

Analyse Electoral Performance of Parliament Using Hadoop

K. Sai Prasad,
Computer Science Engineering,
MLR Institute of Technology,
Hyderabad, India.

K Shekar,
Computer Science Engineering,
MLR Institute of Technology,
Hyderabad, India.

G. Prabhakar Reddy, K Sai Teja
Computer Science Engineering
MLR Institute of Technology,
Hyderabad, India.

Abstract - Bigdata is a buzz word in market and industry for its wide applications. Hadoop a framework which is considered as the solution for many problems of Bigdata. This paper is to analyse the data on the election results for and criminal charges against all candidates contesting elections to the lower house of the Parliament of India, the Lok Sabha, in the 2004 and 2009 elections using hadoop. These were the first national elections conducted after the 2002 Supreme Court ruling mandating that all candidates running for public office file affidavits with the Election Commission of India prior to the election. In these affidavits, candidates report their criminal histories or pending criminal charges. We present graphical results of the election analysed with the surveyed data and predict the winner of the elections i.e; The Member of Parliament. This is done by collecting data of the survey of elections.

Keywords — Elections, Bigdata, Hadoop. Parliament, Loksabha.

I. INTRODUCTION

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. Big Data [1] is a phrase used to mean a massive volume of both structure and unstructured data. But it's not the amount of data that's important. It's what organizations do with the data that matters. Electoral Performance and Criminal Status of Candidates Contesting the 2004 and 2009 Parliamentary Elections to the Lok Sabha (India) (ICPSR 35512). The understand the effectiveness of visualization of historical data we conducted a case study to determine which candidate has got how many votes. So, this can be used in any elections across the globe, small or big. Candidates and voters will determine to what extent analysis plays a major role in the visualization of elections. Visualization functions as a convenient way to display the results. Visualizing these patterns and attaching them to our memories could make the difference between winning and losing.

II. PROPOSED WORK

The traditional process used to solve structured data is using "Relational Database Management System" (RDBMS). In RDBMS the data is stored in the form of relational structural format. Relational databases are powerful because they require few assumptions about how data is related or how it will be extracted from the

database. As a result, the same database can be viewed in many different ways. The proposed work is by using "Hadoop" [2]. Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate.

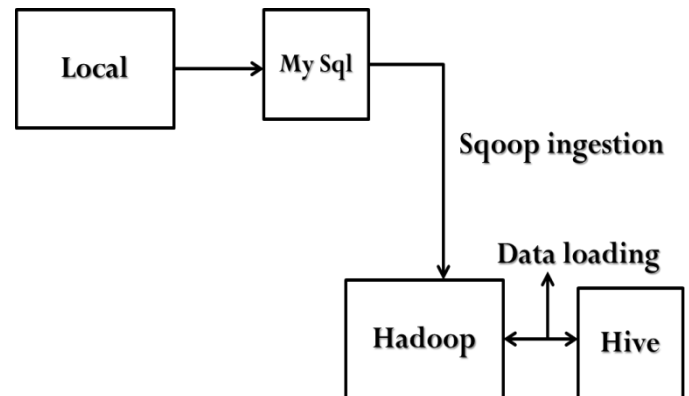


Fig:1 Functional flow of processing in Hadoop.

Fig.1. describes the process data loading and processing in Hadoop.

- The Required datasets are collected and loaded into MySQL server.
- Using Sqoop data from MySQL server is been loaded into Hadoop supported data query language i.e Hive.
- Using Hive query language and hue which is an open source Web interface for analysing data, we analyse the data generate required results.

III. EXPERIENTIAL RESULTS

Collecting Data Sets

- The below fig:2 shows the samples of the data sets collected and stored with pipe separation in HDFS.
- The above figure depicts the datasets collected for analysing electoral performance.
- The step by step procedure involved in our process is described as below:
- We have saved the data sets in csv format
- After arranging the data sets, we uploaded these data sets into the MySQL data base in cloudera.

