Big Data and Business Analytics: Trends, Success Factors, Challenges and Applications

Dr. Manish Kumar

Assistant Professor,University Department of Computer Applications,
Vinoba Bhave University, Hazaribag, Jharkhand-825301, India
Manish14.kk@gmail.com

Dr. Ashish Kumar Saha

Associate Professor,University Department of Chemistry,
Vinoba Bhave University, Hazaribag, Jharkhand-825301, India
Kumar_a_saha@yahoo.com

ARTICLE DETAILS	ABSTRACT
	Massive volumes of data have been transformed and utilized for decision-making in the information age. The word "big data" describe data sets that are tough to handle with usual tools and method because they are not only huge but also diverse and vibrant. The rapid growth of these data sets necessitates the study and provision of solutions for managing and extracting them with knowledge and value. Additionally, decision makers must extract useful information from such varied and quickly changing data, including social media data, daily transactions, and customer interactions. Big data analysis, or the use of sophisticated big data analysis techniques, can offer this value. This paper's goal is to provide a concise synopsis of big data analytics' functions, advantages, difficulties, and applications.

I. Introduction

The introduction of new devices, technologies, and communication channels, such as web logs, Smartphone's, social media site, satellite imagery, human genomics, customer communication,

astronomical and biological report, and more, has led to a rapid annual increase in the amount of data created by humans in the digital world. A collection of enormous data sets known as "big data" are inaccessible using standard computational techniques. It has developed into an integral theme encompassing a range of tools, techniques, and frameworks rather than being a single technique or tool. A computer can operate with quantities, letters or symbols that can be recorded on magnetic, optical or mechanical media and stored and transmitted as electrical signals [1]. Big Data is a huge collection of data that continues to grow over time. It is a fact that the data set is so large and complex that it cannot be processed or stored using any standard data management technology. Regular data and big data are comparable, but big data is much larger. Large amounts of unstructured, semi-structured or structured data are called "big data". We cannot identify a single source of origin for the data because it is collected from multiple sources, including social networks, websites, applications, commercial transactions, images, videos and search engines. To obtain valuable insights that will aid in the organization's growth, this data is gathered, documented, saved, and examined [2]. Big data is a lot.

Big data is characterized by 3 V's [3].

Data Volume: The term "data volume" describes the unfathomable volumes of information produced every second by various sources, including social media, mobile phones, automobiles, credit cards, M2M sensors, photographs, videos, and more. We currently store data in multiple locations using distributed systems, which are connected by a software framework such as Hadoop. On Facebook alone, over 350 million new posts are uploaded every day, 42.5 billion "like" buttons are recorded, and roughly billion messages are generated.

Data Rate: Data rate is the rate at which data is generated. Only Big Data technologies can manage such large amounts of data. Taking social media as an example, every day, 900 million photos are uploaded to Facebook, 500 million tweets are posted to Twitter, 0.4 million hours of video are uploaded to YouTube, and 3.5 billion searches are performed on Google. It's comparable to a nuclear data explosion. To avoid bottlenecks, big data helps the organization manage this explosion, accept the incoming data flow and process it quickly.

Data variety: The term "big data variety" describes semi-structured, unstructured, and structured data collected from various sources. Previously, data could only be collected from databases and

