Dr. Monika Rathore, International Journal of Computer Science and Mobile Applications,
Vol.6 Issue. 1, January- 2018, pg. 23-30 **ISSN: 2321-8363**
Impact Factor: 5.515

Hadoop Map Reduce Framework for Automated Banking System

Dr. Monika Rathore
Associate Professor, Computer science Department, IIIM College, Jaipur
monikadarathore@gmail.com

Abstract: Automation plays a vital role in global industry and in Indian economy. The automated banking system is quite effective automated system which manages the automation of customer information. The banking system is quite challenging in existing system as the customer information cannot be track easily, volume of data cannot store easily. The volume of data is large than risk to handle of that particular data is also high. This paper propose a framework of large volume of data by using the parallel system and a map-reduce algorithm with the help of big data in Hadoop. Big data is a huge word to solve this kind of problem. Data in banking sector increased rapidly which will affect more to the database so big data help to process the database faster and also avoiding the problems. This big data helps to manage the data which is inefficient and also handles data easily.
KETWORDS: ABS, Hadoop, Big Data, Map Reduce.

I. INTRODUCTION

Automation is very important in Indian economy and also rapidly expanding. Many applications are now combining with organization with mathematical tools. The automated banking system is also one of them, which are required in the markets. This banking system is used to manage the customer information. All the information is retrieving is real time. The automated banking management system will store the large amount of data easily which is commonly known as big data.

A big data is a buzz work or a catch phrase which can easily structure, unstructured and semi-structured data. A big data is a unique term as it involves five different V's i.e. volume, variety, velocity, value and veracity. This is much less expensive to access and store. [1]

Big data is very popular and today widely available. The term big data is a word which is mostly searched on Google for implementation in different sectors. The latest observation will acknowledge that big data is quite familiar word for today's world.

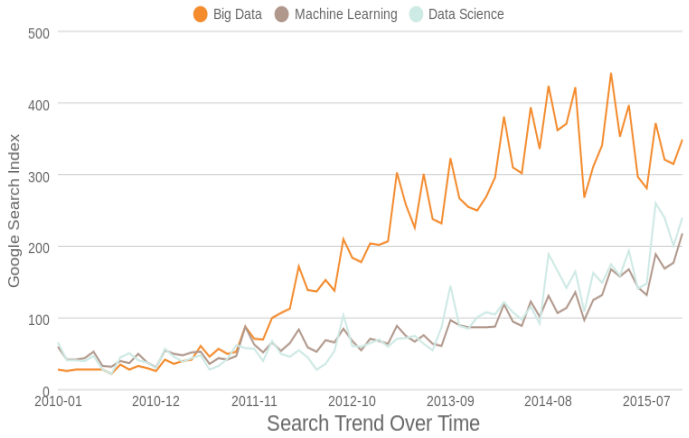

Fig. 1 Big data on Google search index [2]

The latest observation also acknowledges that the big data is to be used with the latest technology commencing in today's world is Apache Hadoop. Apache Hadoop is the most popular on Google search.

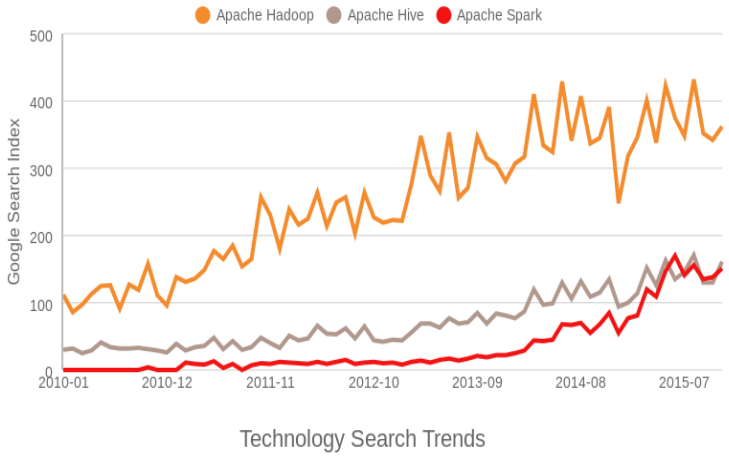

Fig. 2 Apache Hadoop on Google search index [2]

II. EXISTING SYSTEM

In existing banking management system the large scale computation is very difficult. If the single machine is used to store data and system machine will crashed than there is no other way to recover the data.

