

Big data analytics with applications: A Literature Survey

Karan Gautam*, Krishna Kumar Gupta

Department of Computer Science and Engineering & Department of Mathematics, Shri Ram Murti Smarak College of Engineering and Technology, Bareilly, Uttar Pradesh, India

*Corresponding author Email: karangautam1417@gmail.com

Received: 27 November, 2021; Revised: 15 July, 2022; Accepted: 28 November, 2022; Published: 24 May, 2023

DOI: 10.29218/srmsmaths.v6i1.03

Abstract

Keywords:

Big data, big data analytics, analysis, data mining.

In this era of information, the data is the most precious thing in the world. But now this data becomes bigger day by day, so a new term is coined that is big data. Big data refers to various datasets which are very high in volume, variety and velocity. The traditional tools are not able to analyse this huge amount of data. Moreover, the decision makers need some insights from this big data to increase their performance. So new solutions are required that can analyse this big data and can get conclusions and insights from it. This process is called big data analytics. In this paper, an analysis of available tools which can be applied on big data and application areas of big data analytics.

1. Introduction

Envision a world without data storage; a place where every information about an organization or a person, every transaction performed, or every material which can be saved is lost directly after use. Associations would consequently lose the capacity to separate significant data and information, perform detailed analyses, as well as provide new opportunities and advantages. Anything starting from customer names and addresses, to goods available, to purchases made at market, to employees hired, etc. has become essential part of everyday life. Data is the structure block whereupon any association flourishes. Every second, a huge data is being created and needs to be stored and analysed in order to extract value. Furthermore, data has become so cheaper to store, so organizations need to get as much value as possible from this huge amount of stored data. The size, variety, and rapid change of such data require a new type of analytics, as well as different storage methods.

This review aims to provide a brief overview of big data, big data analytics, some of the various big data tools and the benefits of big data analytics across various sectors will be explored.

2. Survey Sources

We analysed various research papers for this survey. The papers ranging from 2008 to 2022 with most literature collected from the range of years 2014-2022. Furthermore we searched various resources and conference papers for literature.

3. Characteristics of Big Data

The Big data is data whose scale, distribution, diversity require the use of latest technical architectures, analytics, and tools so one can permit insights that that unlock new sources of business value. Three main features of big data are volume, variety, and velocity, or the three V's. The volume of the data is its size, and how large it is. Velocity refers to the rate on which data changes or how often the data is generated. Variety includes different data formats and types, and different uses and methods of data analysis.[20]

Data volume is the primary feature of big data. Big data can be measured in TBs or PBs, as well as the number of records, transactions, tables, files or assets.

Also, one of the reasons big data is really growing is that it comes from more diverse sources than ever before, including magazines, visitor streams, and social media. Using these sources for analysis means now that plain structured data is combined with unstructured data such as text or human language, and semi-structured data such as Extensible Markup Language (XML) and Rich Site Summary (RSS) feeds. Means, data in audio and video formats is difficult to classify. As such, the diversity of big data is as vast as its quantity.

Furthermore, big records may be described with the aid of its velocity or pace. This is the frequency of data generation or the frequency of data transfer. The main component of big data is streaming data that is accumulated in actual-time from the web sites.[21]

A few researchers and groups have mentioned the addition of a fourth v, or veracity. Veracity means quality of data. This characterizes big records quality as Good, horrific, or undefined due to information inconsistency, incompleteness, ambiguity, latency, deception, and approximations.[22]

Few researchers mention two more V's As-Variability and Volume. All 6V's can be represented in Figure 1.

4. Big Data Analytics: Definition

Big data analysis refers to a strategy to analyse a large amount of data or big data. This big data is gathered from a variety of sources, including social media, video, digital images, sensors, and sales transaction records.

5. Tools for Big Data Analytics:

As with the analytical techniques, there are several software products and available technologies to facilitate big data analytics. Some of the most common will be discussed here.

EDWs- Enterprise information warehouses are databases used in Fact are analysis. For lots Agencies which might be trying to start dealing with big records the massive question is can the modern-day or planned organization Records warehouse (EDW) deal with huge statistics and superior analytics without affecting performance of other Workloads for reporting and online analytic processing? [21] A few institutions manage their analytic records inside the EDW Itself while others use a separate platform, which allows lowering some of the stress at the server as a result of Coping with your data on the EDW. [14]

