

# Scrutinizing Punjab Elections Scenario via Big Data and Map Reduce Technology

Kulwinder Singh

Yadavindra College of Engineering  
Talwandi Sabo Bathinda, Punjab (India)

Balkrishan Jindal

Assistant Professor, CE Department  
Yadavindra College of Engineering  
Talwandi Sabo Bathinda, Punjab (India)

**Abstract**—In this paper, structured database related to different political parties and political leaders of Punjab state elections in Comma separated value (CSV) or Tap separated value (TSV) is created. This database is undergone a mining process using map-reduce algorithm using Apache Hadoop framework. No one can get desired result by writing different scripts and passing numerous queries on the database and can get final results in graphical form or any different visualization. It will help the voters to select the right party and candidate in their assembly in lok Sabha elections.

**Keywords**—Big Data, Hadoop, MapReduce, structured data, unstructured data.

## I. INTRODUCTION

Big data refers to the data sets that are too big to handle using the offered database management tools in many significant applications, such as Internet search, industry informatics, social networks and social medium and genomics and meteorology. In simple words it can be thought that some data which challenges the currently existing techniques for handling data is referred as big data. Big data present a grand challenge for folder and data analytics research. Gone are the days when remembrance was used to be considered in conditions of GigaBytes, TeraBytes or PetaBytes. Today even larger units are been used to measure memory like ExaBytes, ZettaBytes and Yotta Bytes. Big Data is not a single technology, technique or initiative. Rather, it is a trend across many areas of business and technology [2, 3]. Talking about technologies enabling the use of Big Data, there are three fundamental technological strategies for storing and providing fast access to large data sets [1, 7, 13, 14].

- Superior hardware presentation and capacity: make use of faster CPUs, make use of more CPU cores (require parallel/threaded operations to take advantage of multi-core CPUs), increase disk capacity and data transfer throughput, increased network throughput Massively Parallel Processing (MPP) [14].
- Reducing the size of data accessed: Data compression and data structures that, by design, boundary the amount of data required for queries. E.g. bitmaps and column-oriented databases (NoSQL) Not Only SQL [15].
- Distributing data and parallel processing: putting data on more disks on the way to parallelize disk I/O, set slices of data on separate work out nodes that can work on these less important slices in equivalent, use extremely distributed architectures with importance on fault lenience and presentation monitoring with higher-throughput network to improve data transfer among nodes Hadoop and Map Reduce.

### A. Challenges of Big Data

Big data also has its own unique set of obstacles such as: [2, 4, 6]

- Information Growth-Over 80 percent of the data in the activity consists of shapeless data, which tends to be growing at a much faster pace than traditional relational information. This massive information threatens to swamp all but the well prepared IT organizations.
- Dispensation power- The expected approach of using a single, expensive and powerful computer to moment of truth information just does not scale for Big Data. Because we soon see the method to go is dividing and conquers using commoditized hardware and software via scale out.
- Physical storage- Capturing and managing all this information can chomp through enormous resources, outstripping all budgetary prospects. Data issues: Lack of data mobility, proprietary formats, and interoperability obstacles can all make working with Big Data complicated.
- Cost-Extract, change, and load (ETL) processes for Big Data can be expensive and time consuming, particularly in the absence of specialized well-designed software.

### B. Big Data Use in Politics

The President Barack Obama is the first man on the planet earth to use the Big Data in elections [2]. In U.S.A., right here 2008 the elected birthday party used huge statistics just earlier than analyze the general public feeling which helped it with appropriate outcomes inside the election. It analyzed big public records and engineered social television and different media retailers to create a focused operation to win over younger voters for the elections. The movement proved beneficial in grasp states anywhere Democrats won a booming success. Within the big information evaluation additionally certified Democrats to fix to marketing campaign electorate which enabled them to generate over \$1 billion in profits. Data was not effectively shared to be truly effective in analyzing potential voters. The fund raising lists differed from the get out the vote lists causing problems for the movement office as well as voters During in the establishing stages of the operation the records analytics team understand that the diverse departments together with the manner workplace website department and location departments have been running considering the fact that exclusive units of facts. The analytics group helped to create a great single gadget that can act as a central shop for information.

This records shop enabled the Democrats to gather facts from fieldworker's or fundraisers and public consumer databases for evaluation. This centralized save up helped the marketing campaign office to locate citizens and to create centered campaigns to get their recognition. Analytics on the large datasets certified the fight workplace to find out what exactly appealed to the voter in a particular section. It allowed campaigner to expect which citizens had been likely to provide online. It enabled them to peer who had cancelled their subscription from their campaign lists. This indicated voters which may also have switched to their political rival. It also allowed them to evaluate a thing which includes how the human beings could react to a nearby unpaid helper making a call as distinct to a person from a non-swing nation. The facts indicated that the individuals who had signed up for the short bestow application were 4 instances much more likely to offer than others. This record enabled them to create a stepped forward gift machine where people could make a contribution with lots much less annoy main to higher finances.

