

Study on Challenges in Big Data

Sandra Krishnan¹, Bijin Joseph², Hajira Hazeena³, Pavan Johny⁴, Dhanya J⁵

^{1,2,3,4}UG- Kristu Jyoti College of Management And Technology, Chethipuzha, Kerala

⁵Assistant Professor-Kristu Jyoti College of Management and Technology, Chethipuzha, Kerala

ABSTRACT

Big data is having a greater influence in cloud storage now a days. Big data is a data which contains huge varieties, arriving in massive volumes and with greater speed. Big data is very complex and is very huge in size. Any one of the traditional management tools cannot be applied in big data. It plays a foundation for next generation of advances in medicine, business, science. This paper presents a study on challenges of big data. Big data analytics a process that include research of massive amount of data to reveal hidden patterns and secret correlations. This paper consists of big data content, scope, methods, advantages, challenges etc.

KEYWORDS: Complex, Advantages, Challenges, Huge.

1 Introduction

The definition of big data is the data with large size, huge variety, high volumes and have greater velocity. The bigdata is known for the three v's. These data cannot be processed by the traditional data processing application. The analysis of big data can find new correlation in business trends, prevent diseases etc. Nowadays there is an increase in the data usage as the data is collected by the devices such as mobile phones, software logs, cameras, microphones. Big data requires a set of technologies with new forms of integration to reveal insights from data sets that are complex and large. Example of big data Workstack Exchange, social media etc.



fig(1) : Bigdata

2 Characteristics of Big data

The characteristics of big data are as follows:

- Volume
- Variety
- Velocity
- Variability
- Exhaustive
- Relational
- Scalability
- Relational

Volume- volume refers to the quantity or stored data as the size of data determines the value, potential etc. The size of big data is usually in terabyte, petabyte or larger than that.

Variety- variety refers to the different types of data. The data can be in the forms of mail, photos, videos, PDFs, audio, etc.

Velocity-velocity refers to the speed at which the data is being generated.

Variability-variability refers to the inconsistency in data

Exhaustive-All the data in the available source may or may not be included in big data. Whether the data may record captured or not.

Relational -Relational refers whether those data that are collected has common field that would enable conjoining.

Scalability- Scalability refers that whether the size of the big data can be expanded or not.

3 Types of Big data

The different types of big data are:

Structured

Semi-Structured

Unstructured

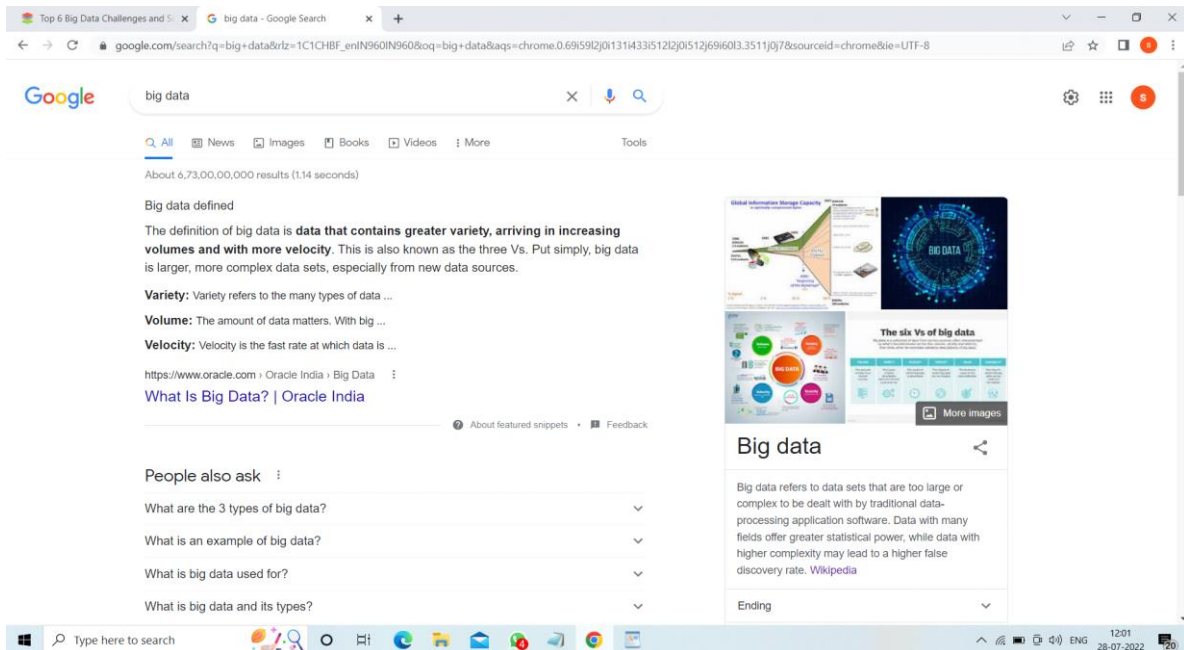
3.1 Structured

Those data that are collected accessed in a fixed format is called structured big data. Examples of structured data are data stored in a database

