International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-2, March 2022
https://doi.org/10.55524/ijircst.2022.10.2.37
Article ID IRPV1058, Pages 185-188
www.ijircst.org

A Review Paper on Big Data Analytics

Ankur Gupta

Assistant Professor, Department of Computer Science & Engineering, RIMT University, Mandi Gobindgarh, Punjab, India

Correspondence should be addressed to Ankur Gupta; ankurgupta@rimt.ac.in

ABSTRACT- The information revolution gave decision makers huge quantities of data available. Big data refers to datasets that are not just large, but also significant in diversity and speed and that make standard tools and approaches hard to handle. Because such data is fast growing, solutions should be explored and supplied so that these datasets are able to handle and extract value and information. In moreover, policymakers need to be able to get useful understanding into such diverse and fast changing data, from everyday transactions to contacts with customers and social network information. The Big Data Analytics, the deployment of advanced Big Data Analytics methodologies can deliver this value. The purpose of the article is to evaluate several analytical methods and means that may be used for Big Data and the potential that Big Data Analytics offers for the use of diverse decision-making areas.

KEYWORDS- Analytics, Big data, Data mining, Datasets, Network.

I. INTRODUCTION

Envision a world lacking data storage; a situation whereby each detail about an individual or group, every transaction made or every documented element is gone right after usage. Organizations might lose the capacity to get important information and understanding, do comprehensive analyses, and also provide new possibilities and benefits. Anything from client names and addresses, availability items, transactions made, hiring staff, etc. is becoming crucial to continuity on even a daily basis. Data is the fundamental basis that every company prospers [1-3].

Consider now the depth of the specifics and the volume of data and information that technological and Internet advances have supplied. With increased storage and data collecting technologies, enormous volumes of data are easily available. More and more data are produced every second, and in order for value extraction it has to be stored and evaluated. In addition, data are less expensive to keep, allowing companies to take full use of the large volumes of data collected.

The input of this study is to analyses existing Big Data analytics publications. Some of the many large data tools, methodologies and technologies available will thus be examined, and their uses and potential in a number of decision fields will be described.

A. *Big Data Analytics*

Nowadays, the phrase "Big Data" has referred to datasets which grow to be difficult to use standard techniques for database administration. Those are data sets which have the capacity to gather, save, store and analyze data over an acceptable span of time beyond the usual software and storage devices.

Big data sizes increase steadily from a few dozen terabytes to several petabytes of data in a single piece of data. As a result, the acquisition, storage, discovery, sharing, processing and analysis of large data are some of the problems.

In this part, we begin by examining the features and relevance of big data. Business advantage, of course, can usually be extracted from the analysis of huge, more complicated data sets which need real-time or near-real time; nevertheless, it requires new data structures, analysis methodologies and tools. The next section will therefore explain the strategies and methodologies of big data analytics, especially from big data storage and administration through big data analysis processing.

1) Characteristics of Big Data

Big data means data that required new technical architectures, analysis and instruments to provide insights into a new source of corporate value to be utilized in terms of size, spread, variety and time. Big data is distinguished by three major features: volume, variety and speed, or three Vs. The volume and quantity of the data are its size. Speed relates to the rate during which or how much the data is changed. Finally, the diversity covers the many data formats and categories and the various purposes and forms of data analysis [4]. Big data, as also the quantity of records, operations, tables or files, may be measured by size in TBs and PBs. Moreover, it comes from a broader range of sources, including logs, clip streams and social media, since one of the events that directly create big data [5-8].

2) Big Data Analytics Tools and Methods

Conventional data management and analysis methodologies and technologies can no longer readily examine these data sets. Accordingly, new tools and methodologies for big data analysis and the systems needed to store and manage that data are needed. The advent of big data hence has an impact on all aspects, from data gathering and processing to final derived choices.

The initial attempts the many tools, analytical tools and methodologies, visualization and assessment tools for big data storage, management and processing into the various phases of the decision-making processes. Throughout this part, each area will be addressed further.

