

# Big Data Analytics: A Review

<sup>1</sup>Srikant Singh, <sup>2</sup>Akhilesh Kumar Shrivastava

<sup>1,2</sup>Dr. C. V. Raman University, Bilaspur, CG, India

## Abstract

In the information era, large amount of data have become available to the decision makers to decide the facts. Big data refers to datasets that are not only enormous, but also big in variety and velocity, which makes them difficult to work with traditional tools and techniques. Due to the rapid growth of this type of data, various results are needed to be followed and provided in order to check and extract value and data from the datasets. On the other hand, the executive systems need to be able to look for valuable insights from this kind of varied and frequently changing data, ranging from daily transactions to customer service and social network data. Such value can be supplied using big data analytics, which is the basic result of advanced data analytics techniques on big data. This paper focus on the analysis of the different data analytics methods and tools which can be applied on big data, as well as the data provided by the results and software that helps in analysis of big data in different decision domains.

## Keywords

Big Data Analytics, Data Analytics Methods

## I. Introduction

Now days the big data is one of the important area for researcher to analysis of data, because data is increasing every day in every organization in the world. The storage and management of data is one of the important issue for every organization. Let us suppose a world of data where we lose our data after use , then what will be the benefits of such data that was used previously. Every transaction performed and gives details of data but what will happen if the data is lose. So the organizations would take information and knowledge of lost data, perform data analyses, and provide new opportunities and advantages. The amount, variety, and frequent change of data need a new type of big data analytics, as well as various storage and data analysis methods. Such amounts of big data need to be analyzed, and resulting information should be extracted. The contribution of this paper is to provide an analysis of big data available on literature. Accordingly, some of the different big data tools, functions, applications and the technologies which can be applied are discussed.

## A. Big Data Analytics

The keyword “Big Data” has been applied with the datasets and data that grow so enormous that it is impossible to use them with the traditional Database Management System. The data sets whose size is beyond the ability of commonly used software and devices to capture, store, manage, as well as process the data within a tolerable elapsed time [12].

Big data amount and sizes are consistently enlarging, currently from a few dozen terabytes (TB) to a large amount petabytes (PB) of data in a one data set. Similarly, some of the problems related to big data involves sharing, storage, capture, search, analytics, and locating. Today, large number of enterprises are exploring large volumes of highly detailed data so as to search facts they didn't know before [17]. Hence, big data analytics is the place where we work on enormous amount of data.

## 1. Characteristics of Big Data

Big data is the data whose amount, use, diversity, and timeliness needs the use of new technical designs, data analytics, and devices in order to look into insights that shows new sources of business value. We have three distinguished features of big data: volume, variety, and velocity, or the three V's. The volume of the data is its amount, and how large it is. Velocity refers to the amount of rate with which data is converting, or how frequently it is created. Finally, variety gives the different formats and types of data, as well as the various kinds of uses and processes of analyzing the data .Data volume is the preliminary attribute of big data. Big data can be calculated by size in Terabytes or Petabytes, as well as the number of records, transactions, tables, or files. Similarly, the different things that makes big data really large is that it is coming from a different variety of sources than before, including data logs, click streams, and social networking websites. Using these data sources for data analytics means that simple structured data is now clubbed with unstructured data, such as text and language, and semi-structured data. There is also data, which is difficult to characterize since it comes from different peripheral devices. On the other hand, multi-dimensional data can be extracted from a data warehouse to add analytics context to big data. Thus, with the use of big data, variety is as big as volume. Furthermore, big data can be defined by its velocity or speed. This is basically the frequency of data generation or the frequency of data delivery. The topmost edge of big data is running or streaming data, which is collected in real-time from the websites [17].

## II. Big Data Analytics Tools and Methods

With the discovery of various technologies and the increased use of data flowing in and out of different organizations daily, so it has become a need for fast and more accurate and efficient ways of data analytics on data .In the Same era, proposed the Big – Data, Analytics, and Decisions (B-DAD) framework which involves the big data analytics tools and methods into the decision making process. The framework describes the various big data storage, processing, and management tools, analytics tools and methods to the different phases of the decision making process. So, the changes in big data analytics are focused in three main areas: data storage and architecture, big data and analytics processing, and, at last, data analyses which can be used for decision making.