Employee_ID	Employee_Name	Department	Salary
235	Rajesh	Finance	650
398	Pratibha	Admin	650
465	Shushil	Admin	500
500	Shubhojit	Finance	500

3.2 Unstructured Data

Those data that are in unknown is called unstructured data. Besides the size of the unstructured data those data are very complex possess many challenges in terms of its processing. Examples of unstructured data are the data containing the combination of simple text files. The output returned by the google search.



fig(2) : example of unstructured data

3.3 Semi-structured

The semi-structured data are those data that includes both form of data. The semi-structured data can be found both in structured and semi-structured form. Example of semi-structured data is the file stored in XML file.

```
<rec><name>ABC</name><sex>Male</sex><age>30</age></rec>
<rec><name>DEF</name><sex>Female</sex><age>40</age></rec>
<rec><name>XYZ</name><sex>Male</sex><age>26</age></rec>
<rec><name>LMN</name><sex>Male</sex><age>20</age></rec>
<rec><name>WXY</name><sex>Male</sex><age>30</age></rec>
```

4.Applications:

In this growing world there is an increase in the use of data. Big data can be applied in different sectors to promote the development. The various applications are:

4.1.In government sector: -

The use of bigdata processes enhance the cost, productivity etc. The most common government organization that use the big data is National Security Administration (NSA) that look after the internet activities

4.2.In Healthcare: -

Big data is used in the healthcare for making personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries. With the increase in the technologies there is a random increase in the data. As the information in healthcare is now electronic, it comes under big data and it is unstructured. As we started using the bigdata in healthcare it also started raising challenges. A subarea related to this application is computer aided diagnosis

4.3.Education: -

As we know that education institution needs the data to store the details of the students studying in an institution.it is difficult to store it in a place. These data should be regularly updated and cleared.

Here we use bigdata to perform our process. Some of the plans are

Civilta's learning-It helps students to succeed in professionally. Here the gather the data from all the universities and help students to plan their career.

12Twenty-This application helps the student to keep in contact with their future employers. It makes the process of data collection easier.

Gradebook Pro-It is an innovative class management tool. They help the student to monitor their performance, behaviour in real-time. This application help to keep the record of the attendance and scores. This app is free of any subscription.

Apex learning-This application is used to keep the records of the students throughout the year.

Knower-It is an app that help student to study maths's problem in a easy way. Here they recognise the weakness and strength of each and every student. There is a dashboard for students and teachers.

4.4 Media

Now a days it has been changed from traditional approach such as using newspaper, television shows magazines etc to electronic form such as WhatsApp, Facebook, twitter, YouTube etc. This has led to increase in production of data. Now media has become an important part of the human life. Some cases where the big data are being used:

Spotify one of the famous music providing app uses bigdata to collect the data and uses these data to provide the recommended music to the users.

Amazon Prime which provides a great variety of videos,audios,movies etc also uses the bigdata .

The virtual personal assistant tools like Siri, Cortana uses big data to collect the data related to the questions asked and by analysing the data the responds are made.

4.5 Internet of Things (IoT)

Big data place an important role in IoT those data from IoT devices will create a mapping of inter connectivity of devices. The industry uses such mapping to increase efficiency. IoT is a way of

collection of sensory data and later these data are being used in different fields like manufacturing, transportation etc.

4.6 Information Technology (IT)

Big data has started to help the employer's work more efficiently and the process of using big data for collecting data within an enterprise is called IT operations analytics. potential uses can be predicted and prevented by using the concept of bigdata in deep computing, machine intelligence in the IT department. Platforms for system management is being generated by Iota businesses

4.7 Insurance

Health insurance is a sector which requires the collection of data about every client about their details, so that they can come to a prediction about the health costs. Later these predictions are being used in the process pricing each client.

Energy Sector

Here the consumed power is read by the smart electric meter a during an interval of every 15-minutest read data is being send to the server and here the data is analysed. With the help of this application, it has made to reduce the electricity bill to a certain range.

Banking

In banking sectors, the bigdata is used to track the financial activities. The natural language process is being used to track the financial activities. In big banks the trade analytics are being managed by using bigdata.

4.8 Transportation

Bigdata has been used in traffic control, route planning, also some strategies to save the fuel and time.

4.9 Sports

The sport sensors use big data technologies.it is used to improve the process of training. The prediction of the winners can be made using bigdata. The sensors collect the data about the players and decision about who will win the race from the data.

5 Tools used in bigdata

The fast-growing world is completely flourished with the data .as the data can bring efficiency. Some of the tools used by bigdata are:

5.1 Hadoop

The Apache Hadoop is a type of framework. If we have clusters of computers then using this framework, we can process all the large set of data.