Visualization Products: one of the problems with huge Records analytics is finding methods to visually represent Effects. Many new visualization merchandise intention to fill this Want, devising strategies for representing statistics points numbering up into the tens of millions. [21] lists this Field as one of those having the most capacity, announcing it Is “poised for competitive adoption.” beyond easy Representation and visualization can also help within the Records seek. Hansen, Johnson, pascucci, and silva Wrote a piece of writing blanketed in [26] discussing Visualization in information-intensive technological in which they Provide an explanation for that visualization products allow us to evaluate Models and datasets and “permits quantitative and Qualitative choice-making.” their article stresses Scalability in visualization technologies and their capability to music provenance in real-time.[14]

Hadoop and MapReduce: MapReduce is a programming Version used to deal with numerous facts simultaneously and Hadoop is one of the extra popular open-source Implementations of that model. Szalay and Blakeley Wrote an article in [26] wherein they speak this specific Software program. They give an explanation for that the principles MapReduce Makes use of are similar to the “allotted grouping and Aggregation skills which have existed in parallel Relational database structures for some time” but they’re Capable of scale very well to deal with for fantastically large information units. They move on to give an explanation for that Hadoop Implements a “records-crawling strategy over hugely Scaled-out, proportion-not anything information partitions” wherein numerous Nodes within the gadget are able to carry out distinctive components of A question on specific components of the records simultaneously. This works thoroughly for large facts, but for smaller Initiatives they remind their readers that this product isn’t as powerful “when an amazing index might provide higher Overall performance via orders of magnitudes.” These are some common tools that are used for big data analytics.

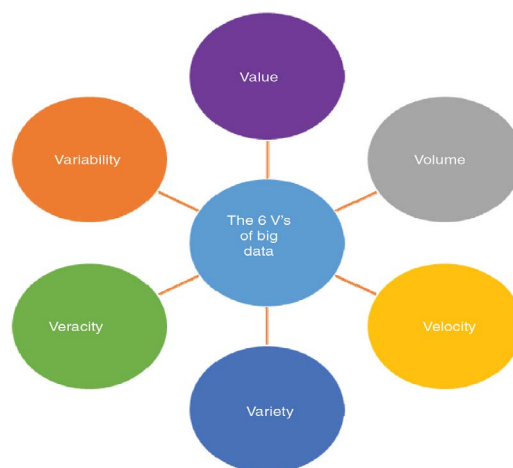


Figure 1: The 6 Vs of big data.

6. Applications of Big Data Analytics

This section of the survey is all about the benefits on Big Data analytics in Different fields. Some fields that are using Big Data Analytics are as follows:

In Healthcare: Applications of Big Data analytics can improve the patient based totally carrier, to come across spreading illnesses earlier, Generate new insights into disorder mechanisms, display the first-class of the clinical and healthcare institutions In addition to offer higher remedy methods.[16-18] Records mining strategies employed on EHRs, web and social media facts enable identifying the optimal Realistic suggestions inside the hospitals, identifying the association regulations in the EHRs [19] and revealing the disorder Monitoring and health based trends. Furthermore, integration and evaluation of the data with extraordinary nature, such as social and clinical, can cause new knowledge and intelligence, exploring new speculation, figuring out Hidden styles.[15] In recent times, smart telephones are remarkable structures to supply private messages to sufferers to involve them in Behavioural adjustments to enhance their wellbeing and fitness conditions. The mobile telephone messages can substitute Delivering of scientific and motivational advices to the patients.[15][3]

Transportation: Big data is the subject of research in the field of Intelligent Transportation Systems (ITS), which can be seen in many projects around the world. Intelligent transportation systems generate a lot of data. Results Big data will have a significant impact on the development and application of intelligent transportation systems, which will make ITS safer, more efficient and more profitable. Big data analytics research at ITS is a thriving field.[5]

Public Sector: [24] asserts that the public area doesn't Stand to advantage as plenty from large facts analytics because it Does now not hold as an awful lot facts as different sectors in order that they do now not Have as an awful lot on which to perform analytics. Anyone that Has crammed out a tax return or applied for a passport might also Disagree with that assessment, but [24] Also Claims that the price of huge data analytics to Europe's public sector may be as plenty as 250 billion euros. No matter the quantity of available records all of us has the same opinion That the authorities can gain by means of utilising the Possibilities available in large records analytics. Authorities has already commenced to tap the beneficial Useful resource of large data. [23] listing Campaign advertising, voter-mobilization, coverage Dialogue, donations, and extra as regions in which the general public Region has already all started the usage of internet 2.0 and the use of web Analytics and social media analytics to similarly their causes. They upload that most of the work in this place is executed via the governments themselves; little of it is being accomplished in Academia.[14]

Higher Education: Education is utilising available technologies more and Extra each year. [25] tells us that one third of the better education populace in 2010 enrolled in at least One completely online course and plenty of extra enrolled in blended Courses (a combination of on line and face to face teaching). Due to the fact Teachers and students are trending toward accelerated use of Technology for coaching there are a number of recorded data Generated. In truth, [27] declare that big data and analytics are going to be the most important factors in What's going to shape the future of higher training. Higher education will be the source of Big Data and an application of Big Data.

