



Analysis of Big Data Challenges and Different Analytical Methods

Gurram Bhaskar¹ & Motati Dinesh Reddy²

¹⁻²Department of Computer Science and Engineering,

SRM Institute of Science and Technology

Chennai, India

ABSTRACT

Big Data holds a great potential to gain the insights of various academics and practitioners to enhance the decision-making process. Big Data Analytics is one of the most trending practices in the present time that organizations are adopting to construct valuable information from the big data. The data analytics methods help improve the operational efficiency of the company and drive new revenue streams and gain competitive advantages. The paper will shed light on the different challenges related to big data and different methods of data analysis. Further, it will talk about the method used for the research process, analysis of the data, and at last conclusion.

Key Words: Big Data, Challenges in Big data, Analytical methods.

1. INTRODUCTION

In today's era, no firm can work without data. An enormous amount of data being produced every minute from numerous trade dealings, customer logs, figures of sales, and data of the stakeholders is the fuel that drives the firms. This entire data gets heaped up in a vast place of data which is known as Big Data. Big Data can be explained as the compilation of data that has a vast volume yet increasing exponentially with the instance. In simple words, it can be said that it is the data with colossal size and intricacy that not any of the tools of data administration can stock up or process in an efficient manner (Sivarajah, Kamal, Irani & Weerakkody, 2017). This can be explained well with the help of an example of social media. According to statistics, more than 500 terabytes of new data can get consumed in the database of social networking sites such as Facebook on a daily basis. The data is mainly generated in terms of uploads of pictures and videos, exchange of messages, inserting comments, and many more. Big Data can mainly be classified into three types which are structured, unstructured, and semi-structured. The data which can be protected, accessed, and practiced in any unchanging setup is referred to as structured data. The data whose form or structure is not known is known as unstructured data, and semi-structured data is the one that may have both forms of data. The present report demonstrates the challenges of big data along with how it can be solved. In addition to this, the report will elaborate on the challenges of extensive data analysis, and at last, a meaningful conclusion will be derived.

2. LITERATURE REVIEW

Numerous challenges are involved in big data. Several researchers are working in the field of big data, and they have identified numerous challenges which are present in the analysis of big data. It is often noticed that in the initial stage, numerous firms get stuck in their Big Data projects because they are not aware of these challenges, and also, they do not have adequate resources to tackle those challenges. Below are some of the disputes of Big Data, along with how they can be solved.



Fig 2. Challenges Involved in Big Data

2.1 Lack of proper understanding

The companies are unable to attain success in Big Data initiatives because they do not have adequate understanding. The employees do not have any knowledge about what data is, its storage capacity, dispensation, significance, and also do not know about its sources (Sharma, 2020). Even though the data professionals have an awareness of what is going on but other people may not have a clear picture. For example: If the recruits do not understand the significance of data storage, they may not keep back up of sensitive data. They might not correctly make use of the database. Therefore, when this essential data is needed, it cannot be recovered quickly.

Solution:

Seminars and workshops on Big Data must be detained at the firm for every employee. Basic programs of training must be prearranged for all the human resources who handle data on a regular basis and are also part of big data projects (Sharma, 2020). In addition to this, a basic comprehension of concepts of data must be instilled by every level of the firm.

2.2 Confusion while a selection of tool of Big Data

Most of the time, it is noticed that firms get confused at the time of choosing the best tool for analysis and storage of Big Data. As there are many tools, therefore, the firms get too much confusion due to which they make poor decisions and choose technology that is not appropriate. Therefore, it can be said that time, working hours, and money get wasted.

Solution:

The most acceptable means to solve this is to take help from professionals. The company can either recruit a professional who has vast experience and also has complete knowledge about these tools. Another way can be Big Data Consulting (Sharma, 2020). Under this method, the consultants will suggest about best tools following the scenario of the firm. Based on their advice, the company can make some tactics and then choose the tool that is most appropriate for the company.

2.3 Issues of growth of data

Another major challenge of big data is hoarding these gigantic data sets appropriately. It has been noticed that there is a rapid increase in the data that is stored in databases and data centers. As these data sets grow at a quicker pace with time, it becomes complicated to grip. The maximum amount of data is not structured, and it comes from videos, files, audios, documents, and varied sources. In simple words, it can be said that a person cannot look for them in databases (Sharma, 2020).