Features are: it has made data processing faster, it is more flexible

5.2 Atlas.ti

It is a research software. We can use this tool for analysis of qualitative data and also in research purposes. The features on atlas's are: the information of source of data can be exported, it allows to work with our data in an integrated manne, it helps to rename a code, help to handle large files.

5.3 HPCC

Developed by LexisNexis Risk solution. It has provided a single platform for data processing. Some of the features are: it is highly efficient, highly reductant and available, it is used for, complex data processing, enhance scalability and performance.

5.4 Storm

Storm is an open-source computational system. It processes some of the features such as distributed real time, fault tolerant processing system. Some of the features of the tools are: it is highly efficient, it uses parallel calculations, If any node dies then the system will restart automatically.

5.5 Cassandra

This tool is best for its better performance in the effective management of data. Some features of this tool are: in case of any fault tolerance the data will be replicated to multiple nodes,even when the entire data centre is down this tool is used to process data.

6 Challenges in Bigdata

Lack of knowledge professionals

Lack of understanding of data

Increase in data

Problem in selecting tools

Integrating the data from different sources

Securing data.

Lack of knowledge in bigdata-There is a need of well experienced professionals to work on large data sets. We require data analyst, data engineering, data scientist to work on these large set of data. The main problem faced by the companies nowadays are the lack of the professionals who are well experienced in the subjects they need to work with.

One of the solutions to this problem was by investing more money for teaching the employers the subject they need to work so that the company will get the benefit from them. This actually expensive and also the training the employers is time consuming.

Lack of Understanding of data-Here the employers don't know have the knowledge about the data, only the data analyst or data scientist have clear knowledge about the data. Others do not know about the data processing, data storing etc. So, when there is a need to access a particular data, it become difficult for the employers to retrieve them.

Increase in data-As we that the data is increasing day by day and has become a huge challenge to store these data. The data in the data centres are also increasing day by day. So, it has become difficult to search for a data as most of the data are in district form. Many modern technologies have been used by the companies to organise these data's such as compression, tiering and deduplication.

compression-This process is used to reduce the size of the data by reducing its bits.

Tiering-here the data is stored in indecent duplicate storage tiers

Deduplication-This is the process of removing the duplicate files and unknown files from the database.

Problem in selecting tools-As we know that big data is a large data and need tools to analyse these huge data. The professionals are in a great dilemma to choose the tools that is to be used. They all will forward to choose a tool that make their process simple.

Integrating the data from different sources-In a company the data comes from the different sources like the emails,presentation,social medias, customer logs etc. The employers have combined the data from each log to make the report is a challenging task. The data integration tools are used to perform these tasks. Some of the data integration tools are ArcESB, Microsoft SQL Qlik View, CloverDX.aq1

Securing data-We know big data is a huge collection of data securing these data is a challenging process. The companies set this process for later stage. But if don't secure our data there will be loss of data at the future stage. So, the company started to appoint the data cybersecurity professionals to safeguard the data.

7 CONCLUSIONS

The scope of the bigdata is increasing day by day. As we know that amount of data usage will grow exponentially and the technology and tools related to bigdata will also increase rapidly.Eventhough there are many challenges data scientist find solution to all the problems. In the forthcoming year the scope of bigdata will increase and there we can see many opportunities in the following year.

8 References

1. <https://bigdataanalyticsnews.com/big-data-applications-in-education/>
2. <https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications> HYPERLINK "<https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications>" HYPERLINK "<https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications>" HYPERLINK "<https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications#GoTop>



3. <https://www.guru99.com/big-data-tools.html>
4. <https://www.xenonstack.com/insights/big-data-challenges>
[HYPERLINK "https://www.xenonstack.com/insights/big-data-challenges"](https://www.xenonstack.com/insights/big-data-challenges)
[HYPERLINK "https://www.xenonstack.com/insights/big-data-challenges"](https://www.xenonstack.com/insights/big-data-challenges)
[w.xenonstack.com/insights/big-data-challenges](https://www.xenonstack.com/insights/big-data-challenges)
5. <https://academic.oup.com/nsr/article/1/2/293/1397586>
6. <https://www.itransition.com/blog/the-future-of-big-data>
7. https://en.wikipedia.org/wiki/Big_data
8. <https://www.upgrad.com/blog/big-data-applications-in-education/>
[HYPERLINK "https://www.upgrad.com/blog/big-data-applications-in-education/"](https://www.upgrad.com/blog/big-data-applications-in-education/)
[HYPERLINK "https://www.upgrad.com/blog/big-data-applications-in-education/"](https://www.upgrad.com/blog/big-data-applications-in-education/)
[om/blog/big-data-applications-in-education/](https://www.upgrad.com/blog/big-data-applications-in-education/)