Health Project

Use Case

edureka!

Version 2.0

Health Project

Table of Contents

Health Project

Problem Statements:

Below are few of the problem statement that we have chosen to work on this dataset.

- How many hospital centres got more than 60% patient satisfaction regarding cleanliness?
- Which hospital centre got maximum overall rating between 9-10?

Important Links:

- **Dataset:**

  https://data.cityofnewyork.us/Health/New-York-City-Health-and-Hospitals-Corporation-HHC/hi3x-y76v

- **Link for all the codes:**

  https://edureka.wistia.com/medias/9huydjttce/download?media_file_id=64181019

- **Hive Installation:**

  http://www.edureka.co/blog/apache-hive-installation-on-ubuntu

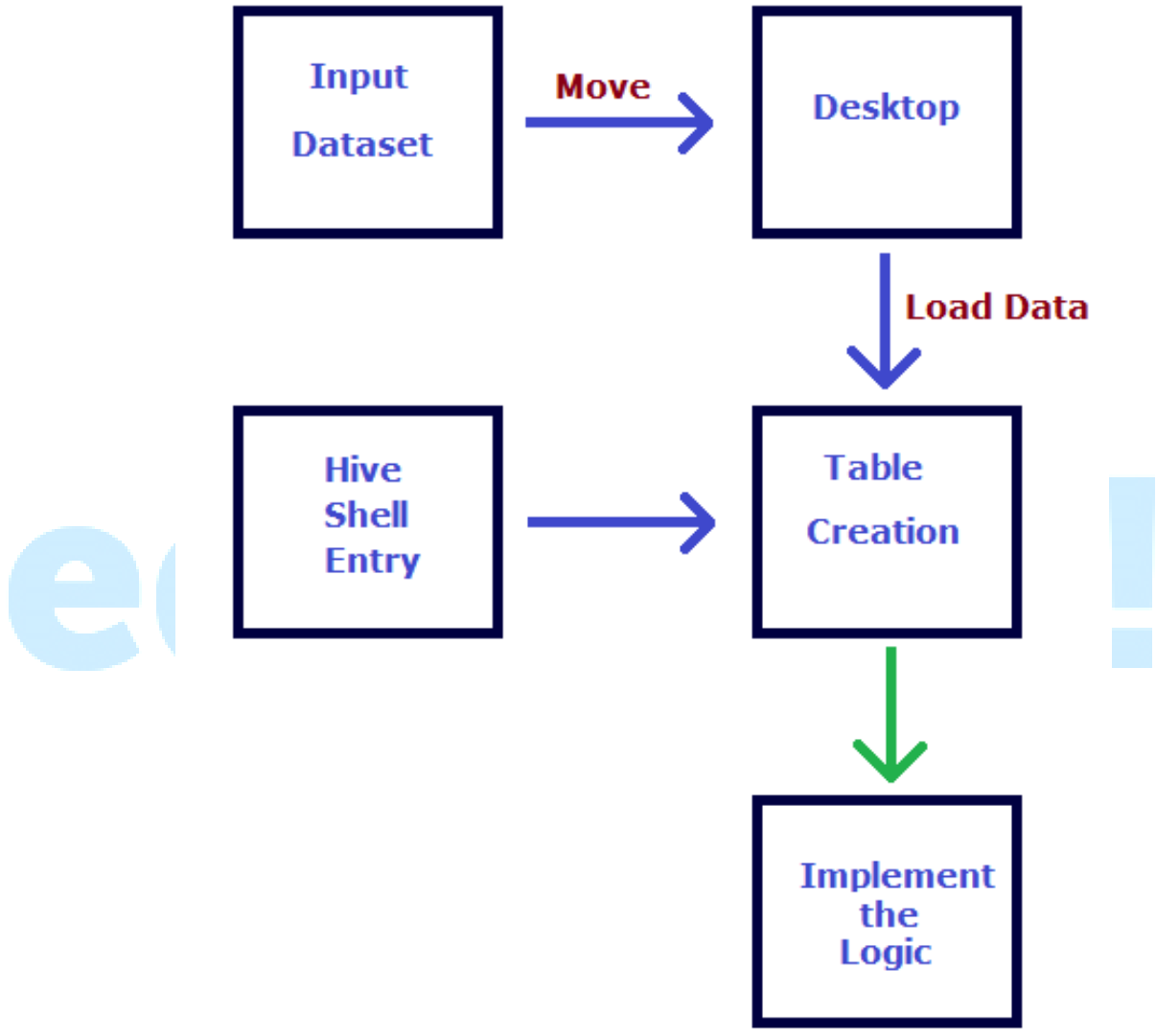
Technology/Software Used:

