IJEE 0973-7383 Volume-9 • Number-1 Jan -June 2017 pp. 231-239 available online at www.csjournals.com

Big Data Analytics: Prospects, Challenges and Barriers

Er. Seema Rani
Assistant Professor, Computer Science and Engineering, CDLSIET, Sirsa.

Abstract

In this information age, huge amounts of data have become available to decision makers. These days, big data is gaining a lot of attention in the IT industry. Large, rapidly changing, and volatile data sets are referred to as "big data," which makes it challenging to handle those using conventional tools and methods. The demand for data storage and analysis has skyrocketed due to the internets and the digital economy's rapid growth, and the IT department is faced with a formidable task in protecting and analyzing these larger amounts of data. Owing to the exponential expansion of this type of data, it is necessary to conduct research and offer solutions for handling and obtaining information and values from these data sets. Produced data now comprises documents, photos, audio, video, and social media information; this data is known as unstructured data or big data instead of the conventional database data, or structured data. By removing value from these enormous amounts of data, big data analytics can increase client retention and open up new marketing avenues. Big data analytics, or the application of sophisticated analytical tools to large amounts of data, can offer such value. In order to assist anticipate future volumes, obtain insights, take preemptive actions, and open the door for better strategic decisions, this research study primarily focuses on discussing the many characteristics, challenges, and barriers of Big Data Analytics systems. Furthermore, in order to increase a company's competitiveness, this article examines the adoption, application, and effects of big data analysis on its commercial value. advantage.

Keywords: Big Data, Analytics, Decision making, Performance, Enterprises, Knowledge management.

INTRODUCTION

Imagine a world without data storage, where all recorded facts, completed transactions, and personal or organizational details vanish the moment they are used. Data serves as both the basis for accountability and the essential component of decision-making. It becomes nearly hard to establish, monitor, and evaluate effective policies in the absence of superior data that offer accurate information at appropriate times [1].The phrase "big data" is significant when referring to information that deviates from a traditional database's typical structure. Data is the foundation that makes every organization successful. Now consider how much more information and detail there is now due to technological advancements and the Internet. Big data is now easily accessible due to advancements in data collection techniques and storage capacity. Every second, a greater amount of data is produced. This data must be kept and examined to be valued. Organizations must extract as much value as they can from the massive volumes of data they

retain because data storage has also gotten more affordable [2]. The problems of handling data that is too big, too unstructured, and moving too quickly for conventional approaches to handle are reflected in big data analytics. Organizations of all kinds, from corporations to academic institutions to governmental bodies, routinely produce data of previously unheard-of size and complexity. Through the years, the internet and mobile devices (such as tablets and smartphones) will continue to present special research opportunities and problems [3]. Getting valuable insights and a competitive edge from vast volumes of data has become critical for businesses all over the world. Gaining rapid and easy access to relevant information from these kinds of data sources is challenging. As a result, analytics is now incorporated to fully utilize big data to boost market share and enhance business performance. The tools are now accessible to tackle amount, rapidity, and diversity of big data have advanced dramatically during last several years. Owing to its volume, diversity, and speed of change, big data analysis has to evolve, necessitating new techniques for both storage and analysis. Such enormous amounts of data need to be carefully examined in order to obtain pertinent information. A number of significant technologies are referred to as "big data"; these include Hadoop, HDFS, NoSQL, MapReduce, MongoDB, Cassandra, PIG, HIVE, and HBASE. Together, these technologies enable the ultimate goal of extracting value from data that was previously deemed worthless. Less than 5% of the data that firms have is really used, according to Forrester Research. This is due to the fact that handling the remainder would be extremely costly. Big data is derived from numerous sources. This covers all kinds of unstructured data sources that are rapidly expanding, not only conventional relational data. For instance, content that is rich and diversified that has to be found is contained in machine-derived data, which is growing quickly. Human social media data is another example; although it is primarily textual, insightful insights are frequently buried behind a mountain of potential interpretations.

These technologies are generally not too costly, and a large portion of the software is available under open source. Open source software and commodity hardware are combined in Hadoop, the most popular framework [4]. It includes capabilities for data analysis in addition to receiving incoming data streams and distributing them to inexpensive storage. To integrate all pertinent internal and external data sources, these strategies necessitate expertise that most IT departments lack. Technology attention is always required as part of a big data strategy, even while it is insufficient on its own. This study analyzes the literature on big data analysis that is currently available. The research paper is organized as, First section explain introduction of big data analytics. Second section provides related work. Third section describes characteristics, Fourth section explain challenges of bid data analytics. Next section tells about different barriers and last section concludes the paper.

