

# A Review on Big Data Analytics and Its Tools

Soni Kumari, Dona Chakraborty

Assistant Professor, Gandhi Institute for Education & Technology, Baniyatangi, Khordha, Odisha

## ARTICLE INFO

### Article History:

Accepted: 10 Dec 2023

Published: 26 dec 2023

### Publication Issue

Volume 10, Issue 6

November-December-2023

### Page Number

414-417

## ABSTRACT

Big data is a term which refers to the huge volume of data and this data is continuously growing with the time. This data is very large and complex which is difficult to store and process by the traditional software or applications. We can say that big data is an updated or upgraded version of traditional data. Big data deals with structured, unstructured data set and can be a mixture of both structured and unstructured data generally called as semi structured data which is enormous and complex to manage.

**Keywords :** Big Data Analytics, Volume, Veracity, Value, Variety, Velocity.

## I. INTRODUCTION

Big data is huge data set which is consisting of Various types of data. Big data is a combination of structured and unstructured data that can be used for decision making in the business environment. This data is collected from different sources in a certain time period.

## II. LITERATURE REVIEW

S. Singh and N. Singh focuses on the features and need of big data analytics in their paper 'Big Data analytics', 2012. T. Garg and S. Khullar describes it is complex to convert that big data in an information. This helps the organizations for future prediction in 'Big Data Analytics: Applications, Challenges & Future Directions', 2020.

According to the 'Challenges in Big Data Analytics Techniques: A Survey', 2019 by C. Komalavalli and C. Laroiya, the big data is rise as per the growth in the

field of health, social media and many more which results huge data. The information extracting from the huge data is very challenging. The techniques which are used previously to manage data is not capable to handle and manage the big data.

D. P. Acharjya and Kauser Ahmed in their paper 'A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools' 2016, focuses on the big data challenges and tools. challenges in big data analytics includes data storage and analysis, computational complexities, visualization of data, information security. big data processing tools includes the apache hadoop, mahout, spark, splunk.

## III.OBJECTIVES OF THE STUDY

- ❖ To study why big data analytics is used?
- ❖ To know the types and tools of big data analytics.
- ❖ To gain the knowledge about the features of big data analytics.

#### IV. BIG DATA ANALYTICS

BDA stands for Big Data Analytics. Data is a collection of raw facts and figures. Information is a set or collection of data that carries a logical meaning. BDA procedure includes collect, process, clean and analyzing huge data to help organizations to achieve their goals.

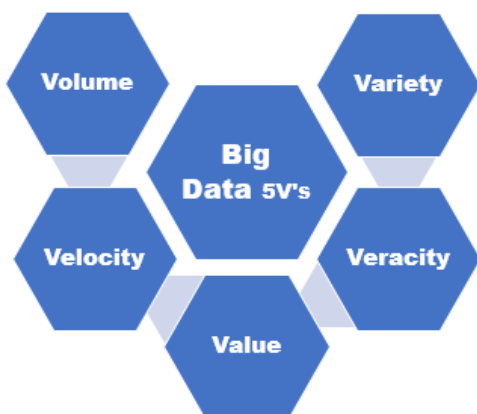
The big data analytics term includes **Big Data** and **Analytics**.



**Big data:** As the name suggest big data means the data in large size which is growing continuously within seconds.

**Analytics:** Analytics is the procedure of the systematic computational analysis of data. It helps to extract meaningful information from bulk data.

#### Big Data Characteristics in terms of V's



**Volume** – Volume of the data includes the size of the data. Big data is generally huge data which comes with more than terabyte.

