

# Big Data: Key Implementation Challenges

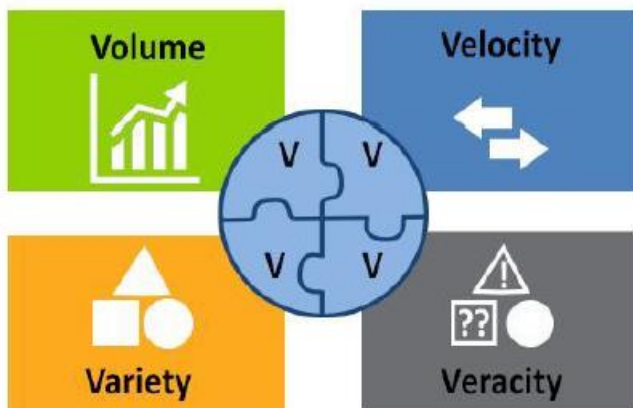
Srinija Sirobhusanam,  
 Computer Science and Engineering Department, GITAM  
 University, Rudraram, Hyderabad,  
 Telangana, India  
 Srinijasiro2595@gmail.com

Dheeraj Chandra  
 Computer Science and Engineering Department, GITAM  
 University, Rudraram, Hyderabad,  
 Telangana, India  
 Chandra\_dheeraj@yahoo.com

**Abstract-**As a breakthrough in this modern era, big data is considered as one of the most important concept in improving the business decisions and also in providing the adequate results in complex situations. The research is still in progress to analyze various effective methodologies which will help in flourishing of the business. Though it sounds easy, but the practical implementation of it deals with many challenges to obtain the desired results. But then, much development is in progress to deal with these challenges and big data is moving ahead sustaining its place being the top in the recent history.

## I. INTRODUCTION

In the digital world today, data is generated from various sources and the rapid transition led to the growth of big data. Big data in general refers to the collection of large and complex data which are actually very difficult to process using the traditional data base management systems. Data is found in various kinds of formats like structured data, unstructured data, and semi structured data. All these kinds of data led to a humongous amount of data to be processed and analysed for the benefit of the organisations to take up good decisions. The big data is usually refers to 4V's which is volume, velocity, variety, veracity. The volume here refers to the huge amount of data which is being generated everyday across the globe. Velocity refers to the rate at which the data is growing day by day. Variety provides the different types of data like structured, unstructured, semi structured data. Veracity refers to the availability and accountability of the data.

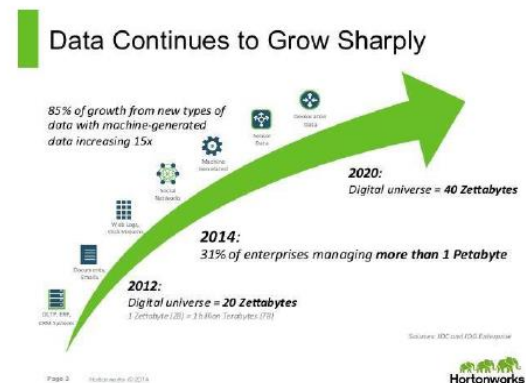


## II. BIG DATA GROWTH AND ITS IMPORTANCE

As the growth of big data is expected to reach very high, much research is going on how to use the big data in a constructive manner to obtain the high output and provide efficient solutions to the users. Analysing huge content of data is a tedious task. Though we have traditional algorithms and analytical tools like Data mining and statistical analysis, but to drive these efficiently we require proper coordination between the database systems and the analytical tools. These challenges generally arise when we wish to perform discovery and practical applications of a business problem.

Though huge amounts of data are being collected by the business organizations, not all the data is useful for the analysis for making proper business decisions. In this paper we would like to mention about the challenges which big data faces during the analysis.

In the recent years, big data has been accumulated across many domains such as retail, biotechnology, public administration, ecommerce, and banking, health care, scientific research etc. Many of the web based applications encounter big data frequently like social computing, internet documents, search indexing, social Network recommendations, online communities, prediction markets etc.



The big data being so popular is because; it provides new opportunities in knowledge processing and analysis. But this cannot be done easily. It includes many of the challenges. To

handle these challenges we need to know about various computational complexities and should keep in mind about the information security and computational methods to analyze big data. Though there are computational methods to deal with, they don't give best results when the data becomes huge. Similarly many computational techniques which work well for small data but face challenges in analyzing big data. The priority task for big data is to understand which data will give the best results out of the whole data which is present in the database system. This includes various mathematical computational technologies to deal with complex data such as machine learning algorithms.

