International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

Big Data Analysis : Challenges and Prospects in the Current Era

Dr. D. V. Bhavsagar
Assistant professor
Seth Kesarimal Porwal College of Arts and Science and Commerce,
Kamptee, Nagpur (MS)

ABSTRACT :

Big Data is a collection of huge volumes of data and it is continuously growing exponentially with time. It is data with such large size and difficulty that none of traditional data management tools can store it or process it efficiently. Big data is also data but with huge size. Big data is an ever growing field in the current era. There are various challenges and opportunities associated with big data analysis. This paper envisages the challenges and opportunities in big data analysis in the current era.

Keyword : big data, data analysis, data life cycle

INTRODUCTION :

Big Data is popular today and has gained much attention from academia and the IT industry. In the world of digital and computing, information is generated and collected at a rate that rapidly exceeds the boundary range. Presently, over 2 billion people worldwide are connected to the Internet, and over 5 billion individuals have mobile phones. By 2022, 6 billion devices are expected to be connected to the Internet. At this point, predicted data production will be 45 times greater than that in 2010. As information is transferred and shared at light speed on optic fiber and wireless networks, the volume of data and the speed of market growth increase. Though, the fast growth rate of such large data generates numerous challenges, such as the rapid growth of data, transfer speed, diverse data, and security. However, Big Data is still in its infancy stage, and the domain has not been reviewed in general. Hence, this study widely surveys and classifies the various attributes of Big Data, including its nature, definitions, rapid growth rate, volume, management, analysis, and security.

The international population is more than 7.2 billion, and more than 2 billion of these people are connected to the Internet. Still, 5 billion individuals are using different mobile devices, according to McKinsey. As a result of this technological revolution, these millions of people are generating more amounts of data through the increased use of such devices. In particular, remote sensors regularly produce much heterogeneous data that are either structured or unstructured. This data is known as Big Data. Big Data is characterized by three aspects: (a) the data are numerous, (b) the data cannot be categorized into regular relational databases, and (c) data are generated, captured, and processed very quickly. Big Data is promising for business application and is rapidly increasing as a segment of the IT industry. It has generated significant interest in various fields, including the manufacture of healthcare machines, banking

transactions, social media, and satellite imaging. Traditionally, data is stored in a highly structured format to maximize its informational contents. However, current data volumes are driven by both unstructured and semi structured data. Therefore, end-to-end processing can be impeded by the translation between structured data in relational systems of database management and unstructured data for analytics.

The staggering growth rate of the amount of collected data generates numerous critical issues and challenges described by, such as rapid data growth, transfer speed, diverse data, and security issues. Nonetheless, the advancements in data storage and mining technologies enable the preservation of these increased amounts of data. In this preservation process, the nature of the data generated by organizations is modified. However, Big Data is still in its infancy stage and has not been reviewed in general. Hence, this study comprehensively surveys and classifies the various attributes of Big Data, including its volume, management, analysis, security, nature, definitions, and rapid growth rate. The study also proposes a data life cycle that uses the technologies and terminologies of Big Data. Future research directions in this field are determined by opportunities and several open issues in Big Data domination.

ABOUT BIG DATA :

The definition of big data is data that contains greater variety, arriving in increasing volumes and with more velocity. This is also known as the three Vs.

Put simply, big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before.

Big data sizes are constantly increasing, currently ranging from a few dozen tera-bytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

The three Vs of big data

Volume	The amount of data matters. With big data, you'll have to process high volumes of low-density, unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a web page or a mobile app, or sensor-enabled equipment. For some organizations, this might be tens of terabytes of data. For others, it may be hundreds of petabytes.
Velocity	Velocity is the fast rate at which data is received and (perhaps) acted on. Normally, the highest velocity of data streams directly into memory versus being written to disk. Some internet-enabled smart products operate in real time or near real time and will require real-time evaluation and action.
Variety	Variety refers to the many types of data that are available. Traditional data types were structured and fit neatly in a relational database. With the rise of big data, data comes in new unstructured data types. Unstructured and semistructured data types, such as text, audio, and video, require additional preprocessing to derive meaning and support metadata.

