

Big Data Analytics: A Review

Anuja K. Tayde¹, Dr. P. M. Jawandhiya²

Computer Science And Engineering Department^{1,2}

Email: taydeanuja27@gmail.com,

Abstract - Based on characteristics of big data, this paper describes the development of technologies in big data analysis and storage and analyses the trends and different values in commercial applications, manufacturing, biomedical science and other applications. It provides a new way of thinking and approaches to analyze and solve problems, which gradually becomes a hot research. In the big data approach, large amounts of data affect our work, life and study, even national economic development. It provides a new way of techniques and approaches to analyze and solve problems. The big data techniques and methods also helps for decision making process. The information obtained from various sources such as scanners, mobile phones, loyalty cards, the web, and social media platforms provide the opportunity to deliver significant benefits to organizations. The big data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions.

Keywords: big data, analytics, characteristics, data analytics tools and methods.

1. INTRODUCTION

A big data is the place where every detail about a person or organization, every transaction performed, or every aspect is documented properly in the form of information. Big data is the hottest words in the IT industry, followed by data warehouse, data analysis and data mining. Big data helps people to acquire knowledge from the massive, complex data, and become another focus after integrated circuit and Internet information technology[4]. This paper provides an analysis of the available literature on big data analytics. Accordingly, some of the various big data tools, methods, and technologies which can be applied are discussed, and their applications and techniques provided in several decision domains are portrayed. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value.

When big data analytics is being researched in academia, several of the industrial advancements and new technologies provided were mostly discussed in industry papers. The idea of Big Data is having the way that the datasets are large to the point that run of the traditional database frameworks are not ready to store and investigate the datasets[2]. The datasets are vast in light of the fact that the information is no more customary organized information, however information from numerous new sources, including email, online networking, and Internet-available sensors. In the past years, Big Data has shown the ability to make more educated and auspicious expectations of business sector patterns, extortion control, training, business exchange, national security, and human

ces spare money, help effectiveness and enhance basic rship in fields as different as activity control, climate ating, catastrophe avoidance, money.

2. BIG DATA ANALYTICS

data analytics is the technique of collecting, organizing urge sets of data and to discover and analyze patterns and useful information. The big data analytics can help organizations to better understand the information contained n the data and will also help identify the data that is most rtant to the business and future business decisions. ysts who are working with Big Data typically want the dledge that comes from analyzing the data. The concept g data is analyzed by using different software tools and cations for predictive analytics, data mining, text mining, forecasting and data optimization. These processes are ctively separate but highly integrated functions of performance analytics[1]. The size of big data is antly increasing, currently ranging from a few dozen ytes (TB) to many pet bytes (PB) of data in a single data The difficulties related to big data include capture, ge, search, sharing, analytics, and visualizing. Apache op is an open source software platform for distributed ge and distributed processing of very large data sets on ounter clusters built from commodity hardware. Hadoop ces provide for data storage, data processing, data is, data governance, security, and operations[5].

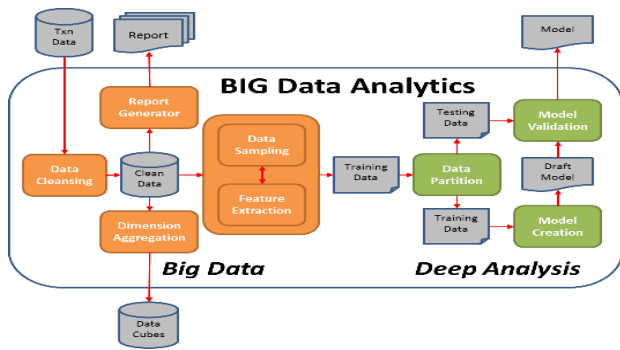


Fig (1): Big data analytics process

2.1 Characteristics of Big Data

2.1.1 Volume:

The volume refers to the quantity of data gathered by a company. This data must be used further to gain important knowledge. The volume dimension of big data is not defined in specific quantitative terms. The term big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze[3].

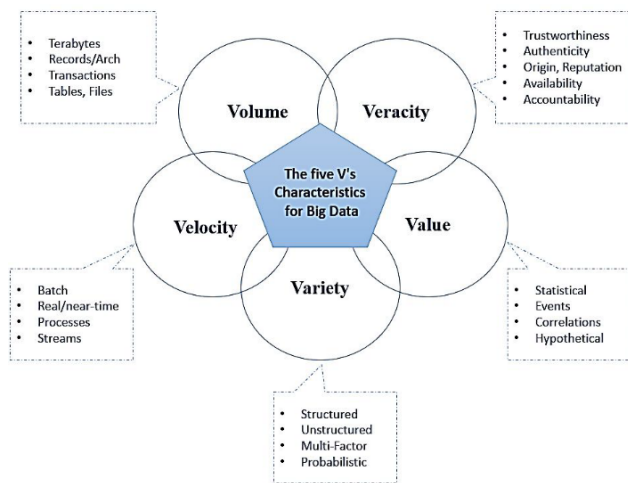


Fig (2): Characteristics of big data

2.1.2 Variety:

Variety refers to large sources and the nature of data, both structured and unstructured. During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications.