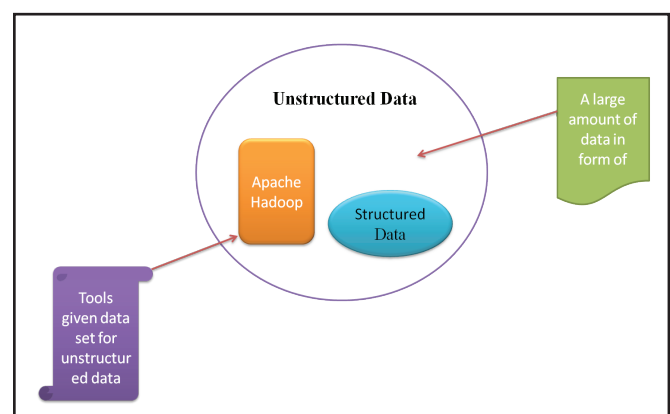


Fig. 1: The Big Data

The above figure shows the amount of data present around us. We have two concentric circles in the figure. The large circle shows the total amount of data while the small circle is showing the amount of structured data. Out of the hundred percent of the data we have 80% of unstructured data that need to be analyzed. The Apache Hadoop is a tool used for the analysis of the unstructured data.

### III. Big Data Analytics and Decision Making

From the decision maker's view, the importance of big data is in its ability to give the information and knowledge of value, upon which to base decisions. The executive decision making process has been an important and thoroughly covered topic in research throughout the years. Big data is becoming an increasingly important for decision making. Large amounts of clearly described data from different sources such as software and devices like mobile phones, loyalty cards, the web, and social media platforms provide the opportunity to give important benefits to institutions. This is possible only if the data is correctly analyzed to give valuable insights, allowing for decision making to capitalize upon the resulting data from the amount of data of historic and real-time data generated data through supply chains, production processes, customer behaviors, etc. [4].

Similarly, the development of the B-DAD framework which shows big data and techniques, into the decision making process. Such a framework is given to enlarge the quality of the decision making process in dealing with big data. The primary phase of the decision making process is intelligence, where the data which can be used to show problems is collected from internal and external data houses. In this phase, the houses of big data need to be identified, and the data needs to be collected from various sources, processed correctly, stored efficiently, and then migrated to the end user. Such type big data needs to be treated correctly, so that after the data sources and types of data required can be easily defined, the picked data is acquired and stored in any of the big data storage devices and management tools. After the big data is collected and stored, it is then organized, prepared, and processed, This is done using a high-speed network using ETL/ELT or big data analytics and processing tools.

The second phase in the decision making process is the design phase, where possible courses of designs are developed and analyzed through a clarification and concept on big data or a representative model of the problem.

Similarly, the next phase in the decision making process is the choice phase, where methods are used to calculate the impacts of the proposed solutions, or courses of action, from the design phase. Finally, the last phase in the decision making process is the implementation phase, where the proposed solution from the previous phase is implemented [8].

### IV. Review on Big Data Analysis

Yanpei C. et.al. [1], have described the degree of renewable energy integration depends on renewable source variability and intermittence, as well as scheduling slack in the data analytic workload. The technology used is data analytics traces. T.K. Das et.al. [2], have discussed public tweets of Twitter and the XML data is parsed to store in a NOSQL database like HBASE. Darrell M. West [3] has suggested the Bringing teachers into the "big data" discussion is crucial because they are the ones, along with parents and students, who will benefit from advances in research and analysis. He also included that the teachings should be spread to the persons in detailed way. Prashant Kumar et.al. [4] have discussed

