Big data: issues, trends, problems, controversies in ASEAN perspective

Haviluddin [a,1,*], Rayner Alfred [b,2]

[a] Faculty of Computer Science and Information Technology, Universitas Mulawarman, Samarinda, Indonesia
[b] Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kinibalu, Sabah, Malaysia
[1] haviluddin@gmail.com*; [2] ralfred@ums.edu.my
* corresponding author

ARTICLE INFO	ABSTRACT
	Big Data has a characteristics is size, new opportunities and have the potential to transform corporations and government and its interactions with the public. This paper attempts to offer a broader definition of Big Data that captures it is other unique and defining characteristics. This paper presents a consolidated description of Big Data by integrating definitions from practitioners and academics. In addition, we summarize the issues, trends, problems and controversies related to Big Data (technology, applications, and people) from infrastructure (i.e., hardware and software), technology for Big Data Analytics (BDA), management, educational and scientists, and government-related to policies perspectives in order to support the Economic Community ASEAN (AEC) era.

1. Introduction

An utilization of a Big Data has changed the process of decision making in an organization or government. Therefore, among the three support each other's ability to fully utilize the data and analysis: an ability to identify and manage a variety of data sources, a capacity to build sophisticated analytic models, and a role of management in order to change the organization are indispensable. As the growing Big Data, across the organization, including governments around the world begin to take advantage of Big Data as part of the decision making [1]. Before applying it, first have to understand Big Data itself, so that in practice no trouble and can run properly as expected, especially in facing the ASEAN Economic Community (AEC) era.

The 12th The Association of Southeast Asian Nations (ASEAN) Summit in Cebu, the Philippines in January 2007 made an agreement a "Declaration on the Acceleration of the Establishment of an ASEAN Community by 2015", marked by the establishment of the AEC Blueprint. However, the implementation of the AEC postponed until 2016. Furthermore, the AEC is an implementation of the ASEAN blueprint and agreement of its members.

Based on ASEAN blueprint, there are three main pillars that are the focus of cooperation among ASEAN countries includes related to security, economy, and social and cultural. First, related security community aims to realize the principle of comprehensive security which recognizes the interrelationship between aspects of political, economic, social and cultural. Second, related economic community aims to realize the economy and trade among ASEAN countries that can compete fairly and to prevent monopoly. Third, related socio cultural community aims to create harmony relationship between ASEAN countries which could create cultural conditions of the people who know each ASEAN country. In support of the three main pillars above, there are five pillars supporting AEC in 2015 consists of: 1. free flow of goods, 2. free flow of services, 3. free flow of investments, 4. free flow of skilled labors, and 5. free flow of capitals [2].

In the AEC era, all ASEAN countries (Indonesia, Malaysia, Thailand, Philippines, Singapore, Brunei Darussalam, Vietnam, Myanmar, Laos and Cambodia) should be able to deliver highly competitive conditions of competition. Each country must have arrangements in various sectors, including an information technology (IT). A sector that has significant potential for growth in the ASEAN, electronics sector including the IT industry and electronics among 11 other business sectors, such as the free flow of goods, services, investment, capitals and labor, wood-based products and auto-motives, healthcare and e-ASEAN, rubber products and textile, agro-based products and fisheries, travel and tourism and logistics [2], [3].

By leveraging IT, a country can make production more efficient, to innovate the work process, which in turn will accelerate the economy. Therefore, strengthening the construction of information communication and technology (ICT) infrastructure such as sustainable ICT infrastructure development between government-to-enterprise, and government-to-government are required by all ASEAN members. Nowadays, every nation around the world including ASEAN nations are fighting a war without gunfire to dominate the next generation communication market [4]. Thus, to support the three pillars of ASEAN Community, a data or information plays an important role and as a strategic commodities.

This paper aims to provide a comprehensive overview and trends of Big Data to readers of this exciting area. This paper is also examined with a discussion of the trends, challenges and prospects of Big Data in the commitments to support the AEC. Section two understands the evolution of data becomes a Big Data. Section three presents the components of Big Data. Section four looks at the solutions and recommendations in Big Data, and finally section five provides concluding remarks.

2. Data/Information is power

"How to understand the data" has actually been around since ancient times. Along with a technological developments or 'Information Explosion' and a growth of data or 'Data Deluge' is increasing [5]–[7]. Since the 1950s, as computing technology has gradually been used in commercial applications, many corporations have developed databases. Hence, most corporations cannot store, process, and analyze these data. In the 1990s the database community started using the term data mining (intersects human intervention, machine learning, mathematical modeling, and databases), that interchangeable with the term knowledge discovery in databases is now the common approach to data analysis [8]–[10].

