WWJMRD 2019; 5(1): 110-115
www.wwjmrd.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
Impact Factor MJIF: 4.25
E-ISSN: 2454-6615

Jeffrey John R. Yasay
Tarlac Agricultural University
Camiling, Tarlac, Philippines.

Big Data Analytics: A Gear towards Technology Advancement

Jeffrey John R. Yasay

Abstract
The purpose of this paper is to highlight the usefulness of big data analytics. The design of highly scalable algorithms and systems to integrate data and uncover large hidden values from diverse, complex, and massive datasets poses a significant challenge in big data analytics. It is also noted that there is no single technology that encompasses big data analytics. Moreover, it is worth noting that the benefits and challenges of big data analytics are sophisticated.

Keywords: Big Data Analytics, Scalable, Innovations.

1. Introduction

The Big data research has generated a lot of interest in science, technology, industry, and society. It is emerging in tandem with the growing science paradigm, technological transformative innovation, the development of new technologies, and societal transformation. In the Big Data age, it is becoming increasingly critical and difficult to safely store and analyze vast amounts of data. Big Data is often processed and queried using a cloud computing platform. Cloud computing can offer a variety of elastic and scalable IT services on a pay-as-you-go basis, but it also introduces privacy and security concerns. Adversaries can also use Big Data correlation and background knowledge to steal critical information. However, data correlations in Big Data are extremely complicated and distinct from standard interactions between data contained in relational databases [1].

Big data technology is commonly used to refer to three aspects of technological innovation and super-large datasets: automated parallel processing, data management techniques, and data mining [2].

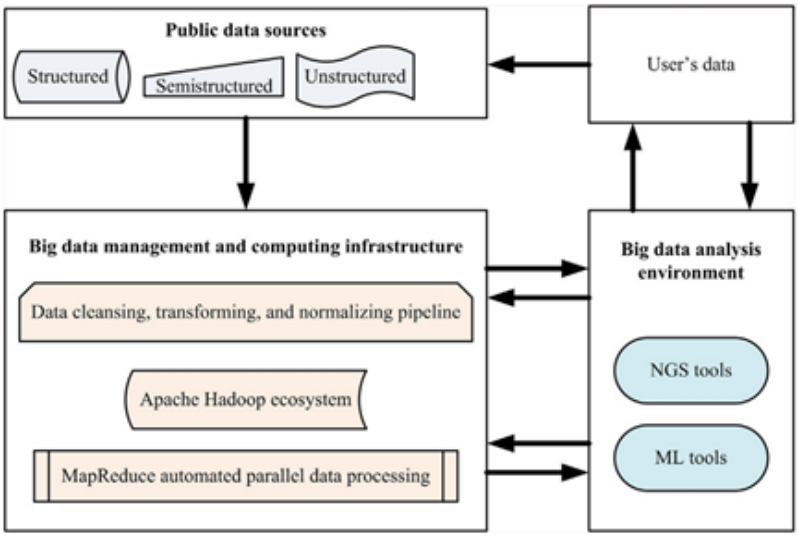

Fig. 1: General Components of the Big Data Technology [3]

Correspondence:
Jeffrey John R. Yasay
Tarlac Agricultural University
Camiling, Tarlac, Philippines.

Big Data is described as amounts of data with variable degrees of complexity, created at varying rates, and with varying degrees of ambiguity that cannot be handled using existing technologies, processing methods, algorithms, or commercial off-the-shelf solutions [4].

2. Advantages of Big Data Analytics

a. Risk Management

The Federal Information Security Management Act defines information security as "the protection of information and information systems against illegal access, use, disclosure, interruption, alteration, or destruction" in order to ensure their confidentiality, integrity, and availability" [5]. Risk management is a field that helps organizations detect, analyze, and respond to risk. Information security is one technique to decrease risk, and in the larger context of risk management, information security management is concerned with lowering information system-related risk to an acceptable level for the company [6].

Fig. 2: Information Security Risk Management Must Occur.