B. Big Data Storage and Management

The conventional techniques to store and retrieve structured data include relationship databases, data marts, and data storages. The data can be uploaded to the storage of OSs with extraction, transformation, loading or extraction, loading or transforming, tools for extracting data from external sources and transforming the data in order to suit operational requirements. Data is therefore cleansed, processed and catalogued prior pattern discovery and analytical capabilities are available online [9]. The big data environment, meanwhile, requires the capability of Magnetic, Agile and profound analytics that differ from the feature of a typical data warehouse. First and foremost, typical EDW methods prohibit the inclusion of new sources of data until they have been cleaned up. As large data settings nowadays have to be magnetic, so that all data sources are attracted irrespective of the information quality.

C. Big Data Analytic Processing

The analytical processing happens after large data storage. Thus according to four major data processing needs exist. Initially, rapid data loading is necessary. Because traffic on the disc and network conflicts with queries during data loading, data loading time must be reduced. Next, quick query processing is necessary. Many questions are important to answer times to fulfil the needs of large workloads and in-time requests. In furthermore, extremely efficient use of storage space is the third need for Big Data Processing. Because the rapid increase of user activity might need scalable storage capacities and computer power, there must be limited disc space for the proper management of data storage during processing and problems with storing the data in order to optimize the space usage. The fourth condition, in conclusion, is that the workload patterns would be very dynamic. Big data sets should be highly adapted and not specified for unanticipated dynamics in data processing, since they are examined by different platforms and applications for different goals and for different methods.

MapReduce is the initial phase of mapping data input to a series of value pairs as output. As a result, the "Map" function splits huge computer jobs into smaller work and allocates them to the relevant key/value pairs. Unstructured data, for example, text, may therefore be mapped to a structured key-value pair in which the key, for instance, could have been the word in the manuscript. This outcome is then the "Reduce" feature input. Reduce then collects and combines the output to produce the final result of the computing work by merging all values with the same key value [10].

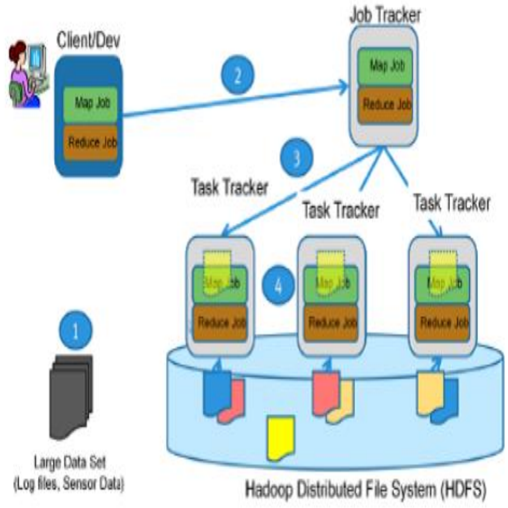

Figure 1: illustrate the diagram shows how the MapReduce nodes and how HDFS work together

Figure 1 demonstrates the collaboration between MapReduce nodes and HDFS. In step 1, a huge dataset is available comprising log files, sensor data or whatever. The HDFS contains replicas of the data shown on the Data Nodes by blue, yellow, beige and rose symbols. In step 2, the customer defines and runs a task on the map and reduces work in a given data set. The Job Tracker subsequently assigns jobs in step 3 out across task trackers. The task tracker operates the mapper and creates output, then maintained in the system of HDFS files. In step 4, the reduction work runs on the mapped data to generate the result.

D. Big Data Analytics and Decision Making

The importance of big data from the decision-point makers of view is its capacity to give information as well as to recognize the cost from which to base judgments. The management decision-making process seems to have been a major issue of research all throughout period. Big information for policy-makers is becoming an essential asset. Widespread amounts of very specific information from many sources, also including scanners, mobile telephoning, loyalty cards, the Internet and social media channels, provide companies important benefits [11]. This is only feasible if the data is correctly analyzed and useful insights are revealed, allowing decision makers to exploit the opportunities coming from the richness of historical and real time data created by supply chains, industrial processes, consumption habits, etc.