spreadsheets, but now it can be found in a wide variety of formats, including emails, PDFs, images, videos, audio files, social media posts and much more. One of the crucial aspects of Big Data is variety.
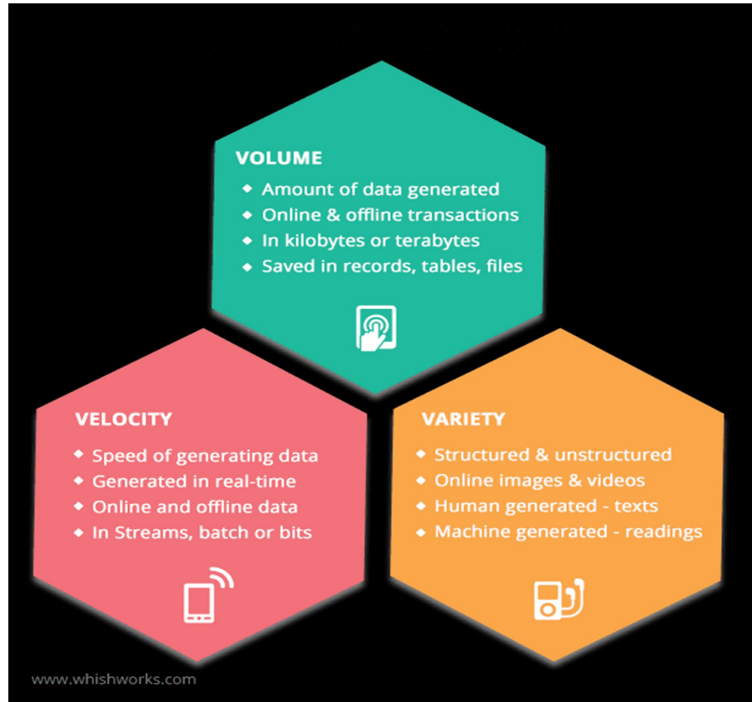

Figure-1 V's of Big Data

Key enablers for the appearance and growth of 'Big-Data' are:

- Increasing storage capacity

- processing power

- data availability

A. Big Data Analytics Technologies and Tools [4].

Big Data Analytics is based on several technologies and tools that may sound complex:

Hadoop: Think of Hadoop as a massive digital warehouse. Businesses such as Amazon utilize it to effectively store vast amounts of data. For instance, Hadoop is used by Amazon to manage your past purchases and make product recommendations that you might find interesting.

Spark: Consider Spark to be your extremely quick data cook. It is used by Netflix to swiftly evaluate your viewing habits and suggest your next show that is worth bingeing on.

NoSQL databases: Airbnb uses NoSQL databases, such as MongoDB, to store user information and booking details. Because of the speed and adaptability of these databases, the platform can provide you with the appropriate information at the appropriate time.

Tableau: Tableau is like an artist who creates stunning visuals out of data. In order to make interactive graphs and charts that make complex economic data easier to understand, the World Bank uses it. **R and Python:** For data scientists, R and Python are like magic tools. To solve challenging issues, they employ these languages. For instance, Kaggle uses them to forecast, using historical data, things like home prices.

II. Literature Review

The K. In Vengatesan (2021) [5], some important ideas regarding privacy and security issues unique to stock market study using time series data analytics technique are presented. According to the document Big Data Security/Privacy Issues and challenge, many big data users encounter difficulties managing daily big data operation at different phases of the big data ecological unit. The research done to address the primary issues and difficulties of big data security is explained in this article, which also clarifies some things to think about when working with big data. In order to bring privacy and security issues unique to big data back into the spotlight, Vishal Joshi 2020 [6] offers some important ideas. A thorough analysis of the problems and difficulties related to security, privacy, and confidentiality in big data and cloud computing is provided by R. Sumithra 2018 [7]. The advantages of a data warehouse, the extra data analysis capabilities of distributed file systems, and sound decision-making are all provided by big data. Big data has financial benefits. The big data code of conduct, cyber security, data privacy and integrity, and intellectual property rights were all addressed legally in this document. Jose Mura 2014 [1] Security and privacy concerns with big data. The security and privacy issues of today are covered in this document. Numerous sources of unstructured data exist, such as social media, sensors, scientific applications, surveillance, image and video files, Internet search indexing, and medical records. P. Kamakshi Dec 2014 [8] Privacy and big data research. The power and uses of big data, along with a number of privacy concerns, are covered in this article. Lo'ai A. In this article [9], Tawalbeh and Gokay Saldamli discuss current multi-tier cloud architectures, a big data storage system, the P2P Cloud System (P2PCS) for big data, and a hybrid mobile cloud computing model built on the cloudlet concept. They also use this model to process and analyze case studies in healthcare systems. Minit Arora and Dr. Himanshu Bahuguna (2016) [10] found that businesses used a range of data anonymization methods to ensure data security and privacy. The most widely used methods for ensuring security and privacy are verbal and written commitments. History has shown, however, that this strategy is not flawless. When sharing and aggregating data in dynamic, distributed data systems, low-level but widely used technical solutions like passwords, controlled access, and two-factor authentication can enhance security and