At and Between All Levels of the Organization to Enable Pervasive Risk Awareness and to Help Ensure Consistent Risk-Based Decision Making Throughout the Organization [7]

b. Product Development and Innovations

This involves the design of a high-quality product as well as manufacturing methods to ensure consistent manufacture of a high-quality product. Product development documentation describes all of the methods and requirements associated with the development of the product. The documentation procedure provides reviewers and inspectors with information about the product and production processes. Documentation regulates or reduces the risks, misunderstandings, and potential errors that may arise throughout the product development process [8].

While innovation is an iterative process aiming at the development of new goods, processes, information, or services via the application of new or even current knowledge [9]. Data-driven innovation is the use of any type of data to generate value during the innovation process [10]. The increasing trend of big data-driven innovation is enabling data-driven planning, data-driven marketing, and data-driven operations across all industrial sectors and domains [11].

c. Quicker and Better Decision making

In today's world, corporate leaders face high consumer expectations, fierce competition, rising labor and material prices, and shorter product lifecycles. Globalization is blurring national boundaries. Location and distance from the market are no longer impediments to market entry. Firms must continually monitor for risks and possibilities in such a turbulent market and make business choices fast based on available facts [12].

When business transaction data is mined for association rules, it can give critical insights for decision makers about products purchased together or forecast demand for certain things [13].

Fig. 3: Conceptual framework on big data and decision making [12].

3. Big Data Challenges

a. Synchronization across Disparate Data Sources

As data sets get larger and more diversified, incorporating them into an analytical platform becomes increasingly difficult. Complete synchronization and incremental synchronization are the two major data synchronization strategies. Complete synchronization will transfer the data from all departments' databases to the central database

once, and the central database will update transmitted data synchronously, which has simple logic but needs a large network overhead. In contrast, incremental synchronization only requires the delivery of updated data to the target database to maintain data synchronization, which can reduce network overhead and therefore improve synchronization efficiency. The key to incremental synchronization is how to gather incremental data [14].

b. Deficiency of Specialists Who Understand Big Data Analysis

Data analysis is essential for making this massive quantity of data created per minute usable. With the exponential increase of data, the market has generated a large need for big data scientists and big data analysts. Because the role of a data scientist is interdisciplinary, it is critical for business companies to recruit a data scientist with a diverse set of abilities. Another significant difficulty that organizations confront is a scarcity of personnel that understand Big Data analysis. In compared to the vast amount of data being created, there is a severe scarcity of data scientists [15].

c. Data Storage and Quality

High-quality data are required for evaluating and utilizing big data, as well as ensuring the data's value. Currently, detailed study and research of quality standards and quality evaluation methodologies for large data are absent [16].

d. Security and Privacy of Data

Big data is frequently comprised of aggregated data from various sources that are not always identifiable. As a result, there is no process for requesting a person's consent for the resulting data, which is frequently more personal than the set of data the person would consent to give [17].

Big data is more likely to have a negative impact on the welfare of unsophisticated, vulnerable, and technologically illiterate consumers. Such consumers may be unaware of multiple information sources and are less likely to receive current and accurate information about multiple suppliers in a way that allows for effective search and comparisons. They are also unable to assess the sensitivity of their online actions, making them more susceptible to deception by malicious actors. A number of big data applications are currently hampered by regulatory uncertainty. Because regulatory institutions are underdeveloped, there is a need for a firm-level big data policy that considers the degree of sensitivity of information used in predictive modeling. Most organizations, however, have not developed best practices to ensure the privacy and security of customer data [18].

Information privacy also includes the regulations required for companies to protect data. And as more data protection regulation grows worldwide, global privacy requirements and demands will also expand and change. However, the one constant is adequate data protection: it’s the best way to ensure that companies are both complying with the law and guaranteeing information privacy [19].