RELATED WORK

Big data analysis is a technique made possible by new technologies that enable quick data gathering, storing, and analysis. Emails, mobile outputs, and sensor-generated data are examples of data sources outside of the conventional company database [5]. There isn't a single definition

for big data analytics because they are new and developing concepts. Different parties gave varying and sometimes contradictory definitions. A widely-cited definition of big data was first proposed based on the 2001 Gartner Report. Gartner suggested using the three V's of volume, velocity, and variability to define big data. Plausibility, which stands for the dependability and uncertainty requirements of data and the outcomes of data analysis, was added to Gartner's definition in 2012. IDC identified the 4th V as value in a 2012 research. Big Data analytics handles unstructured data that may be processed into useful business intelligence through the use of computing technology. Examples of this data include phone calls, mobile transactions, user-generated content on the internet (e.g., blog posts, tweets), web searches, and photos patterns and trends in the data.

Technology plays a big role in arranging this data. Large and complicated, big data also necessitates cutting-edge technology for processing and analysis. Big data surpasses the capability of existing & conventional systems, allowing for new approaches to cross-border problems that were previously unachievable or impracticable with existing or traditional methods. Business challenges are rarely really data problems, and practitioners often find it difficult to integrate data into their intricate business development decision-making processes, even in situations where data is plentiful. A study of 1,469 executives from significant companies, the industry, and regional areas was carried out in 2012 by McKinsey & Company. According to the poll, 49% of participants stated that their organizations are using big data to understand, segment, and target customers to improve overall performance [6]. Even more respondents—60,444 percent—said that their organizations ought to concentrate on leveraging data and analytics to acquire these kinds of insights. These findings prove that how a corporation may maximize the benefits of big data has replaced the question of whether big data can benefit it.

Utilizing past data to forecast customer behavior and patterns is known as predictive analytics [7]. It forecasts future strategies using past and previous data. In order to find trends and draw conclusions from past data, this research makes use of statistical models and machine learning algorithms [8]. Another way to describe predictive analytics is as a method that makes predictions about data by analyzing it with machine learning. Predictive analytics is expected to help businesses develop more smart marketing campaigns in the future, according to 67% of businesses, and its primary advantage is seen by 68% of respondents as being competitive advantage [9]. Predictive search, price management, and product recommendations are often possible when using predictive analysis for online purchasing.

On the other hand, an online shopping site or application can quickly find or forecast products that closely match user likes by creating a recommendation system [10]. The method of creating a database of past user preferences is termed Collaborative Filtering. A preference database is consulted when a new consumer visits an online store in order to determine which discount class best suits their preferences. The customer is then given recommendations for those products. Online purchasing also makes advantage of the clustering process. Users with similar

preferences are identified by groups that the clustering algorithm finds. After then, these users are put into a single group with a special identification number. The average similarity of each cluster member is used to forecast new customer groupings. As a result, a user's membership in many clusters can vary based on how much weight their average opinion carries. The scientific process of turning data into insights to help in decision-making is known as advanced analytics. Within operations research, advanced analytics has become a recognized subject of study. Analytics and certain other fields have a lot in common, and there are various recognized categories for different kinds of analytics.

BIG DATA KEY CHARACTERISTICS

Big data are those that require the use of cutting-edge technical infrastructures, analyses, and tools due to their volume, dispersion, diversity, and timeliness in order to generate new sources of insights that can create value for businesses. The primary factors, also known as the three V's, define big data: volume, variety, and velocity.