privacy. K. P. Maheswari, P. A study and analysis of the security threat levels in big data and cloud security is presented by Ramya and S.Nirmala Devi (2017) [11]. Big data-related issues are more severe in some industries and government initiatives. Because the network that links the systems is crucial, the security concerns with big data technologies and systems also apply to cloud computing. Jonesst. C., S. Maria Celestin Vigila, and JL Jonesston Dhas. Ezhil Star (2017) [12] created a framework for protecting privacy and security when archiving health data using big data. There are numerous immediate issues with storing medical records as big data. One of these is cloud data protection. Identifying the record and safeguarding health information from unauthorized users are explained here.

III. Methodology

An important tool for uncovering potential in sizable and intricate data sets is data analytics. The following primary steps comprise the big data methodology [13], [14]:

1) Data Collection: The basis of big data analytics is data. It is the congregation of information from various sources, including social network, sensors, survey, and customer opinion. Accurate data collection is the primary objective of data collection. Knowledge will increase with more data.

2) Data Cleaning (Data Preprocessing): dealing out this data comes next. It often needs to be cleaned. This entails deleting duplicates, fixing errors, and replacing missing data. It's similar to sorting through a wealth trove, removing the rocks and other debris to reveal only the priceless jewels.

3) Data processing: This will be our next topic. Inscription, organize, and format the data so that it can be used for analysis are crucial steps in this process. It's similar to a chef assembling ingredients before preparing meals. Transforming data into a format that can be processed with analytical tools is known as data processing.

4) Data analysis: To take out the most important findings from the process data, data study is carried out using arithmetical, numerical, and machine learning techniques.

5) Data visualization: Interactive tables, graph, and dashboard are a few examples of how data analysis is typically displayed visually. Large volumes of data could be made simpler with the help of visualizations.

IV. Big Data Analytics Challenges and Solutions

The security of private data is receiving more attention every day. This helps us to understand that security is a top priority for any business. However, it is now simple to access confidential data across various platforms due to the widespread adoption of web-based, smartphone, and cloud-based applications. Particularly if they are not adequately maintained, these platforms are extremely susceptible to hackers. In contrast to earlier times, businesses now gather and utilize a lot of customer data. A company's reputation could be jeopardized and major security problems could result from inadequate data security. A company's reputation and financial standing can drastically shift in the face of Big Data.

A. Big Data Analytics Challenges [15]

Despite the enormous advantages of big data analytics, there are a number of drawbacks as well.

1) Data Overload: Take Twitter, which posts about 6,000 tweets every second. Sorting through this deluge of data to uncover insightful information is the difficult part.

2) Data quality: The information produced by Big Data Analytics may be inaccurate if the input data is imprecise or lacking. Inaccurate sensor readings, for instance, may result in inaccurate weather forecast conclusions.

3) Privacy issues: Given the volume of personal information utilized, for example in Facebook ad targeting, there is a thin line separating creating individualized experiences from infringing on privacy.

4) Security risks: Safeguarding sensitive information is essential given the rise in cyberthreats. Banks, for instance, employ big data analytics to identify fraudulent activity, but they must also safeguard this data from security breaches.

5) Costs: Big Data Analytics system implementation and upkeep can be costly. Airlines like Delta must make sure the advantages of using analytics to optimize flight schedules outweigh the drawbacks.

