Issues, Challenges and Solutions of Big Data in Information Management: An Overview

Syaiful Hisyam Saleh, Raihan Ismail, Zaharuddin Ibrahim, Norhayati Hussin

Issues, Challenges and Solutions of Big Data in Information Management: An Overview

Syaiful Hisyam Saleh, Raihan Ismail, Zaharuddin Ibrahim, Norhayati Hussin

Faculty of Information Management, Universiti Teknologi MARA, UiTM Selangor, Malaysia

Abstract

Big data is too voluminous and requires the use of computer technology and the right data processing application to capture, store effectively, analyse, and present big data to enable the business to have clearer visibility of trends, make plans and decision for future direction. The data revolution is reshaping the way knowledge is produced, business is conducted, humanitarian assistance is handled, public officials are elected, and governance is enacted. Big data provides unprecedented insights and opportunities across all industries, and it raises concerns that must be addressed.

This article discusses the challenges and solutions for big data as an important tool for the benefit of the public. It suggests that big data and data analytics if used properly, can provide real-time actionable information that can be used to identify problems and needs, offer services, and provide feedback on the effectiveness of policy action.

Introduction

The world is experiencing a huge data revolution. The explosion of data is directly connected to the arrival of the digital age. The term "Big Data" refers to the vast amounts of data in which traditional data processing procedures and tools would not be able to handle. It emerged in the 1990s and gained momentum in the early 2000s and has been variously defined and operationalized. Clearly, size often comes to mind when referring to big data. It is commonly defined as the astonishing amount of structured and unstructured data that are being generated, captured, and stored at an amazing speed.

It is obvious that data comes from literally everywhere, every time and from all sorts of devices. Data is often produced and accessible in real time, and it arises from the merging of different sources. Organizations have been relying on these sources of data to describe, interpret, and forecast and provision economic and business activities and to decide for the next direction. Today, various efficient and intelligent techniques are available to help the organization in providing the best interpretation of this large volume of data from different types of heterogeneous sources, to be

processed and analysed and presented in an understandable, visual and decent manner to suit the business language and stakeholders' objectives.

Big data is categorized as having three features, the 3Vs; which are value, velocity and variety. Volume refers to the sheer amount of data, velocity is the speed with which it is being delivered in real-time, and variety is the number of different sources. With the combination of increased volume, velocity and variety, organizations need to reconsider, rethink, and repurpose their ways of working with innovation, the marketplace and communication. Their entire business processes may need to change as well.

Characteristics of Big Data Management
As one of the current trend terms in the world today, there is no exact way to define big data. The term is often used with related concepts such as Business Intelligence (BI) and data mining. These three terms are about analysing data, but big data concepts differed from the others two concept where the data volumes, number of transactions and the number of data sources is bigger and much complex and acquired special methods and technologies to analysed the data. The three major characteristics that define big data are volume, variety and velocity.

The first characteristic is volume, and it became as the first and important characteristic because of in every organization, they have many large archived data along the way from the starts of their business but they can't process the data to help them make decisions and plan their future business. This has represented as the most immediate challenge to the conventional IT structures. The said situation faced by the organization what inspired the IT researchers to come out with big data. The ability of big data to process large amounts of information is the main attraction of big data analytics to create value from relevant data.

The second major characteristic in big data is velocity and it refers to the pace of increasing speed of the new data is generated and the pace at which data been transfer around. In the current electronic and digital era like business processes, machines, networks and human interaction with things like social media sites, mobile devices, etc. Social media messages going to viral in seconds in 1999, Wal-Mart's data warehouse stored 1,000 terabytes (1,000,000 gigabytes) of data. In the year 2012, it had access to over 2.5 petabytes (2,500,000 gigabytes) of data. Every minute of every day, we upload hundreds of hours of videos on YouTube, we send over 200 million emails through Gmail. Reacting quickly enough to deal with data generate pace is a challenge for most organizations. Twitter is one of the most used online diary worldwide and it can be as the best example of the velocity of big data. It produces around 6,000 tweets every second that corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and approximately 200 billion tweets per year.

The third characteristic of big data is variety as the sources and types of data is different and big data is not only a structured data and for sure is not an easy task to put big data into a relational database. Today data comes in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. This variation of unstructured data creates problems for storage, mining and analysing data. Dealing with

a variation of structured and unstructured data significantly increases the difficulty of storing and analysing big data. Around 90% of data generated is data in unstructured form.