III. LITERATURE SURVEY

The existing management system informed that the procedure is quite long, more paper work, less efficiency and turnaround time is also not effective. Banks do not bothered about their performance so they move bit slow in their process. As one part is related to the other part so the time is also increased. The process is quite long to responds to the customer need.

This data will show the bad loan scene of both private and public sector.

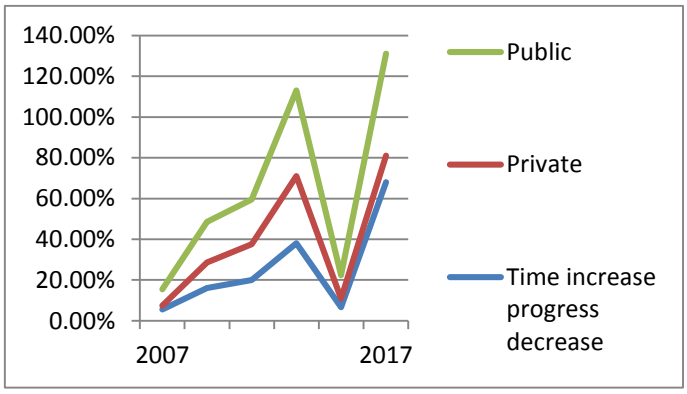

Fig. 3 Bank progress [3]

The banks are inadequate facilities and equipment required to provide modern banking services, absence of frequent training programs for the staff to shape up their attitude towards customers. The following weaknesses of the current system at the banks are:-

The bank staff finds it time consuming when it's figuring the customer data, all the formalities and verification leads delay in passing the loan to a customer. [4]

The bank staffs are quite slow in progress because maximum work done manually as they are not aware about the new technology.

Paper work also reduces the efficiency of the banking system. [5]

Earlier, most banks have failed to make use of data. However, these days, banks have starts using data to reach their main

objectives of marketing. By using the data, many personal data can be revealed like money movements, thefts, disasters. Due to advance in data technology, all the banking operations and procedures square measure automatic. The use of IT technology is evolving in banking sector. Nowadays, the banking system is centric on electronic customer records. Now, the banking sector gears up to process huge volumes of data generated. Some industry experts expect a large increase in the size of data, before 2020. Big Data is huge step towards the development of banking industries, and will forward it into the 21st century. [6]

IV. BIG DATA IN BANKING SYSTEM

Big Data is vast step towards the development of banking industries, and will boost it into the 21st century.

Prime challenges in banks: - [7]

1. Scattered data
2. Fraud identification
3. Targeting/customer analytics
4. No single view of customer
5. Governance

Key areas of application in banks: - [7]

1. Risk analytics
2. Customer experience
3. Operations optimization

Bank analytics says that: - [7]

1. 71% of the financial firms globally believe analytic creates comparative advantages in banking sector.
2. 26% increase in big data spends estimated for 2015-2019 in banking sectors.
3. 37% banks have first-hand experience implementing big data technology for operation and consumer efficiency.
4. 41% banks use the advance big data tools (predictive/ real-time analysis & data visualization).
5. 90% of the people believe that big data initiatives determine winners of the futures.
6. 72% respond in financial service sectors believe that big data has positive impact on rate of innovation.
7. Big data value to retail banking industry estimated at £6 billion over next 5 years.