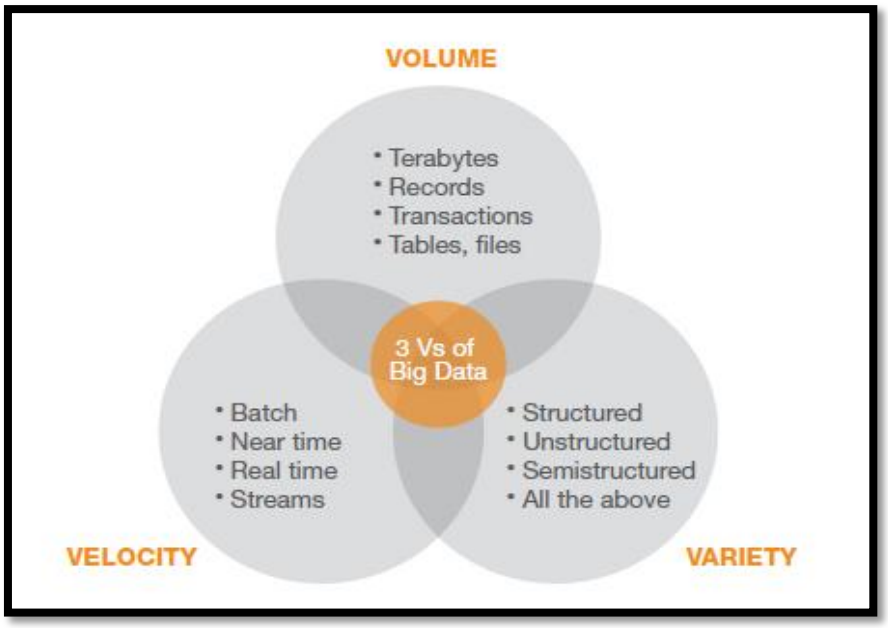

Figure 1. The Three V's of Big Data Volume

1.Volume

It's clear that the main characteristic of big data is data volume. Given this, the majority of people define huge data in terms of terabytes, occasionally petabytes. For instance, several users TDWI spoke with manage three to ten terabytes (TB) of data for analytics. Nevertheless, transactions, files, tables, and records can all be counted in order to quantify large data. Big data can be more usefully quantified in terms of time for certain businesses. Because of the US statute

of limitations, which is seven years, many businesses, for example, decide to keep data that is available for risk, compliance, and legal analysis for seven years. The quantification of big data is also influenced by its extent. For instance, the data gathered for general data warehousing and the data gathered for other purposes are different in many firms.

2. Variety

The fact that big data is coming from more sources than ever before is one of the factors that makes it so massive. Logs, clickstreams, social media, and other Web sources include a large number of the more recent ones. Yes, user groups have been gathering data on the Web for a long time. However, it's been a form of hoarding for the majority of corporations. The shift lies in the fact that a greater number of people are now examining massive data rather than just storing it. These days, the few firms have been studying this data at a degree of sophistication and complexity. While big data itself is not new, its useful analytical use is. Additionally, data from audio, video, and other sources can be difficult to classify. Variety in big data is therefore just as important as volume. Furthermore, volume and variety often feed off one another.

3. Velocity

One way to characterize big data is by its speed or velocity. It might be better to consider the frequency at which data is sent and generated. Consider the stream of data that emerges from any device like cameras searching a particular face in a crowd, microphones listening for activity in a safe area, thermometers measuring temperature, or robotic manufacturing processes. Big data collection now is not new. Numerous businesses are gathering and using streaming data from websites to suggest products to website users. Data quantities quickly increase when real-time, persistently streamed sensor and Web data is presented to you. The analytics that accompany streaming data present a much greater challenge since they must interpret the data and maybe take action in real time.

CHALLENGES IN DATA LIFECYCLE

Large datasets that are impossible for standard software tools to gather, store, manage, and analyze are known as "big data." This massive data is enormous in terms of both quantity and complexity. Additionally, big data includes data in a variety of formats, including text, audio, video, images, and more. Unstructured data currently makes up 90% of all data and is growing at a higher rate than organized data. As a result, new processing methods are required to acquire data insights that enhance decision-making. Three categories can be used to group the issues related to the data life cycle: process challenges, management challenges and data challenges (Fig. 1).