Barack Obama campaign had a devoted team of workers of over hundred people with over 50% of them in a special analytics department to research the information and 30% in the area to take the results. With reference to era they rate Hadoop for using the huge records analytics engine but they have been not able to do so because it required notably specialized skills to increase packages to apprehend the big records. Some other trouble that they facade turned into that Hadoop in its first variations changed into now not designed to deal with actual-time query. The crew sooner or later used straight up a huge records appliance that turned into scalable and easy to put into effect. Vertical is a column orientated database that gives a standardized interface and sq. equipment to get right of entry to the facts, as a result existing gear and users can effortlessly work with it without specialized skill units. The crucial statistics storehouse for the campaign was created on directly up, which enabled the analysts to get a 360-diploma view.

The problems that the Democratic motion faced with era are very commonplace with different following campaigns around the world. Humans may not have the assets to assemble an analytical engine much like that used by the Democratic campaign. Everywhere in the global humans may have records that they hope to apply now not simply to influence electorate but also to perceive problem regions for their own constituencies. Lease a large analytics group and developing a records computation facility isn't feasible in maximum instances. Elections in India until currently comprised heat, dirt, theatre, fixed expertise, opinion polls, speech, procession, door-to-door visit or sweat and toil. Two irreparable tendencies that had been sign in 2014 parliamentary elections were dreadfully huge younger voter base and use of era to its excellent. In 2014 Lok Sabha elections in India are well lead via virtual social media technologies. In 2008 USA presidential drive for Barack Obama is started with the use of Social Media and 2012 bring huge data Analytics to front role. One method for predicting the results of upcoming elections is via exit poll. The most valuable information regarding campaigns and their affect on general public is provided by citizens themselves.

Data analysts develop models based on this information and perform predictions regarding winning and losing chances of any political party and any political leader. If such results are properly harnessed, they could gain sizeable gains. Elections in India have always comprised issues based on caste, religion, sentiments, traditional wisdom, opinion polls and rallies. But 2014 Lok Sabha elections witnessed the use of technology to its very best by political parties. All this idea was actually borrowed by the way Barack Obama contested his elections in America and raise to power in 2008 and 2012. In an extraordinary attempt to engage digitally literate electorates of India Google and some other social platforms started a forceful digital information campaign. Google India launched one such hub related to elections where electorates can search for political candidates or political parties and election platforms and voting related information in their regions. They even launched one site on the counting date which updated about live status of results on the day of counting. It changed into discovered that Narendra Modi constantly topped the quest traits while as compared to different applicants. For conduct 2014 Lok Sabha elections 543 Parliamentary constituency and 4120 assembly constituency were set up. All over India total of 9 lakh 30 thousand polling booths be set up for conducting just elections. Voter rolls were ready in 12 different languages and total of 9 lakh pdf files which amounted to 2.5 crore pages were translate. The genuine challenge was removal of voter info from these 2.5 crore PDF pages and transliteration of the similar into English to merge with other source.

## II. HADOOP AND MAP-REDUCE TECHNIQUE

Hadoop is a java based framework that is well-organized for processing large data sets in a distributed computing environment [11, 12]. Hadoop is sponsored by Apache Software Foundation. The maker of Hadoop was Doug Wounding and he named the framework after his child's swollen toy elephant. Applications be made run on systems with thousands of nodes making employ of thousands of terabytes via Hadoop. Dispersed file system in Hadoop facilitate fast data transfer among nodes and allows continuous operations of the system even if node failure occurs. This concept lowers the risk of disastrous system breakdown even if multiple nodes become out of action. The inspiration behind working of Hadoop is Google's Map reduce which is a software framework in which application under consideration is busted behind into number of small parts [5, 6]. Hadoop is a framework which comprised of six components [4]. Every component is assigned a particular job to be performed.

- HDFS – Hadoop distributed file system are distributed cages where all animals live i.e. where data resides in a distributed format.
- Apache HBase – It is a well-groomed and large database.
- Zookeeper- Zookeeper is the person responsible for managing animals play.

Pig – Pig allows playing with data from HDFS cages.

Hive- Hive allows data analysts play with HDFS and makes use of SQL.

- HCatalog helps to upload the database file and automatically create table for the user.

Map Reduce is a framework originally developed at Google that allows for easy large scale disseminated computing from corner to corner a number of domains [8]. The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets from corner to turn clusters of computers with easy programming models. It is planned to range up from only servers to thousands of machines, every offering local computation and storage space. Hadoop MapReduce includes several stages, all with a key put of operation selection to get to your purpose of getting the answer you need from big data. The method starts with a user request to run a MapReduce program and continues until the results are written back to the HDFS.

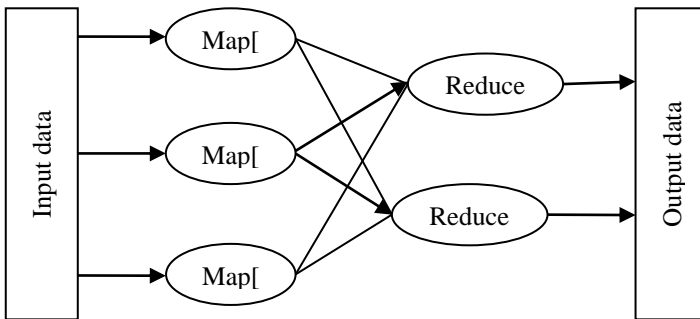


Fig. 1 Working of Map Reduce Technology

MapReduce is an architectural model for parallel processing of tasks on a distributed computing system. This algorithm is first described inside a paper "MapReduce easy Data Processing going on Large Clusters," by Jeffery Dean and Sanjay Ghemwat from Google. This algorithm allows split of a single computation task to various nodes or computers for distributed processing.