Big Data is unavoidable. In the last twenty years, the data is increasing day-by-day in structured, semi structured or even in unstructured forms, called data deluge from across the world. The emergence of Big Data is more because of the increasingly widespread use of social media and supported by the growth of social media platforms quite rapidly. Government and corporations also need social media to improve performance and as brand awareness. In addition, increasingly more devices connected to the internet, also mobile phones which is getting a lot of data collected. Wherein, the data come from a variety of sources (i.e., Google processes 100 Petabyte (PB)/day, YouTube processes 1000 Petabyte (PB) video storage; 4 billion views/day, Email, and Social media (i.e., Facebook processes 300 Petabyte (PB) plus 500 Terabyte (TB)/day, Twitter processes 124 billion tweets/year and LinkedIn, etc.)) for 7 billion people around the world were created every day with the data size Petabyte (PB) to Yottabyte (YB), called information explosion. A survey institute, McKinsey Global Institute has stated that by 2018, demand for the needed talent to capitalize on Big Data is estimated to exceed the supply by 50-60%. Then, the total amount of digital data will grow exponentially from 1 zettabyte (ZB) in 2010 to 45 zettabytes (ZB) in 2020 [11]–[14]. These data has an impact on the storage media. At this time, the storage media is increasingly easy and inexpensive to provide, e.g., in 2000 a Gigabyte (GB) of disk \$17, now < \$0.07, in 2000 a Gigabyte (GB) of random access memory (RAM) \$1800, now < \$1, and in 2009 a Terabyte (TB) of relational database management system (RDBMS) was %70K, now < \$ 20K [15].
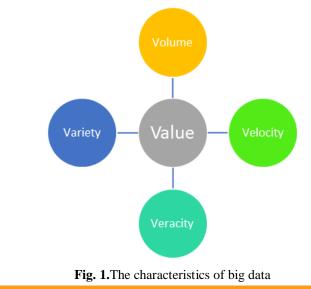
Currently, the concept of data management becomes a trend in order to understand the data. This concept comes from the desire in order to integrate all the data into one place, making it easy to do the analysis, called the integration agent, and need for insights real-time [15]. Then, Roger Mougalas of O'Reilly Media coined the term Big Data in 2005. What is the Big Data or data science? Big Data definitions have evolved rapidly, which has raised some confusion. In some academic communities, Big Data is a term called data science, which is associated with a collection of data size, diverse form

or format (structured, unstructured and semi-structured), interdisciplinary efforts from mathematics, sociology, economics, computational science, and management science. In some professional communities, Big Data called business intelligence and business analytics refers to Big Data analytics (BDA) or Big Data mining in handling the amount of data that can reach Gigabytes (GB) and continues to grow rapidly.

The definition of Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of database architectures" [15], [16]. The National Science Foundation describes Big Data as: "Large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future" [17]. In other opinion, Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation" [18].

Big Data can be described as the data that exceeds the processing capacity of conventional database systems. Data that are too big, moving too fast, or not in accordance with the structure of the database architecture. In order to obtain the value of big data, IT must choose an alternative way to process it. Actually, Big Data have a characteristics is size. However, there is no unified definition of Big Data and it has been defined differently from corporation, technological, research or academic perspectives [10], [17], [19]–[21]. The Big Data, in management perspectives, called Big Data management. The most recent definition of Big Data is characteristics by three parameters called 3V's includes; volume (refers to a massive amount of data that is ever growing every day), variety (refers to the data comes from in various formats and any type; structured, semi-structured, and unstructured), velocity (refers to the data comes at high speed e.g., arriving and stored data) [22], [23]. In addition to the 3V's, other dimensions of Big Data have also been mentioned by IBM, variability (refers to the variation in the data flow rates in some sources of main buzz data) [23]–[25], by SAS, veracity (refers to unreliability, provenance, and accurateness inherent in some dirty sources of data) [15], [26], and by Oracle, value (refers to the economic value contained in the various of data) [9], [11], [22]. The Big Data of characteristics can be seen in Figure 1.

In technological perspectives, refers to the Big Data initiative should be the responsibility of specialized information technology (IT) departments to support the technology in other parts of the government or corporations such as banking, marketing, demographic, or health. Then, Big Data is collected, analyzed, and reapplied for the purposes of government and corporate activities to be more valuable and useful. Therefore, the right technology platform to collect, maintain, analyze and implement it must be appropriate. Therefore, the right technology platform to collect, maintain, analyze and implement must be appropriate, such as hardware being used, software being used, as well as the right person to do it. Therefore, many applications that have been created by the company in order to manage the volumes and complexity of information involved, such as Microsoft, Intel Big Data, Amazon Web Services, Dell Big Data analytics, Teradata, Google BigQuery, VMware, Redhat, IBM, SAS, Tableau, Oracle, EMC2, and others.

Fig. 1. The characteristics of big data

In research perspectives, refers to in an effort to overcome the 3Vs, many early innovators in Big Data technologies focused on the field; computing and storage. However, along with the development of Big Data, the researchers not only focus on two things but Big Data demands more, so researchers continue to develop; computational, analyzes techniques, also security. The focus of research to provide quality hardware, software analysis which corresponds to specific needs. In fact, education is held by higher education, as well as specialized institutes are already implementing learning materials on the use of Big Data.

Based on the phenomenon above, a strategy of smart-intelligence in order to harness Big Data to be translated (business analytics) and assist (advanced analytics) in improving the performance and worth, also have new values is required. However, these large data sets is almost impossible to manage and process using traditional business intelligence tools or based on databases traditional such as relational database management system (RDBMS), database management system (DBMS), and object-relational database management system (ORDBMS) [20], [21]. Nevertheless, through science and knowledge and the latest technology in the management of data science, it is possible to create a system that will automatically to be able to process these large data [4].