This will conclude that the "big" in big data is not solely about volume because big data involves of a large amount of data. But, big data not only refer to data volume because it does not only mean you have a lot of data but the data also coming to you at a very fast pace and it comes with a variation of forms and variety of sources. It is important to say that today's big data may not be tomorrow's big data because the technologies always evolve. As a perspective, if the organization facing a significant challenge to manage with data's volume, velocity and variety, the organization need to have data management with the assistance of technologies and techniques.

Big Data Presents Challenges and Opportunities
Big data is not just about volume and from various sources; it is about its other characteristics such as size, speed of data, structure and quality and new-generation analytic technologies that help organizations get more value from their information assets. It helps organizations to provide business insights on customer behaviour and patterns that can be used to improve operations, anticipate opportunities and business growth, and to detect for any possibility of issues or problems.

Today, organizations are generating, receiving, processing, and storing remarkable amount of data to and from a wide range of resources such as databases and Internet. The process of managing, handling and storing huge amount of data is known as big data management. Big data has many features and varies in nature, they can be simple or complex, structured or unstructured, secure or with a very minimum security. This makes the management and storage of big data becomes more challenging and extremely important, and would definitely require the help of technology and techniques.

No matter how challenging it can be, organizations must put in the necessary control measures to solve or lower the risk imposed so that the data stored are available, retrievable, and can be used to make ad-hock decisions and plans for the future. The main purposes of the management of big data are to store and handle available data in a simple and understandable manner with easier and flexible retrieval methods to make the right and strategic decisions for enhancing their businesses.

Issues and Challenges of Big Data in Information Management
New technology comes with uncertainty as it may not be understood well by all organizations especially during its early introduction, adoption, and implementation. Big data is not an exceptional. Organizations lack of information and understanding on the fundamental of big data such as what is actually is, the benefits, and the infrastructure requirements. Uncertainties are the outcome from unknown opportunities, challenges and emerging risks. While opportunities create new values, risks create threats to the organization.

Lack of Business Sponsorship and Management's Support

When introducing any new technology in the organization, it is important to obtain top management's buy-in. This is very essential if it involves changes in the organizational culture and structure as well as amount of time and money to invest. Hence, big data as any new technology needs support from top management and stakeholders. Since investing on big data wisely would give organizations competitive advantage, top management should have the vision to see the impact of big data in the organization's future. Big data is not only a technology, but it involves innovation, cultural change, analytical mindset and new skillset. It also requires more effort to educate people on how to treat the data.

Data Privacy and Security

Organizations benefit from many conveniences and breakthroughs due to Big Data-powered applications and services. However, one of the most sensitive issues that organizations feel reluctant to adopt big data is privacy. It involves the disclosure of personal and sensitive information to people who should not have access to it, whether it is done intentionally or mistakenly. Privacy breaches occur when there are insufficient security measures in place. Safety and diversity can be compromised with the absence of privacy. Since most of the data in the data sets are important, it should be a top priority for organizations to secure the data from security breaches as it may cause serious privacy concerns and safety of employees and customers could be compromised.

Data Discrimination

Tapping into big data can make discrimination more prevalent. While Big Data helps businesses become better marketers and service providers, as well as making consumers accept that they are being analysed in detail hence their better experience, it makes them feel discriminated. The ability to access electronic information on users' behaviour on Internet activities and preferences could negatively affect a person's opportunity, such as on his or her bank loan application without giving a chance for him or her to justify or defend himself or herself. It is unfair and not acceptable to discriminate against people based on data that organizations have collected on peoples' lives. Therefore, decisions should not be made solely out of electronic data, especially those that have an adverse impact on someone.

Inaccurate and Incomplete Data

Big data is meaningless unless it is used for improved decision-making. For that, organizations must take necessary actions to manage data such as data acquisition, extraction and recording, data cleansing, data integration and aggregation, as well as data representation and analytics including modelling, analysis and interpretations. Data that will be used to analyse comes from diverse sources and of different formats. It may contain wrong information, duplication and contradictions. It is unlikely that data of extremely inferior quality can bring any useful insights or promising opportunities to organization's precision-demanding business tasks. Carefully structured data is essential for efficient and accurate data analysis. Incomplete data can lead to incorrect data analysis resulting in poor result, judgement and decision. Data cleansing or data scrubbing involves inconsistent data, data derived from heterogeneous sources, and data that are not up-to-date or

obsolete. Data needs to be cleansed to be ready for contemporary use and available for discovery and reuse.