Solution:

To handle these vast sets of data, the firms can make use of modern methods such as deduplication, compression, tiering, and many more. Through reduplication, firms can remove replica and unnecessary data from the data set. With the help of compression, the firm can decrease the number of bits in data which thereby helps in the reduction of the overall size. Data tiering helps companies to stock up data in different storage tiers. It makes sure that data inhabits in most suitable storage space. The data tiers can be public cloud, flash storage, and classified cloud that is reliant entirely on the size and significance of data (Sharma, 2020).

2.4 Securing Data

Securing massive data sets is one of the most frightening challenges of Big Data. It is often observed that firms are so hectic in comprehending, storing, and investigating their data sets that they push the procedure of data security for upcoming phases. However, this is not taken into consideration as a smart move because the data storehouses that are not protected can be breeding grounds for malevolent hackers.

Solution:

To solve this issue, these days the companies are hiring more cybersecurity professionals to safeguard their data. Other steps that are taken for protecting this data are inclusive of segregation of data, data encryption, implementation of endpoint security, identity, and access control, Etc. (Sharma, 2020).

2.5 Lack of data professionals

To run tools and the latest technologies of Big Data, the firms require skilled data professionals. They will be inclusive of data analysts, data engineers, data scientists who have a high degree of experience in working with big data tools and know-how to make sense of enormous data sets (L'heureux, Grolinger, Elyamany & Capretz, 2017). The organizations face the issue of Big Data professionals because there is a quick evolvement of data handling tools, but this is not observed in the case of data professionals. It is essential to take significant steps to bridge this gap.

Solution:

For solving this issue, the firms are making investments in hiring skilled professionals. In addition to this, they also have training programs for current employees so that they can get the most from them. Another crucial step that is taken by the firm is the purchase of solutions of data analytics which are powered with the help of machine learning. These tools can be run through experts that do not have expertise in data science but have fundamental acquaintance. With the help of this step, the firms can save a considerable amount of money for hiring (Sharma, 2020).

2.6 Expensive

A considerable amount of expenses is incurred for adopting projects of Big Data. If the company chooses a premise solution, then the company will have to keep the cost of hardware, electricity, hiring of new recruitment in mind. Even though the frameworks required are open source, but the company will have to pay for setup, enlargement, configuration, and continuation of the new software (Bekker, 2018).

Solution

The specific deliverance of the wallet of the firm will be completely reliant on the firm's specific technological requirements and business objectives. The company can save money through data lakes. This can give cheap storage opportunities to the firm for the data that it does not need to analyze now.

2.7 Big Data Security Holes

It is often noticed that big data implementation projects push safety for many stages. This is not at all a smart move. The technologies of Big Data do evolve, but their security features are avoided as it is assumed that safekeeping will be given at the time of request (Bekker, 2018).

Solution

This problem can be solved by giving priority to security. It is specifically essential in the phase of designing the architecture of the solutions. The main reason behind this is that if the person does not get well with significant data security from the beginning, it will be risky when it is least expected.

2.8 Upscaling troubles

The design of the solution may be thought, and adjustments can be made to upscale without making any extra efforts. However, the real issue is not in the actual procedure of commencing new hoarding capacities, but it is in the complication of scaling up that the performance system does not decline and reside within the budget (Bekker, 2018).

Solution

The best solution for this problem is to have a decent architecture of big data solutions. If a big data solution can do this, then there will be fewer problems in the upcoming duration. In addition to this, businesses must design big data algorithms while keeping hope upscaling in mind.

3. SCOPE AND OBJECTIVE

The research objectives are: To understand the challenges related to big data and To understand the different Analytical methods.

3.1 Research Question

1. What are the Big Data Challenges?
2. What are the different methods of Data Analysis?

4. METHODOLOGY

4.1 Data Collection Tools and Techniques

This part of the research paper interprets and analyzes the data gathered from secondary sources of data. The primary purpose of the research is to determine the critical challenges related to Big data and different data analytical methods (Johnston, 2017). The data for the research purpose have been collected keeping in mind the requirement of the project and focusing more on the issues related to Big data. A straightforward data collection tool and technique have been used for collecting information, referring to scholarly journals and articles, books, newspapers, websites, and magazines that directly link with the research theme and will help to conclude [4].