First step of decision making processes is the intelligence phase during which data are collected through internal and external data sources which may have been utilized to identify challenges and opportunities. During this period, large data sources should be identified, data acquired, processed, stored and transferred to the end-user from diverse sources [12]. Following the identification of the origins and kinds of information required for evaluation, the selected information is gathered, saved, and utilized in one of the already stated distributed databases and administration software. Big data is organized, produced, and analyzed once it has been gathered and kept.

When alternative action sequences are generated and assessed through a conceptualization or a representational model of the challenge, the next step in decision-making is the preliminary design. Each phase is subdivided into three phases, model planning, data analysis and analysis. Therefore, a model is picked, planned, and then implemented for data analytics, including those previously described, and assessed.

1) Risk Management and Fraud detection

Big data analytics in the field of risk management may be used by industries also including investments and retail banking, as well as reinsurance. Big data analytics can assist to identify investments by comparing the chance of returns versus likelihood of damages since risk assessment and carrying is a crucial component for the financial services industry. The comprehensive and dynamic risk assessment may also be analyzed for internally and externally big data.

Big data analytics can also be used to reduce fraud, in particular in the public, banking and insurance industries. Although analytics have already been frequently employed in the field of automated detection of fraud, companies and industries want to exploit the potential of big data to optimize their systems [13].

2) Improvement and Quality Management

In order to improve profitability and decrease expenses, Big Data may especially be utilized for manufacturing, energy, utilities and telecommunications industries for quality management by increasing the quality of goods and services offered. For instance, predictive analytics may be utilized in the production process in order to decrease the variability in performance and prevent quality problems with early warning warnings. This can improve scrap rates and reduce the amount of time for marketing as any manufacturing process interruptions can save a considerable amount of money before they arise [14].

Big data may also be employed to better understand variations in location, frequency and weather and climatic intensity. Citizens and businesses, also including farmers and tourists and transportation firms, can profit from this. Furthermore, weather related natural disasters may be foreseen using new sensors and analytic tools for long term climate models and closer weather predictions and preventative or adaptive actions can be implemented in advance.

II. LITERATURE REVIEW

S. Kumari studied the phrase "Big Data" refers to novel approaches and technology for capturing, storing, delivering, managing and evaluating high-velocity and various structures of petabytes or bigger datasets. Big data might be formatted, unstructured or semi-structured, and standard data management approaches cannot thus be used. Parallelism is applied for the economic and efficient processing of very huge quantities of data. Big data is a data which scale, diversity and complexity demand the development and extraction of value and information from it of new architecture, methods, algorithms and insights. Hadoop is the main platform for big data processing and addresses the difficulty of building information relevant for analysis. Hadoop is a software-project for open source processing of large-scale data collections across multiple server clusters. It is meant to scale thousands of machines from one server with a high tolerance of faults [15].

Tessa Van Der Valk et al. studied the application of sociological networks in policies design and evaluation is fairly limited. The goal of this paper is to highlight study areas in innovations that may benefits through the utilization of SNA, as well as to investigate planning and organizational ramifications derived from the use of SNA in the academic community. Three important study topics have been identified: cooperation networks; network infrastructure; and technology networks. The managerial and regulatory consequences and possible guidance are addressed [16].

Sitaram Asur studied the social media have grown omnipresent and crucial for social networking and sharing of content in recent years. Yet the material produced on these blogs remains mostly unexplored. On this article, we show how material in media platforms may be utilized to predict real results. We are using Twitter.com buzz in particular to estimate film box office income. We show that a basic model built on a rate of tweeting can be superior to market-based predictions for certain subjects. We also show how Twitter emotions could be further used to increase social media forecasting power [17].

III. DISCUSSION

The literature has thus been examined to give analysis of and significance to decision making in the ideas of Big Data Analytics that are under study. Big data and its properties and significance were therefore examined. In addition, several of the tools and methodologies for large data analysis have been studied. Big data storage and administration were therefore specified, as were the processing of Big Data Analytics. Some of the other sophisticated approaches of data analytics have also been examined further. This analysis has thus offered people and companies with examples of the different tools, techniques and technology that may be used for the use of big data. This offers consumers a sense of the technology they need and developers and idea of what they can do to deliver better solutions for Big Data Analytics for decision-making. So the help to decision making from Big Data Analytics has been shown.