7. Conclusion

In this survey, we examined the basics of big data, its Characteristics and big data Analytics with its benefits across various sectors. We are living in the era of technology and big data is among the top technologies of this time. So, in this survey we studied and researched the various contents available on this topic. We conclude this survey with a thought that big data will be the field of research for most of the researchers in next 25-30 years and its area will be expanded exponentially.

References

1. Zhou H, Ren H, Royer P, Hou H, Yu X-Y. Big Data Analytics for Long-Term Meteorological Observations at Hanford Site. *Atmosphere*. 2022; 13(1):136. <https://doi.org/10.3390/atmos13010136>
2. Norjihan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, Ejaz Ahmed, Big Social Media Data Analytics: A Survey, *Computers in Human Behavior* (2018), <https://doi.org/10.1016/j.chb.2018.08.039>
3. Ristevski, Blagoj and Chen, Ming. "Big Data Analytics in Medicine and Healthcare" *Journal of Integrative Bioinformatics*, vol. 15, no. 3, 2018, pp. 20170030. <https://doi.org/10.1515/jib-2017-0030>
4. Choi Tsan-Ming., et al., Big Data Analytics in Operations Management. *Production and Operations Management* (2017). <https://doi.org/doi:10.1111/poms.12838>
5. Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2018). Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 383-398.
6. Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. *Computer Networks*, 129, 459-471.
7. Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: promise and challenges.

- Nature Reviews Cardiology, 13(6), 350-359.
8. Peter Baumann, Paolo Mazzetti, Joachim Ungar, Roberto Barbera, Damiano Barboni, Alan Beccati, Lorenzo Bigagli, Enrico Boldrini, Riccardo Bruno, Antonio Calanducci, Piero Campalani, Oliver Clements, Alex Dumitru, Mike Grant, Pasquale Herzig, George Kakaletris, John Laxton, Panagiota Koltsida, Kinga Lipskoch, Alireza Rezaei Mahdiraji, Simone Mantovani, Vlad Merticariu, Antonio Messina, Dimitar Misev, Stefano Natali, Stefano Nativi, Jelmer Oosthoek, Marco Pappalardo, James Passmore, Angelo Pio Rossi, Francesco Rundo, Marcus Sen, Vittorio Sorbera, Don Sullivan, Mario Torrisi, Leonardo Trovato, Maria Grazia Veratelli & Sebastian Wagner (2016) Big Data Analytics for Earth Sciences: the EarthServer approach, *International Journal of Digital Earth*, 9:1, 3-29. DOI: 10.1080/17538947.2014.1003106
 9. Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of big data*, 2(1), 1-21.
 10. Elgendy, N., & Elragal, A. (2014, July). Big data analytics: a literature review paper. In *Industrial conference on data mining* (pp. 214-227). Springer, cham.
 11. Cardenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy*, 11(6), 74-76.
 12. Chan, J. O. (2013). An architecture for big data analytics. *Communications of the IIMA*, 13(2), 1.
 13. Ularu, E. G., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on big data and big data analytics. *Database Systems Journal*, 3(4), 3-14.
 14. Maltby, D. (2011, October). Big data analytics. In *74th Annual Meeting of the Association for Information Science and Technology (ASIST)* (pp. 1-6).
 15. Archena J, Anita EM. A survey of big data analytics in healthcare and government. *Procedia Comput Sci* 2015; 50:408–13.
 16. Agarwal M, Adhil M, Talukder AK. Multi-omics multi-scale big data analytics for cancer genomics. In: *International Conference on Big Data Analytics*. Cham, Switzerland: Springer International Publishing; 2015: 228–43.
 17. He KY, Ge D, He MM. Big data analytics for genomic medicine. *Int J Mol Sci* 2017; 18:412.
 18. Tan SL, Gao G, Koch S. Big data and analytics in healthcare. *Methods Inf Med* 2015; 54:546–7.
 19. Dinov ID, Heavner B, Tang M, Glusman G, Chard K, Darcy M, et al. Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS One* 2016;11:e0157077.
 20. EMC: Data Science and Big Data Analytics. In: *EMC Education Services*, pp. 1–508 (2012).
 21. Russom, P.: Big Data Analytics. In: *TDWI Best Practices Report*, pp. 1–40 (2011).
 22. TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: *TechAmerica Reports*, pp. 1–40 (2012).
 23. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
 24. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity (pp. 1–143). The McKinsey Global Institute.
 25. Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9–20.
 26. Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research.
 27. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–32.