Big Data Concepts, Opportunities, Challenges, and Paradoxes

Stepani Sisca Wulandari[1]

[1]Accounting Department, Faculty of Economics and Business, Perbanas Institute, Jakarta, Indonesia
Email: sisca.wulandari@perbanas.id

Abstract: Big data is becoming increasingly important because it is starting to be widely used for analysis in business and government sectors. Apart from that, big data has also started to be utilized in new fields such as Artificial Intelligence, eCommerce, and Fintech. Analysis using big data produces more complete and accurate information than using less data. Although the opportunities opened up by big data are becoming more numerous, big data management is different and more difficult than traditional data management. Apart from that, there is big data paradoxes that must be understood to be more careful in using big data. The contribution of this paper is to provide big data literature, namely the concept of big data, the opportunities obtained by utilizing big data, the challenges of using big data, and the paradoxes of big data.

Keywords: big data; concepts; opportunities; challenges; paradoxes

1. BACKGROUND

In the information era, data is generated not only by companies and governments, but also by individual in society. It creates huge amounts of data, and with the increasing in storage capabilities and methods of data collection, that huge amounts of data have become easily available to end users (Chong and Shi, 2015). Accuracy in big data may lead to more confident decision making, and better decisions can result in greater operational efficiency, cost reduction, and reduced risk (Jain, 2017).

The contribution of this paper is to provide the literature on big data (the concepts, opportunities, challenges, and paradox of big data). The use of big data for analysis in the business and government sectors is becoming increasingly significant due to its growing prevalence. More and more institutions need knowledge and insight regarding big data, so it is necessary to provide big data literature. The literature on big data mostly were found to be journal papers, conference papers, and books.

The remainder of the paper is structured as follows: first, the description about big data. Second, opportunities that can be obtained from the use of big data. Third, the challenges in managing big data. Fourth, the paradoxes that accompanies the journey of big data. At the last section, a research conclusion.

2. LITERATURE REVIEW

I. Description of Big Data

Big data is a term to describe a very large volume of data that cannot be processed by conventional database systems. The availability of abundant digitally recorded data will be needed for decision making in various companies.

Big Data consists of a collection of data that has various forms, such as text, numbers, audio, video, images, and others that originate from facts but do not have meaning. These data will be processed into an information. The definition of Big Data refers to three characteristics, namely the 3V: volume, variety, velocity. In addition, there are those who add other V elements such as veracity and value (Chandarana and Vijayalakshmi, 2014).

- Data volume or capacity is the size of data storage media that can accommodate up to zettabytes. The amount of digital data generated every second through websites, portals, and online applications is called volume (Bagiwa, 2017).
- Furthermore, there is a variety or variation of data which refers to the diversity of data types. Variety in big data refers to structured, semi-structured, and unstructured data created by humans or machines (Misra and Bera, 2018).

- Velocity is the speed of processing data in real time. While veracity and value are truth and value related to data uncertainty and the value of the benefits of the information (Rajaraman, 2016).
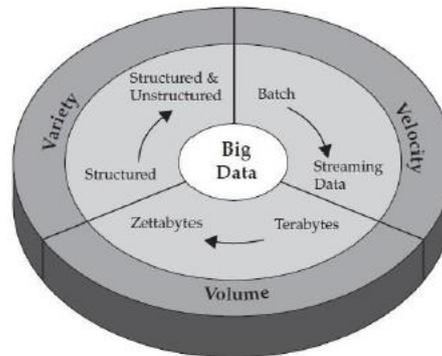

Figure 1. Three characteristics of big data

II. Opportunities from Big Data

The benefits of Big Data in companies include knowing consumer responses and behaviour towards products, helping to make more valid and accurate decisions from lots of structured data obtained, and helping to improve corporate image by knowing consumer desires and market trends through data. Companies can serve consumers better by using existing data in the company.

Big Data is of course also going to change various aspects of life, especially in the field of marketing. Marketing, which so far has focused more on attracting market attention (attention seeking), will change its focus to finding customers (subscriptions). Companies will form and maintain relationships with customers and use the data they collect to sell products and services according to the customer needs. So that companies can sell products and deliver services in accordance with the preferences, habits, or needs of consumers and create a personalized product or service.

The opportunity to utilize Big Data is very potential in various sectors, both the business sector as well as the public. Using better analysis of large volumes of data makes progress faster in many areas, including:

- In biology, scientific data can be stored in a public data warehouse, and creates a public database for use by other scientists (Ammu and Irfanuddin, 2013).
- Using big data is revolutionizing urban planning, smart transportation, environmental modelling, as well as energy saving (Ammu and Irfanuddin, 2013).
- Health sector: in clinical data, pharmaceutical data, data on personal practices (including dietary habits, exercise patterns) (Kambatla et al., 2014). Advance analytical techniques can also gain insight into genetic and environmental causes of diseases (Chandarana and Vijayalakshmi, 2014).
- New models of governance, such as mobile electronic voting will become viable. The convenience of such models of governance motivates stronger involvement of the public in government decisions and processes. Fraud detection is another significant analytics application (Kambatla et al., 2014).

