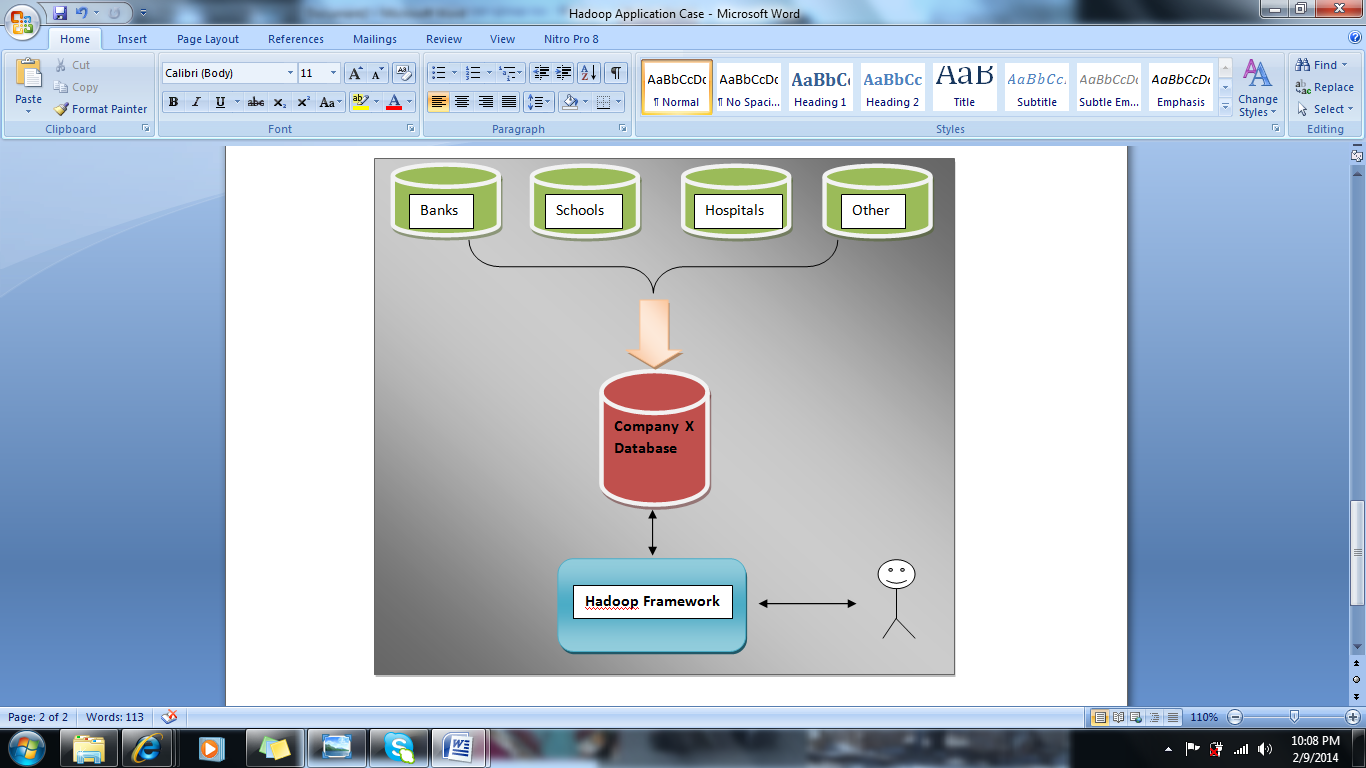
**Hadoop Application Case for “Business Services Search”**

Company X promotes business through the application of the search engine. It provides the details of services in the locality. The services include the usual necessity and utilities like Banks, Schools, Restaurants, Shopping Mall, and Hospitals etc.

The CTO of the company X realized that the data will grow over time as the company will cover more and more cities and as more local businesses will be added in the database.

Storing the data will not be an issue but a quick search on real time basis in a huge database will definitely call for the application of Hadoop framework.

**Business Framework:**



The company will receive data from various third parties in different formats like data feed, flat files, spreadsheets, access to internal/external database etc. This data will be cleaned up, formatted and submitted in the database of company X in a specified form. This may include concepts of Relational database to manage and store the data efficiently.