Since data is having valuable, many researchers, corporations including government began to digging and analyzing it, by leveraging existing applications. As well as, many researchers propose several methods to improve performance on large data processing. In business communities, Big Data has been utilized in industrial IT services and solutions [27], economics area; bank [27], [28], vehicles rental [29], [30], manufacturing [30], consumer analytics [31]–[33], marketing [34], [35],energy consumption [36], healthcare domain; healthcare financing, pharmaceutical, Ebola, DNA microarray analysis, DNA, RNA and protein [37]–[45], engineering domain; futuristic applications domain (i.e., robot taxi, city information model (CIM)), geospatial, remote sensing, wireless network [37], [46]–[50]. According to the McKinsey Global Institute study of Big Data on five types of industries, namely the US healthcare industry, sector governance in Europe, the retail industry in the US and global scale manufacturing stated that Big Data can generated considerable value [12]. In academic and professional communities around the world have been established to exchange and disseminate theoretical findings, practical experience on Big Data; Beacon Institute for Rivers and Estuaries, Clarkson University, The Royal Institute of Technology (KTH) Stockholm, Sweden, Marine Institute Ireland, University of Ontario Institute of Technology, Uppsala University, the Swedish Institute of Space Physics [30], China's Research Centers on Fictitious Economy and Data Science and for Dataology and Data Science, the Data Science Consortium in Japan, the International Council for Science: Committee on Data for Science and Technology (CODATA; based in France), and the UK Data Science Institute, and others [17], [51]. In the USA, Boston University to support the Rafik B. Hariri Institute for Computing and Computational Science and Engineering, and the University of Rochester to support the Institute for Data Science on personalized medicine, national security, and online commerce [13].

3. Main Focus of the Chapter

In this chapter we discusses some of Big Data issues and trends especially in Big Data analytics to support the ASEAN Economic Community (AEC). Then, of course, with popularity, comes controversies and problems.

Currently, Big Data is growing in popularity. In general, the challenges in information communication and technology (ICT) is related in order to human resources, infrastructure [52], policies and laws including utilization of Big Data. In other words, Big Data is also about how can take advantage of existing opportunities such as policy and regulation by governments and corporations.

To change the large data into valuable, they should be review how to manage Big Data in the data center. They need a system that is optimized to acquire, manage, and load the data (structural, semi-structured, and unstructured) into their database, so it can be effectively presented and analyzed in depth. Therefore, the scalability of big data solutions in the data center is an important consideration, then the Big Data management planning is important.

The governments play a key role in promoting Big Data including rules and technologies [17]. Currently, the revolution will continue to explore the Big Data has been cross country and overthrow

the existing differences in per capita income, population, and the indicators of socio-economic. Therefore, to answer the Big Data issues and trends, in the ASEAN perspective in order to support the implementation of the AEC, ASEAN ICT Masterplan (AIM) 2015 was launched in January 2015, in Kuala Lumpur, Malaysia. This confirms that ASEAN countries should be prepare for the digitalization era. In line with the characteristics of Big Data, the rules agreed upon by ASEAN in the management of Big Data is required. In order to support AEC, the AIM 2015 is based on six strategic thrusts including (1) economic transformation, (2) people empowerment and engagement, (3) innovation, (4) infrastructure development, (5) human capital development, and (6) bridging the digital divide. With these, the AIM 2015 is designed to deliver the following four key outcomes, (1) ICT as an engine of growth for ASEAN countries, (2) recognition for ASEAN as a global ICT hub, (3) improved quality of life ASEAN, and (4) contribution to ASEAN integration. In other words, all ASEAN countries agreed on an "open government plan" that encourages governments to open their data to the public.
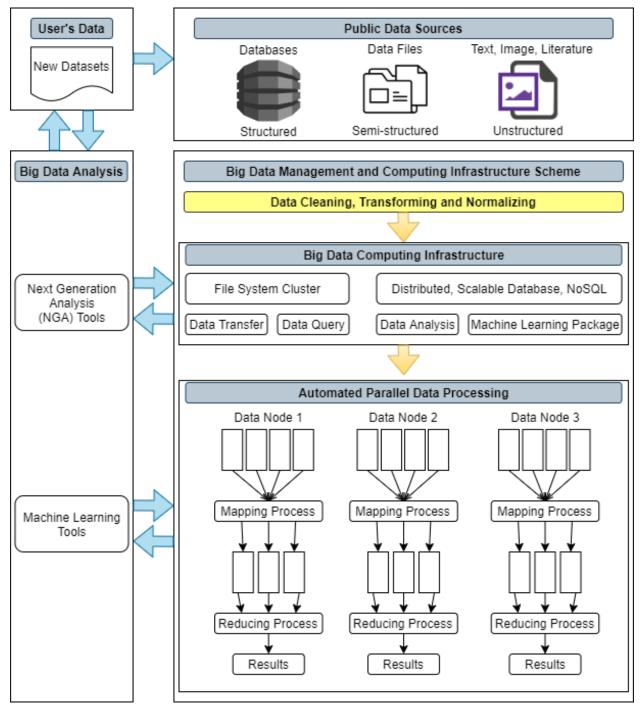

Fig. 2. The schemes of components of big data technology

For implementation of the AIM, for example, a memorandum of understanding (MoU) have been signed on July 9th, 2015, at the launch by The Ministry of Science, Technology and Innovation (MOSTI) and MIMOS, Cyberview, SAS Institute, and Cyber Security for a Smart City IoT trial shows one of the early IoT implementation initiatives as a result of this collaboration.