Source : https://www.oracle.com/in/big-data/what-is-big-data/

THE HISTORY OF BIG DATA

The concept of big data itself is relatively new, the origins of large data sets go back to the 1960s and '70s when the world of data was just getting started with the first data centers and the development of the relational database.

Around 2005, people began to realize that how much data users generated through Facebook, YouTube, and other online services. Hadoop (an open-source framework created specifically to store and analyze big data sets) was developed that same year. NoSQL also began to gain popularity during this time.

The development of open-source frameworks, such as Hadoop (and more recently, Spark) was essential for the growth of big data because they make big data easier to work with and cheaper to store. In the years since then, the volume of big data has skyrocketed. Users are still generating huge amounts of data—but it's not just humans who are doing it.

With the development of Internet of Things (IoT), more objects and devices are connected to the internet, gathering data on customer usage pattern and product performance. The emergence of machine learning has produced still more data.

While big data has come far, its value is only just beginning. Cloud computing has expanded big data possibilities even further. The cloud offers truly expandable scalability, where developers can simply spin up ad hoc clusters to test a subset of data. And graph databases are becoming increasingly important as well, with their ability to display enormous amounts of data in a way that makes analytics fast and comprehensive.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

BIG DATA USE CASES

Big data can help you address a range of business activities, from customer experience to analytics. Here are just a few.

Product development	Companies like Netflix and Procter & Gamble use big data to anticipate customer demand. They build predictive models for new products and services by classifying key attributes of past and current products or services and modeling the relationship between those attributes and the commercial success of the offerings. In addition, P&G uses data and analytics from focus groups, social media, test markets, and early store rollouts to plan, produce, and launch new products.
Predictive maintenance	Factors that can predict mechanical failures may be deeply buried in structured data, such as the year, make, and model of equipment, as well as in unstructured data that covers millions of log entries, sensor data, error messages, and engine temperature. By analyzing these indications of potential issues before the problems happen, organizations can deploy maintenance more cost effectively and maximize parts and equipment uptime.
Customer experience	The race for customers is on. A clearer view of customer experience is more possible now than ever before. Big data enables you to gather data from social media, web visits, call logs, and other sources to improve the interaction experience and maximize the value delivered. Start delivering personalized offers, reduce customer churn, and handle issues proactively.
Fraud and compliance	When it comes to security, it's not just a few rogue hackers—you're up against entire expert teams. Security landscapes and compliance requirements are constantly evolving. Big data helps you identify patterns in data that indicate fraud and aggregate large volumes of information to make regulatory reporting much faster.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

Machine learning	Machine learning is a hot topic right now. And data—specifically big data—is one of the reasons why. We are now able to teach machines instead of program them. The availability of big data to train machine learning models makes that possible.
Operational efficiency	Operational efficiency may not always make the news, but it's an area in which big data is having the most impact. With big data, you can analyze and assess production, customer feedback and returns, and other factors to reduce outages and anticipate future demands. Big data can also be used to improve decision-making in line with current market demand.
Drive innovation	Big data can help you innovate by studying interdependencies among humans, institutions, entities, and process and then determining new ways to use those insights. Use data insights to improve decisions about financial and planning considerations. Examine trends and what customers want to deliver new products and services. Implement dynamic pricing. There are endless possibilities.

Source : https://www.oracle.com/in/big-data/what-is-big-data/

CHALLENGES OF BIG DATA :

1. Big data is so big. Even though new technologies have been developed for data storage, data volumes are doubling-up in size about every two years. Organizations are still struggling to keep pace with their data and find ways to effectively store it.
2. We don't have to think only about the storage the data. Data must be used to be valuable and that depends on cleaning. Clean data, or data that's applicable to the client and organized in a way that enables meaningful analysis, requires a lot of work. Data scientists spend 50% to 80% of their time cleaning and preparing data before it can actually be used.
3. Also the big data technology is changing at a rapid pace. A few years ago, Apache Hadoop was the preferred technology used to handle big data. Then Apache Spark was introduced in 2014. Today, a combination of the two frameworks are used popularly. Keeping up with big data technology is an ongoing challenge.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

