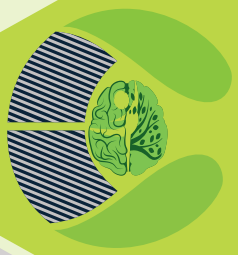
Volume - 2
Issue - 3
March 2020
e-ISSN 2582-5755

3



# Big Data: A Problem and Mother of many Inventions

– **Rohit Gupta**
Student, BCA, AMITY, NOIDA
https://orcid.org/0000-0002-7779-2670 rohitgupta10020@gmail.com

– **Akshansh Kumar**
Student, BCA, AMITY, NOIDA
https://orcid.org/0000-0001-8758-1132 akshanshkmr821@gmail.com

ARTICLE HISTORY
**Paper Nomenclature:** Scrutiny Tip (ST)
**Paper Code:** CYBNMV2N3MAR2020ST1
**Submission Online:** 03-Mar-2020
**Manuscript Acknowledged:** 04-Mar-2020
**Originality Check:** 10-Mar-2020
**Originality Test Ratio:** 01% (Turnitin)
**Peer Reviewers Comment:** 14-Mar-2020
**Blind Reviewers Remarks:** 15-Mar-2020
**Author Revert:** 20-Mar-2020
**Camera-Ready-Copy:** 21-Mar-2020
**Editorial Board Citation:** 23-Mar-2020
**Published Online First:** 25-Mar-2020

In this world,where we are known as humans do a lot of things daily, in every hour, in each minute and in each second with this each moment of time we perform several actions and we know each action of ours has data which was processed by our mind. Think If we didn't have this mind would it be possible to process this huge amount of data. So here comes the concept of Big data in our real life, the data of our actions which can't be processed using any part of the body without our mind is Big data for us. Similarly, in digital terms we produce a lot of data each day (in TBs) which is having a lot of important information which cannot be processed by the traditional computers. So, we call this Data "Big Data ''. In this we will see all about Big Data Technology.

## Keywords

- Legal
- Misinformation
- Disinformation
- Technology
- Information

## Introduction

In today's world, the technology is evolving everyday like telephone become mobile, Desktop data became cloud, car becameSmart/self-driving cars but one thing we can't deny is that with the increase in Technology the amount of data we produce daily is also increasing at a rapid rate(Fact: The data produced in 2019 is more than the data produced in the all history of digital world), so it is very difficult for any traditional systems to process or store this data. So, to solve this problem we discovered a technology called Big Data which consists of sub technologies like "Big data Analytics' ', "Data Science", "Hadoop' '.

When we talk about Big Data Technology there is a lot of confusion due to its name **"BIG"** data, people think that here we are talking about TBs, PBs, Ex..Etc., Yes the big data works on huge amount of data but we also call a data Big Data relatively to the system in which it is going to be processed. For example: We know our g-mail can only attach a file of size 25MB or less but if we have a doc of 100MB, g-mail can't attach it directly so this 100Mb data will be the Big Data in terms of g-mail. One more example is, take a airlines brand Indian Airlines, we know a single journey of plane produces a huge amount of data(data of passengers, staff, route details, security data and many more) which is in TBs so for a normal computer is not possible to process this data, this data will be called Big Data.

"**Big Data Refers** to the large amount of data that can't be stored or processed using the traditional form of technology within a given time frame."

According to edureka! "**Big Data** is the term for collection of data sets so large and complex that it becomes difficult to process using on-hand database system tools or traditional data processing applications".

## History

Digital data is produced since centuries about from 1600s, but size of this data is not a problem till 18th century, it can be stored and processed easily, but after 18th century the problem with data monitoring has started during 1880 data became a very huge problem for U.S. Census Bureau,

during the same year the amount of data collected will take about 8 years for processing according to their calculations, and 10 years for next year data. Fortunately, A young man Herman Hollerith who was working in this organization created a Machine called Hollerith Tabulating Machine, which is a solution to process this huge data and reduced the processing time of 10 years to months.

So, after this incident the huge data became a problem and called big data.

## Sources of DATA:

**Weather forecast:**

Daily weather forecasts need a lot of calculations which produce and need a lot of data.

**Networking and telecommunication:**

All the network and telecommunication devices data on daily bases which needs to be stored for long term.

**IoT:**

All the devices use the internet as IoT devices which produce a huge amount of data.

**Sensors:**

Sensors like Camera, mic produce the data.

**Social Media:**

Social sites like Facebook, tweeter receive TBs of data in minutes.

**Census, share market and E-commerce:**

The largest hum of importance and data related to economics is produced by these organizations of platforms which need a lot of processing for value.

## Characteristics of Big Data:

### 5 V's of big data:

**Volume:**

As we know the volume of data is increasing at a very rapid rate, according to a report there are chances that by 2020 the size of digital data will be about 44 ZBs, which is very huge.

**Variety:**

All this huge amount of data is coming in the number of forms like audio, video, text, json etc.