on the data on edge networks. It also analyses the data present on the networks. Herodotos Herodotou et.al. [5] have introduced the study of starfish technology in hadoop that help in business and data analytics in business over big data. Xueyuan Su et.al. [6] have used Java 8 Stream and add a distributable Stream abstraction that supports federated query execution over an extensible set of distributed compute engines. Each query eventually results in the creation of a materialized result that is returned either as a local object or as an engine designed distributed Java Collection that can be saved and/or used as a source for future queries. He suggested that we can use the Big Data Concept to handle the data with java. [6] Domenico Talia [7] have discussed on the use of cloud technology in data analytics. Harrine Freeman [8] has introduced the integration of Big Data and Artificial intelligence. Marcos D. Assuncao et. al., [9] have represented the solutions for carrying out analytics and Big Data on Clouds. Identification of gaps in technology for Cloud-based analytics. Recommendations of research directions for Cloud-based analytics and Big Data. Nathaniel Schenker, [10] have suggested use of Big Data, and described its benefits and advantages. Big Data Working Group, has suggested the use of Big data and Network Security. It describes how we can access the Big Data to show the use of Networks. [11]. Emmanuel Letouzé, has described how to recognize that Big Data and real-time analytics are no modern panacea for age-old development challenges. In this paper we found the proper analysis of the data given for real time analysis. [12]. Hsinchun Chen et.al. [13] have discussed on business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations. Philip Sheridan Buffum et. al [14], have presented on the in-progress development of a new middle school curricular module for Big Data, situated as part of a new CS Principles-based middle school Curriculum. Amber Dennis [15] represented the use of Data Monetization Techniques which can be further advanced to find out better results. Jie ZHANG [16] has described goal of an effective course management system for assisting course managers to make informed decisions about what materials should be most appropriate to be presented to students (learners) and what learning strategies or methods should be used for the students. John Gantz et.al. [17] have deals with the expansion of data from over one decade. He shows the detailed description of data that is present around us in the form of logs and that can be analyzed. Dan Zhang, [18] has deals with the use of data over internet data. The amount of the vast amount of data that is present over the internet which describes the use of data to find out the structured information. Karthik Kambatla et. al. [19] have suggested Map Reduce-based analytics are particularly synergistic with the pay-as-you-go model of a cloud platform. Gali Halevi et.al. [20] has given the details on the basis of computer science data that is present around us for the analysis process. He shows how we can extract the data and how we can work on the data to find out better results. Julia Lane, [21] have presented the Big Data offers an enormous opportunity to advance the science of science policy. Making the links, so that science funders have new understanding of what is needed to foster science. Ray Harris [22] has present about the big data presents with many challenges, but at the same time presents many opportunities to influence how science grows and develops for the better, not least by adding data-driven science to hypothesis-driven science. Improvements in professional data management will result in better science. Daniel S. Katz. et.al. [23] has proposed a broad spectrum of the university

administration and faculty; it cannot be just one person's vision, though it may start with one person. The funding needs to be stable over a number of years; activities need to be sustained to be successful, and this needs to be clear to the community from the beginning. Kalev H. Leetaru, [24] has shown the report with the use of data in Data Analytics. Fabian M. Suchanek, et al. [25] have given an overview on state-of-the-art methods for the automatic construction of large knowledge bases and harnessing them for data and text analytics. It covers both big-data methods for building Knowledge bases and knowledge bases being assets for big-data applications. The tutorial also points out challenges and research opportunities.

## V. Conclusion

Big data is one of the important and challenging issues for every organization. This paper focused on the advanced topic of big data, which has recently gained lots of interest due to its opportunities and benefits. In the information era, we are currently living in, large varieties of high velocity data are being developed daily, and within them shows details and patterns of hidden knowledge which should be extracted and utilized. Hence, big data analytics can be used to enhance business change and enhance decision making, by applying advanced analytic techniques on big data, and revealing hidden insights and valuable knowledge.