The Big Data technology generally refers to three aspects of technical innovation that cope with super-large datasets: namely, automated parallel computation (storage capacity as standard tools and the processing to find), data management schemes include people, and data mining analytics (analyze massive datasets), see in Figure 2.

First, automated parallel computation, refers to every systems will be generate more and various public data sources (structured, semi-structured, and unstructured forms). Then, of course data become larger, more complex and more inexplicable. As a human being, of course, will be difficulties in deciphering and interpreting the large data including data storage constraints. Therefore, the computational role in Big Data to define and interpret large data including data storage. In principle, information technology (IT) infrastructures in Big Data must meet the following criteria [13], [19], [53]:

- Data acquisition, referring to the flow of data at high speed and a wide variety. The information technology (IT) infrastructures to support large data acquisition must be delivered slowly, predictable both in capture and process data quickly and simply, can handle very high transaction volumes, in a distributed environment, flexible, and has a dynamic data structure is needed;

- Data access, referring to data service called platform as a service (PaaS) such as Google, VMware, Amazon, Microsoft, HP, and Oracle. The PaaS includes computing platform and a solution stack with software as a service (SaaS) and infrastructure as a service (IaaS), right now associated with cloud computing;

- Data organization, referring to data integration. The IT infrastructures to manage large data must be able to process and manipulate the data in the original storage location is required; and

- Data analysis, referring to the data presentation. The IT infrastructures for analyzing large data must be able to support a deeper analysis, such as statistics, machine learning and data mining analysis. Most importantly, the IT infrastructures must be able to integrate the analysis on a combination of big data and traditional data.

Second, data management schemes, refers to the processing and analysis of large amount of data. The Big Data would be meaningless if it is not treated. Therefore, it is necessary to analyze the scheme to process large data such. No matter what the purpose of processing and analyzing data, it is critical that the data contain the insights or answers being sought, for example, data processing models are Samza, Tez, and MapReduce. The working principle of these models are to separate the data into the machine and each processor, but it works using the same algorithm [9].

Third, data mining analytics, refers to integrate popular applications software and machine-learning-based computing infrastructure that supports Big Data such as an environment analysis is provided as a service use-on-demand by integrating popular software and machine-learning-based supporting computing infrastructures. Currently, many popular tools that can be used by governments and corporations in order to analyze their data, such as WEKA, Java and JVM languages, R, Python, SAS, MATLAB, and others. However, each of these tools has its own advantages and disadvantages. Many statistical and machine learning techniques are averaging processes in Big Data Analytics (BDA) in order to improve the performance of these tools. The application of these techniques have been widely applied to solve problems, such as classification, regression, clustering, prediction, summarization, segmentation, feature selection, dimensionality reduction, ensemble learning, network analysis, and density estimation. Furthermore, these analytics process in order to meet and support the policy makers is needed [7], [17], [19], [54]–[58]. In order to utilization of Big Data in both the government and corporate organizations, of course, will be able to get a better view to the public service (for the government) as well as on their business sustainability (for corporations).

Behind a great opportunities, some researchers also stated that Big Data also has controversies that can affect the success of the management, such as bigger data are not always better (confusing and solving complex), because real-world data are incomplete and noisy, some technologies of Big Data

analytics are not always the best tools, and human life becomes less privacy (human daily activities will be a "record" and collected in storage). For the purpose of analysis of huge amount of data which is in structured, semi-structured, and unstructured forms. In this manner, Big Data Analytics (BDA) that in handling of the complex and massive datasets is needed [19], [55], [59]–[61].

4. Challenges and Prospects

The unique characteristic of Big Data is the process in which value is discovered. Thus, the biggest challenge in the Big Data of era itself is the amount of data generated from various sources (data capture, data lifecycle and variability), called technology, people, and process. To process Big Data into valuable process is a long term process that cannot be done in hurriedly. Therefore, successful implementation of Big Data is at least dependent on technology, people, and processes. In order to challenges, this paper we categorize the challenges of Big Data related to technology (i.e., security and privacy), people, and process (i.e., guidance) into two categories: infrastructures (i.e., hardware and software) and data analytics.

In infrastructure perspectives of hardware is a combination of components that produce an architecture that delivers business outcomes. Build an infrastructure designed to enable new levels of insights derived from exploiting all relevant data. The platform should be fluent in all forms of data and analytics or transactional data including local area network (LAN), wide area network (WAN), internet, and so forth. Build a highly flexible, scalable information technology (IT) infrastructures, tuned for today's Big Data environment and designed to capitalize on the integration of social, mobile, and cloud technologies. In infrastructure perspectives of software, there are some of technologies in order to manipulating, analyzing and visualizing the Big Data as an advance analytics including Lucene, Solr, Hadoop and HBase. Social Media such as Facebook, Twitter and LinkedIn are going a step further thereby publishing open source projects for Big Data like Cassandra, Hive, Pig, Voldemort, Storm, IndexTank. Back to the fundamental question, whether these applications capable of processing large data accurate? Although, it may be easy to use but needs desired by the management is not necessarily appropriate [19], [39], [51].