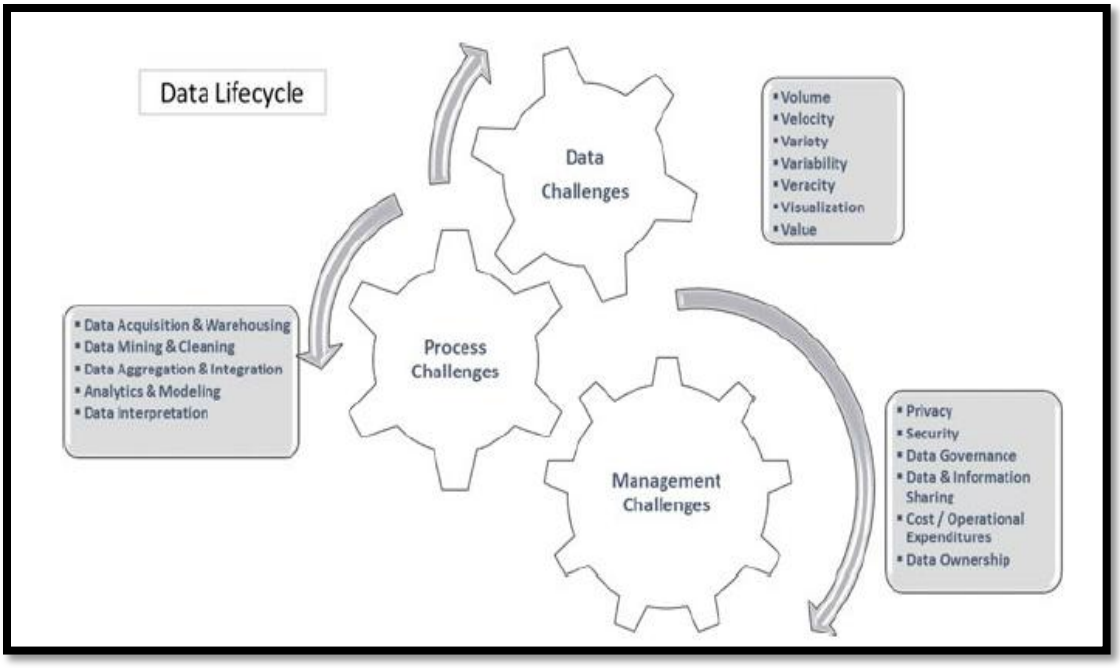

Figure 2: Big Data Life Cycle Challenges

1. Process challenges: Process issues are associated with the methods required for Data warehousing, Data aggregation, Data transformation, Data mining & cleaning, Data analytics and modeling, Data interpretation in order to use the big data to obtain insights [11].

2. Data management challenges: Data security, privacy, governance, cost/operational expenses, and data ownership are among the issues that face data management.

3. Data challenges: are related to features of big data. Big data qualities including volume, velocity, variety, veracity, visualization, value and variability are referred to as data challenges [12].

BIG DATA ANALYTICS BARRIERS

According to research from industry sectors, Organizations use less than half of their structured data for decision-making, analyze and spend 80% time in finding and preparing data [13]. Additionally, 70% of employees have access to data that they shouldn't.

1. Leadership: Management issues state that successful businesses in the data-driven era have leadership teams that set goals, adjust for success, and pose the appropriate queries. Even with its technical approach, without human vision big data's potential cannot be fully realized. As a result, company executives who possess a clear idea to identify emerging opportunities will be able to take initiative, inspire teams to work hard, and accomplish goals.

2. Talent management: Businesses need highly skilled human resources to operate and exploit latest technologies to obtain usable information and to harness information from big data

analytics. Individuals possess abilities in machine learning, mining, statistics, and mastery of visualization technologies. These are necessary to extract insightful information from big data that aids the decision-making process [14]. But the demand for these professionals—data scientists, data analysts, etc.—is enormous because they are very hard to find. To Find data scientists with expertise in both analytics and subject awareness is difficult. There are generally not as many data scientists as are required.

3. Decision-making process: Knowledge obtained from data exploitation and decision-makers are located in the same location in effective organizations. Nevertheless, handling massive data is a challenge. Expertise must be skilled at solving problems and able to offer solutions using the appropriate data or the collaboration of various parties when using big data to solve problems [15].

4. Quality of Decision Making. The quality of decision-making is a crucial factor in taking advantage of the opportunity. In that regard, ensuring the quality of decision-making is associated with elements such as staff, capabilities, and data quality. The accuracy of huge data sources is critical to making high-value decisions with no regrets. However, big data analytics capabilities depend on using the right methods and resources from big data analytics professionals.

5. Data-driven culture: One major barrier to putting a data-driven plan into practice is enterprise culture. A data-driven culture is built on the capacity to quickly gather, assess, and provide vital company information to decision makers. Businesses are empowered when that expertise is developed and enhanced, which benefits all company sectors and increases return on investment. This idea is essential for enhancing business performance. In that regard, businesses must cease acting purely on emotion and hunches and start using data to inform all of their decisions.