2.1.3 velocity:

The term 'velocity' refers to the speed of data i.e. how fast the data is generated and processed to meet the demands, determines real potential in the data. Big Data

Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks and social media sites, sensors, Mobile devices, etc.

2.1.4 Veracity:

The term veracity refers to the degree at which the decision can be taken. Therefore, finding the right correlations in Big Data is very important for the business future[5].

2.1.5 Value:

Value refers to the important feature of the data which is defined by the added-value that the collected data can bring to the intended process, activity or predictive analysis. Data value will depend on the events or processes they represent such as stochastic, probabilistic, regular or random.

2.2 Big Data Analytics Methods and tools:

With the evolution of technology and the increased in data volume there has become a need for faster and more efficient ways of analyzing such data. Having a large amount of data on hand is no longer enough to make efficient decisions at the right time. Therefore, some new tools and methods are specialized for big data analytics, as well as the required architectures for storing and managing such data.

The Big - Data, Analytics, and Decisions (B-DAD) framework which incorporates the big data analytics tools and methods into the decision making process. The framework maps the different big data storage, management, and processing tools, analytics tools and methods, and visualization and evaluation tools to the different phases of the decision making process[4]. Hence, the changes associated with big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the big data analyses which can be applied for knowledge discovery and informed decision making. Each area will be further discussed in this section. However, since big data is still evolving as an important field of research, and new findings and tools are constantly developing, this section is not exhaustive of all the possibilities, and focuses on providing a general idea, rather than a list of all potential opportunities and technologies.

Big Data Storage and Management

While managing the data first thing for organizations is to deal with big data, is where and how this data will be stored once it is acquired. The traditional methods of

structured data storage and retrieval include relational databases, data marts, and data warehouses.

Big Data Analytic Processing

The analytic process is main after storing large amount of data. Accordingly there are some critical requirements for big data processing.

The first requirement is fast data loading. It is necessary to reduce the data loading time to increase the processing time of the data[2].

The next requirement is fast query processing. In order to satisfy the requirements of heavy workloads and real-time requests, many queries are response-time critical.

The final requirement for big data processing is to manage storage space properly to store large amount of data properly which is also efficient at the time of performing operations on them.

3. BIG DATA ANALYTICS AND DECISION MAKING

The decision making process perform research throughout the years it requires huge amount of data and information related to all fields. Big data is important asset for decision makers. From the decision maker's perspective, the significance of big data lies in its ability to provide information and knowledge of value, upon which to base decisions. The information obtained from various sources such as scanners, mobile phones, loyalty cards, the web, and social media platforms provide the opportunity to deliver significant benefits to organizations[2].

3.1 Customer Intelligence:

Big data analytics can provide organizations with the ability to profile and segment customers based on different socioeconomic characteristics, as well as increase levels of customer satisfaction and retention. Big data can create transparency, and make relevant data more easily accessible to stakeholders in a timely manner . Big data analytics holds much potential for customer intelligence, and can highly benefit industries such as retail, banking, and telecommunications.

3.2 Supply Chain and Performance Management:

In supply chain management, big data analytics can be used to forecast demand changes, and accordingly match their supply. This can increasingly benefit the manufacturing, retail, as well as transport and logistics industries. By analyzing stock utilization and geospatial data on deliveries, organizations can automate

replenishment decisions, which will reduce lead times and minimize costs and delays, as well as Process interruptions.

3.3 Quality Management and Improvement:

Especially for the manufacturing, energy and utilities, and telecommunications industries, big data can be used for quality management, in order to increase profitability and reduce costs by improving the quality of goods and services provided[3]. While developing any new model it should be necessary that it maintain the quality of product which properly meet the market conditions. Every time it is necessary that quality of product should maintain accordingly.

3.4 Risk Management and Fraud Detection:

Industries such as investment or retail banking, as well as insurance, can benefit from big data analytics in the area of risk management. Since the evaluation and bearing of risk is a critical aspect for the financial services sector, big data analytics can help in selecting investments by analyzing the likelihood of gains against the likelihood of losses. Fraud detection helps to continue the smooth working of our program.

4. BENEFITS OF BIG DATA

4.1 Errors within the organization are known instantly:

Real-time view into errors helps companies to perform operation quickly to mitigate the effects of an operational problem. This can save the operation from falling behind or failing completely or it can save your customers from having to stop using your products.

4.2 Cost savings:

The Real-Time Big Data Analytics tools may be expensive; its implementation will eventually save a lot of money. There is no waiting time for business leaders and in-memory databases (useful for real-time analytics) also reduce the burden on a company's overall IT landscape, freeing up resources previously devoted to responding to requests for reports[4].

4.3 Cost savings:

The Real-Time Big Data Analytics tools may be expensive; its implementation will eventually save a lot of money. There is no waiting time for business leaders and in-memory databases (useful for real-time analytics) also reduce the burden on a company's overall IT landscape, freeing up resources previously devoted to responding to requests for reports.