B. Big Data Analytics Strengthening Mechanism [16]

Here are some ways to strengthen Big Data security:

1) Data encryption: Big data security depends on it. By preventing unwanted access or manipulation, it improves data security during processing, transmission, and storage. It also transforms data into code that needs to be decrypted.

2) Access Control: Makes sure that only authorized individuals can interact with particular data by managing access to data and actions through user roles, permissions, and authentication.

3) Data Masking: Replace private information with encrypted or fake data to protect sensitive data. This helps preserve the confidentiality of sensitive data by preventing misuse and illegal access.

4) Data Loss Prevention (DLP): DLP uses technologies like network monitoring and data loss prevention software, as well as policy enforcement and monitoring, to stop data loss or leakage, whether it is unintentional or deliberate.

5) Safe data storage: Use encryption, backups, disaster recovery plans, and secure systems to safeguard data while it's at rest.

6) Big Data Encryption: Data encryption is another popular method for protecting data. It is employed to protect Big Data's privacy. It should be mentioned that, in contrast to conventional encryption methods, homographic encryption enables you to encrypt data computation.

As a result, this approach guarantees the confidentiality of information while permitting its extraction through computation and analysis. It is crucial to remember that big data security architecture is a complicated and dynamic area that requires constant evaluation by organizations.

V. Benefits of Big Data Analytics

Big Data Analytics offers a series of concrete advantages [17][18]:

1) Well-informed Choices: Consider a Wal-Mart-type store. When deciding which products to stock, they use big data analytics to make informed choices. This not only lowers waste but also maintains customer satisfaction and increases profits.

 2) Think Amazon for better customer experiences. Big Data analytics is the source of the accuracy of manufactured goods recommendations. Having a personal shopper who understands your preferences and assists you in finding what you're looking for is similar to that.

3) Fraud Detection: Big Data Analytics are used by credit card companies, such as MasterCard, to identify and stop fraudulent activity. Having a guard who looks after and protects your money is similar to that.

4) Streamlined logistics: FedEx, for instance, uses Big Data Analytics to move its packages more quickly and with less environmental impact. It's similar to being environmentally conscious while pursuing the quickest path to your goals.

5) Targeted Promotions: Businesses can offer customized products to their target market using big data, saving a lot of money on unsuccessful marketing campaigns. Big data allows businesses to track in-store and online sales to analyze consumer trends. After that, this data is utilized to develop focused advertising campaigns that help companies connect with potential consumers and foster brand loyalty.

VI. Usage of Big Data Analytics

Big Data Analytics offers a number of tangible benefits [17] [18]:

1) Informed Choices: Consider a retailer like Walmart. They use big data analytics to help them decide which products to stock. This keeps customers happy, increases profits and reduces waste.

2) Increased customer satisfaction: consider Amazon. Big Data Analytics is the source of high accuracy in product recommendations. Having a personal shopper who understands your preferences and helps you find what you're looking for is similar to this.

3) Fraud detection: Credit card companies like MasterCard use Big Data Analytics to identify and block fraudulent transactions. It's like having a guardian who keeps your money safe.

4) Optimized logistics: FedEx, for example, uses Big Data Analytics to reduce its environmental impact and deliver packages faster. It's how to use the fastest way possible while preserving the environment.

5) Targeted and targeted promotions: Companies can use big data to offer personalized products to their target market without having to spend a fortune on failed marketing campaigns. Big Data allows companies to monitor in-person and online transactions to analyze consumer trends. Then, using this data, brands can develop targeted advertising campaigns that help them delight consumers and build brand loyalty

VII. Conclusion

Big Data is the driving force behind a business's expansion and prosperity, but it also presents significant privacy issues. The risk of privacy invasion increases with data volume, and protecting it requires more work. In the event that a data breach occurs, such as when a cyber attack targets an airline database and the attacker looks for user information, all of the credit card information ends up in the hands of the attacker and could cause significant harm to the airline. However, there are concerns about the legitimacy of such an airline. Due to the harm this attack will do to the airline's reputation, its operations may come to an end. No matter how big or small, how important or not, this incident makes it abundantly evident that protecting big data by overcoming challenges is crucial. This article also provides a concise synopsis of big data analytics' operation, advantages, difficulties, and applications. Based on numerous studies conducted in recent years, it would be accurate to state that Big Data technology is becoming more prevalent in practically every industry.