In infrastructure perspectives of data analytics. It refers to be able to generation parallel and distributed systems in process large data (i.e., captured, manipulated/communicated, analyzed, stored and displayed), there are quite a lot of technical and non-technical challenges then displayed them in advance and in real-time. At first, Big Data is seen as an objective to manage to reduce the cost of data management. Now, governments and corporations are focused on creation potential values. In order to benefit from the additional insights gained there is a need to assess the analytical capabilities and the implementation of Big Data [12], [19], [55].

To get the benefits of Big Data are required, among others, the large investment required to build the infrastructures (i.e., hardware and software) of the system, governance, culture, skills, and business processes. The latest mathematics or statistical and machine learning models for large data also remains a hot research topic that are still being developed by researchers in many fields related Big Data Analytics (BDA) such as classification, regression, clustering, prediction, summarization, feature selection, segmentation, dimensionality reduction, ensemble learning, network analysis, density estimation, and security [53], [55], [62]–[68]. Nevertheless, it should be a concern that the failure of the technology development often begins with the assumption that new technology will be able to generate value or insight by itself. This assumption is certainly a danger, due to heavy investment (both in terms of material and time) that have been issued in the form of these technologies produces only a small return value. To avoid danger, so before investing in new technology, a company must begin with the issue of Big Data Analytics, and then identify the data is required in order to solve the issue. Each issue in the government and corporations are sometimes require different technologies. Thereby, the government and corporations could be learn in order to implement organizational solutions and receive information on the operational challenges and limitations of the data, thus helping governments and corporations determine the conditions which must be owned technology solutions that would be invested.

In addition, another problem of the biggest challenge related to Big Data are privacy and security. How to maintain data security, both of infrastructures (i.e., hardware and software) and data analysis results terms [10], [69]–[71]. Hence, the challenges and prospects for each of the ASEAN countries are linking data such as legislation and policies, policy, culture and other values, processes and

procedures, organizations and people, transparency/privacy, information quality, information architectures categories (i.e., health care and techno-biology, informatics and risk, and sustainable and smart systems). Currently, new legislation to deal with various issues related to the law in the digital world including Big Data technologies have been proposed, discussed and endorsed the government in various world.

5. Solutions and Recommendations

In this section, we provide solutions and recommendations related Big Data technology in order to supporting management of government and corporations. Based on the analysis of IDC (2012), the investment for Big Data and analytics has increased significantly. Where, in 2015 an estimated expenditure of Big Data in the Asia Pacific region will reach US $ 1.61 billion, up 355% from a year ago [72]. The impact of the development of Big Data, governments and corporations willing to process the data with investing. In the view of ASEAN, the use of Big Data will be a factor supporting the success of the ASEAN Economic Community (AEC). Therefore, the following factors of common concern. In order to obtain the benefits of Big Data, we classified based on infrastructure (i.e., hardware and software) and analysis [16], [19], [73].

In infrastructure perspectives of hardware, a good infrastructure will be impact on the management of Big Data, especially in speed, availability, and access. Speed refers to understanding how data and insights flow and how consistently they flow at speed is an important essential in planning. You need to balance cost versus speed and the value of business outcomes. In addition, also consider future growth. A well architected solution should achieve the best balance among all requirements. Then, availability refers to the ability of a system to absorb or avoid damage without suffering complete failure (disaster recovery). Lastly, access refers to data locality, data life cycle management, and data deduplication. It is important to remember that the goal of building information technology (IT) infrastructures platform is to easily integrate these large data sets, and provide facilities to conduct in-depth analysis of the combined data set [16], [19], [50], [74].

In infrastructure perspectives of software, Big Data Analytics (BDA) is indispensable. In order to support Big Data Analytics (BDA), the smart based applications in data management for various purposes are required. Not enough to rely on existing applications, in addition to expensive, the facilities provided are also difficult to meet the needs of corporations and government. This creates an opportunity for human resources are still used

In infrastructure perspectives of data analytics, for the purpose of processing large amounts of data, the Big Data Analytics (BDA) technologies is extraordinary indispensable. The Big Data Analytics (BDA) are necessarily involves the computer, the application of computer technology, operational research, and statistics to solve problems in business and industry. Of course, mathematics underpins the methods used in analytics, and the science of analytics is concerned with extracting useful insights or properties of the data using computable functions. Furthermore, from the perspective of technology, machine learning will continue to play a role in Big Data Analytics (BDA). Currently the technique is capable of working in a large data analysis (i.e., diagnostic analysis, predictive analysis and prescriptive analysis), with a wide variety of data types. In other words, the key value of Big Data analytics or data mining is to obtain intelligent knowledge [7], [17], [19], [55], [75], [76].

The most frequent obstacle encountered is the lack of information about things to be known, so that the processing of Big Data requires a fairly long time, and human resources that are used in large quantities. Therefore, the framework for analyzing the data is needed, such as data types, applications, algorithms, and the target analysis.