4. The analytics does not have enough data to generate new insights. This may either be caused by the lack of data integrations or poor data organization. In this case, it makes sense to run a data audit and ensure that accessible data integrations can provide the required insights. The integration of new data sources can eradicate the lack of data as well. It's also worth checking how raw data comes into the system and make sure that all possible dimensions and metrics are exposed for analytics. Finally, data storage variety might also be a problem. One can cope with this issue by introducing a Data Lake.
5. If our system relies on data that has defects, errors, or is incomplete we'll get poor results. Data quality management and an obligatory data validation process covering every stage of your ETL process can help ensure the quality of incoming data at different levels (syntactic, semantic, grammatical, business, etc.). It will enable us to identify and weed out the errors and guarantee that a modification in one area immediately shows itself across the board, making data pure and accurate.
6. The next problem may bring all the efforts invested in creating an efficient solution to naught. If using data analytics becomes more complicated, you may find it difficult to extract value from your data. The complexity issue frequently boils down either to the UX (when it's difficult for users to navigate the system and grasp info from its reports) or to technical aspects (when the system is over-engineered).
7. The system takes much time to analyze the data even though the input data is already available. It may not be so critical for batch processing, but for real-time systems such delay can cost high.
8. Any system requires regular investment in its maintenance and infrastructure. And every business owner wants to minimize these investments. The maintenance cost is too high in case of bid data analytics.

OPPORTUNITIES OF BIG DATA

1. Opportunities to national development : At present, the world has completely entered the era of the information age. The widespread use of Internet, Internet of Things, Cloud Computing, and other emerging IT technologies has made various data sources increasing at an unprecedented rate, while making the structures and types of data increasingly complex. Depth analysis and utilization of big data will play an important role in promoting continued economic growth of countries and enhance the competitiveness of companies.
2. Opportunities in industrial upgrades : Big data is currently a universal problem faced by many industries, and it brings grand challenges to these industries' digitization and bulk information.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

Research on common problems of big data, particularly on breakthroughs of core technologies, will facilitate industries to harness the complexity induced by data interconnection and to master doubts caused by redundancy and/or shortage of data. Everyone hopes to mine from big data demand-driven information, knowledge and even intelligence and eventually taking full advantage of the big value of big data. It will not only be the new engine to sustain the high growth of the information industry, but also the new tool for industries to improve their competitiveness.

3. Opportunities in Scientific Research : The development of big data has spawned a new research pattern; that is, with big data, researchers may only need to find or mine from it the required information, knowledge and intelligence. They even do not need to unswervingly access the objects to be studied.
4. Opportunities in emerging interdisciplinary research : Big data technologies and the corresponding fundamental research have become a research focus in academic world. An emerging interdisciplinary discipline called data science has been steadily coming into place. This takes big data as its research object and aims at generalize the taking out of knowledge from data. It spans across many disciplines, including information science, mathematics, social science, network science, system science, psychology, and economics. It employ a variety of techniques and theories from many fields, including signal processing, probability theory, machine learning, statistical learning, computer programming, data engineering, pattern recognition, visualization, uncertainty modeling, data warehousing, and high performance computing. Many universities and research institutes have set up differnet courses on data analytics for cultivating talents, including data scientists and data engineers.
5. Opportunities in helping people : Big Data, especially big networked data, contains a wealth of information related to the social aspects and can thus be viewed as a network mapped to society. To this end, analyzing big data and further abbreviating and finding clues and laws it unreservedly contains can help us better perceive the present. Deep mining information limited in big data can also help people make better decisions. Significance to helping people better predict the future through effective integration and precise analysis on multi-source heterogeneous big data, better predictions of future trends of events can be achieved. It is possible for big data analysis to even endorse sustainable developments of society and economy and further give birth to new industries related to data services. The aptitude of big network data has been being highly developed and effectively applied in the field of security and military.

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

CONCLUSION :

Big data analytics no matter how advanced they are, does not remove the need for human insights. On the contrary, there is a convincing need for skilled people with the ability to understand data, think from the business point of view and come up with insights. For this very reason technology professionals with Analytics skill are finding themselves in high demand as businesses look to control the power of Big Data.