The use of big data analysis in the business sector is a strategic step for companies, including (Kambatla et al., 2014):

- Companies can take advantage of the information obtained from access to social media data from search engines make it possible to adapt their business strategy to the opinion of consumers.

- Companies can improve their customer service, where big data and data analysis technologies are used to read and evaluate consumer response.
- Early identification if there is a risk to the product or service.
- Knowing the public's response to the products issued through sentiment analysis on social media.
- Helping companies make decisions more precisely and accurately based on data.
- Help to improve the company's image in the eyes of customers.

3. DISCUSSION

The challenges of managing big data

Organization dealing with big data facing a great number of challenges. Meeting challenges presented by big data will be difficult; volume of data increasing every day, velocity of its generation is increasing faster than ever; variety of data is also expanding. Big data has a number of risks as follows:

a. Data Security
    This type of risk is related to the logistics of data collection and analysis. Security data risk is clearly a crime that is increasingly growing with bigger attacks and more destructive. Five of the six most damaging data thefts all time done in the last two years. The more data you have, the more likely the data is sold and misused by irresponsible parties (Jain, 2017).
b. Data Privacy
    Privacy is an issue that is closely related to security issues. So that, it should be ensured that sensitive information stored or collected will not be misused by those who have been given the responsibility to analyse and report it (Jain, 2017).
c. Continuous Availability
    A certain amount of downtime is built-in to RDBMS and NoSQL systems. This is a challenge for big data because business cannot accept downtime, moreover a business that is run online must be live 24/7 continuously (Bagiwa, 2017).
d. Costs
    Data collection, aggregation, storage, analysis, and reporting costs money. Meeting even one of the challenges presented with RDBMS or most NoSQL solutions can be very expensive. These costs can be reduced by doing a careful budgeting during the planning stage. However, an error was made in the planning stage can lead to more expenses. It takes development of a good strategy to define what companies want to achieve and the benefits that can be obtained so that can be adjusted to the resources needed (Bagiwa, 2017).
e. Bad Analytics
    Bad analytics is also known as "getting it wrong". Misinterpreting patterns indicated by aggregated data and incorrectly identifying the causation where there actually a random chance condition. Other than that, bad analytics can form a confirmation bias, when the information obtained conflict with preconceived notions or information (Jain, 2017).
f. Performance
    Big data must move at extremely high velocities, since the world's sales compete in nanosecond. The data handling of RDBMS and most NoSQL solutions put a serious drag on performance (Bagiwa, 2017).
g. Bad Data
    Not a few companies are wrong by collecting irrelevant data, out of date or wrong. This is usually due to time to devising a project strategy is not enough. Bad Data is formed because of the company use the approach of "gathering everything, but analyse it later" which not only results in data storage costs is increasing, but also causes a large amount of data that can become wears out very quickly. If the company does not analyse proper data, then the company will not be able to draw the right insights that gives value to company (Jain, 2017).
h. Rules and Regulations

Regulations have evolved in response to the development of big data. In an increasingly globalized economy, the rules are becoming increasingly complex to financial institutions and for that matter, all organizations with customers or operations in more than one country, must comply with some regulations about the way of the company storing, accessing, and sharing personal information (Jain, 2017).

i.   Manageability

Most NoSQL solutions are trouble by operational complexity and hidden configurations. Even worse if you have to use RDBMS technology, because it is a costly, time-consuming, and often wasted effort (Bagiwa, 2017).

Privacy Threats with the Use of Big Data

Security of costumers' personal data should become a guarantee given by the company to its customers. When costumers feel that their personal data is kept securely, they will feel convenient to use the services provided by the company. Leaked personal data results in loss of convenience, security, and safety to use the company's services. Therefore, personal data leakage poses a privacy threat to consumers who have provide their personal data to the company.

Privacy is crucial for human life. With privacy each individual can maintain his identity within the limits of social life. The role of privacy in the social life of society is to manage information in and out of networks and data sets in society (Richards and King, 2016).

Privacy in big data includes identity protection, equality, security, and trust (Richards & King, 2016). The existence of privacy in big data can make it easier for customers. First, identity protection in big data can prevent recommendations based on behavioural analysis and individual preferences that can limit the costumer's space. Second, equality includes restrictions on collections, algorithmic transparency and accountability, as well as restrictions on the use of analytics for sorting and treat people differently. Third, security, guarantee against the absence of transfer and engineering of costumer's personal data. Fourth, trust, is given by costumers to the company when they feel there is a guarantee of security against privacy and personal data that has been provided to the company.