**Velocity** – Velocity includes the speed of data.

**Value** - Value means the Importance of Data.

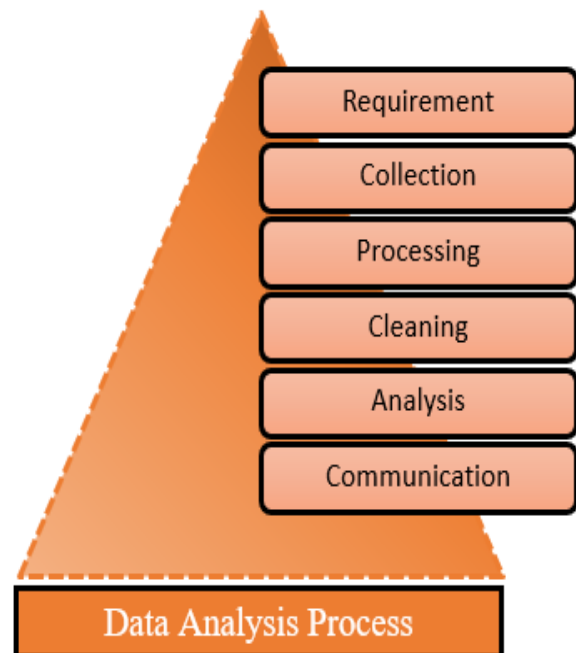
**Veracity** – Veracity includes the quality of data which is depended upon the methods of data collection and analyzation.

**Variety** - Variety means the nature of data which includes structured, semi-structured and

unstructured data. Data may group of structured, semi-structured and unstructured data.

#### Big Data Analysis

Data analysis process consists the following stages that are as follows:



**Data requirements specification:** The Data is required for any analysis. Data can consist of numeracy or categorized.

**Data gathering:** it also known as data collection. This is the process of collecting data which ensures the collected data is accurate. Data is collected from various sources such as interview, questionnaire etc.

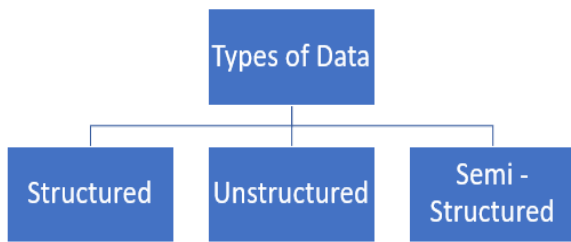
**Data processing:** in this process the collected data is organized and structured as per need to analysis tool.

**Data cleaning:** this includes the correction of the errors and reduce data delicacy in the processed data.

**Data analysis:** Now the cleaned, processed, organized data is ready for analysis. Various Data analysis techniques are used to analysis the data.

**Communication:** in this stage the results of data analysis are maintained in a format as per user requirements for decision making and further action.

## Types of data in Big Data



**Structured data:** This type of data is an organized data.

**Semi- Structured data:** It is semi organized data.

**Unstructured data:** This type of data is unorganized data.

### Big Data Analytics Tools:

**Hadoop** - It helps in storing and analyzing data

**MongoDB** - It is used on datasets that change frequently.

**Talend** - It is used for data integration and management

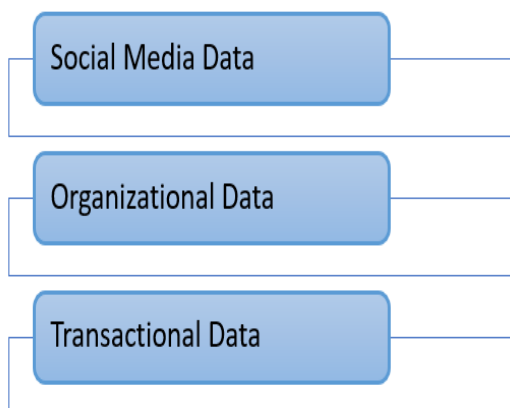
**Cassandra** - A distributed database used to handle chunks of data

**Spark** - For real-time processing and analyzing large amounts of data

**Kafka** - A distributed streaming platform that is used for fault-tolerant storage.

### Sources of Big Data

There are mainly three fields that generates Big data which are as following:



## Applications of Big Data Analytics:

**Ecommerce sectors** - E-commerce sectors are using Big Data analytics for Predicting customer trends and prices.

**Education sectors** - In this sector BDA is Used to develop new and update existing courses based on recent need.

**Healthcare sector** - Big Data analytics is used for predicting the patients' health issues and medical reports.

**Media and entertainment sector** – This sector is used to know the demand of shows, movies, songs etc.

**Telecommunications** – It is Used for improve customer experience.

### Merits of BDA

Merits of big data analytics are as follows:

- Less Time-consuming
- Cost saving
- Accurate result
- Better decision making
- Improve customer service
- Product development
- It makes easy to understand the market conditions.
- It improves public health with availability of record of patients.
- One platform carries unlimited information.