As an only task can be broken down into many subparts, each handled by a separate node the number of nodes determines the processing rule of the system. There are a choice of commercial and open-source technologies that implement the MapReduce algorithm as a part of their internal architecture. A popular implementation of MapReduce is the Apache Hadoop, which is used for data processing in a distributed computing environment. As MapReduce is an algorithm, it can be written in any programming language [17,18].

The initial part of the algorithm is used to split and 'map' the sub tasks to computing nodes. The 'reduce' part takes the results of individual computations and combines them to get the final result.

In the MapReduce algorithm, the mapping function reads the input data and generates a set of intermediate records for the computation. These intermediate records generated by the map function take the form of a key, data pair. As a part of mapping function, these records are distributed to different computing nodes using a hashing function. Individual nodes then perform the computing operation and return the results to the reduce function. The reduce function collects the individual results of the computation to generate a final output.

### III. PROPOSED WORK

Snap shot of database created in the proposed method is shown in Figure 2. It involves 15 different attributes which are related to elections conducted in Punjab 13 Lok Sabah sheets. It involves both string values and integers.

|    | A                         | B   | C           | D      | E      | F           | G                  | H                  | I             | J           | K           | L         | M               | N                | O    |
|----|---------------------------|-----|-------------|--------|--------|-------------|--------------------|--------------------|---------------|-------------|-------------|-----------|-----------------|------------------|------|
| 1  | NAME                      | AGE | EDUCATION   | SEX    | PARTY  | PARTY TYPE  | VOTES IN FAVOUR OF | VOTES IN FAVOUR OF | CRIMINAL CASE | ASSETS      | LIABILITIES | STATUS    | WINNING CHANCES | constituency Nam | Year |
| 2  | Harsimrat Kaur Badal      | 47  | 10th        | Female | SAD    | State       | 534777             | 46.09              | 0             | 100 Crore   | 41 crore    | Crorepati | Bright          | BATHINDA         | 2014 |
| 3  | Ipreet Singh S/O Gurdas S | 52  | Graduate    | Male   | INC    | National    | 495332             | 44.35              | 0             | 42 Crore    | 4 Crore     | Crorepati | Bright          | BATHINDA         | 2014 |
| 4  | Jasraj Singh Longia       | 40  | Graduate    | Male   | AAP    | National    | 87901              | 7.87               | 0             | 31 Lac      | Nil         | Lakhpati  | Average         | BATHINDA         | 2014 |
| 5  | Kuldeep Singh             | 32  | 10th        | Male   | BSP    | National    | 13732              | 1.22               | 0             | 80 Lac      | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 6  | Ashish                    | 25  | 10th        | Male   | IND    | Independent | 6626               | 0.59               | 3             | 10 Thousand | Nil         |           | Poor            | BATHINDA         | 2014 |
| 7  | Bhagwant Singh Samran     | 36  | 8th         | Male   | CPI    | National    | 5984               | 0.53               | 0             | 2 Lac       | 15 Thousand | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 8  | Satish Anora              | 51  | 12th        | Male   | IND    | Independent | 5936               | 0.52               | 0             | 1 Crore     | Nil         | Crorepati | Poor            | BATHINDA         | 2014 |
| 9  | Ipreet Singh S/O Gurdas S | 33  | 10th        | Male   | IND    | Independent | 4618               | 0.413              | 0             | 7 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 10 | Shaminder Singh           | 34  | 12th        | Male   | JKKNPP | National    | 4610               | 0.412              | 0             | 33 Lac      | 2 Lac       | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 11 | Geeta Rani                | 50  | 5th         | Female | ABSR   | National    | 4300               | 0.39               | 0             | Nil         | Nil         |           | Poor            | BATHINDA         | 2014 |
| 12 | Sanjeev Kumar Thapar      | 34  | 10th        | Male   | IND    | Independent | 2833               | 0.253              | 0             | 26 Lac      | 2 Lac       | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 13 | Makhian Lal               | 43  | 8th         | Male   | BSP(A) | National    | 2587               | 0.231              | 0             | 11 Thousand | Nil         |           | Poor            | BATHINDA         | 2014 |
| 14 | Vijay Kumar               | 44  | 10th        | Male   | IND    | Independent | 2258               | 0.202              | 2             | 6 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 15 | Swaran Singh              | 45  | Graduate    | Male   | IND    | Independent | 2077               | 0.185              | 0             | 1 Crore     | Nil         | Crorepati | Poor            | BATHINDA         | 2014 |
| 16 | Rajinder Singh            | 34  | 10th        | Male   | SAD(A) | State       | 1960               | 0.175              | 0             | 26 Lac      | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 17 | Navmeet                   | 42  | 12th        | Male   | SHS    | National    | 1893               | 0.169              | 0             | 7 Thousand  | 5 Lac       |           | Poor            | BATHINDA         | 2014 |
| 18 | Jagdeep Singh Gehri       | 39  | 10th        | Male   | IND    | Independent | 1665               | 0.149              | 0             | 1 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 19 | Bhupesh Kumar             | 36  | Graduate    | Male   | IND    | Independent | 1407               | 0.133              | 0             | 6 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 20 | Gurdeep Singh             | 44  | Illiterate  | Male   | NL     | National    | 1404               | 0.125              | 0             | 5 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 21 | Gurmeet Singh Rangheta    | 44  | 12th        | Male   | PLP    | National    | 1386               | 0.124              | 1             | 35 Lac      | 1 Lac       | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 22 | Sushil Kumar Jindal       | 31  | 10th        | Male   | IND    | Independent | 1330               | 0.119              | 0             | 1 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 23 | Jagdish Rai Sharma        | 57  | 1st Graduat | Male   | NBDP   | National    | 1248               | 0.111              | 1             | 95 Lac      | 3 Lacs      | Lakhpati  | Poor            | BATHINDA         | 2014 |
| 24 | Sukhwinder Singh          | 33  | 12th        | Male   | BMP    | National    | 1086               | 0.097              | 0             | 2 Lac       | Nil         | Lakhpati  | Poor            | BATHINDA         | 2014 |