## References

- [1] Yanpei C., David E. Culler, Randy Katz, "Integrating Renewable Energy Using Data Analytics Systems: Challenges and Opportunities", IEEE, 2011
- [2] Darrel M. West, "Big Data for Education: Data Mining, Data Analytics, and Web Dashboards", Governance studies at Brookings, 2012.
- [3] T.K. Das, P.Mohan Kumar, "BIG Data Analytics: A Framework for Unstructured Data", IJET, Vol. 5, No. 1, 2013.
- [4] Prashant Kumar, Khushboo Pandey, "Big Data and Distributed Data Mining: An Example of Future Networks", Vol. 1, Issue 2, 2013.
- [5] Herodotos Herodotou, Harold Lim, Gang Luo, Nedyalko Borisov, Liang Dong, "Starfish: A Self-tuning System for Big Data Analytics", CIDR, 2011.
- [6] Xueyuan Su, Garret Swart, Brian Goetz, Brian Oliver, Paul Sandoz, "Changing Engines in Midstream: A Java Stream Computational Model for Big Data Processing", Oracle Corporation First.Lastg@oracle.com
- [7] Domenico Talia, "Clouds for Scalable Big Data Analytics", Vol. No. 0018-9162/13, 2013.
- [8] Harrine Freeman, "Big Data and Artificial Intelligence: Advances in Natural Language Processing", Dataversity, 2016.
- [9] Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya, "Big Data computing and clouds: Trends and future directions", J. Parallel Distrib. Comput. 79–80 (2015) 3–15, (2014)
- [10] Nathaniel Schenker, "Big Data: Uses and Limitations", CDC, 2013.
- [11] Big Data Working Group, "Big Data Analytics for Security Intelligence", loud Security Alliance, 2013.
- [12] Emmanuel Letouzé, "Big Data for Development: Challenges & Opportunities", Global Pulse, 2012.
- [13] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, "Business intelligence and analytics: From big data to big impact", MIS Quarterly Vol. 36, No. 4, pp. 1165-1188, 2012.
- [14] Philip Sheridan Buffum, Allison G. Martinez-Arocho, Megan Hardy Frankosky, Fernando J. Rodriguez, Eric N. Wiebe, Kristy Elizabeth Boyer, "CS Principles Goes to Middle School: Learning How to Teach "Big Data", 2014.
- [15] Amber Dennis, Data Monetization: "An Unparalleled Opportunity for Relevance & Value", Dataversity, (2016)
- [16] Jie ZHANG, William Chandra TJHI, Bu Sung LEE, Kee Khoon LEE, Julita VASSILEVA & Chee Kit LOOI, "A Framework of User-Driven Data Analytics in the Cloud for Course Management", S. L. Wong et al. (Eds.), 2010.
- [17] John Gantz David Reinsel, "The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", IDC 1414\_v3, 2012.
- [18] Dan Zhang, "Internet of Things 2016 Trends: Leveraging Actionable Insights", Dataversity, 2016
- [19] Karthik Kambatla, Abhinav Pathak, Himabindu Pucha, "Towards Optimizing Hadoop Provisioning in the Cloud", IBM Research Almaden, 2014.
- [20] Gali Halevi, Dr. Henk Moed, "The Evolution of Big Data as a Research and Scientific Topic", Research Trends, Issue 30, 2012
- [21] Julia Lane, "Big Data Science Metrics and the black box of Science Policy", Research Trends, Issue 30, 2012.
- [22] Professor Ray Harris, "ICSU and the Challenges of Big Data in Science", Research Trends, Issue 30, 2012.
- [23] Daniel S. Katz, Gabrielle Allen, "Computational & Data Science, Infrastructure, & Interdisciplinary Research on University Campuses: Experiences and Lessons from the Center for Computation & Technology", Research Trends, Issue 30, 2012.
- [24] Kalev H. Leetaru, "A Big Data Approach to the Humanities, Arts, and Social Sciences", Research Trends, Issue 30, 2012.
- [25] Fabian M. Suchanek, Gerhard Weikum, "Knowledge Bases in the Age of Big Data Analytics", VLDB Endowment, Vol. 7, No. 13, 2014.
- [26] Radha Shankermani, "Big Data Analytics", wiley text book series, 2016.
- [27] Arun c. Murthy, "Apache Hadoop YARN", Pearson, 2015.
- [28] Radha Shankermani, "Big Data Analytics", Wiley text book series, 2016.
- [29] Arun C. Murthy, "Apache Hadoop YARN", Pearson, 2015.



Srikant Singh received his MCA degree from India Gandhi National Open University, New Delhi, India in 2010, and doing M.Phil degree in Computer Science from C.V. Raman University, Bilaspur, Chhattisgarh, India.