Big data has made a strong impact in almost each and every sector and industry today. The successful applications of big data in industry point to the following essential conditions for a big data project to be successful. There must be very clear necessities, regardless of whether they are technical, social, or economic. To properly work with big data, we will need to explore and find the kernel structure or kernel data to be processed. Finding kernel data and structures, which are small enough and yet can characterize the behavior and properties of the underlying big data, is non-trivial because it is very domain-specific. A top-down management model should adopt. Even though a bottom-up approach may allow us to solve some niche problems, the isolated solutions often cannot be put together into a total solution. Finally, the goal should be to solve the whole problem by an integrated solution, rather than striving for isolated successes in a few aspects. In short, an integrated and efficient approach should be employed in managing a big-data. The big data analysis has huge opportunities which live long as rate of information doubling is constantly declining and we have to manage this big data so that it will helpful to the society and industries at large.

REFERENCES :

1. Cuzzocrea : Privacy and security of big data: current challenges and future research perspectives:Proceedings of the First International Workshop on Privacy and Security of Big Data, PSBD '14 (2014)
2. O'Neil, R. Schutt: Doing Data Science: Straight Talk from the Frontline,O'Reilly Media, Inc. (2013) 7. Big data:http://en.wikipedia.org/wiki/Big_data (2014)
3. Elgendy, N.: Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, p. 164 (2013)
4. Elgendy, Nada & Elragal, Ahmed. (2014). Big Data Analytics: A Literature Review Paper. Lecture Notes in Computer Science. 8557. 214-227. 10.1007/978-3-319-08976-8_16.
5. Mishra Rishabh, Sharma Rakesh, BIG DATA: OPPORTUNITIES AND CHALLENGES, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 4, Issue. 6, June 2015, pg.27 – 35, https://www.ijcsmc.com/docs/papers/June2015/V4I6201508.pdf (2015)
6. Monal Chaudhary, 2017, A Literature Survey on Big Data, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICPCN – 2017 (Volume 5 – Issue 19),
7. Shen, Z., Wei, J., Sundaresan, N., Ma, K.L.: Visual Analysis of Massive Web Session Da-
8. ta. In: Large Data Analysis and Visualization (LDAV), pp. 65–72 (2012)

International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 12 Issue 06, June- 2022
ISSN(O): 2249-3905, ISSN(P) : 2349-6525 | Impact Factor: 8.202 |
Thomson Reuters ID: L-5236-2015

9. V. Mayer-Schonberger, K. CukierBig Data: A Revolution That Will Transform How We Live, Work, and ThinkHoughton Mifflin Harcourt (2013)
10. Zeng, D., Hsinchun, C., Lusch, R., Li, S.H.: Social Media Analytics and Intelligence.
11. IEEE Intelligent Systems 25(6), 13–16 (2010)
12. https://www.guru99.com/what-is-big-data.html
13. https://www.oracle.com/in/big-data/what-is-big-data/
14. https://www.xenonstack.com/insights/big-data-challenges
15. https://towardsdatascience.com/big-data-its-benefits-challenges-and-future-6fddd69ab927
16. https://ieeexplore.ieee.org/document/9213765
17. https://cra.org/ccc/wp-content/uploads/sites/2/2015/05/bigdatawhitepaper.pdf
18. https://elearningindustry.com/big-data-analytics-challenges-faced-business-enterprises-7-top
19. https://www.knowledgehut.com/blog/big-data/big-data-analytics-challenges-and-opportunities
20. https://www.researchgate.net/publication/316956144_Big_Data_Challenges_Opportunities_and_Realities
21. https://www.analyticsinsight.net/5-challenges-of-big-data-analytics-in-2021/
22. https://www.slideshare.net/GeneMooLee/big-data-analytics-challenges-and-opportunities
23. https://www.hindawi.com/journals/tswj/2014/712826/?utm_source=google&utm_medium=cpc&utm_campaign=HDW_MRKT_GBL_SUB_ADWO_PAI_DYNA_JOUR_Complexity_X0000_Geotarget&gclid=Cj0KCQjw8amWBhCYARIsADqZJoUloW14Ng1zzgxuuztUdDoivf4unHY1hXSXjp7EHADIBIKmIKcvCGMaAlkpEALw_wcB