- Hadoop environment
- Apache Hive

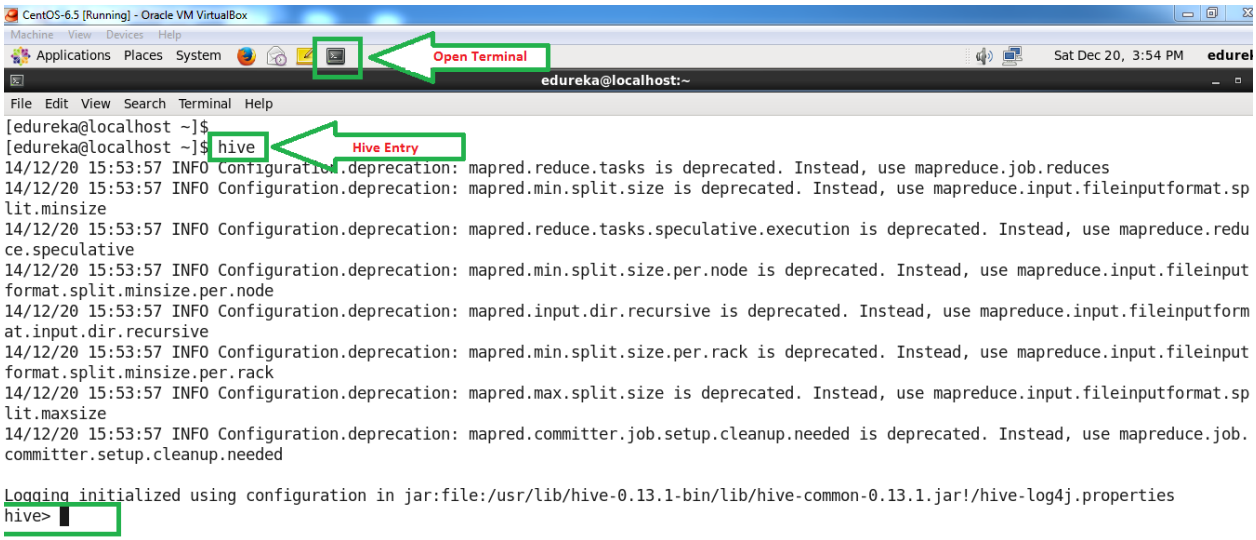
Project Workflow:

Environment Creation for Creating Solution:

To run the commands and to get the output we have done few required steps, which are depicted by the following images:

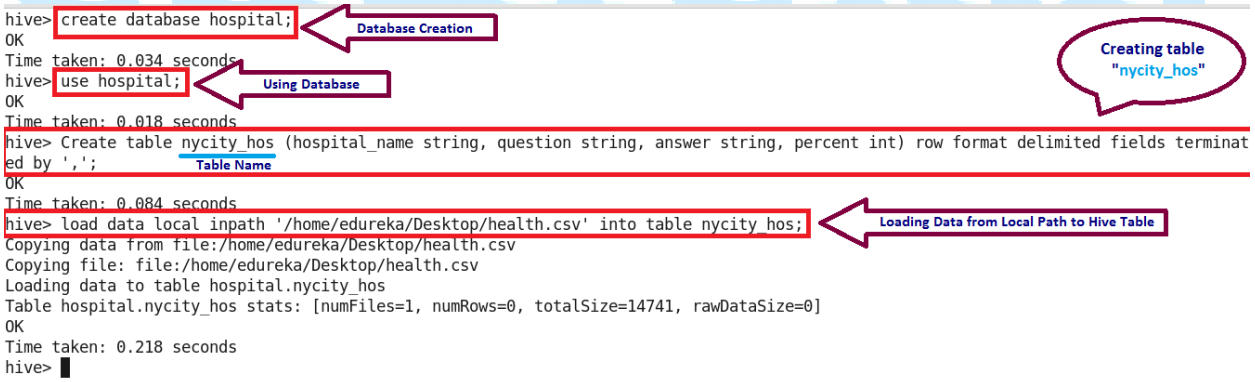


- **HIVE Entry**



- **Creating Database and Table & Loading Data**

We have downloaded the data in local system in .csv file. And loading the data in the created table. Refer below image-



Implementation:

We have chosen HIVE over here because of its built-in functions, it is easy for us to solve complex query without implementing complex codes of Map-Reduce and other environments of Hadoop.

Problem 1: How many hospital centers got more than 60% patient satisfaction regarding cleanliness?

Here, we have used LIKE statement to find the word CLEAN and having percentage more than 60%.

```
hive> SELECT hospital_name, percent FROM nycity_hos WHERE (question LIKE '%clean%') and (percent>60);
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1419051797093_0016, Tracking URL = http://localhost:8088/proxy/application_1419051797093_
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job  -kill job_1419051797093_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2014-12-22 12:52:08,925 Stage-1 map = 0%,  reduce = 0%
2014-12-22 12:52:14,238 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.17 sec
MapReduce Total cumulative CPU time: 1 seconds 170 msec
Ended Job = job_1419051797093_0016
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 1.17 sec   HDFS Read: 14970 HDFS Write: 285 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 170 msec
OK
CONEY ISLAND HOSPITAL   64
HARLEM HOSPITAL CENTER  61
JACOBI MEDICAL CENTER   63
KINGS COUNTY HOSPITAL CENTER    73
LINCOLN MEDICAL & MENTAL HEALTH CENTER  65
METROPOLITAN HOSPITAL CENTER    63
NORTH CENTRAL BRONX HOSPITAL    62
QUEENS HOSPITAL CENTER  68
WOODHULL MEDICAL AND MENTAL HEALTH CENTER       65
Time taken: 11.192 seconds, Fetched: 9 row(s)
```

Problem 2: What percent got maximum overall rating between 9-10?

Calculating the maximum percent number which is having ration between 9-10.

```
hive> SELECT MAX(percent) FROM nycity_hos where question LIKE "%overall%";
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1419051797093_0028, Tracking URL = http://localhost:8088/proxy/application_1419051797093_0028/
Kill Command = /usr/lib/hadoop-2.2.0/bin/hadoop job  -kill job_1419051797093_0028
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2014-12-22 13:28:20,747 Stage-1 map = 0%,  reduce = 0%
2014-12-22 13:28:26,064 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.15 sec
2014-12-22 13:28:31,373 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.06 sec
MapReduce Total cumulative CPU time: 2 seconds 60 msec
Ended Job = job_1419051797093_0028
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 2.06 sec   HDFS Read: 14970 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 60 msec
OK
59
Time taken: 17.498 seconds, Fetched: 1 row(s)
hive>
```

Further Implementation:

Below are few of the Problem Statements which are not solved. Based on the solution for above problem statements try to work on the below ones.

- How many hospital centers got more than 60% patient satisfaction regarding giving information about medicines they have been subscribed?

- How many hospital centers got more than 60% patient satisfaction regarding well communication by nurses?

Solution of above unsolved queries is given under **"Important Links Section"** on **Page 2**.