Further, the data can be maintained at different level which includes storing data in different tables at business level, for example, all the banks data in one table and hospital data in another table.

Data can be stored separately at region level, for example, Mumbai data stored in separate table and Bangalore data in different table.

In case the data size grows tremendously, the search will not be quick enough to meet the customer expectations. Hence, moving the data to Hadoop distributed framework will be useful.

The data from relational database can be moved to Hadoop Framework using SQOOP features available for the Hadoop.

**Steps for DB preparation phase**

**Create database <db\_name>;** // To create a DB in mysql server

**Use <db\_name>;** // use the same DB

**# To create a table**

create table AppData(ID int NOT NULL AUTO\_INCREMENT, State varchar(100), City varchar(100), Location varchar(200), PinCode varchar(8), Bank varchar(40), Restaurant varchar(50), Hospital varchar(60), ShoppingMall varchar(60), School varchar(30), PRIMARY KEY(ID));

**$ sudo chmod 777 -R /var/lib/mysql**

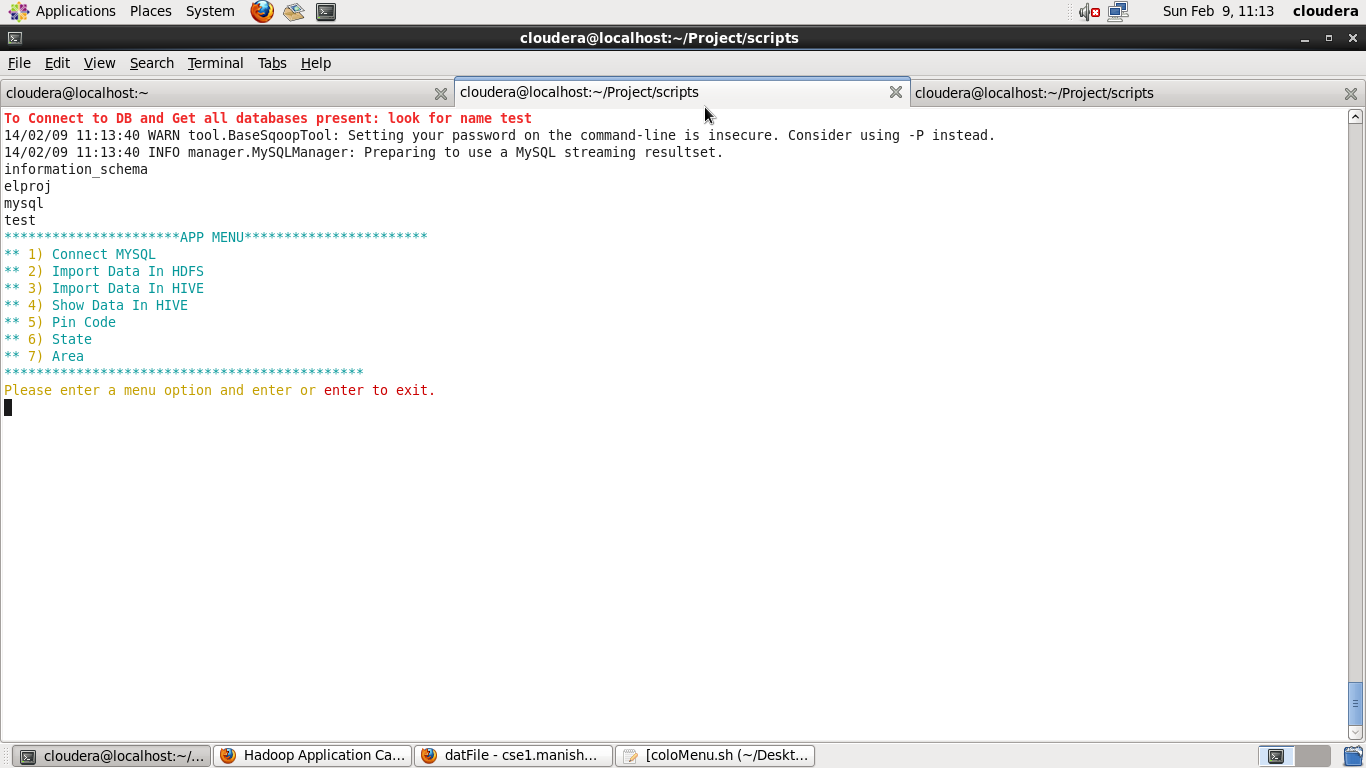
**$ sudo cp /home/hduser/HadoopAppdata.csv /var/lib/mysql/<db\_name>**