Data Availability and Accessibility

This is referring to data being available and accessible much larger and faster in real time and across various industries. This huge amount of data need to be processed and analysed, and the tasks could be time-consuming as it takes longer time to analyse. In this fast moving world, results of the analysis are demanded almost immediately.

Organizations need to have information from different resources. There may be cases when organizations do not have sufficient data to do the analytics and would probably seek or buy data from third-parties that may or may not want to share the data. It is difficult to consider that big data analytics always gives correct results as erroneous data could generate inaccurate results resulting in misleading decision-making.

Interpretations can become an Ethical Issue

A significant share of big data is generated from people's perception, intentions, and desires. The purpose of analysing big data is to help organizations to make decisions. However, relying on electronic data alone without much concern on its impact to people or the environment, can lead to ethical issues. Organizations must be careful when making conclusion and judgment about what the data conveys. This is because perception, intentions, and desires can change rapidly. They must learn that decisions should always be based on how it will affect everyone involved, and not only from the numbers and information shown on papers.

Other ethical issues when handling big data could also include issues of identity, privacy, ownership, and reputation. Intellectual property right issues also arise during the collection, storage, sharing, and processing of big data. Veracity is closely related to trust issues. All these ethical considerations are related to one another. The situation is even more sensitive when it has to do with personal and confidential data such as medical and financial records. From the process of big data cleansing up to analysis, the privacy may be threatened due to the exposure of this information to unauthorized parties.

Technology moves too fast

Decision making about adopting new technologies can often take a long time or to due to necessary process to be followed including levels of approvals. It can also be confusing to choose big data technologies on the market. Choosing a technology itself can be time consuming and it evolves to fast making organizations hard to keep abreast with the latest technologies and trends, resulting in poor decision making even from the beginning of choosing a product or solution to help them with actual issues with big data management.

Big data, being fast data, if its relevance can be obtained, quickly analysed, planned and applied them back into operational systems, then it can affect events as they are still unfolding. The ability to make fast and right decisions are very important too. Data management for computation may be a challenge and will require major investment in information and communication technology.

Lack of Skill Set

As new technologies become available in the market, new skillset are demanded. Accurate and actionable data mining and analysis, particularly in real-time, requires extensive technical skills. It would be a great challenge for organizations to find skilled data analysts to make use of the organizations' data. Although data analysts were always being needed in the organizations, the required analytics skills are different with big data. Organizations need to form a common data analyst team either to equip existing staff with the right skillset by ending them to trainings and obtain necessary certifications, or source out for new employees who are specialized in big data as they can understand data from a scientific perspective, the business and its customers, relate their data findings and apply directly to them. Besides the critical analytical skills, they need to be close to products and processes within organizations.

Scale

Organizations are required to reconsider about the design, build and operate data processing components to meet data volumes scale that is scaling faster than computer resources. Furthermore, as the technology now is moving towards cloud computing, sharing of resources on expensive and large clusters requires new methods to execute data processing jobs to achieve the objectives of workload in a cost effective manner as well as to face with system failures and could adversely impact data processing and recovery methods.

Big investment

There is no denying that the adoption of requires a big investment. Some of the major investment will be on the location of storage whether it is on-premise or on cloud, cost of new *hardware, and people to process and analyze big data*. Notwithstanding on that, organizations must provision and anticipate future *expansions* to cater the growth of big data, and ths again includes on hardware, software, maintenance, subscription, and people to manage.

Storage, Sharing, Transfer
Location to store the big data is another great issue that organizations are dealing with. New storage models are in the form of elastic web services with petabytescale data warehouses and have limitation to keep low input/output latency from large data repositories. It is hard to avoid increased latency while handling high-volume data acquisitions and supporting a variety of mixed data structures.

Different Expectation from Different Audience
Data analysis and results may be time-sensitive, while the accuracy depend on the target audience who may have different objectives out of the information, such as a team looking for trend analysis, another team looking for profit and there may be other teams expecting different information that they can use to make other decision for the business. Latency must be taken into consideration as mass user behaviour with respect to real-life events demands a desired output. As much as the analysis can help, it can also be a big challenge for the data analysts to cater all expectations from the management and different groups of users.

Solutions for Big Data in Management
In managing big data, there are three element involved which is people, process and technology. Every organization generated and use large amount of data and information in the organization environment and the requirement of processing, sharing, storing, securing and displaying information give emphasis to the vital role that information plays in achieving success. People or leader in the organization have on accurate, relevant and timely information to make data driven decisions regarding the current and future goals of the organization. The fact that information is without any doubt is not only a valuable corporate asset but the difference between a successful and unsuccessful organization. Implementing a successful solution for organization management on information in today's environment requires a complete understanding of the organization and the three element which is people, process and technology.