Big data advantages in banking system are:-

1. Big data, banks can now use the information to regularly, monitor their client's transaction performances in real time. This real-time evaluation will boost overall profitability.
2. Big data compress large number of data, but this job is quite simplified. Whenever the name and account number get entered into the system, it will check and examine all the data and then it will show the relevant information. This process will save both time and cost.
3. Fraud Detection & Prevention - Big Data will allow banks to make sure that no unauthorized transactions will be made, this big data will providing the level of safety and security.
4. Sentiment Analytics - Banks have to continuously monitor what customers say for marketing purpose. Banks have to find who the vital customers are and by getting feedback they have to improve those problems to increase productivity and services.

The task of executing Big Data on a large scale is just figure out, with many IT departments worried about the transition to advanced IT organization.

Big data now extract the good information easily and very quickly from the data according to the bank customer. [8]

V. PROPOSED SYSTEM

The proposed system will address the problem of the large data set which is stored in the system, and also not easy to manipulate. This proposed system will report the problem of data quality in electronic customer records using a computerized customer records. This system extracts multiple parameters from the traditional system and manages the respective electronic records. This system deals with the loan department of the banking system. The records are divided into automobile loan module, home loan module, etc.

VI. MAP REDUCE ALGORITHM

Map reduce is a programming model and associated and implemented the large dataset with the parallel distributed algorithms on a cluster. The term Map Reduce made up of two distinct words i.e. map and reduce. The first phase is map phase in which processing the programming filtering and storing operations and a reduce phase perform the summary of operations.
The map reduces are the basic functions in the algorithm contain two most important tasks:-

- Map (), the map or mapper will process the input data, and then input file passed to the mapper function line by line. Later this mapper will process the data into the small sets of data or we can say create small chunks of data. Basically this map () function will perform the filter and sort operation.[9]

Map function is deals into sub steps:-

a. Splitting - Input dataset is to select from the source data and then divide into small sub-datasets.[10]
b. Mapping - Mapping will take those sub-datasets and then performed required action or computation on each sub-datasets.[10]

The key-value pair is set successfully. Now the shuffling (combining) is to be done.

c. Merging - This merging process will combine all key-value pair which have similar key.[10]
d. Sorting - Sorting will sort the key-value with the help of key.[10]

Sorted key pair are returned as an output of this whole map ( ) function.

- Reduce (), the reducer will process the data which is comes from the map phase of mapper, and after processing, it will produce new set of output. Basically reduce () function perform the summery of operation.[9]

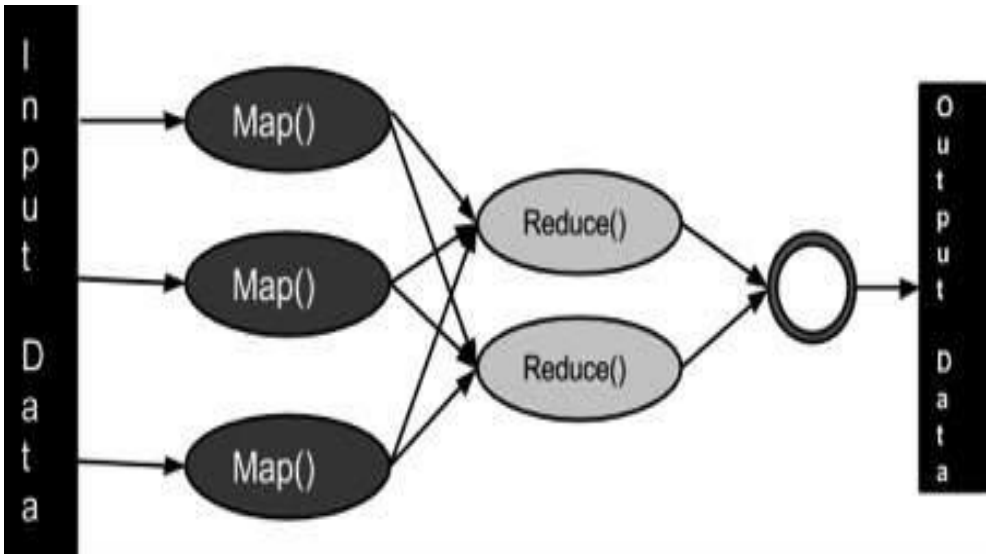

Fig. 4 Map Reduce Diagram [9]