Fig. 2 This Snapshot show the structured database of Punjab

In this snapshot shows the all candidate details like 15 different attributes name, age, education, sex, party, party\_type, votes\_in\_favour, % votes\_in\_favour, criminal\_case, assets, liabilities, status, winning chance, constituency type, year.

#### IV. RESULTS AND DISCUSSION

In this section results of the proposed method are presented and discussed. The result obtained from the database after running appropriate queries are shown in the form of visualization as below using Apache Hadoop Framework. In this database first query Select \* from kulwinder65; is used to display all the attributes in the table.

Table 1. Query result generated in the proposed work on created table 'kulwinder65' using query 1.

| S. no | Name                            | Age | Education        | Sex | Party | Party Type  | Votes In Favor | % of votes | Criminal case | Assets In crore | Liabilities in crore's | status   | Winning chance | Constituency name | Year |
|-------|---------------------------------|-----|------------------|-----|-------|-------------|----------------|------------|---------------|-----------------|------------------------|----------|----------------|-------------------|------|
| 1     | Harsimrat Kaur Badal            | 47  | 10 <sup>th</sup> | F   | SAD   | State       | 514727         | 46.09      | 0             | 100             | 41                     | Crotpati | Bright         | Bathinda          | 2014 |
| 2     | Manpreet Singh S/o Gurdas Singh | 52  | Graduate         | M   | INC   | National    | 495332         | 44.35      | 0             | 42              | 4                      | Crotpati | Bright         | Bathinda          | 2014 |
| 3     | Kuldeep Singh                   | 40  | Graduate         | M   | AAP   | National    | 87901          | 7.87       | 0             | 0.31            | Nil                    | Lakhpati | Average        | Bathinda          | 2014 |
| 4     | Ashish                          | 32  | 10 <sup>th</sup> | M   | BS P  | National    | 13732          | 1.22       | 0             | 0.80            | Nil                    | Lakhpati | Poor           | Bathinda          | 2014 |
| 5     | Bhagwant Singh Samaon           | 25  | 10 <sup>th</sup> | M   | IND   | Independent | 6626           | 0.59       | 3             | 0.80            | Nil                    |          | Poor           | Bathinda          | 2014 |
| 6     | Satish Arora                    | 36  | 8 <sup>th</sup>  | M   | CPI   | National    | 5984           | 0.53       | 0             | 0.2             | 0.15                   | Lakhpati | Poor           | Bathinda          | 2014 |
| 7     | Manpreet Singh S/o Gurdev singh | 51  | 12 <sup>th</sup> | M   | IND   | Independent | 5936           | 0.52       | 0             | 1               | Nil                    | Crotpati | Poor           | Bathinda          | 2014 |

In this table different attributes like name, age, education, sex, party, party\_type, votes\_in\_favour, % of votes in favors, criminal\_case, assets, liabilities, winning chance, constituency\_name, year of all candidate consist in election of Punjab state in show table 1.

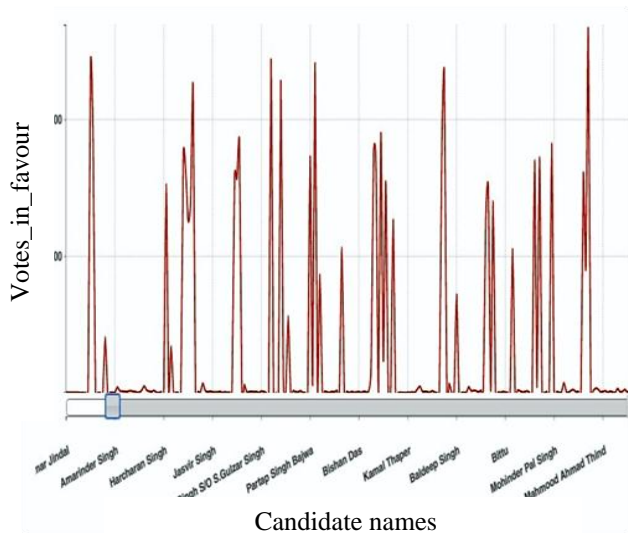


Fig. 3 visualization result of the proposed method

In Fig 3 is show the information about the 473 candidate's information for consists the election in Punjab from the fig. It will help the people to judge which candidate of most lead in past or wanted to elect their leader in future.