**#To load data in table from .csv file from local path -**

Load data local infile 'HadoopAppdata.csv' into table AppData fields terminated by ',' enclosed by '"' lines terminated by '\n' ignore 1 lines (State, City, Location, PinCode, Bank, Restaurant, Hospital, ShoppingMall, School);

After importing data, verify that the data is available in a relational DB

select \* from appdata;

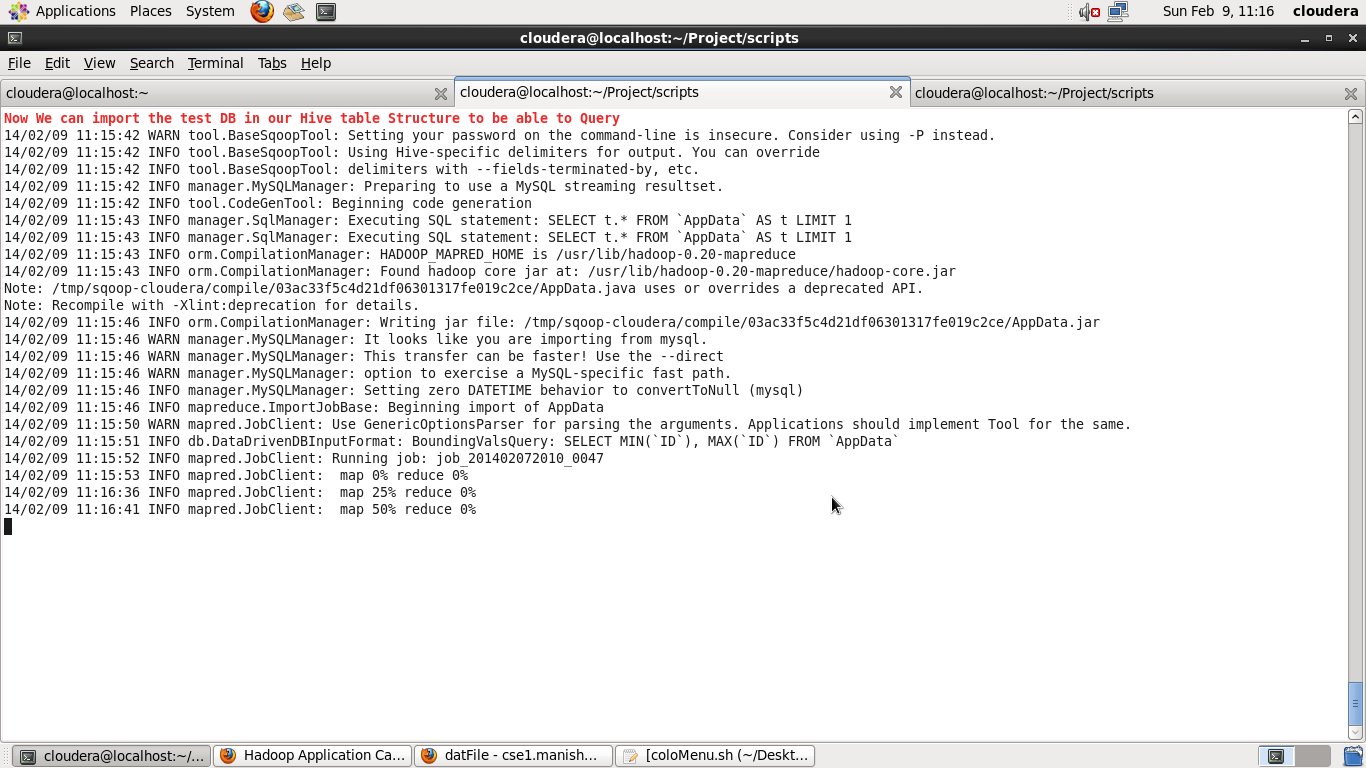


Once the data in relational DB is verified, we can export the data in the Hadoop framework.