In this sense, these three element should be considered as the solution for demanding management of big data in the organization. To improve competitive advantage and decision-making, an organization must consider information as the fundamental key to manage the whole organization. In the information and digital era, information has become an asset necessary for business survival. O'Brien and Marakas (2013) argue that there are three key in information systems which are: 1) supporting processes and operations, 2) supporting decision making by agents of the organization, and 3) supporting strategies for competitive advantage. Basic requirements need to collaborate decision-making and information systems such as the sources of support given, frequency and form of information presented, information format and method used in processing the information.

Big data is considered as an important tool to generate fundamental input to decision-making and competitive advantage. Every leader and manager in any organization want all of their action and decisions is based on accurate and precise information. Therefore, big data analytics is the best solution that distils terabytes of low-value data, transforming them into a single bit of high-value

data. Big data management has the ability to generate information from a single bit of high-value data that present different structures. And, for as the solution to data management of big data, three element mentioned earlier can assist helps the environment where big data is identified. Therefore, the three elements for managing big data are discussed as follows.

People element – in big data management, it is a new trend use in organization, people involved must be prepared with new skills to process the data. Therefore, this people need to have skill to understand and to work with the large amount and different type of data. The skill also not restricted to those who manage data process related with big data technology but also to the people who act as the decision maker as they need to view those large amount of data to gain the needed information for making important decisions in the organizational and cultural changes.

There are a few titles given to people dedicated with big data management and positions as such Data Analyst, Data Architect, BI Manager and Data Scientist which is really related with big data. The demand on professionals with the data-savvy know-how to analyse big data to make effective decisions is really high in the current global technology and business environment. However, big data capability does not remove human factor. We know the most important aspects of big data is the outcome of reports on how decisions are made and who are in charge to make them. Therefore, the capability of managing big data technologically is not correlate with the ability big data gives to the decision maker. People who manage big data in the organization must be seen as the valuable element that give the organization competitive advantage.

Process element –process related to the actions performed in the technological environment. For instance, some process in the technological environment use specific tools and techniques to ensure the business process work properly and it also responsible to generate the data as well as to use them precisely and accurately. However, in big data management, process can be interrelated because it is essential that they simultaneously perform the activities and process and technical activities related to the business.

Technology element – several technologies and techniques are considered in big data management and it is started from collecting, storing, processing and analysing data. After the trend of big data existed in the technological environment, many technologies and techniques have been developed and it capability to the analysis that is important in big data performance. For examples of the techniques used are as follows:
- Data mining – a technique used to extract patterns from large amounts of data by merging statistical methods and machine learning data management.
- Machine learning – a technique used artificial intelligence principles and considers the development of algorithms for recognising complex patterns in large volumes of data and propose intelligent decisions.

Technologies that related to big data environment which include:
- Hadoop – open source framework in processing large volumes of data in distributed systems.

- MapReduce – software framework introduced by Google company to process high volumes of data.
- Business intelligence – a software developed to display and to analyse the data
- Cloud computing – technology refers to a computing paradigm with a high level of computational resources sometimes configured as distributed systems to provide services through digital networks.

Solutions for Business Challenge

In competitive business environment, every organization must utilize the information and manage and process the data in hand to generate a good business activities and better management of the resources. There are six factors of performance management where big data would improve considers by Cokins (2015):

1.   Strategic planning and execution

Strategic planning is the fundamental process in an organization in defining the strategy, direction and decision making by utilising the data and resources. The organization will develop a strategy map and balance scorecard. By using this two, they set the navigation to fulfil the mission and vision strategy. Key performance indicator (KPI) will be outline based on the analysed big data that helps the organization to do better planning and can forecast the trends of their business.

2.   Cost visibility and driver behaviour

With the help of big data, profitability analysis can be process to analyst organization activity-based costing principles by modelling cause-and-effect relationship established on business and cost drivers. This involved the use of technique and technology of big data not the conventional and traditional managerial accounting (Cokins, 2015).

3.   Customer intelligence

Customer intelligence is about generating information regarding customers where powerful marketing and sales can be analysed to find out trends and preferences of the customers. In the organization customers' information consist of large volume of data and with big data, business analytics can help the organization to create more profit from the customers. The most used tools are Customer Relationship Management (CRM) software application for this purpose.