Use 2nd Query Select \* from kulwinder65 where criminal\_case>=2; to show the criminal case retails of all candidates.

Table 2 Result obtained using Query2.

| Name                   | Age | Education             | Party | Criminal Case |
|------------------------|-----|-----------------------|-------|---------------|
| Ashish                 | 25  | 10 <sup>th</sup>      | Ind   | 3             |
| Vijay Kumar            | 44  | 10 <sup>th</sup>      | Ind   | 2             |
| Captain Amrinder Singh | 72  | Graduate              | Inc   | 3             |
| Arun Kumar Joshi       | 40  | 12 <sup>th</sup>      | Ind   | 3             |
| Dr.inder Pal           | 52  | Graduate Professional | Ind   | 2             |
| Balwinder Singh        | 57  | Post Graduate         | Ind   | 2             |
| Gurdeep Singh Kahlon   | 42  | Graduate professional | Ind   | 2             |
| Simrjeet Singh bains   | 43  | Graduate              | Ind   | 2             |

Table 2 shows that number of candidates registered against then 2 or more criminal cases.

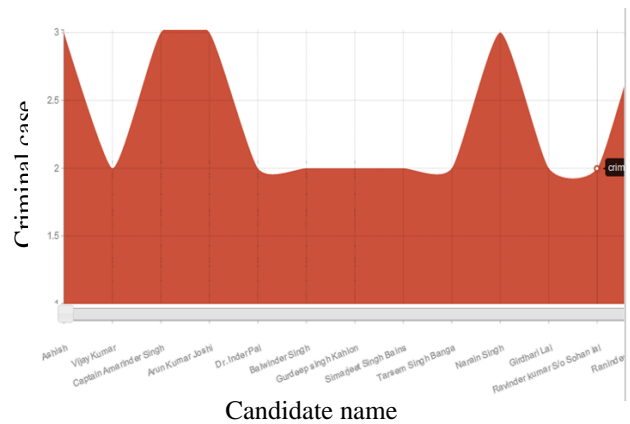


Fig.4 visualization result of the proposed method

The information about the candidate against under who have more than 2 criminal cases registered in past show in fig 4.

Use 3<sup>rd</sup> Query Select \* from Kulwinder65 constituency\_name="Fridkot"; to show all the candidates of district Fridkot.

Table 3 Query generated in this proposed work on created table kulwinder65using query 3.

| S.no | Name                   | Age | Education        | Sex | Party  | Party Type | Votes in Favors |
|------|------------------------|-----|------------------|-----|--------|------------|-----------------|
| 0    | Sukhwinder Singh Danny | 32  | Post Graduate    | M   | INC    | Natio nal  | 39569 2         |
| 1    | KauhalChamanbhaura     | 60  | Graduate         | M   | CPI    | Natio nal  | 19459           |
| 2    | Parmjit Kaur gulshan   | 60  | Post Graduate    | F   | SAD    | State      | 45773 4         |
| 3    | Resham Singh           | 56  | Doctorate        | M   | BSP    | Natio nal  | 34479           |
| 4    | Gurmeet Singh          | 39  | 12 <sup>th</sup> | M   | PLP    | Natio nal  | 1243            |
| 5    | Jasvir Singh           | 35  | 8 <sup>th</sup>  | M   | MB     | Natio nal  | 910             |
| 6    | Pritam Singh           | 70  | 5 <sup>th</sup>  | M   | RPI    | Natio nal  | 812             |
| 7    | Prem Singh             | 63  | 8 <sup>th</sup>  | M   | SP     | Natio nal  | 3133            |
| 8    | Raj kaur               | 60  | Literate         | F   | AID WC | Natio nal  | 1041            |

Table 3 is show of candidate who belongs to district Fridkot and their education, age, sex etc.

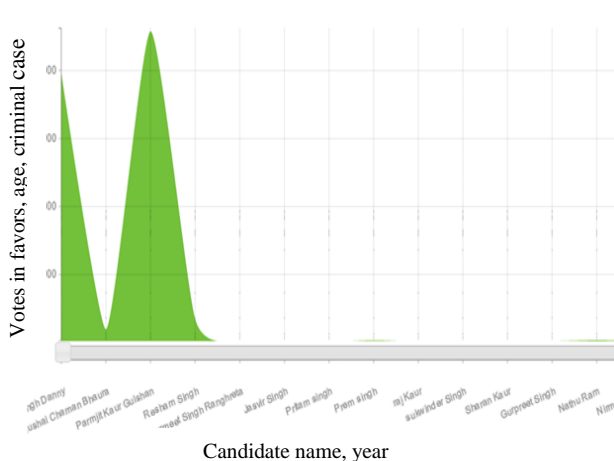


Fig.5 visualization results of the proposed method

In this Fig 5 show the information about the Fridkot district constituency in all candidates.