4. Different Types of Big Data Analytics

a. Descriptive Analytics

Descriptive analytics is primarily used to describe data characteristics such as what, who, when, and how. It is also the traditional form of Business Intelligence and data analysis, seeks to provide a depiction or "summary view" of facts and figures in an understandable format, in order to either inform or prepare data for further analysis. To report on past events, it employs two main techniques: data aggregation and data mining. It presents historical data in an easily digestible format for the benefit of a broad business audience [20].

b. Diagnostic Analytics

Diagnostic analytics is used to figure out why something occurred in the past. Drill-down, data discovery, data mining, and correlations are some of the techniques used. Diagnostic analytics examines data in greater depth to determine the underlying causes of events. It aids in determining what factors and events influenced the outcome. The analysis is primarily based on probabilities, likelihoods, and the distribution of outcomes. Diagnostic analytics would help you understand why sales have decreased or increased for a specific year or so in a time series data of sales. However, the ability of this type of analytics to provide actionable insights is limited. It only provides a backward-looking understanding of causal relationships and sequences [21].

c. Predictive Analytics

Predictive Analytics is a subset of advanced analytics that is used to forecast unknown future events. Predictive analytics investigates current findings to make predictions about the future using many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence. The predictor, which is defined as a variable used to predict future behavior, is the main entity in predictive analytics. The predictors forecast future probabilities with high accuracy. Machine learning techniques and regression techniques are the approaches used to conduct predictive analytics. Because of their outstanding performance in managing large scale datasets with uniform characteristics and noisy data, machine learning techniques have become increasingly popular in predictive analytics [22].

d. Prescriptive Analytics

A collection of mathematical techniques used to compute a set of high-value alternative actions or decisions given a complex set of objectives, requirements, and constraints, with the goal of improving business performance [23]. Prescriptive analytics is thought to be the next frontier in business analytics. It offers organizations adaptive, automated, and time-dependent courses of action for capitalizing on potential business opportunities. Given the objectives of the enterprise, prescriptive analytics helps it maximize its business values while mitigating its potential risks by recommending optimal action sequences [24].

5. Big Data Analytics Tools

a. Hadoop

An open-source software framework for storing and processing large data sets on commodity hardware clusters. Hadoop is a top-level Apache project that is being built and used by a global community of contributors and users. It is distributed under the terms of the Apache License 2.0. [25].

Fig.4: Applications Run Natively in Hadoop [25].

b. Talend

This is used to influence how the world makes decisions. Working with data and making it useful to all organizations the information marketplace was at a crossroads in 2005. While the rest of the industry was gravitating toward complex and expensive integration products, Talend founders Fabrice Bonan and Bertrand Diard saw an alternative: something simple, powerful, flexible, and scalable. Their solution was the first open-source commercial data integration software [26].

c. Cassandra

It is a NoSQL distributed database. By design, NoSQL databases are lightweight, open-source, non-relational, and largely distributed. Counted among their strengths are horizontal scalability, distributed architectures, and a flexible approach to schema definition. NoSQL databases enable rapid, ad-hoc organization and analysis of extremely high-volume, disparate data types. That's become more important in recent years, with the advent of Big Data and the need to rapidly scale databases in the cloud. Cassandra is among the NoSQL databases that have addressed the constraints of previous data management technologies, such as SQL databases. Spark - used for real-time processing and analyzing large amounts of data [27].

d. Storm

This refers to the distributed real-time computation system that is open source and free to use. Apache Storm makes it simple to process unbounded streams of data in a reliable manner, effectively replacing Hadoop in real-time processing. Apache Storm is easy to use, works with any programming language, and is a lot of fun to play with. There are numerous applications for Apache Storm, including real-time analytics, online machine learning, continuous computation, distributed RPC, ETL, and others. Apache Storm is fast: a benchmark found it to process more than a million tuples per second per node. It is scalable, fault-tolerant, ensures data processing, and is simple to set up and use [28].