4.   Forecasting, planning and predictive analytics

Business Intelligence and data mining generated data to help the organization to forecast the future volume of data and information from many variables such as, future workers need and budgeting level for resources spending can be manage and calculated to help the organization to give accurate and timely data in organization strategy and planning. Using big data analytics it is possible and help the management of organization to predict the best scenario or business activities in future.

5.   Enterprise Risk Management (ERM)

Enterprise Risk Management is a process of planning, organising, leading and controlling the activities in the organization to minimise the risk and increase opportunities on the organization resources and earning. Big data analytic help the management in organization to understand risk and plan on how to reduce risk and also evaluate the opportunities that relevant with the mission and vision of the organization.

6.   Process improvement

Using big data analytic, the organization can improve their business process in order to reduce or remove long process that will affect the cost and to accelerate the process to increase the productivity of their business and also will raise the organization personnel as their skills and competence can be use efficiently.

Conclusion

The age of digital and technology has led to the trend of big data. Big data is a low-cost technology and a new information management and analytic software. Huge volumes of data generated from numerous business all over the world which led to many issues and challenges faced by organization and the way of handling the data in the organization is also a challenge. With the use of software such as Hadoop we obtained the capability of managing, processing and providing the analysis in decision making and strategy planning.

By using big data management, organization capable to represent a genuine leap forward and opportunity to actualise efficiency, productivity, revenue and profitability. Big data has the control to transform conventional business to established online business and raise revenue for the organization. with big data management, organization can forecast, strategize and predict their future business and resources, as big data can offer accurate and timely analytics. Big data will change the game of organization manage their information, data and resources and big data will lead the direction of business nature of the organization and high level of data quality and accessibility of business intelligence is the main goal of big data management.

References
Bekker. A. (2018). The 'scary' seven: big data challenges and ways to solve them. Retrieved from https://www.scnsoft.com/blog/big-data-challenges-and-their-solutions

Cokins, G. (2015). "The big data theory" game show for data analysts. Retrieved from https://icrunchdata.com/blog/499/the-big-data-theory-game-show-for-data-analysts/

Halaweh. M., & El Massry, A. (2015). Conceptual model for successful implementation of big data in organizations. Journal of International Technology and Information Management, 24(2), 21- 34. Retrieved from https://scholarworks.lib.csusb.edu/jitim/vol24/iss2/2

Hasnat. B. (2018). Big data: an institutional perspective on opportunities and challenges. *Journal of Economic Issues*, 52(2), 580 – 588. https://doi.org/10.1080/00213624.2018.1469938

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big data for dummies.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Intezari, A., & Gressel, S. (2017). Information and reformation in KM systems: big data and strategic decision-making. *Journal of Knowledge Management*, 21(1), 71-91. Retrieved from https://doi.org/10.1108/JKM-07-2015-0293

Kanchi, S., Sandilya, S., Ramkrishna, S., Manjrekar, S., & Vhadgar, A. (2015). Challenges and solutions in big data management - an overview, presented at 3rd International Conference on Future Internet of Things and Cloud. Rome, Italy, 2015: IEEE

Krishnaveni, C., & Udhayakumar, U. (2018). Big data analytics: challenges and applications for social media data, presented at National Conference on Emerging Trends in Computing Technologies. Villupuram, Tamil Nadu, 2018: International Research Group

Marr, B. (2017). 3 massive big data problems everyone should know about. Retrieved from https://www.forbes.com/sites/bernardmarr/2017/06/15/3-massive-big-data-problems-everyone-should-know-about/#7df0e5606186

Mawed, M., & Al-Hajj, A. (2017). Using big data to improve the performance management: a case study from the UAE FM industry. Facilities, 35(13-14), 746-765. Retrieved from https://doi.org/10.1108/F-01-2016-0006

Rossi, R., & Hirama, K. (2015). Characterizing big data management. Issues in Informing Science and Information Technology, 12, 165-180. Retrieved form http://iisit.org/Vol12/IISITv12p165-180Rossi1921.pdf

Top 10 big data issues and how to solve them (2016). Retrieved from https://datafloq.com/read/top-10-big-data-issues-how-to-solve-them/2335

Top 5 problems with big data (and how to solve them). (2018). Retrieved from https://www.piesync.com/blog/top-5-problems-with-big-data-and-how-to-solve-them/

Waqar, A., & Ameen, K. (2017). Defining big data and measuring its associated trends in the field of information and library management. *Library Hi Tech News*, 34(9), 21-24. Retrieved from https://doi.org/10.1108/LHTN-05-2017-0035