| State | Postal    | County Name | FIPS       | Obama vote | Obamaper    | Romney vote | Romneyper |
|-------|-----------|-------------|------------|------------|-------------|-------------|-----------|
| AK    | Alaska    | 0           | "91,696"   | 41.6       | "121,234"   | 55          |           |
| AK    | Alaska    | 2000        | "91,696"   | 41.6       | "121,234"   | 55          |           |
| AL    | Alabama   | 0           | "793,620"  | 38.4       | "1,252,453" | 60.7        |           |
| AL    | Autauga   | 1001        | "6,354"    | 26.6       | "17,366"    | 72.6        |           |
| AL    | Baldwin   | 1003        | "18,329"   | 21.6       | "65,772"    | 77.4        |           |
| AL    | Barbour   | 1005        | "5,873"    | 51.3       | "5,539"     | 48.3        |           |
| AL    | Bibb      | 1007        | "2,200"    | 26.2       | "6,131"     | 73.1        |           |
| AL    | Blount    | 1009        | "2,961"    | 12.3       | "20,741"    | 86.5        |           |
| AL    | Bullock   | 1011        | "4,058"    | 76.3       | "1,250"     | 23.5        |           |
| AL    | Butler    | 1013        | "4,367"    | 46.1       | "5,081"     | 53.6        |           |
| AL    | Calhoun   | 1015        | "15,500"   | 33.5       | "30,272"    | 65.5        |           |
| AL    | Chambers  | 1017        | "6,853"    | 47.1       | "7,596"     | 52.2        |           |
| AL    | Cherokee  | 1019        | "2,126"    | 21.8       | "7,494"     | 76.8        |           |
| AL    | Chilton   | 1021        | "3,391"    | 19.5       | "13,910"    | 79.8        |           |
| AL    | Choctaw   | 1023        | "3,785"    | 47.5       | "4,150"     | 52.1        |           |
| AL    | Clarke    | 1025        | "6,317"    | 45.7       | "7,463"     | 54          |           |
| AL    | Clay      | 1027        | "1,770"    | 26.7       | "4,802"     | 72.3        |           |
| AL    | Cleburne  | 1029        | "971,15.4" | "5,269"    | 83.6        |             |           |
| AL    | Coffee    | 1031        | "4,899"    | 24.8       | "14,638"    | 74.2        |           |
| AL    | Colbert   | 1033        | "9,160"    | 39.2       | "13,931"    | 59.6        |           |
| AL    | Conecuh   | 1035        | "3,551"    | 50.6       | "3,434"     | 49          |           |
| AL    | Coosa     | 1037        | "2,188"    | 41.7       | "3,028"     | 57.8        |           |
| AL    | Covington | 1039        | "3,155"    | 20.5       | "12,148"    | 78.8        |           |
| AL    | Crenshaw  | 1041        | "2,048"    | 31.9       | "4,326"     | 67.5        |           |
| AL    | Cullman   | 1043        | "4,855"    | 14.6       | "27,930"    | 84.2        |           |
| AL    | Dale      | 1045        | "5,283"    | 28.5       | "13,105"    | 70.6        |           |
| AL    | Dallas    | 1047        | "14,599"   | 69.7       | "6,284"     | 30          |           |
| AL    | DeKalb    | 1049        | "5,235"    | 22         | "18,316"    | 76.8        |           |
| AL    | Elmore    | 1051        | "8,940"    | 25.2       | "26,223"    | 74          |           |
| AL    | Escambia  | 1053        | "5,476"    | 36.9       | "9,263"     | 62.5        |           |
| AL    | Etowah    | 1055        | "12,792"   | 30.1       | "29,102"    | 68.5        |           |
| AL    | Fayette   | 1057        | "1,803"    | 22.8       | "6,034"     | 76.3        |           |
| AL    | Franklin  | 1059        | "3,171"    | 29.2       | "7,567"     | 69.7        |           |
| AL    | Geneva    | 1061        | "2,039"    | 18         | "9,173"     | 81          |           |
| AL    | Greene    | 1063        | "4,514"    | 84.7       | "7,799"     | 15          |           |
| AL    | Hale      | 1065        | "5,401"    | 62.6       | "3,205"     | 37.1        |           |
| AL    | Henry     | 1067        | "3,046"    | 35         | "5,605"     | 64.4        |           |
| AL    | Houston   | 1069        | "12,348"   | 29.5       | "29,214"    | 69.8        |           |
| AL    | Jackson   | 1071        | "6,811"    | 38.3       | "4,123"     | 36.3        |           |

Fig.6 Unstructured database Obama and Romney of USA Elections

We have compare my structured database with unstructured database and found that results are obtained in less time in case of structured database as compared to unstructured database. The four parameters considered are time in milliseconds, number of mappings, number of reductions and average R/W operation done.