**Velocity:**

With time the speed of data production is also increased (in every minute we have 100000 + tweets, 695999+ status update on Facebook), we also need to increase our processing speed for this data.

**Value:**

Now the huge amount of data which is collected at a rapid speed have a lot of important data or data combinations, so we mine for useful content from data and apply our analytics for that data.

**Veracity:**

The value or data which is collected may there be some chances of uncertainty or any time of doubt (like incomplete data) regarding data may be present. It can also occur due to messy structure of data.

## Types of big data:

- **Structured Data:**

In this form, the data is processed and stored in proper format. Data is stored in RDBMS and is one of the examples of "structured data". Structured data make the processing easy and efficient.

- **Semi-Structured Data:**

In this form, the data does not have a proper structure for storing and processing, like tables in RDBMS, but it has some properties which makes it semi structured like tags etc.

- **Unstructured Data:**

In this foCrm, the data is not having any structure for processing or storing, and it can't be stored in RDBMS until it is changed in structured form.

## Issues:

- Huge data problem.
- Storage problem.
- Analysis becomes difficult.
- Processing is not easy for such a huge amount of data.
- Cost for Huge data storage and Processing is very high.

"OOPsss, we only talked about problems occur due to big Data,
Now let's Bring some positivity"

Volume - 2 Issue - 3 | March 2020 | e-ISSN 2582-5755

## Hadoop: Solution of all the problems:

Hadoop is an open source java-based framework which was developed by Doug Cutting and Mike Cafarella in 2002 in apache to solve the problems of storing, processing and analysing the big data. It is being used by many big organizations like Facebook, YouTube, twitter etc.

## Modules of Hadoop:

- **HDFS:**

Hadoop Distributed File System. In this module the files are broken into modules and stored into nodes over the distributed system architecture.

- **YARN**:

Yet Another ResourceNegotiator is the module which is used to manage the clusters of data and job scheduling.

- **Map Reduce:**

This module is a framework which is developed in java for parallel computation.

- **Hadoop Common:**

This module consists of libraries which will be used to start Hadoop and use Hadoop modules.

## Problems solved by Hadoop:

**I. Storage:**

Hadoop uses HDFS (Hadoop Distributed File System) to solve the problem regarding the huge amount of data.

**II. Processing:**

Map Reduce paradigm module of Hadoop solved our problem for processing the big data.

**III. Analyse:**

Hadoop solved our problem for analysing the data and finding the values from data.

**IV. Cost:**

Hadoop is open source free technology so cost is no more any problem.

## Advantages of Hadoop:

- **Scalable:**

The data over the Hadoop network is stored in clusters and it can be extended just by adding nodes in it.

- **Resilient to failure:**

HDFS is having a technique of property which makes the copies of data over the network so in case the node consisting a copy is down then the data can be accessed from another node.

- **Cost Effective:**

As it is an open source hardware and software so there is no cost to store or process the data.

- **Fast:**

HDFS has a distributed system for data storage which makes the data retrieval process very fast. The TBs of data can be processed in minutes.

## Big Data Applications:

- **Smarter Healthcare:**

Helps to store the data of patients, staff and disease to extract important information from them.

- **Network and Telecom:**

Big data helps the big organisations like telecom and network organization to store the huge amount of data and process it for meaningful information.

- **Retail:**

Big data helps the retailers to store the consumers information for beneficial use and understand their likes and dislikes using big data analytics.

- **Traffic control:**

Traffic congestion is one of the major problems in most of the big cities, proper use of data and sensors helped us to reduce this very efficiently.

- **Manufacturing:**

Big data analytics helps the organization to improve their work quality, increase efficiency and reduce time, money and defects in the manufacturing area.

- **Search Quality:**

The websites like amazon, YouTube or search engine like google store our search data to provide related or interest-based information in future.

## Challenges with Big Data:

- **Data Quality:**

Due to the huge amount of data the data gets messy and it becomes very difficult to extract useful information from it or process it using analytics.

- **Discovery:**

It becomes very difficult to find small data from the huge database. It's like finding a needle from a haystack, so we need very powerful algorithms to perform this task.

- **Storage:**

As the size of the data increases, it becomes more difficult and complex to store that data or to process that data.

- **Analytics:**

It becomes very difficult to analyse the data due to its huge size and lack of information about type of data because a big data may consist of many types of data at the same time.

- **Security:**

As the data size is very huge it is difficult to perform authentication and accessibility operations on it due to which there are chances of security breach.

- **Lack of Talent:**

Due to the lack of knowledge among the team working on big data creates a lot of problems and complexity.

## Future Concept:

In the coming future, we all know the Technology will grow more and more at rapid rate but one thing we can't deny is that we the increase in the technology the amount of data will also increase exponentially so it will become more and more difficult to process the data or to store the data so the need of big data concept will never obsolete with time rather we need to work harder in this area to make is more powerful and efficient to maintain balance according to future data.

Soon we will see many technologies will merge with this big data technology, like Artificial Intelligence to process and extract the useful information very efficiently, blockchain technology to make it secure over the network and reliability( there may be problem in merging with blockchain technology due to the huge size of data) for the users, Data analytics using different technologies will grow and merge with the big data concept.