4.2 Ethical Consideration

The research paper was prepared by maintaining the ethical standards that need to be undertaken throughout the research process. The data collected from the different sources were kept confidential without any alteration to the facts and figures. The conclusion obtained with the help of the data is unbiased and were obtained through fair means.

5. FINDINGS AND ANALYSIS

The findings and analysis have been made based on secondary research, which includes the use of scholarly journals, articles, newspapers, magazines, and websites.

5.1 Methods of Data Analysis

There are three central V's through which Big Data is characterized, and they are a massive volume of data, the velocity at which the data gets processed, and the last is the wide variety of data. Due to velocity, data analytics has expanded in the technological fields of artificial learning and machine intelligence (Lee, 2017). The companies that tie together the real power of big data can increase their operating margins up to 60%. This is possible through data analytics, but a skilled team of analysts is required who comprehend the fundamentals of research methodologies. Below are some of the data analysis methods [5].

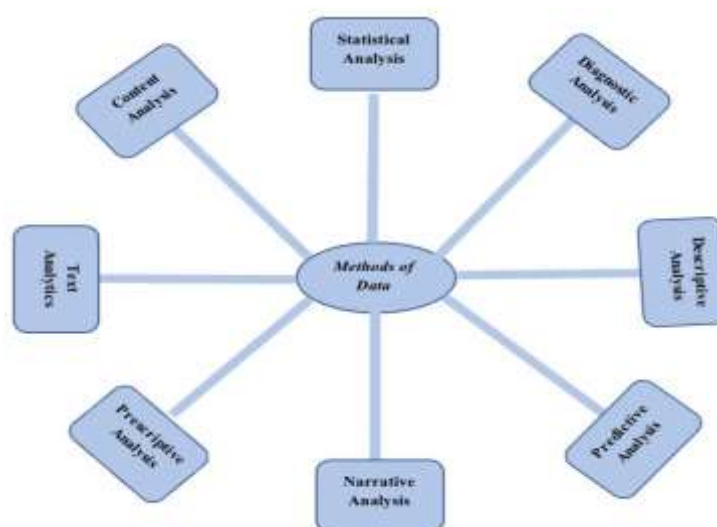


Fig 2 Methods of Data Analysis

5.1.1 Statistical Analysis

This is one of the most common methods as it can be used for both small data sets and massive information in the library. This is a game of numbers, but the investigators define it as a science of gathering, exploring, and demonstrating vast amounts of data to discover fundamental inclination and prototypes. This method can be used everywhere. For example, the person who writes a research paper can make use of statistical analysis to analyze the data gathered from customers and improve the assignment (Big Data, 2020).

5.1.2 Diagnostic Analysis

The ultimate purpose of this method is to get a detailed analysis of the matter and get to know about the causes of provided event, concept, or phenomenon. With the help of diagnostics, the firms can detect associations between parameters of the business. For example, it is often noticed that content writing firms are dependent on diagnostics analysis to learn why top resume reviews are good or evil than expected (Big Data, 2020).

5.1.3 Descriptive Analysis

This can be explained as the term that is used for analyzing the data, which helps in describing, showcasing, or summarizing data in a meaningful manner so that patterns can emerge from data. This method is often used for tracking KPI. Under this, the goal is highly simply meaning the person has a clear set of KPI's and descriptive statistics which showcase result based on the actual recital of the firm (Big Data, 2020).

5.1.4 Predictive Analysis

This is another method of data analysis that is mainly used for identifying trends and future business results. Instead of relying on self-intuition for making decisions, the person can take the benefit of this analysis to make good decisions that are driven through data. It is a kind of statistical modeling that helps in deriving meaningful conclusions on a particular subject. This method can be used for undertaking risk assessment and forecasting sales (Big Data, 2020).

5.1.5 Narrative Analysis

If a person is willing to do business analysis with the help of words instead of numbers, then this can be the best method for the company. The main idea behind this analysis is to investigate ideas, opinions, attitudes, and stories. In this way, the company can discover essential predilections among the employees and their HR software and again consider the entire culture of the firm (Big Data, 2020).