5-step, **flow** of parallel and distributed computation of Map Reduce [9]:

1. **Map () input** – First of all the input is divided into the possible number of chunks or division based on available data and also the processing capacity of individual unit.
2. **Mapper function** – All the chunks are processed simultaneously at the same time and passed to mapper function.
3. **Shuffling** – Later shuffling, aggregation and sorting take place.
4. **Reduce () output** – Reducer () combine all the patterns to make the possible final output as per the logic.

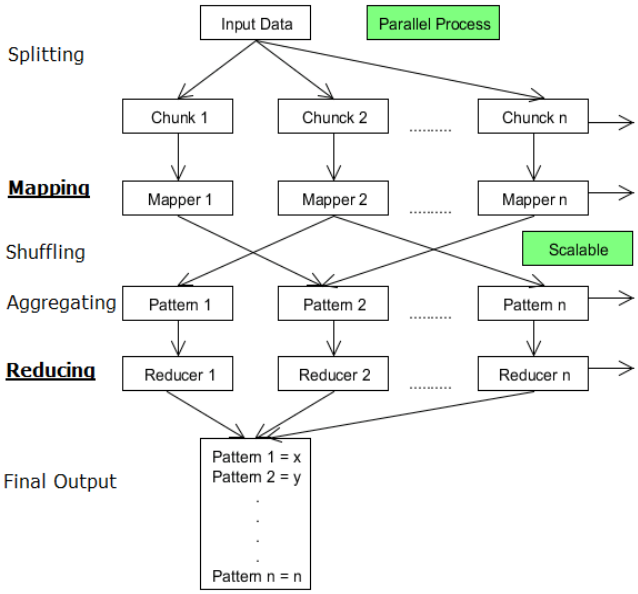

Fig. 5 Map Reduce Block Diagram [11]

VII. IMPLEMENTATION

The automobile loan module, home loan module are implemented.
Map reduce is a programming model and associated and implemented the large dataset with the parallel distributed algorithms on a cluster.

1. **Automobile loan module:**

CREATE DATA ACCOUNT- A user ID is assigned to the new user.

Account includes –
User ID proof
Address proofs
Current address
Aadhar card
Driving license
Income profile
Stability profile

VERIFIES USER - A verification of the user is to be checked. If the user exists already than check the validity of the account.

CHEQUE IS GENERATED – The cheque is generated which is to be send to the dealer or the amount is transferred to the dealers account.

DETAILS REQUIRED –
Firm duty
Bank approval letter
Sanction letter

Different steps –
Assessment portfolio (account info, personal info, residence info, salary)
Check the credit history.
Check the score if there is any pending loan or not.

Sign the paper form 16A/B.
Repaying capacity
Margin norm (margin repaying income)
Generate the cheque.

2. **Home loan module:**

CREATE DATA ACCOUNT- A user ID is assigned to the new user.

Account includes –
User ID proof
Address proofs
Current address
Aadhar card
Driving license
Income profile
Stability profile

VERIFIES USER - A verification of the user is to be checked. If the user exists already than check the validity of the account.

REQUIRED – work experience, loan amount which user want (Min-Max).

Now the loan data of every department is divided and stored into many cities. So for example, the big data technology will collect only the data of the automobile loan department which is located on different cities. Split into different modules map those module verify the details and then finally reduce the data. The generated output is very effective as the large amount of data gets spread all over and to collect a specific data according to the requirement.