In order to ensure the success of Big Data projects, governments and companies must make the following points with the precautionary principle, among others:

- Think carefully every purpose business intelligence and insight that is most valuable to the organization;

- To avoid combining data from various sources without sorting and tidying up the data first (data cleansing);

- The most valuable insights on business performance can be achieved by determining early on what information is required and then ask specific questions related to it;

- To achieve the best results is very important to cooperate with reliable analytics provider by involving stakeholders from relevant business in the beginning of the period;

- To allocate a budget to develop solutions 'self-service' which allows users to build their own reports when needed; and

- Debriefing skills needed to ensure that governments and corporations will receive a return on investment in the management of Big Data faster.

6. Future Research Directions

In this section, we provide insights and an overview of the importance of Big Data related security, privacy, and guidance for further research. Based on perspectives (i.e., infrastructures (hardware and software) and Big Data Analytics (BDA)) in Big Data in order to obtain the benefits of Big Data. The challenges and opportunities for study of the future utilization of Big Data that we propose, among others:

6.1 In Infrastructures (i.e., hardware and software)

There are three main challenges and opportunities in infrastructure related to the hardware and software. These three concerns are (1) How to develop new infrastructure and tools, create data dynamicity (volume, variety, velocity, value, and veracity), storage, search, linkage, and retrieval of Big Data including speed data capture, (2) How does merging data from multiple sources with traditional data structured?, and (3) How to make the rule of law in order to avoid the misuse of digital data.

6.2 In Infrastructure Perspectives of Data Analytics

In the coming years, all measures taken in the governments and corporations will be based on data. The algorithm will determine what action should be taken. Artificial Intelligence (AI) would be a priority in the choice of completing the task of analyzing Big Data. Several opportunities for further research related algorithms into Big Data Analytics (BDA) to reveal some kind of existing data, such as text analytics (i.e., information extraction (IE), text summarization, question answering (QA), and sentiment analysis); audio analytics (i.e., speech analytics); video analytics; social media analytics (i.e., community detection, analysis of social influence, and link prediction); and predictive analytics (i.e., heterogeneity, noise accumulation, spurious correlation, and incidental endogeneity), especially in order to supporting the implementation of the ASEAN Economic Community (AEC). Of course, mathematical and machine learning methods are still an option in the BDA.

- How to analyze the data in real-time/streaming analytics, interactive analytics and machine learning;

- How to determine and select the right algorithm in Big Data will affect the results of the analysis;

- How to transform data into insight to solve business challenges and create competitive advantages;

- How can people pay attention to the results of data analysis or statistics that have been produced through story telling; and

- How to design problems and violations of the law concerning privacy in Big Data technologies.

6.3 In Education

Government and skilled manpower is essential in the era of Big Data. There are still many students who do not understand career paths and career benefits of data analytics. Scientists in particular have a critical role to play to educate the public about how to interpret and use their data in a way that is responsible, and also to ask the right questions about the re-use of data, such as questioning the possibility of bias in the sample of data, applying predictive models compared to only consider causality, and algorithms principal policy-making process. A study on the evolution of Big Data as research and scientific topic has shown began in 1970s, then crowded published in 2008. Today the concept of Big Data is treated from a different point of covering implications in various fields.

- How to educational will be improved Big Data across education levels (curriculum reform) in order to produce professionals in data analytics with a high value in the world market.

- How to provide education related to Big Data to the communities, especially in ASEAN region.

6.4 In Government

Most governments are still slow in adapting to new trends, including Big Data. In other words, the data revolution is being challenged social scientific models and its ability to inform policy and make it available to the government in the new forms of evidence that can inform the design, implementation and evaluation of public policy.

- How to development of 'smart government' is an opportunity in order to provide the public services with good quality;

- How to the existing system of government will be work to allow corporations relate more easily;

- How to increase the interest of the government and corporate innovation projects based on data in the ASEAN region specifically, particularly in the social investment approach, such as legislation and policies, policy, culture and other values; and

- How to design and create a code of conduct to regulate and control the big data technology without losing any of its characteristics.

7. Conclusion

In this article, an overview of Big Data's concept, tools, techniques, applications, advantages and challenges have been reviewed. The results have given away that regardless of the fact that accessible information, tools and techniques available in the literature, there are numerous focuses to be viewed as, discussed, analyzed, developed, and improved, and so on.

The term Big Data is emerging due to the fact that indicates that data growth continues to multiply over time has exceeded the limits of storage media and database systems that exist today. Because of these impacts, governments and corporations need tools and techniques for managing big data is to get the value of benefits.

In future infrastructure perspectives of hardware and software, data centers are the physical locus of big data in all its forms. How large collection of data that is replicated in multiple data centers for improved performance and durability. Moreover, software technology that is specific to analyze Big Data across multiple data centers (and spread over tens of thousands of processors and hard-disk drive), thereby providing both scaling and better performance.

In future infrastructure perspectives of data analytics, the role of machine learning to produce smart system still remains an option and the trend of Big Data processing. Moreover, in the role of business, in which every business should have the right strategies for managing Big Data, especially in the information business that continues to grow. How to maintain data security is also an important part in keeping Big Data. Thus, business leaders are able to obtain maximum results. Furthermore, the government and employers can collect large amount of data from their operational systems and found the value of the benefits from the data stored. Analysis of large data with the correct method will help the government and employers to gain a true insight into the decision making.

In education perspectives. Academics, scientists in particular have a very important role in order to data innovation. When applying 3V's principle of Big Data related to giving people greater control over the data, they actually allow the democratization of research capabilities.