6. Statistics of Big Data Analytics in Industry

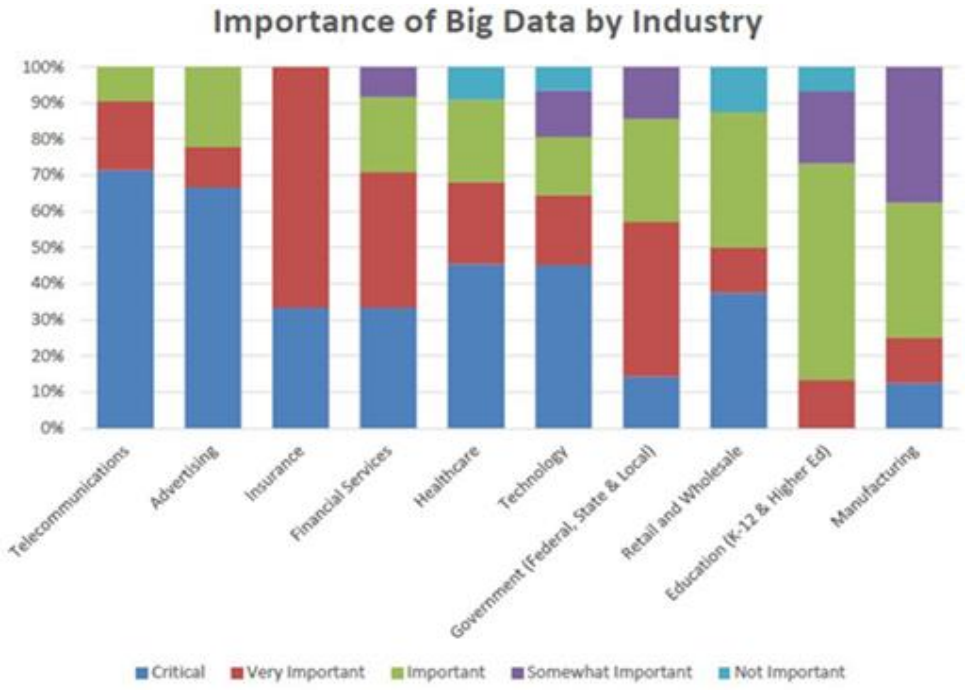

Fig. 5: Importance of Big Data by Industry [29].

The three industries that claim big data analytics is the most crucial to their business intelligence projects are telecommunications, advertising, and insurance. Big data analytics is also important to the healthcare and retail and wholesale industries' continuous business intelligence efforts. Big data is vital to 80 % of all businesses' business intelligence activities, at the very least [29].

Conclusion

The field of big data analytics is gaining traction. Enterprises have realized that their big data repositories are a largely untapped gold mine that can assist them in cutting costs, increasing revenue, and becoming more competitive. They do not just want to collect massive amounts of data; they want to transform that data into meaningful insights that will help them improve their businesses. As more data is generated and collected, data analysis requires scalable, flexible, and high-performance tools to provide timely insights. Organizations, on the other hand, are confronted with a growing big data ecosystem in which new tools emerge and quickly become obsolete. As a result, keeping up with the pace and selecting the appropriate tools can be difficult.