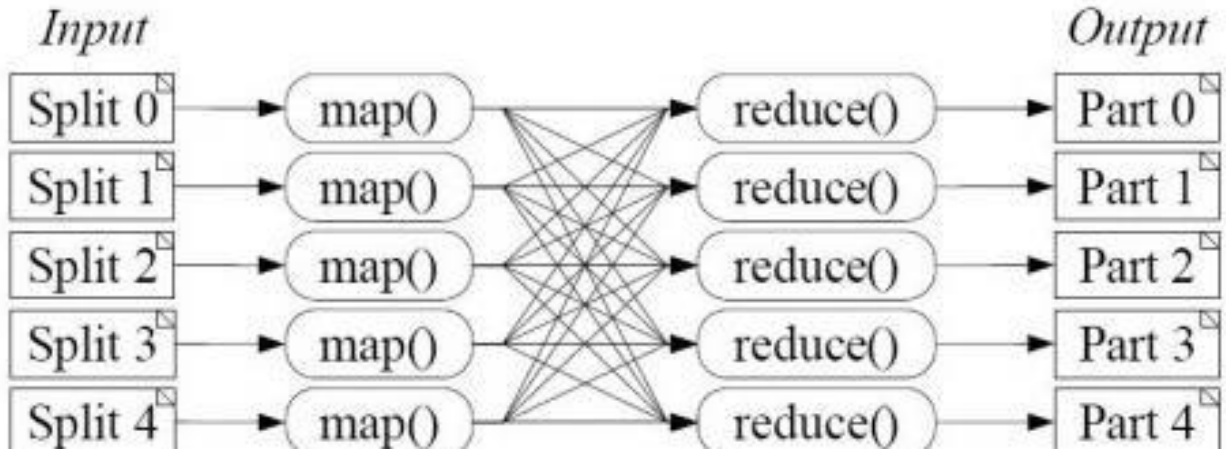

Fig 6 Mapping and Reducing [12]

MAP REDUCE WORKS

Run a map reduce job by clicking on a single method call submit () on a job object. These are the steps that Hadoop takes to run a job.