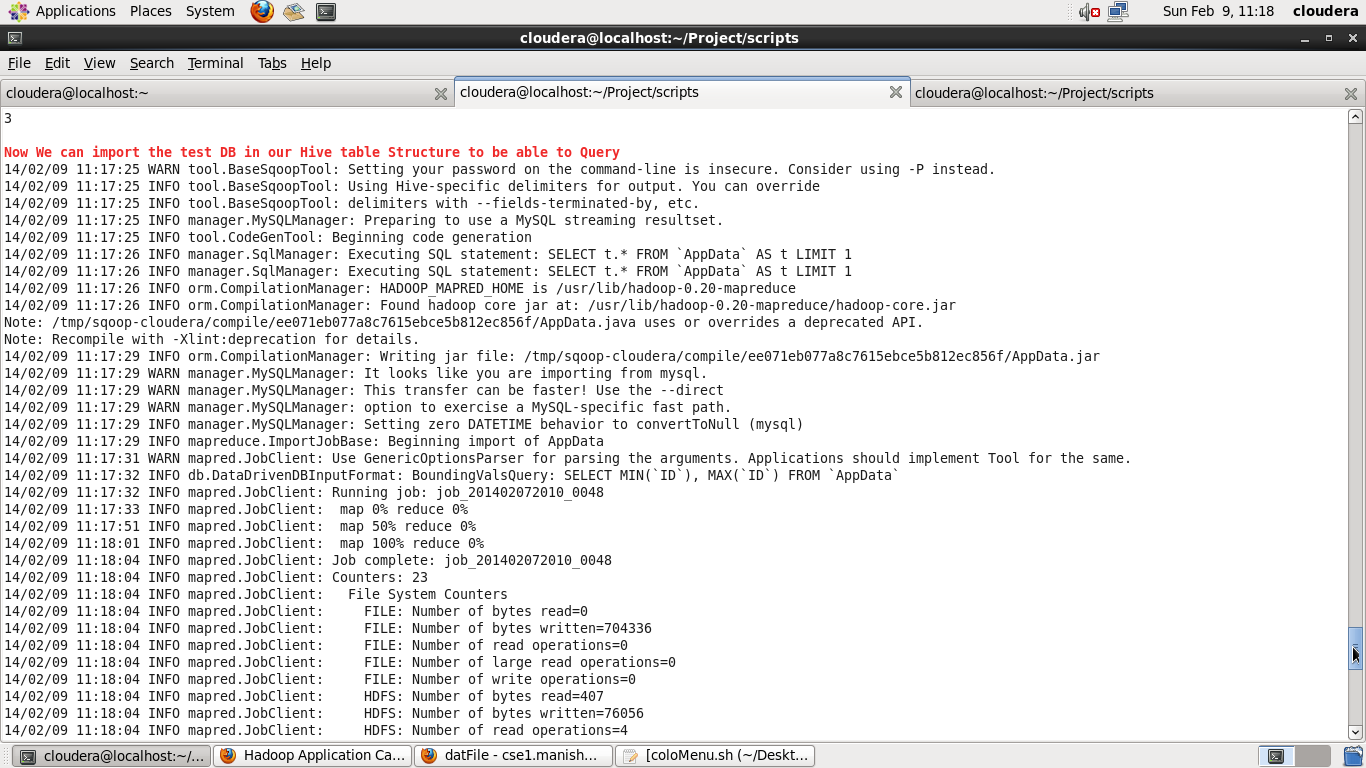
Data can be imported to HDFS

* For HBASE queries
* For Hive (direct queries)