### Challenges of big data analytics

- With huge amounts of data, storage and processing become more complicated.
- Complex to maintain big data quality.
- Choosing big data analytics tools and platforms in the marketplace can be difficult.

## V. CONCLUSION

Big data is a set of various data types which is growing day by day. Since, data is generated by us for each purpose from a house to big organizations. This paper

involves the big data definition, its types, sources, merit/challenges of big data analytics, big data characteristics in term of V's and big data analysis, tools and applications with pictorial representation.

## VI. REFERENCES

- [1]. M. S. Mahmud, J. Z. Huang, S. Salloum, T. Z. Emara and K. Sadatdiyev, "A survey of data partitioning and sampling methods to support big data analysis," in *Big Data Mining and Analytics*, vol. 3, no. 2, pp. 85-101, June 2020, doi: 10.26599/BDMA.2019.9020015.
- [2]. G. Zhai, Y. Yang, H. Wang and S. Du, "Multi-attention fusion modeling for sentiment analysis of educational big data," in *Big Data Mining and Analytics*, vol. 3, no. 4, pp. 311-319, Dec. 2020, doi: 10.26599/BDMA.2020.9020024.
- [3]. S. Singh and N. Singh, "Big Data analytics," 2012 International Conference on Communication, Information & Computing Technology (ICCICT), 2012, pp. 1-4, doi: 10.1109/ICCICT.2012.6398180.
- [4]. T. Garg and S. Khullar, "Big Data Analytics: Applications, Challenges & Future Directions," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 923-928, doi: 10.1109/ICRITO48877.2020.9197797.
- [5]. C. Komalavalli and C. Laroia, "Challenges in Big Data Analytics Techniques: A Survey," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019, pp. 223-228, doi: 10.1109/CONFLUENCE.2019.8776932.
- [6]. R. Vatrapu, R. R. Mukkamala, A. Hussain and B. Flesch, "Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," in *IEEE Access*, vol. 4, pp. 2542-2571, 2016, doi: 10.1109/ACCESS.2016.2559584.
- [7]. M. Li, H. Wang and J. Li, "Mining conditional functional dependency rules on big data," in *Big Data Mining and Analytics*, vol. 3, no. 1, pp. 68-84, March 2020, doi:10.26599/BDMA.2019.902009
- [8]. Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao and R. C. Qiu, "Big Data Analytics in Mobile Cellular Networks," in *IEEE Access*, vol. 4, pp. 1985-1996, 2016, doi:10.1109/ACCESS.2016.2540520.
- [9]. De Mauro, A., Greco, M. and Grimaldi, M. (2016), "A formal definition of Big Data based on its essential features", *Library Review*, Vol. 65 No. 3, pp. 122-135. <https://doi.org/10.1108/LR-06-2015-0061>.
- [10]. W. M. Al-Rahmi et al., "Big Data Adoption and Knowledge Management Sharing: An Empirical Investigation on Their Adoption and Sustainability as a Purpose of Education," in *IEEE Access*, vol. 7, pp. 47245-47258, 2019, doi: 10.1109/ACCESS.2019.2906668.
- [11]. D. P. Acharjya and Kauser Ahmed, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools" in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.

### Cite this article as :

Soni Kumari, Dona Chakraborty, "A Review on Big Data Analytics and Its Tools", *International Journal of Scientific Research in Science and Technology (IJSRST)*, Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 6, pp. 414-417, November-December 2023. Available at doi : <https://doi.org/10.32628/IJSRST52310684>  
Journal URL : <https://ijsrst.com/IJSRST52310684>