In the ASEAN perspectives. The ASEAN members are mostly agricultural countries. Therefore, Big Data could be used to help farmers to increase crop production, predicting the right time for planting, analyze the quality of the soil, etc. In healthcare, Big Data could be used for patient data collection e.g., medical history, similarity disease, pharmaceuticals, and financial, etc. In taxes, Big Data can also be utilized to increase the government tax, such as raising awareness of paying taxes, and reducing the tax fraud. In banking, Big Data could be used to collect information regarding the

transaction, customer's relationships, choosing products in the right time. In telecommunications, Big Data could be used to understand behavior of customers, and so forth.

Finally, in principle, each of the ASEAN countries are ready to implement the ASEAN Economic Community (AEC), of course, with the strengths and weaknesses of the member countries of ASEAN, especially in order to realize the ASEAN as "One Windows" in the field of information communication and technology (ICT) includes data integration, data security and connectivity infrastructure lines to share administrative data are integrated. In other words, Big Data account for a growing share of countries' economy.

References

[1] F. Dzikrullah and M. A. Rinjani, "A framework design to develop integrated data system for smart e-government based on big data technology," *Bull. Soc. Informatics Theory Appl.*, vol. 1, no. 2, pp. 41–51, 2017.

[2] ASEAN, "ASEAN Economic Community Blueprint," 2015.

[3] ASEAN, "ASEAN Economic Community Blueprint," 2008.

[4] S.-T. Park, Y.-R. Kim, S.-P. Jeong, C.-I. Hong, and T.-G. Kang, "A Case Study on Effective Technique of Distributed Data Storage for Big Data Processing in the Wireless Internet Environment," *Wirel. Pers Commun*, vol. 86, no. 2016, pp. 239–253, 2016.

[5] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. networks Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

[6] S. M. Zobaed and M. A. Salehi, "Big Data in the Cloud," in *Encyclopedia of Big Data Technologies*, 2019.

[7] N. D. Shah, E. W. Steyerberg, and D. M. Kent, "Big Data and Predictive Analytics," *JAMA*, 2018.

[8] U. Fayyad and P. Piatetsky-Shapiro, G. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI MAGAZINE*, pp. 37–54, 1996.

[9] A. Mohebi, S. Aghabozorgi, T. Y. Wah, T. Herawan, and R. Yahyapour, "Iterative big data clustering algorithms: a review," *Softw. Pract. Exp.*, vol. 46, no. 2016, pp. 107–129, 2016.

[10] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2015, pp. 137–144, 2015.

[11] J.-P. Dijcks, *Oracle: Big Data for the Enterprise*. Redwood Shores, CA, USA: Oracle Corporation, 2013.

[12] J. Manyika *et al.*, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[13] J. M. Tien, "Overview of Big Data: A US Perspective," *Natl. Acad. Eng.*, vol. 44, no. 4, pp. 12–19, 2014.

[14] N. Trifunovic, V. Milutinovic, J. Salom, and A. Kos, "Paradigm Shift in Big Data SuperComputing: DataFlow vs. ControlFlow," *J. Big Data*, 2015.

[15] M. Troester, "Big Data Meets Big Data Analytics," 2012.

[16] E. Dumbill, *Big Data Now: 2012*. 2012.

[17] Y. Shi, "Big Data History, Current Status, and Challenges going Forward," *Bridg.*, 2014.

[18] G. IT Glossary, "Big Data," 2016.

[19] J. Dean, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*. 2014.

[20] Y. Wang, L. A. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technol. Forecast. Soc. Change*, 2018.

[21] D. Rajasekar, C. Dhanamani, and S. K. Sandhya, "A Survey on Big Data Concepts and Tools," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 5, no. 2 February 2015, pp. 80–84, 2015.

[22] L. Doug, "3D Data Management: Controlling Data Volume, Velocity, and Variety," *Appl. Deliv. Strateg.*, pp. 1–4, 2001.

[23] IBM, "IBM big data platform - Bringing big data to the Enterprise," *Big data at the speed of business*, 2016. .

[24] P. Zikopoulos, D. DeRoos, C. Bienko, R. Buglio, and M. Andrews, *Big Data Beyond the Hype: A Guide to Conversations for Today's Data Center*. McGraw-Hill Education, 2015.

[25] P. C. Zikopoulos, D. DeRoos, K. Parasuraman, T. Deutsch, D. Corrigan, and J. Giles, *Harness the Power of Big Data: The IBM Big Data Platform*. 2012.

[26] P. Russom, "Big Data Analytics," 2011.

[27] G. CSC, "Big Data Tools Connect The Dots For CSC's Digital Marketing," 2016.

[28] G. CSC, "Global bank secures risk and compliance with CSC," 2016.

[29] G. CSC, "Avis Budget's Road to Customer Value," 2016.

[30] IBM, "IBM Big Data Success Stories," 2011.

[31] S. Erevelles, N. Fukawa, and L. Swayne, "Big Data consumer analytics and the transformation of marketing," *J. Bus. Res.*, vol. 69, no. 2016, pp. 897–904, 2016.

[32] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics," *Decis. Support Syst.*, vol. 81, no. 2016, pp. 30–40, 2016.