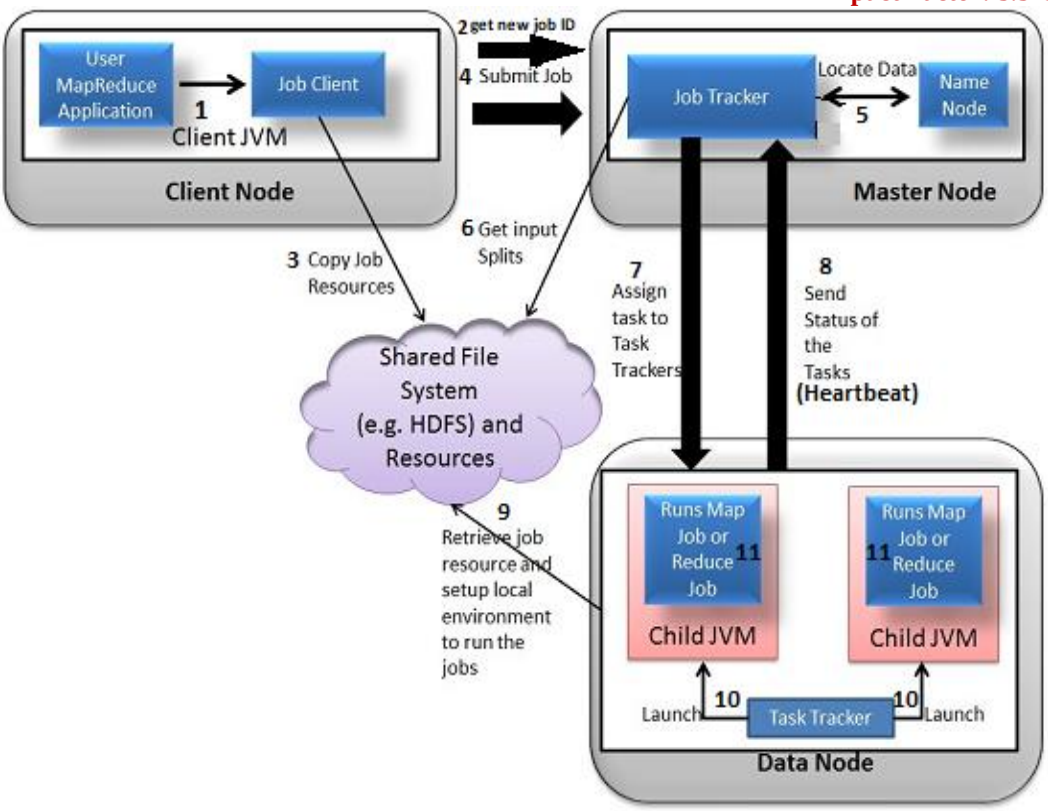

Fig 7 Map Reduce Works

VIII. SYSTEM ARCHITECTURE

The banking systems architect design founds the basic structure of the system, describing the important basic design features and elements that provide the framework for all that follows. The banking systems architect provides the architects view of the users vision, and the paths along which it must be able to progress, and it strives to maintain the integrity of that vision as it grows during detailed design and implementation.

IX. CONCLUSION

In this paper we have presented the automated banking management system which is developed on abstraction of Hadoop map reduce framework. The date which is to be analyzed is semi-structured data. The banking system will solve the problems associated with the existing system. This system can find the total records inserted. The banking systems are now more secured as they required authentication. This paper has solved the problem of storing the large amount of data and retrieving the required reduced data accordingly.

X. FUTURE ENHANCEMENTS

In 2006, when Hadoop came in the existence it will change the life of the whole markets.

The rise in the apache Hadoop, a next generation data architecture is developing that connects the system powering business intelligence and business transaction. Hadoop is especially a capable of storing, aggregating and also used for refining the multi-structure data sources.

In the near future more enhancements is to be done in the field of banking system, the large data files get compressed without eliminating the actual data.

Hadoop started relatively a humble beginning and now evolve to meet more workloads.

REFERENCES

[1] Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2), 77-84.

[2] "Big Data" reaches plateau while interest in Machine Learning grows 12th Jan 2016 http://eliteanalytics.com/big-data-reaches-plateau-while-interest-in-machine-learning-grows/

[3] "Why India's bad loan problem is really bad" Jun 12 2017 http://www.livemint.com/Opinion/o883yxDtd5D4OAVHGMWXTK/Why-Indias-bad-loan-problem-is-really-bad.html

[4] Jain, A., & Bhatnagar, V. (2016). Analysis of Grievances in the Banking Sector through Big Data. International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), 7(4), 21-36.

[5] 3dec 2015 "www.am22tech.com/gov-bank-vs-private-bank-loan-india/ "

[6] 3rd may 2016 "Role of big data in banking industry" http://bigdata-madesimple.com/role-big-data-banking-industry/

[7] 16th Dec `15, 12:36 PM in Banking / Finance "How banking sector makes use of big data analytics" http://bigdata-madesimple.com/how-banking-sector-makes-use-of-big-data-analytics/

[8] Srivastava, U., & Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. Procedia Computer Science, 50, 643-652.

[9] tutorialspoint.com/hadoop/hadoop_mapreduce.htm

[10] https://www.journaldev.com/8848/mapreduce-algorithm-example

[11] http://www.thegeekstuff.com/2014/05/map-reduce-algorithm/

[12] Ch, R., Rajesh, G., Annapurna, G., Swetha, C., Reddy, M. A., & Krishna, G. G. (2016). Automated Health Care Management System Using Big Data Technology. Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org, 6(4).

[13] Agarwal, R., Baltassis, E., Brock, J., & Platt, J. (2014). Enabling Big Data: Building the capabilities that really matter. Recuperado de http://www. the-digital-insurer. com/wp-content/uploads/2014/09/327-Enabling_Big_Data_Building_Capabilities_Matter_May_2014_tcm80-160519.

