\_14.txt' into table temp;

Loading data to table custom.temp

Table custom.temp stats: [numFiles=1, totalSize=437]

OK

Time taken: 7.121 seconds

hive> insert into table temperature\_data select from\_unixtime(unix\_timestamp(date,'dd-MM-yyyy'),'MM-dd-yyyy'),zipcode,temperature from temp;

Query ID = acadgild\_20170401160000\_f7a78087-5b23-44ca-9a03-64d26ec47989

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1491017400247\_0003, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0003/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2017-04-01 16:02:20,500 Stage-1 map = 0%, reduce = 0%

2017-04-01 16:03:21,919 Stage-1 map = 0%, reduce = 0%

2017-04-01 16:03:43,879 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 11.31 sec

MapReduce Total cumulative CPU time: 11 seconds 310 msec

Ended Job = job\_1491017400247\_0003

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://localhost:9000/tmp/hive/acadgild/6da24097-3f7f-4b8b-818b-4e5f50670de6/hive\_2017-04-01\_16-00-43\_107\_5846781257916991543-1/-ext-10000

Loading data to table custom.temperature\_data

Table custom.temperature\_data stats: [numFiles=1, numRows=20, totalSize=459, rawDataSize=439]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 12.64 sec HDFS Read: 670 HDFS Write: 539 SUCCESS

Total MapReduce CPU Time Spent: 12 seconds 640 msec

OK

Time taken: 192.217 seconds