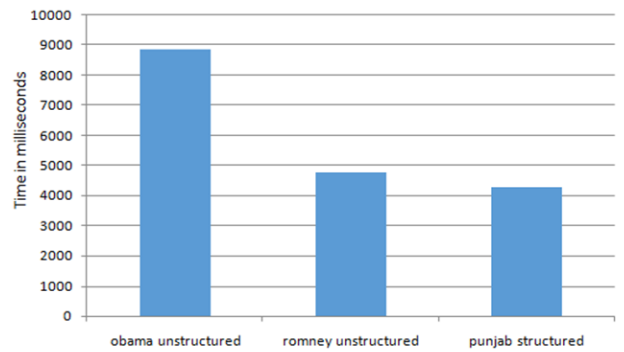


Fig.7 Comparison Obama unstructured, Romney unstructured and Punjab structured

The figure 7 below shows the pictorial representation of the comparison conducted. The two parameters considered are Database and Time in milliseconds for creation of the graph below. I have considered database taken from US elections. The first database has been related to Barrack Obama status and second to Romney status during elections. Later third database is that of Punjab elections. The third database is structured database and the results obtained after executing these databases are shown in the figure below.

Table 4 Time in milliseconds

| Database            | Time in milliseconds |
|---------------------|----------------------|
| Obama Unstructured  | 8830                 |
| Romney Unstructured | 4750                 |
| Punjab Structured   | 4260                 |

In this table 4 show of the comparison of structured database Punjab and unstructured database of USA elections.

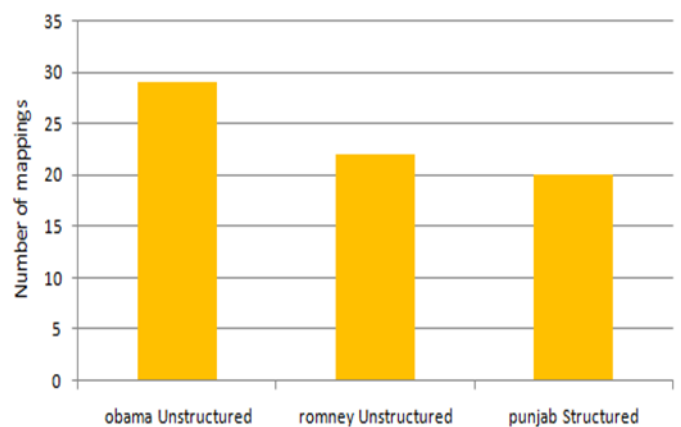


Fig.8 Comparison Obama unstructured, Romney unstructured and Punjab structured chart number of mappings.

In this proposed work the figure 8 shows the pictorial representation of the comparison conducted. The two parameters considered are Database and Number of mappings for creation of the graph below. I have considered Punjab election database taken from US elections. The first database has been related to Barrack Obama status and second to Romney status during elections. Later third database is that of Punjab elections. The third database is structured database and the results obtained after executing these databases are shown in the figure below.

Table 5 Number of mappings

| Database            | Number of Mappings |
|---------------------|--------------------|
| Obama Unstructured  | 29                 |
| Romney Unstructured | 22                 |
| Punjab Structured   | 20                 |

In this table show of the comparison of structured database Punjab and unstructured database of USA.in number of mappings and show the values.

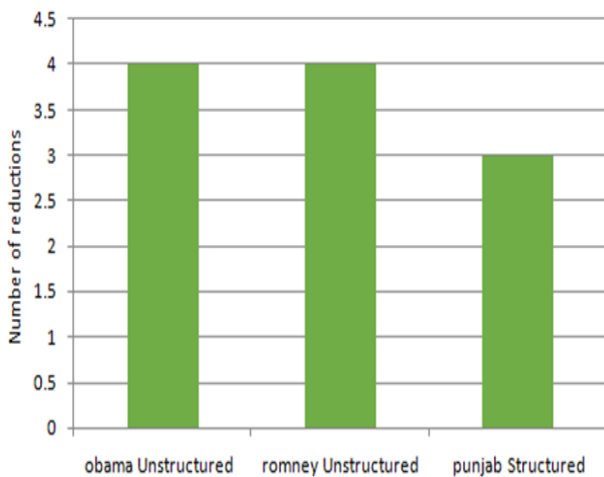


Fig.9 Comparison of Obama unstructured, Romney unstructured and Punjab structured database show the number of reductions.

The pictorial representation of the comparison conducted. The two parameters considered are Database and Number of reductions for creation of the graph below. I have considered database taken from US elections. The first database has been related to Barrack Obama status and second to Romney status during elections. Later third database is that of Punjab elections. The third database is structured database and the results obtained after executing these databases are shown in the figure 9 below.

Table 6 Number of reductions

| Database            | Number of reductions |
|---------------------|----------------------|
| Obama Unstructured  | 4                    |
| Romney Unstructured | 4                    |
| Punjab Structured   | 3                    |

This table show of the comparison of Obama unstructured, Romney unstructured database of USA elections and Punjab structured database .in number of reductions and show the values.