- Later copied the csv files from local system to cloudera.

| | | | | |
|------|---------------------------|------|------------------------|-----------|
| 5 | Shri Ziaur Rahman Ansari | Male | Uttar Pradesh | Congress |
| 6204 | 1 | 64 | 40 | |
| 5 | Shri Frank Anthony | Male | Nominated Anglo-Indian | Nominated |
| 6206 | 4 | 62 | 30 | |
| 5 | Shri S.R.A.S. Appalanaidu | Male | Andhra Pradesh | Congress |
| 7540 | 1 | 13 | 41 | |
| 5 | Shri K.M. Asghar Husain | Male | Maharashtra | Congress |
| 7563 | 0 | 54 | 56 | |
| 5 | Dr. Henry Austin | Male | Kerala | Congress |
| 2548 | 3 | 945 | 75 | |
| 5 | Shri Bhaqet Jha Azad | Male | Bihar | Congress |
| 6950 | 0 | 23 | 62 | |
| 5 | Shri Aziz Imam | Male | Uttar Pradesh | Congress |
| 3256 | 6 | 52 | 54 | |
| 5 | Shri Gurbas Singh Badal | Male | Punjab | Akali Dal |
| 1296 | 5 | 36 | 43 | |
| 5 | Shri Ran Chandra Badi | Male | Madhya Pradesh | Jai Sangh |
| 9774 | 0 | 466 | 56 | |

Fig:2 Sample data format for processing

Login into MySQL

Now create databases and tables for the corresponding csv files in MySQL.

```
cloudera@quickstart ~]$ mysql -u root -p;
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \.
Your MySQL connection id is 278
Server version: 5.1.73 Source distribution
```

Creating table in MySQL

- The command used for creating table is: > **create database elections;**
- Loading data from csv file into MySQL

Input:

```
mysql> load data local infile 'home/cloudera/elections/elections.csv' into table parliament fields terminated by ',' lines terminated by '\r\n';
```

When we run the above command we get the output as below.

result:

```
Query OK, 368 rows affected, 15 warnings (0.04 sec)
Records: 368 Deleted: 0 Skipped : 0 Warnings: 15
```

Now exit from MySQL using “EXIT” command.

- *Import data to Hadoop from MySQL using SQOOP*

> sqoop import --connect jdbc:mysql://localhost/elections --username xxxx -- password xxxx --table tournament -- fields-terminated-by ‘,’ --hive-import --m 1”.After executing the sqoop command the following results are generated as shown in fig:3.