### III. FEW CHALLENGES DEALING WITH BIG DATA

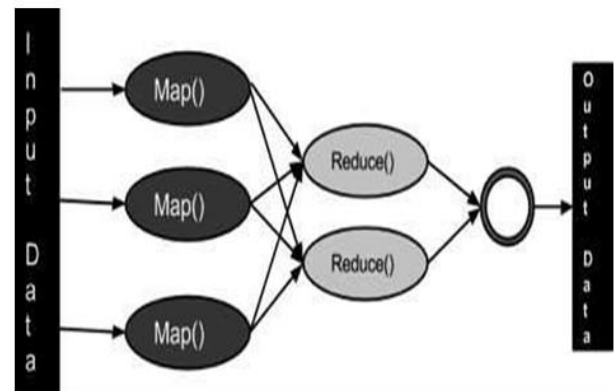
#### 1. Storage of data and analyzing the data

As there is a tremendous amount of data being generated, the challenge is where to store that data. Though all the data may not be useful analyzing but still the storage of the data needs to be available for future. Previously the storage was mostly done on hard disk drives but it is slower in random input/output performance than sequential input/output. To overcome this limitation, the concept of solid state drive (SSD) and Another big challenge with big data analysis is the diversity of the data. As the number of large datasets increases day by day, data mining tasks complexity has been increased. During the data mining tasks additionally data reduction, data selection, feature selection are included in it and is considered as an essential task especially when dealing with complex data sets.

This is the greatest challenge as the existing algorithms many not always respond in the adequate time when dealing with such complex data sets. New machine learning algorithms are being developed in order to deal with such complex data in order to analyse and get the adequate results of out of the data. Map reduce and Hadoop make it possible to collect the unstructured and semi structured data in a very less time. But the real challenge lies in analysing the data pulled by them and produce an effective result which helps the organizations to take up decisions.

#### 2. Map Reduce:

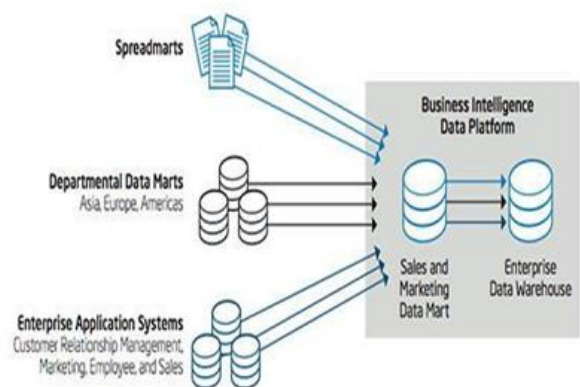
Map reduce is one of the main process technique which deals with two tasks, namely map and reduce. Map will take the set of data and the converts it to another set of data where the individual elements are broken down into key/value pairs. The reduce task will take output from map as its input and combines those tuples into a smaller set of tuples. The reduce task is always applied after the map function is done.



#### 2. Knowledge Discovery and computing complexities

Knowledge of data is obtained by processing the facts present in the data. Knowledge discovery is one of the primary issues in big data. It includes various fields like authentication, information gathering, archiving, management, preserving and representation.

There are several tools which are used for the knowledge representation like fuzzy sets, principal component analysis etc. There are many other techniques which are implemented for knowledge discovery but then they are dependent on the problem which is being dealt with. Some techniques may scale to smaller datasets and some may scale up to larger datasets. The most popular approach of data management is data warehouses and data marts. Data ware houses collect the information from different data sources and which helps in the easy analysis of the large datasets. Data marts general deal with the specific domains whereas the data ware house deals with the whole.



There are mainly two kinds of approaches dealing with the data ware houses. One is the top down approach where the data ware house is created first and then the data marts are created. This is suitable for well-established organizations. The other approach is bottom up approach where the data

markets are created first and then data ware house is created. This is suitable for small business organizations. Analysis of large dataset requires more computational complexities.

The major issue is to handle the inconsistencies and uncertainty which is present in the datasets. We need to understand the domain specific complexities in order to deal with the implementation. The major research is carried out on machine learning algorithms for the latest driving requirements and to deal with the complexities.