4.4 Understand the market conditions:

To understand the current market condition you need to

analyze the available big data. For example, by analyzing customers' purchasing behaviors, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors. At every time of research it is necessary that to properly understand all the market situations for proper development.

4.5 Development of new product:

By knowing the trends of customer needs and satisfaction through analytics you can create products according to the wants of customers. It is necessary that our new product can definitely match all the customer requirements and it should stand in market for long period of time.

4.6 Notice new strategies:

By analyzing the available big data and performing different operations on them we can easily understand the new market strategies. While we need to generate any type of program the market strategy should be analyzed properly and decisions should be taken accordingly.

5. PROBLEMS OF BIG DATA:

5.1 Privacy and security:

When we combine personal information of a person with external large data sets, leads to the inference of new facts about that person and it's possible that these kinds of facts about the person are secretive and the person might not want the data owner to know or any person to about them. While we need to store any type of data the first consideration is privacy which is very important all types of constraints.

5.2 Data access and sharing:

The large set of data is collected and accessed to take any final decision regarding to any big data challenges therefore we need to perform the sharing of some important data for proper operation. The data sharing process helps to take lot of decisions properly by accessing the required amount of data.

5.3 Analytical challenges:

The big data have huge analytical challenge. To perform this type of analysis it requires lot of knowledge and skill. This type of analysis is depends upon what type of data is available for process[1]. When we need to perform analysis on data sets it is necessary that we access all the required information with proper set and to generate only authentic output.

5.4 Human resource and manpower:

Big data is very emerging technology therefore it requires proper skill. We need to perform different training program for improving the required skills. For performing operations we need to adopt the skilled of expertise. In any type of department there is necessary need for the intelligent and skilled manpower to operate all procedure.

5.5 Technical challenges:

While dealing with huge amount of big data we need to also deal with different type technical problems. Scalability, quality of data, heterogeneous data, fault tolerance these are the common types of technical problem occurred while dealing with big data. When the technical problem it is necessary that to solve that as early as possible for smooth work of our program[5].

6. ACKNOWLEDGMENT:

A great pleasure to submit the technical paper topic titled 'Big Data Analytics: A Review', to Pankaj Ladhhad Institute of Technology and Management Studies, Buldana, which gave we an opportunity in fulfilling our goals and also help to achieving our desire goals. We very much thankful to our Principal Dr. P. M. Jawandhiya, for providing all the facilities needed for successful completion of this paper and providing necessary assistance while preparing for this project work[2].

We thankful to our guide Dr. P. M. Jawandhiya,, because without his valuable guidance this work would not have a success. His timely suggestions and encouragement in every time helped we to carry out my paper work. We would also like to thank our Head of Department Dr. P. M. Jawandhiya, for co-operating we. We very much thankful to all Professors, Lecturers and Staff member of Computer Science and Engineering Department.

7. CONCLUSION:

In this research paper, we have examined the innovative topic of big data, which has recently gained lots of interest due to its recent advancement. In the information era new data being produced daily therefore different types of operations should be performed on them. The big data concept should be applied on business change for decision making, by applying different types of analytical challenges. The big data was discussed with its characteristics and importance. Some of the big data analytics tools and methods were also examined. Big data benefits and challenges were also discussed in detail.

The different types of analytics methods helps to big data to access valuable information for decision making and support informed decisions. Thus, some of the different areas where big data analytics can support in decision making were examined. It was found that big data analytics can provide number of opportunities in various applications and areas. Accordingly, this research has provided the people and the organizations with examples of the various big data tools, methods, and technologies which can be applied. This gives users an idea of the necessary technologies required, as well as developers an idea of what they can do to provide more enhanced solutions for big data analytics in support of decision making.

If any new technology, if applied correctly can bring with it several potential benefits and innovations,

let alone big data, which is a remarkable field with a bright future, if approached correctly. To deal with big data, it requires proper storage, management, integration, federation, cleansing, processing, analyzing, etc. With all the problems faced with traditional data management, big data exponentially increases these difficulties due to additional volumes, velocities, and varieties of data and sources which have to be dealt with.

We believe that big data analytics is of great significance in this era of data overflow, and can provide unforeseen insights and benefits to decision makers in various areas. If properly technique is applied, big data analytics has the potential to provide a basis for advancements, on the scientific, technological, and humanitarian levels.

8. REFERENCES:

- [1] Asur, S., Huberman, B.A.: Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499(2010).
- [2] Elena Geanina Ularu, Florina Camelia Puican, Anca Apostu, Manole Velicanu, "Perspectives on Big Data and Big Data Analytics", Database Systems Journal, Volume 3, No. 4, 2012.
- [3] "Challenges and Opportunities with Big Data", A Community White Paper Developed by Leading Researchers Across United States, 2012.
- [4] Changqing Ji, Yu Li, Wenming Qiu, Uchechukwu Awada, Keqiu Li, "Big Data Processing in Cloud Computing Environments", 2012 International Symposium on Pervasive Systems, Algorithms and Networks.
- [5] Economist Intelligence Unit: The Deciding Factor: Big Data & Decision Making. In Capgemini Reports, pp. 1–24 (2012).