If we talk about the jobs and employment in the area of Big Data, so currently a data scientist (which works on big data) earns about $95000-$250000 yearly. With time in future the demand of Data Scientists will increase and due to increase in demand their pay will always remain high.

## Some Ending words:

As we all know, "The problem is mother is the mother of invention". Similarly, Big Data came as a problem in our lives but with this problem we invented the techniques, tools and many more things which not only solved this problem but also changed the way we use to see our data or work on our data. We use to ignore many important and useful things due to lack of resources and knowledge to solve this problem of reading and processing such a huge amount of data but now we use the Big Data concept and use this data very beneficially and efficiently which not only benefited for organizations or works but also helped them to improve the consumer and personal experience, security and reliability.

**Rohit Gupta** is a final year graduate student. Pursuing Bachelor in Computer Application from Amity University (Noida). He is a scholarship holder for his academic excellence, he has been a consistently top scorer in his batch, he is having great interest in new and trending technologies. Blockchain, Big Data, Data Science are some of the new technology he recently learned about. He is also an avid reader, and a casual content writer. He has worked on many projects ranging from VB.NET to node.js and is keen to work on more. His father is his inspiration as well as his role model. Quote of his life is "I love my family, I will be a great son".

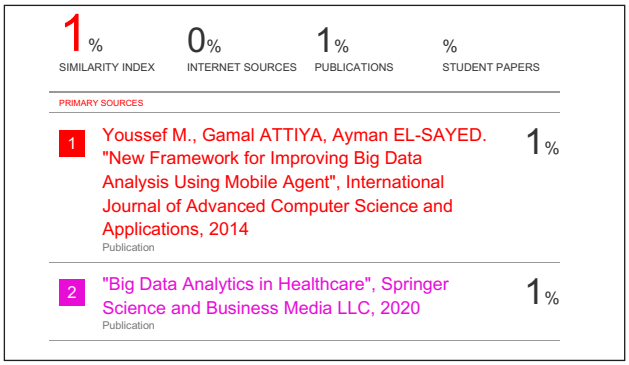
rohitgupta10020@gmail.com

**Akshansh Kumar** is a final year graduation Student, pursuing BCA from Amity University, he has a keen interest in technology and is eager to learn new things, he is particularly good in programming languages like C++ and Python, apart from academics he believes in utilizing classroom knowledge in real world, he is an also a member of Google bug hunter Hall of Fame. Special thanks to Dr. Rajbala Simon for trusting him and giving him the opportunity to write an article for the Cybernomics 2020 with the title "Big Data: A Problem and Mother of many Inventions.

akshanshkmr821@gmail.com

## Annexure I

| Submission Date | Submission Id | Word Count | Character Count |
|---|---|---|---|
| 10-Mar-2020 | 1314616292 | 2039 | 9681 |




*Note: The Cybernomics had used the turnitin plagiarism [https://www.turnitin.com/] tool to check the originality.*

## Reviewers Comment

**Reviewer's comment 1:** The concept is explained well in an understandable way. The author has stated some of the sub-technologies under Big Data and given brief knowledge about future usage of technology and employment. The paper can be weighted more if it is connected to current usage of big data technology and its applications

**Reviewer's comment 2:** The article is organized in a well structured manner by taking various examples that improves the clarity of the concept undertaken.

**Reviewer's comment 3:** The article covers various aspects such as sources, types, advantages, characteristics and applications of Big Data. The study has been supported by the very recent and updated data and facts & figures from various sources for the analysis in the study which is commendable.

## Editorial Excerpt

The article has 01% of plagiarism which is accepted percentage for publication the finding related to this manuscript Big data. It is true that Big Data and given brief knowledge about future usage of technology and employment It has been earmarked finalized for publication under the category of "**Scrutiny Tip** (ST)". This scrutiny tip can throw a light on a real time data and its gigantic nature. It can also emphasize on the data analytics too.

## Acknowledgement

I specially thanks to Dr. Rajbala Simon for trusting me and giving him the opportunity to write an article for the Cybernomics 2020 with the title "**Big Data: A Problem and Mother of many Inventions**". And I would also like to thank my friend Akshansh Kumar for always being there with me like my family and for also helping me in writing this article

## Disclaimer

All views expressed in this article are my own. References for relevant sections can be cited for more understanding. I, as an author have cited my own work along with content from other referenced sources in this particular article. All contents are provided in good faith and make no representation Or warranty of any kind regarding validity and completeness of the content.

## Citation

Rohit Gupta and Akshansh Kumar
"Big Data: A Problem and
Mother of many Inventions"
Volume-2, Issue-3, March 2020.
(www.cybernomics.in)

Frequency: Monthly, Published: 2020
**Conflict of Interest**: Author of a Paper had no conflict neither financially nor academically.

Scholastic Seed Inc.
www.scholasticseed.in
https://orcid.org/0000-0002-7952-096X