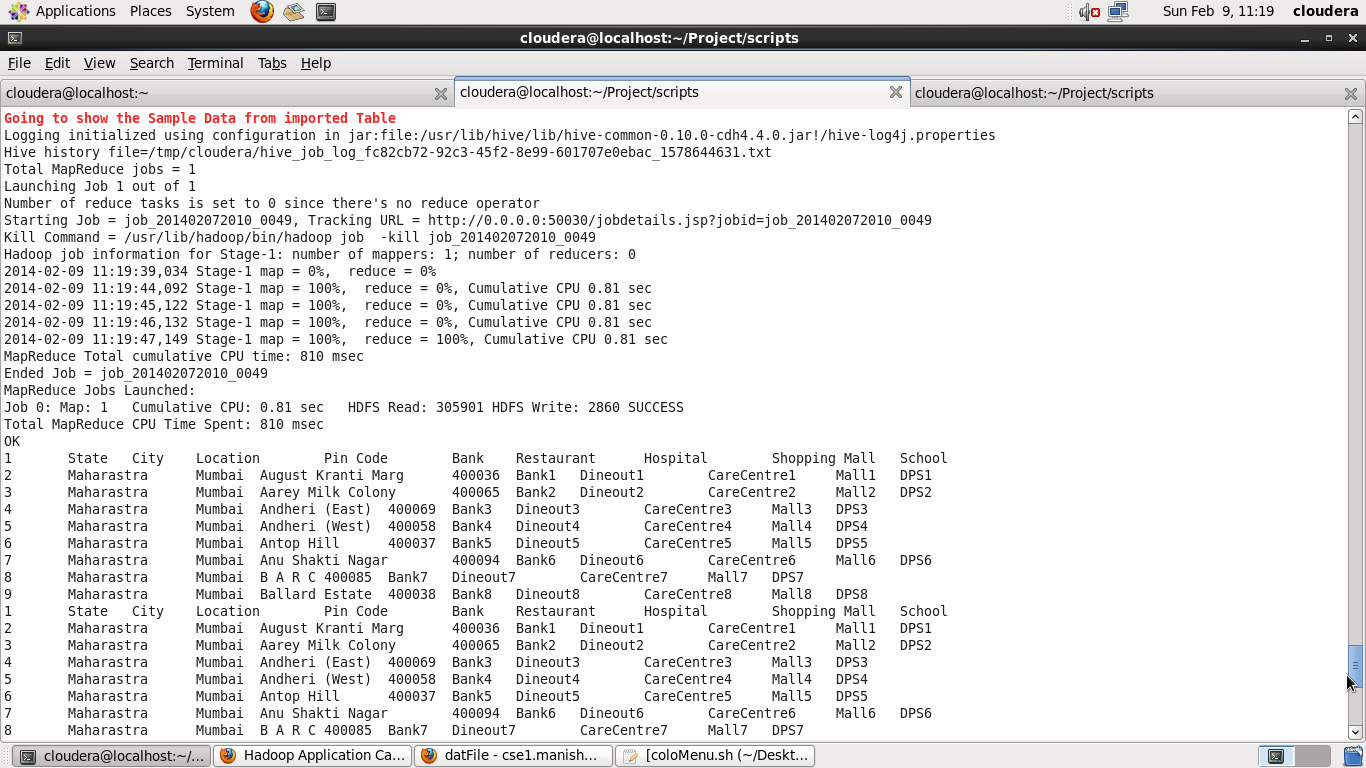
**Export To HDFS:**



**Export To Hive:**



**Verify that the data has moved to Hive**



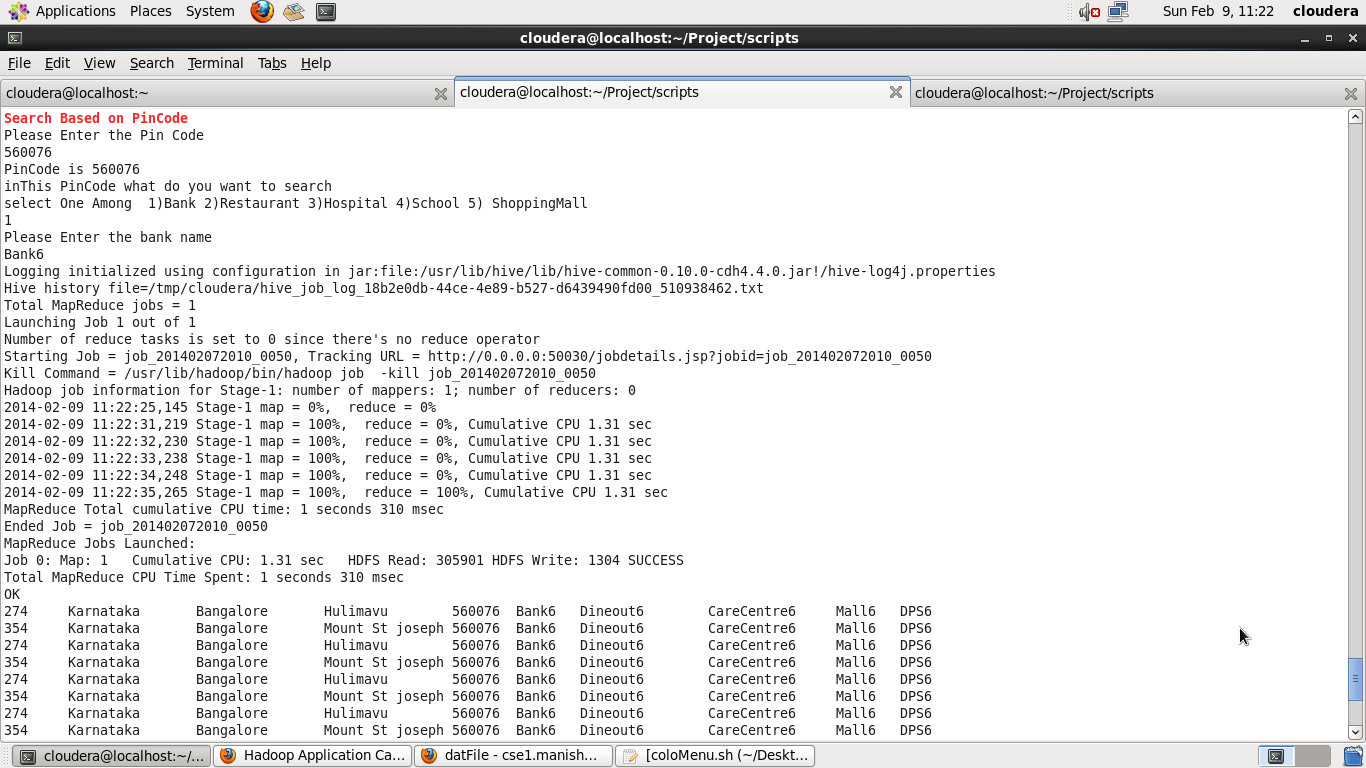
With data available with different components of the Hadoop, we can handle the users request through multiple approaches.

**Using PIG/HIVE and Map-Reduce to fetch responses for user’s request.**

Pig internally converts all the queries into Map-Reduce jobs. The application will fetch results based on user’s selection on the menu.

The user will access the system using command window.

Application provides option to enter the pin code which is followed by Multiple Criterion.



Many Such combinations for Search are possible. The exact search is implemented due to technical issues in implementing **“LIKE”** operator.

**PS:** There is an existing bug with the **“LIKE”** operator implementation for Hive.

Like operator assists in search where input isn’t the exact match of the data stored in the database.

Details can be found at the following link:

social.msdn.microsoft.com/Forums/windowsazure/en-US/dbb54b05-157c-44bb-ac2e-160bbc0b7e2f/hive-console-using-operator-like-in-hive-query?forum=hdinsight

**Summary of work flow for application:**

* Data is imported into MYSQL from flat file/csv for further cleaning and manipulation
* SQOOP interface is used to export data from MYSQL to HIVE/HDFS file system
* Date is queried over for multiple user selection using hive
* Linux shell scripting is used for console based UI

Code for shell scripting is available in the **MyMenu.sh** file.

**Steps to run the shell script:**

Paste the shell file in the working directory of your system.

Give permission like **chmod u+x MyMenu.sh**

Run the script using **./MyMenu.sh**

**Snapshot of the Console based UI:**