References

1. Ou, L. (2016). Big Data || Security and Privacy in Big Data. , (), 285–308. doi:10.1016/b978-0-12-805394-2.00012-x
2. [Gharajeh, Mohammad Samadi (2017). [Advances in Computers] || Biological Big Data Analytics. , (), –. doi: 10.1016/bs.adcom.2017.08.002
3. C. Ma, H.H. Zhang, X. Wang, Machine learning for big data analytics in plants, Trends Plant Sci. 19 (2014) 798–808.
4. Zhuge, Hai (2016). Multi-Dimensional Summarization in Cyber-Physical Society || Introduction. , (), 1–19. doi:10.1016/B978-0-12-803455-2.00001-9
5. Federal Information Security Management Act of 2002, Pub. L. No. 107-347, 116 Stat.2946. §301, codified at §3542(b)(1).
6. [Gantz, Stephen D. (2013). FISMA and the Risk Management Framework || Risk Management. , (), 329–365. doi:10.1016/B978-1-59-749641-4.00013-8
7. Managing information security risk: organization, mission, and information system view. Gaithersburg, MD: National Institute of Standards and Technology, Computer Security Division; March 2011. Special Publication 800-39. p. 9.
8. Soni, Vandana (2018). Dosage Form Design Parameters || Documentation Protocol in Product Development Including Clinical Records. , (), 403–440. doi:10.1016/B978-0-12-814421-3.00012-9
9. Kusiak, A. (2009). Innovation: A data-driven approach. International Journal of Production Economics, 122(1), 440–448. doi: 10.1016/j.ijpe.2009.06.025.
10. Stone, D., & Wang, R. (2014). Deciding with data – How data-driven innovation is fuelling Australia's economic growth. PricewaterhouseCoopers (PwC). http://www.pwc.com.au/consulting/assets/publications/Data-drive-innovation-Sep14.pdf
11. Zillner S. et al. (2016) Big Data-Driven Innovation in Industrial Sectors. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_9
12. S. Jeble, S. Kumari, and Y. Patil, "Role of big data in decisionmaking," Operations and Supply Chain Management: Inter-national Journal, vol. 11, no. 1, pp. 36–44, 2017.
13. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. Decision support systems, 31(1), pp. 127-137.
14. Z.Y. Lu, Z.B. Guo (2015) A Method of Data Synchronization Based on Message Oriented Middleware and Xml in Distributed Heterogeneous Environments. AIIE 2015
15. Kumar, A. (2018, June 8). 7 Top Big Data Analytics Challenges Faced By Business Enterprises. eLearning Industry. Retrieved June 10, 2018, from https://elearningindustry.com/big-data-analytics-challenges-faced-business-enterprises-7-top
16. Cai, L. and Zhu, Y., 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14, p.2. DOI: http://doi.org/10.5334/dsj-2015-002
17. Pirlot, A. (2014, January 21). Big data: A tool for development or threat to privacy?. Retrieved from ⟨https://www.privacyinternational.org/blog/ big-data-a-tool-for-development-or-threat-to-privacy⟩.
18. Kshetri, Nir (2014). Big data's impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134–1145. doi: 10.1016/j.telpol.2014.10.002
19. Petters, J. (2018, September 28). Data Privacy Guide: Definitions, explanations and legislation. Retrieved October 15, 2018, from https://www.varonis.com/blog/data-privacy
20. Ghosh, P. (2017, July 06). Fundamentals of Descriptive Analytics. Retrieved November 28, 2018, from https://www.dataversity.net/fundamentals-descriptive-analytics/
21. Mehta, A. (2017, October 13). Four types of business analytics to know. Retrieved August 10, 2018, from https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/
22. B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 492-499, doi: 10.1109/ICCONS.2017.8250771.
23. Song S., Jeong DH., Kim J., Hwang M., Gim J., Jung H. (2014) Research Advising System Based on Prescriptive Analytics. In: Park J., Pan Y., Kim CS., Yang Y. (eds) Future Information Technology. Lecture Notes in Electrical Engineering, vol 309. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55038-6_89
24. Soltanpoor R., Sellis T. (2016) Prescriptive Analytics for Big Data. In: Cheema M., Zhang W., Chang L. (eds) Databases Theory and Applications. ADC 2016. Lecture Notes in Computer Science, vol 9877. Springer, Cham. https://doi.org/10.1007/978-3-319-46922-5_19
25. Bappalige, S. P. (2014, August 26). An introduction to apache hadoop for big data, from https://opensource.com/life/14/8/intro-apache-hadoop-big-data
26. Healthy Data, healthy business - modern cloud ETL. (n.d.). Retrieved September 19, 2018, from https://www.talend.com/
27. Cassandra Basics. (n.d.). Retrieved August 20, 2018, from https://cassandra.apache.org/
28. Apache Storm. (n.d.). Retrieved September 6, 2018, from https://storm.apache.org/index.html

29. Columbus, L. (2018, December 23). Big Data Analytics adoption soared in the enterprise in 2018. Retrieved December 23, 2018, from https://www.forbes.com/sites/louiscolumbus/2018/12/23/big-data-analytics-adoption-soared-in-the-enterprise-in-2018/?sh=1d105097332f