IV. CONCLUSION

In this study, we studied the new Big Data subject that has attracted a great deal of attention due to its exceptional potential and advantages. In this information era, large kinds of high speed data are created every day and hidden knowledge patterns and inherent features are laid inside the high-speed data. Large data analysis may thus be used with improved analyses of big data and uncover hidden insights and important knowledge in the use of sophisticated analytical technologies to make business changes easier. Furthermore, if implemented appropriately, every new technology may provide numerous potential advantages and improvements, let alone big data, who's, if properly addressed, is an important sector with a promising future. It includes sufficient storage, administration, integration, federation, purification, processing, analysis, etc. Big data multiplies these challenges enormously with all the difficulties in traditional data management because of greater volumes, speeds and variety of data and sources. Continued studies might thus concentrate on developing a Big Data Management roadmap or framework that can incorporate prior problems. In this era of data overflow we think that big data analytics are of tremendous importance and can bring unanticipated insights and advantages for policymakers in numerous sectors. Big data analytics may offer the framework for the study, technical and humanitarian advances if properly exploited and implemented.

REFERENCES

[1] Siddiqui MHF, Kumar R. Interpreting the Nature of Rainfall with AI and Big Data Models. In: Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020. 2020.

[2] Sehgal D, Agarwal AK. Real-time sentiment analysis of big data applications using twitter data with Hadoop framework. In: Advances in Intelligent Systems and Computing. 2018.

[3] Sehgal D, Agarwal AK. Sentiment analysis of big data applications using Twitter Data with the help of HADOOP framework. In: Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016. 2017.

[4] TechAmerica Foundation's Federal Big Data Commission. Demystifying Big Data: A Practical Guide To Transforming The Business of Government Listing of Leadership and Commissioners Global Executive Vice President and General Manager. UNICOM Gov. 2012;1–40.

[5] Jain N, Awasthi Y. WSN-AI based Cloud computing architectures for energy efficient climate smart agriculture with big data analysis. Int J Adv Trends Comput Sci Eng. 2019;

[6] Gupta P, Tyagi N. An approach towards big data - A review. In: International Conference on Computing, Communication and Automation, ICCCA 2015. 2015.

[7] Al-Bahri B, Noronha H, Pandey J, Singh AV, Rana A. Evaluate the Role of Big Data in Enhancing Strategic Decision Making for E-governance in E-Oman Portal. In: ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions). 2020.

[8] Gupta D, Rana A, Tyagi S. A novel representative dataset generation approach for big data using hybrid Cuckoo search. Int J Adv Soft Comput its Appl. 2018;

[9] Cuzzocrea A, Song IY, Davis KC. Analytics over large-scale multidimensional data: The big data revolution! Int Conf Inf Knowl Manag Proc. 2011;101–3.

[10] Herodotou H, Lim H, Luo G, Borisov N, Dong L, Cetin FB, et al. Starfish: A self-tuning system for big data analytics. CIDR 2011 - 5th Bienn Conf Innov Data Syst Res Conf Proc. 2011;261–72.

[11] Elgendy N, Elragal A. Big Data Analytics in Support of the Decision Making Process. Procedia Comput Sci. 2016;100:1071–84.

[12] Tiwari S, Wee HM, Daryanto Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Comput Ind Eng. 2018;

[13] SAS. The Value of Big Data and the Internet of Things to the UK Economy. Rep SAS by Cent Econ reforms. 2016;(February):54.

[14] Wamba SF, Gunasekaran A, Akter S, Ren SJ fan, Dubey R, Childe SJ. Big data analytics and firm performance: Effects of dynamic capabilities. J Bus Res. 2017;

[15] Choi TM, Wallace SW, Wang Y. Big Data Analytics in Operations Management. Prod Oper Manag. 2018;

[16] Van Der Valk T, Gijsbers G. The use of social network analysis in innovation studies: Mapping actors and technologies. Innov Manag Policy Pract. 2010;12(1):5–17.

[17] Asur S, Huberman BA. Predicting the future with social media. In: Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010. 2010.