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/elections --username root --password cloudera --table elections --fields-terminated-by ',' --hive-import -m 1;
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
16/12/15 22:11:35 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.7.0
16/12/15 22:11:35 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
16/12/15 22:11:36 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
16/12/15 22:11:36 INFO tool.CodeGenTool: Beginning code generation
16/12/15 22:11:37 ERROR manager.SqlManager: Error executing statement: com.mysql.jdbc.exceptions.jdbc4.MySQLSyntaxErrorException: Unknown database 'elections'
com.mysql.jdbc.exceptions.jdbc4.MySQLSyntaxErrorException: Unknown database 'elections'
at sun.reflect.NativeConstructorAccessorImpl.newInstance(Native Method)
at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:57)
at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
at java.lang.reflect.Constructor.newInstance(Constructor.java:526)
at com.mysql.jdbc.Util.handleNewInstance(Util.java:377)
at com.mysql.jdbc.Util.getInstance(Util.java:360)
at com.mysql.jdbc.SQLError.createSQLException(SQLError.java:978)
at com.mysql.jdbc.MySQLIO.checkErrorPacket(MySQLIO.java:3887)
at com.mysql.jdbc.MySQLIO.checkErrorPacket(MySQLIO.java:3823)
at com.mysql.jdbc.MySQLIO.checkErrorPacket(MySQLIO.java:870)
at com.mysql.jdbc.MySQLIO.secureAuth411(MySQLIO.java:4332)
at com.mysql.jdbc.MySQLIO.doHandshake(MySQLIO.java:1258)
at com.mysql.jdbc.ConnectionImpl.coreConnect(ConnectionImpl.java:2234)
at com.mysql.jdbc.ConnectionImpl.connectOneTryOnly(ConnectionImpl.java:2265)
at com.mysql.jdbc.ConnectionImpl.createNewIO(ConnectionImpl.java:2664)
at com.mysql.jdbc.ConnectionImpl.<init>(ConnectionImpl.java:798)
```

Fig 3: Result of sqoop command

- After the successful completion of sqoop import open web browser and click on hue interface to interact with data which is loaded.
- Later go to query editor and choose hive editor in HUE browser as shown in fig:4.

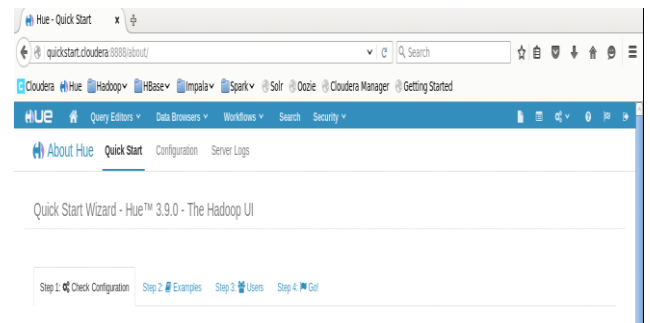


Fig: 4 HUE browser for generating results using hadoop.

- Now to do the analysis part select the chart option to see the relation between the different columns in terms of

1. Pie charts
2. Bars etc.

- To check the relationships between the different columns in a data set change the columns on X and Y axis respectively.
- Let us understand the analysed results as shown in fig : 5

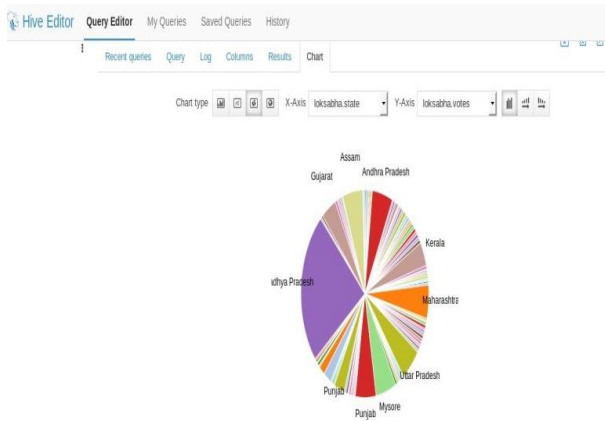


Fig: 5 Results obtained using hive query in HUE.

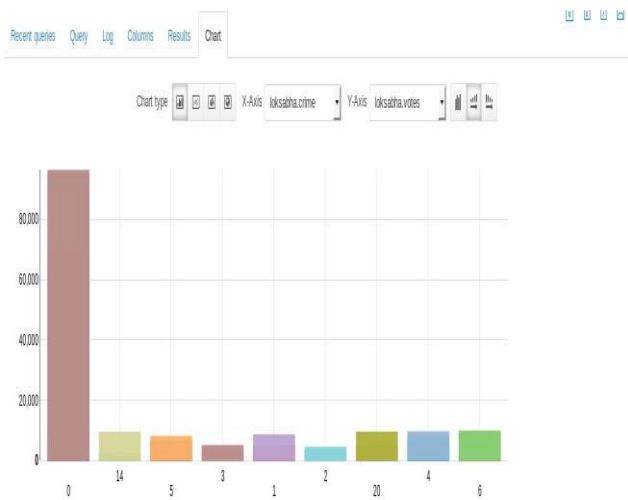


Fig: 6 Final result of analysis

4. CONCLUSION

By using the Hadoop framework tools we performed analysis on historical dataset of elections and obtained results on those data sets in the forms of graphs and pie charts. By using this process, we can work with HDFS, hive, sqoop, hue, impala, cloudera. A novice user of Hadoop can easily understand the process which we have followed to get the results and analyzing any types of data. Our results clearly shows the crime status of various places for parliament contested members in different locations. The architecture clearly shows how data collected can be processed and analyzed. For a very huge data of structured or unstructured data we can get results effectively using the hadoop.

Scope of Future Work:

Based on this framework this we can create a platform to predict election results or any analytical results for all the countries and even make up a data base for world population. This can be made into standard system for all the analysis procedures for heavy datasets for organizations to make the works easier.

REFERENCES

1. HAN HU1, YONGGANG WEN2, (Senior Member, IEEE), TAT-SENG CHUA1, AND XUELONG LI3, (Fellow, IEEE) Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, 2169-3536 2014 IEEE.
2. Jonathan R. Owens, Brian Femiano, Jon Lentz Hadoop Real World Solutions Cookbook ISBN:1849519129 9781849519120

BIBLIOGRAPHY:

- [1] <https://www.tutorialspoint.com/>
- [2] <http://Hadooptutorials.co.in/tutorials/Hadoop/internals-of-HDFS-file-read-operations.html>
- [3] <http://www.Hadooptpoint.com/Hadoop-hive-architecture/>
- [4] http://downloads.vmware.com/d/info/desktop_downloads/vmware_workstation/7_0
- [5] <http://www.cloudera.com/>