5.1.6 Prescriptive Analysis

The main objective of business intelligence is to look for ways to make enhanced and more correct decisions. This makes perspective analysis extremely valuable, but it is also not easy. It is an amalgamation of predictive and descriptive analysis. This can be explained as that area of business analytics that is dedicated to looking for the best course of action for a particular situation. The drawback of this methodology is logistics because it necessitates a huge workforce and extensive budgeting.

5.1.7 Text Analytics

This is a mechanized procedure for decoding vast volumes of unstructured data in quantitative data to discover insights, patterns, and trends. This method is also known as text mining. This method often works in amalgamation with data visualization stands, thus resulting in a more consequential indication of the targeted procedure. As the internet is completely packed with all kinds of textual posts, a firm can make use of this method to comprehend and evaluate the content that is associated with the brand. This helps lots of firms to keep their online goodwill irrespective of dealing with numerous customers all over the world (Khan and Vorley, 2017, Ref [6]).

5.1.8 Content Analysis

If an individual is involved in analyzing qualitative data, then he or she must give this method a try. It is a research tool that is used for determining the existence of certain words, concepts, themes within some provided qualitative data. This tactic is usually used in textual analysis. Therefore, it is well known amongst customer paper and assignment help services. This method is excellent for those firms that have an eagerness to comprehend their consumers because it helps decision-makers to know about the actual meaning of surveys, reviews, and other kinds of feedback [7].

5.1.9 Hypothesis Testing

This method is also known as the T testing method. With the help of this technique, the individual is trying to prove or disprove the outcomes of the research. In simple words, it can be said that a person is trying to test whether the outcomes are valid or not by figuring the odds that outcomes have happened by chance. If an individual is willing to prove a correlation between certain

business aspects, then it is the best method. For example, an individual can claim that if a person spends less amount of time on social networking sites, then it leads to a higher level of employee engagement. However, before a person limits the utilization of social media at the workplace, individuals must test it and prove the point.

5.1.10 Monte Carlo Simulation

The businesses that are interested in risk alleviation can entirely depend on this method. The basic concept of this method is to make use of randomness to solve the issues that may be deterministic in principle. This is a simple method, but business managers do not frequently make use of this method, but it is valuable to the model data analytics if other methods do not go well with the current requirements of the business (Torres and Tobar, 2019).

6. CONCLUSION

The paper represents the holistic view of Big Data practices, and the challenges related to these are represented in the part of the literature review. Based on the findings, the paper analyzes and presents a structured analysis of Big data to support future research. The Big data analytics are very beneficial for the organizations as it helps to analyze a large amount of data from various sources and make better-informed decisions. It also provides a better understanding of the needs of the customers that can provide better marketing insight and help in product development.

REFERENCES

1. Sivarajah, Kamal, Irani & Weerakkody, "Critical analysis of Big Data challenges and analytical methods". *Journal of Business Research*, 2017, vol. 70, issue C, 263-286
2. Sharma, 2020, "Augmenting Data Warehouses with Big Data", *Information Systems Management Volume 32, Issue 3: Business Intelligence*.
3. Alex Bekker, "The Scary Seven big data challenges", *Science soft*, 2017.
4. Johnston, M. P. "Secondary data analysis: A method of which the time has come." *Qualitative and quantitative methods in libraries*, 3(3): 619-626, (2017).
5. Lee, I. "Big data: Dimensions, evolution, impacts, and challenges." *Business Horizons*, 60(3): 293-303, (2017).
6. Khan, Z. and Vorley, T., 2017. "Big data text analytics: an enabler of knowledge management" *Journal of Knowledge Management*
7. Krippendorff, K., "Content analysis: An introduction to its methodology." Sage publications, (2017).
8. Cevallos-Torres, L. and Botto-Tobar, M., Monte Carlo simulation method In *Problem-Based Learning: A Didactic Strategy in the Teaching of System Simulation*, (pp. 87-96). Springer Cham, 2019
9. L'heureux, A., Grolinger, K., Elyamany, H. F., and Capretz, M. A. (2017), "Machine learning with big data: Challenges and approaches". *IEEE Access*, 5: 7776-7797.