Privacy and legal issues pose challenges in the use of big data. Although big data has many benefits for companies, but on the other hand big data also threatens privacy. Big Data allows companies to better understand competitors and their customers as well as enabling the government to be more transparent. However, this kind of surveillance can cause misuse of information to influence, prevent or control.

In a digital society, privacy cannot exist without security and security cannot exist without privacy (Richards and King, 2016). With ease of moving data from one party to another, there is a need for precaution of data engineering. In addition, data quality also needs to be maintained. So, a mechanism is needed to maintain data integrity and prevent it from individual information engineering.

Big Data Paradoxes

According to Richards and King (2019), although proponents of big data say that by using big data analysis, people get insights that can help overcome many problems, but there has never been an analysis of the potential risks. In their article entitled "Three Paradoxes of Big Data", Richards and King (2019) outline the potential risks of big data critically. The paradox they explain regarding big data is:

➢   The Transparency Paradox

Big data consists of a lot of data that is collected invisibly, because the data is collected from sensors, mobile phones, click patterns, and others. The data obtained comes from connected equipment in the Internet of Things (IoT), where the equipment and sensors drive mobile data traffic that grows exponentially.

Big data promises to use these data to make the world more transparent, but it turns out that the data collection is not transparent because it takes private data from IoT.

➢ The Identity Paradox

Application users who utilize big data cannot independently choose the desired path from the application used. For example, Netflix users cannot choose which films to watch themselves because they are always influenced by films recommended by Netflix using big data analysis results (usually recommendations come from the history of films that users have watched, then big data analysis will recommend films the same genre as the movies the user has watched). This means that Netflix users (or other applications that use big data) are influenced by the recommendations from big data analysis, so they are unable to choose independently. The power of big data analysis to influence and persuade limits the independent identity of application users.

➢ The Power Paradox

Big data is widely described as a powerful tool that helps users see the world in a sharper and clearer way. However, big data sensors are mostly controlled by dominant institutions, not ordinary people. Big data creates winners, unfortunately it's the institutions that own the tools that benefit.

4. CONCLUSION

Big data is a term to describe a very large volume of data that cannot be processed by conventional database systems. Big Data consists of a collection of data that has various forms, such as text, numbers, audio, video, images, and others that originate from facts but do not have meaning. These data will be processed into an information. The opportunity to utilize Big Data is very potential in various sectors, both the business sector as well as the public. Organization dealing with big data facing a great number of challenges. Meeting challenges presented by big data will be difficult; volume of data increasing every day, velocity of its generation is increasing faster than ever; variety of data is also expanding. Privacy and legal issues pose challenges in the use of big data. Although big data has many benefits for companies, but on the other hand big data also threatens privacy. In their article entitled "Three Paradoxes of Big Data", Richards and King (2019) outline the potential risks of big data critically. The paradox they explain regarding big data is the Transparency Paradox, the Identity Paradox, and the Power Paradox.

REFERENCES

Ammu, N., & Irfanuddin, M. (2013). Big data challenge. *International Journal of Advanced Trends in Computer Science and Engineering*, *2*(1), 613–615.

Chandarana, P., & Vijayalakshmi, M. (2014). Big Data analytics frameworks. *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, 430–434. https://doi.org/10.1109/CSCITA.2014.6839299

Chong, D., & Shi, H. (2015). Big data analytics: a literature review. *Journal of Management Analytics*, *2*(3), 175–201. https://doi.org/10.1080/23270012.2015.1082449

Idris Bagiwa, L. (2017). Big Data: Concepts, Approaches and Challenges. *International Journal of Computer Networks and Communications Security*, *5*(8), 181–187. www.ijcncs.org

Jain, V. K. (2017). *Big Data and Hadoop*. Khanna Book Publishing. https://books.google.co.id/books?id=i6NODQAAQBAJ

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573. https://doi.org/10.1016/j.jpdc.2014.01.003

Misra, S., & Bera, S. (2018). Introduction to Big Data Analytics. *Smart Grid Technology*, 38–48. https://doi.org/10.1017/9781108566506.005

Rajaraman, V. (2016). Big Data Analytics (articulo en ingles). *Resonance*, *August*, 695–716.

Richards, N. M., & King, J. H. (2016). Big Data and the Future For Privacy. In *Research Handbooks on Digital Transformations*. Edward Elgar Publishing.

Richards, N. M., & King, J. H. (2019, April 30). *Three paradoxes of Big Data*. Stanford Law Review. Retrieved December 6, 2022, from https://www.stanfordlawreview.org/online/privacy-and-big-data-three-paradoxes-of-big-data/