3). Scalability and Visualization of Data

Scalability plays a very important role these days because the users tend to have a very fast responding applications. To increase the scalability many techniques are being used such as the parallel computing and graph theory, increasing the number of cores which helps in faster retrieval of data.

Big markets like Amazon, Flipkart etc. which has millions of users and the corresponding data. They use the big data analytics and visualize the data using tools such as tableau, power BI to understand where their business stands and what are the key challenges they are facing and how to take better decisions to improve their company.

In order to meet the complexity big data has been developed with multiple computational methods and also the mathematical computational techniques. Few of them include cloud computing, distributed computing etc.

4). Information Security

In the big data, as the analysis is dealt with the real time data of the organizations, the information must be secured. This is one of the biggest challenges of big data to safeguard the security of the data which is being used.

Many techniques such as authentication, encryption, and multi-level authorization techniques have to be implemented to double check with the user who is dealing with the data. As the security is considered as an important aspect, big data research is widely carried out on protecting the data security and abstraction of the data.

5. Finding the right use cases

When to make the decisions about the big data analytics the use cases play a key role. We need to come up with the business use cases which are actually leading them to business success. Finding out the right business use cases which suits the organisation and benefits the organisations in moving towards the success rate. For this to happen it is important to understand the business, their pros and cons of the changes made to the business, how can we improve the business keeping in mind the competitors and how innovative is the approach to drive the business. Understanding the entire business and the open market and the challenges which is the business face in the market proactively will help the business to grow though predictive analysis using big data.

6. Quality of data

Using of humongous amount of data will actually help in the better analysis. Even the business leaders want the huge data whereas IT leaders will want the technical data and useful data to drag the statistics before taking up decisions. Big data analytics play a vital role in choosing the relevant data to ensure the quality of the output. Out of the large data only the data which provide qualitative results should be taken into consideration which is a big challenge.

7. Finding the right resources and skills

Last but not least, we need to build a right team with the right skills to achieve success. By now data scientists have become popular as good decision makers. The basic skills that should be possessed is the business analysis, communication abilities, creativity, statistical analysis, understanding business intelligence.

Few other skill set is also treated as necessary for driving big data analysis as follows:

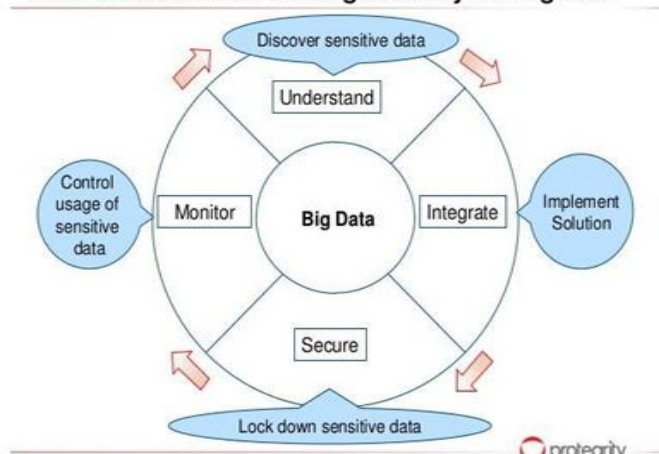
**Analysis Architect:** This person is responsible for driving the analysis map and be a use case lead, understand the business and would play a role as business analyst.

**Data Engineer:** This person is actually an expert data guru, expert in data warehousing and ETL concepts.

IV. FUTURE SCOPE

Big Data analysis would always be a potential concept to drive

The Process of Automating Security for Big Data



the business in a successful way. In spite of the major challenges in its implementation Big Data concepts play a vital role in future and the analytical expertise would help in analyzing the business and keep tracking the trends of the business to take better decisions for better tomorrow.

## V. CONCLUSION

In spite of the challenges in implementing the big data concepts, big data has always been very helpful in taking better decisions and improve the business to stand in the competition ahead. Huge research is being done in this area encouraging the newer enhancements to the existing technology for the future demands.

## REFERENCES

1. <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
2. <http://searchbusinessanalytics.techtarget.com/essentialguide/Guide-to-big-data-analytics-tools-trends-and-best-practices>
3. <https://knowledgegen.com/whitepaper/2015-big-data-survey-current-implementation-challenges/>
4. <http://blog.onapproach.com/7-challenges-of-implementing-a-big-data-analytics-solution>
5. <https://www.linkedin.com/pulse/7-implementation-challenges-big-data-analytics-laurent-fayet>