References

[1] He, B., Patel, M., Zhang, Z., Chang, K. C.-C. (2007). Accessing the Deep Web. Communications of the ACM 50(5), 94–101.

[2]. Koch, C. (2001). Data Integration against Multiple Evolving Autonomous Schemata. CERN-THESIS-2001-036.

[3]. Michel, F., Montagnat, J., Faron-Zucker. C. (2014). A survey of RDB to RDF translation approaches and tools.

[4]. Ray, S., Bandyopadhyay, S. and Pal, S. (2009). Combining Multisource Information Through Functional-Annotation-Based Weighting: Gene Function Prediction in Yeast. IEEE Transactions on Biomedical Engineering, 56(2), pp.229-236.

[5] K.Vengatesan, A. Kumar, A.K, Kharade, & R.K, Kamat, (2021). Stock Market Analysis using Time Series Data Analytics Techniques. 2021 International Conference on Computing, Communication and Green Engineering (CCGE), 1–5. https://doi.org/10.1109/CCGE50943.2021.9776372

[6] M. V. Joshi, "Security/Privacy Issues and Challenges in Big Data,"International Research Journal of Engineering and Technology (IRJET),vol. 07, no. 06, 2020.

[7] R. Sumithra, "Security, Privacy Issues and Challenges in Big Data and Cloud," Special Issue based on Proceedings of 4th International Conference on Cyber Security (ICCS), 2018.

[8] P.Kamakshi, "SURVEY ON BIG DATA AND RELATED PRIVACY ISSUES," International Journal of Research in Engineering and Tech- nology, vol. 03, no. 12, Dec 2014.

[9] L. A. T. a. G. Saldamli, "Reconsidering big data security and privacyin cloud and mobile cloud systems," Journal of King Saud University – Computer and Information Science, 2019.

[10] M. A. a. D. H. Bahuguna, "Big Data Security – The Big Challenge," International Journal of Scientific Engineering Research, vol. 7, no. 12, Dec 2016.

[11] P. a. S. D. K.P.Maheswari, "STUDY AND ANALYSES OF SECURITY LEVELS IN BIG DATA AND CLOUD COMPUTING," International Journal of Innvative Research in Science and Engineering, vol. 3, no.02, 2017.

[12] S. M. C. V. a. C. E. S. J.L. Joneston Dhas, "A Framework on Security and PrivacyPreserving for Storage of Health Information Using Big Data," International Science Press, 2017.

[13] Ruggles, S., Hacker, J. and Sobek, M. (1995). General Design of the Integrated Public Use Microdata Series. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 28(1), pp.33-39

[14] Sequeda, J., Tirmizi, S. H., Corcho, Ó., Miranker, D. P. (2011). Survey of directly mapping SQL databases to the Semantic Web. Knowledge Eng. Review 26(4), 445–486.

[15] Smith, J., Bernstein, P., Dayal, U., Goodman, N., Landers, T., Lin, K. and Wong, E. (1981).Multibase. Proceedings of the May 4-7, 1981, national computer conference on - AFIPS '81.

[16] Spanos, D. E., Stavrou, P. Mitrou, N. (2012). Bringing Relational Databases into the Semantic Web: A survey. Semantic Web Journal 3(2), 169–209.

[17] J. C. Ogbonna, F. O. Nwokoma and A. Ejem, "Database Security Issues: A Review," International Journal of Science and Research, vol. 6, no. 8, 2015.

[18]P. Praveen , B. Rama ,Ch. Jayanth Babu 2016," Big dataenvironment for geospatial data analysis" International Conference on Communication and Electronics Systems(ICCES2016),DOI: 10.1109/CESYS.2016.7889816.