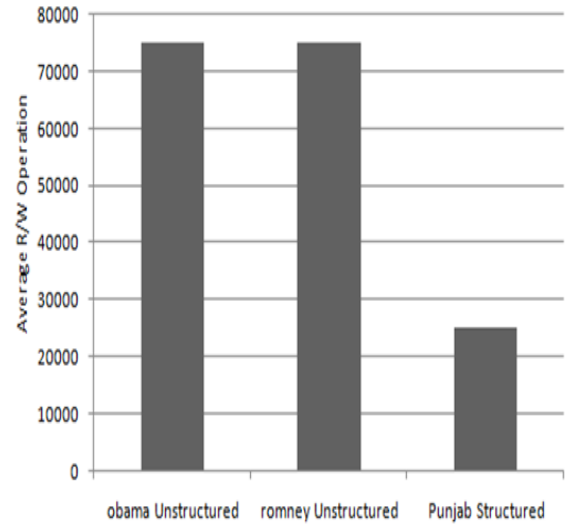


Fig.10 Comparison of Obama unstructured, Romney unstructured And Punjab structured database.

In this proposed work figure10 shows the two parameters considered are database and average read and writes operations for creation of the graph below.

Table 7 Average R/W operation

| Database            | Average R/w operation |
|---------------------|-----------------------|
| Obama Unstructured  | 75153                 |
| Romney Unstructured | 75153                 |
| Punjab Structured   | 24891                 |

In this proposed work table 7 shows the two data base comparison values.

## V. CONCLUSION

It has been concluded that big data will act as a backbone for the next elections and may be path breaker in the way it's fought. It could turn into a massive data gathering exercise where unique databases (for e.g. voter registration, social media, subscription data, transaction profile, mobile records, television viewership and channel bouquet, work profile, location, etc.) are integrated together and analyzed with zeal to find correlations and patterns. It has been predicted that about 160 million of those unsure about who to vote could be reached through mobile phones and about a 100 million through television. These people are waiting to hear the right message to make that choice of which party to vote for and may be the right message is hidden somewhere waiting to be uncovered. Advanced big data analytics could be the key to uncover the winning mantra for the candidate as well as political party.

## REFERENCES

- [1] Marz, N and Warren, J. "Big Data: Principles and best practices of scalable realtime data systems". Manning Publications, pp.1-32, 2013
- [2] Smolan, R and Erwitte, D. "The Human Face of Big Data. Sterling" Publishing Company Incorporated, 2012.
- [3] Gantz, J. and Reinsel, D. IDC: The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. pp. 1-11, 2012.
- [4] Taylor, R., "An overview of the Hadoop/Map Reduce/HBase framework and its current applications in bioinformatics," BMC Bioinformatics article, 11(12):pp.1-7, 2010.
- [5] Jagdev, G., Singh, B. and Mann, M., "Big Data Proposes an Innovative Concept for Contesting Elections in Indian Subcontinent. International journal of scientific and Technical Advancement, 1(3): pp. 23-28, 2015
- [6] Krzywinski, M., Birol, I., Jones, S. and Mesarra, M., "Hive plots-rational approach to visualizing networks," Briefings in Bioinformatics, View · 13(5):pp. 627-644, 2012.
- [7] Katal, A., Wazid, M. and Goudar, R., "Big Data: Issues, Challenges, Tools and Good Practices" Proc. of the IEEE 24<sup>th</sup> conference on contemporary computing, 33(2):pp. 404-409, Noida, 2013.
- [8] Dean, Jeffery, and Ghemawat Sanjay. "MapReduce: Simplified Data Processing on Large Clusters." pp.1-14, 2004.
- [9] Kaisler, S., Armour, F., Espinosa, J.A., and Mony, W., "Big Data issues and Challenges Moving Forward. International conference on system Sciences" Proc. of the IEEE 46<sup>th</sup> Hawaii international conference on system science: pp.995-1004, 2013
- [10] Umasri.M.L, Shyamalagowri.D ,Suresh Kumar.S"Mining Big Data:- Current status and forecast to the future" 4 (1): pp. 5-11,2014
- [11] Mukherjee, A., Datta, J., Jorapur, R., Singhvi, R., Haloi, S., Akram, W., "Shared disk big data analytics with Apache Hadoop" IEEE 18-22 Dec, 2012.
- [12] Aditya B, Manashvi Birla and Ushma Nair, "Addressing Big Data Problem Using Hadoop and Map Reduce", 2012
- [13] Harshwardhan and Prof. Devendra Gadekar, JSPM's Imperial College of Engineering & Research, 4(10):pp.1-7, Wagholi, 2014.
- [14] Mrigank.M, Akashdeep. K, Snehasish. D and Kumar. N "Analysis of Big data using Apache Hadoop and Map Reduce" 4 (5):pp.67-78, 2014.
- [15] Kyong.H "Parallel Data Processing with Map Reduce: A Survey" SIGMOD Record, 40(4):pp.445-478, 2011.
- [16] Chen. H, "Matchmaking: A New Map Reduce Scheduling" in 10th IEEE International Conference on Computer and Information Technology, pp. 2736-2743, 2010.
- [17] R. Ahmed and G. Karypis, "Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks," Knowledge and Information Systems, 33(3): pp. 603-630, Dec. 2012.
- [18] M.H. Alam, J.W. Ha, and S.K. Lee, "Novel Approaches to Crawling Important Pages Early," Knowledge and Information Systems, 33(3): pp.707-734, Dec. 2012.