6. New technology utilization: While business intelligence and/or data warehousing tools have helped many organizations realize the power of data, technologies are unique. To extract value from big data, businesses must use tools that are now accessible. Given how quickly these techniques are developing, IT sector ought to grow and stay abreast of new developments. For example, issues will arise if big data analytics choices are not supported by database software.

7. Data privacy: A lot of people find the collecting of data to be highly dubious. Big data is seen by them as an invasion of privacy. When it comes to consumers' perceptions of data, marketers are facing challenges because 71% believe that companies that have their personal information use it unethically, and 58% have never used a digital service due to privacy concerns that affect their decisions about which apps to download, which email addresses to share, and which social media platforms to use for website links. Consequently, businesses must implement security measures to guarantee that consumer privacy is maintained when using data. To fully utilize the

value of big data, data policies pertaining to intellectual property, security, and privacy should be taken into consideration.

CONCLUSIONS

Introduction of Web 3.0 era further expanded the potential for external data collecting. Businesses across industries started concentrating on using data to their advantage in order to obtain a competitive edge. These days, almost every company has subtly welcomed the big data era, realizing that data-driven decisions are almost always smarter and more accurate. Not every business employing business analytics, including big data analytics, is benefiting from it, even though many companies in a variety of industries use it. A corporation can become data-driven by doing more than just employing analytical methods and resources. Being able to analyze data well is crucial for success in the data-driven business environment of today. Analysis and domain knowledge are not distinct fields because of the ever-increasing volume of data. Analytical abilities and an understanding of business procedures are needed of both academic and applied experts working for the company. Businesses with successful challenge management and data-driven cultures ought to look forward to a bright future. There is a lot of evidence to support the claim that data-driven decision making, big data technology, and analytical tools and processes may improve business performance.As more businesses pick up the fundamentals of exploiting big data and interact with the ever-evolving technology of today, they might soon differentiate themselves from the competition and gain a clear competitive edge.

REFRENCES

1. Villars R. L, Olofson C. W, Eastwood M, “Big data: What it is and why you should care”, IDC, White Paper, Framingham, June 2011.
2. Chae B, Sheu C, Yang C and Olson, “The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective”, Decision Support Systems, Vol 59, Pp 119-126, 2014.
3. Chen H, Chiang R, Storey V.C, “Business intelligence and analytics: from big data to big impact”, Miss. Q. 36(4), Pp 1165–1188, 2012.
4. Provost F, Fawcett T, “Data science and its relationship to big data and data-driven decision making” , Big Data Vol 1, Pp 51–59, 2013.
5. DATASTAX Corporation, “Big Data: Beyond the Hype - Why Big data Matters to you” , White paper, October 2013.
6. Manyika J, “Big data: the next frontier for innovation, competition, and productivity”, 2011.
7. Mosavi A, Vaezipour A, “Developing Effective Tools for Predictive Analytics and Informed Decisions”, Technical Report, University of Tallinn, 2013.
8. Shmueli G, Koppius O, “Predictive Analytics in Information Systems Research”, MIS Quarterly, Vol 35(3), Pp 553-72, 2011.
9. Millard S, “Big Data Brewing Value in Human Capital Management”, Ventana Research, 2013.

10. Sarwar B, Karypis G, Konstan J, Riedl J, "Recommendation systems for large e-commerce: Scalable neighborhood formation using clustering", in Proceedings of the fifth international conference on computer and information technology, 1, 2002.
11. Friendly M, "The golden age of statistical graphics", Stat. Sci, Vol 23(4), Pp 502–535, 2008.
12. Power D.J, "Understanding data-driven decision support systems", Inf. Syst. Manag, Vol 25(2), Pp 149–154, 2008.
13. McAfee A, Brynjolfsson E, "Big data: the management revolution", Harv. Bus. Rev, Vol 90(10), Pp 60–66, 68, 128, 2012.
14. Waller M.A, Fawcett S.E, "Data science, predictive analytics, and big data: a revolution", J. Bus. Logist, Vol 34(2), Pp 77–84, 2013.
15. Richard L. Villars, Carl W. Olofson, Matthew Eastwood, "Big data: What it is and why you should care", IDC White Paper, June 2011.