[33] V. E. Mariana, A. P. Wibawa, J. Hammad, R. Herdianto, and Haviluddin, "Customer trust of online shopping using social media," in *IOP Conference Series: Materials Science and Engineering*, 2018.

[34] S. Fan, R. Y. K. Lau, and J. L. Zhao, "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix," *Big Data Res.*, vol. 2, no. 2015, pp. 28–32, 2015.

[35] Z. Xu, G. L. Frankwick, and E. Ramirez, "Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective," *J. Bus. Res.*, 2016.

[36] K. Zhou and S. Yang, "Understanding household energy consumption behavior: The contribution of energy big data analytics," *Renewable and Sustainable Energy Reviews*. 2016.

[37] H. Aly, M. Elmogy, and S. Barakat, "Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions," *Int. J. Comput. Sci. Eng.*, vol. 4, no. 6, pp. 300–3013, 2015.

[38] Amankwah-Amoah, "Emerging economies, emerging challenges: Mobilising and capturing value from big data," *Technol. Forecast. Soc. Change*, 2016.

[39] M. V. Boland, "Big Data, Big Challenges," *Am. Acad. Ophthalmol.*, vol. 123, no. 1 January 2016, pp. 7–8, 2016.

[40] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Recent advances and emerging challenges of feature selection in the context of big data," *Knowledge-Based Syst.*, vol. 86, no. 2015, pp. 33–45, 2015.

[41] G. CSC, "World's Largest Healthcare Informatics Data Warehouse," *CSC, G.*, 2016. .

[42] B. Mounia and C. Habiba, "Big Data Privacy in Healthcare Moroccan context," in *The 2nd International Workshop on Privacy and Security in HealthCare (PSCare15)*, 2015.

[43] C. Seebode, M. Ort, C. Regenbrecht, and M. Peuker, "BIG DATA infrastructures for pharmaceutical research," in *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 2013.

[44] S. D. Young, "A 'big data' approach to HIV epidemiology and prevention," *Preventive Medicine*. 2015.

[45] C. Church, H. Young, G. Parraga, and G. N. Maksym, "Improvements in Functional Image Impedance Modelling of Asthmatic Lungs Using K-Means Clustering," in *B20. ASTHMA: NEW FINDINGS IN AIRWAY SMOOTH MUSCLE PATHOPHYSIOLOGY AND POTENTIAL THERAPY. American Thoracic Society*, 2018, pp. A2746–A2746.

[46] S. Li *et al.*, "Geospatial big data handling theory and methods: A review and research challenges," *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016.

[47] Y. Ma *et al.*, "Remote sensing big data computing: Challenges and opportunities," *Futur. Gener. Comput. Syst.*, vol. 51, no. 2015, pp. 47–60, 2015.

[48] Y. Qin, H. K. Yalamanchili, J. Qin, B. Yan, and J. Wang, "The Current Status and Challenges in Computational Analysis of Genomic Big Data," *Big Data Res.*, vol. 2, no. 2015, pp. 12–18, 2015.

[49] C. Yang, "Learning methodologies for wireless big data networks: A Markovian game-theoretic perspective," *Neurocomputing*, 2016.

[50] Haviluddin, F. Agus, M. Azhari, and A. Saleh Ahmar, "Artificial Neural Network Optimized Approach for Improving Spatial Cluster Quality of Land Value Zone," *Int. J. Eng. Technol.*, 2018.

[51] E. R. Mardis, "The challenges of big data," *Dis. Model. Mech.*, 2016.

[52] B. Prasetyo, F. S. Aziz, K. Faqih, W. Primadi, R. Herdianto, and W. Febriantoro, "A review: evolution of big data in developing country," *Bull. Soc. Informatics Theory Appl.*, vol. 3, no. 1, pp. 30–37, 2019.

[53] B. Schoenborn, *Big Data Analytics Infrastructure For Dummies*. 2014.

[54] R. Alfred, "Summarizing relational data using semi-supervised genetic algorithm-based clustering techniques," *J. Comput. Sci.*, 2010.

[55] E. Alpaydin, *Introduction to Machine Learning Second Edition*. 2010.

[56] Shilpa and M. Kaur, "BIG Data and Methodology-A review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 2013.

[57] Purnawansyah, Haviluddin, A. F. O. Gafar, and I. Tahyudin, "Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities," in *Proceedings of the Eleventh International Conference on Management Science and Engineering Management*, 2017.

[58] Marselina, E. Hiswati, A. F. O. Gaffar, Rihartanto, and Haviluddin, "Minimum wage prediction based on K-Mean clustering using neural based optimized Minkowski Distance Weighting," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 90–93, 2018.

[59] J. P. Holdren and E. S. Lander, "Big Data and Privacy: A Technological Perspective," 2014.

[60] S. Roy *et al.*, "IoT, big data science & analytics, cloud computing and mobile app based hybrid system for smart agriculture," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, 2017, pp. 303–304.

[61] M. Ge, H. Bangui, and B. Buhnova, "Big Data for Internet of Things: A Survey," *Futur. Gener. Comput. Syst.*, 2018.

[62] R. Alfred, "Discovering knowledge from multi-relational data based on information retrieval theory," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009.

[63] H. Haviluddin and A. Jawahir, "Comparing of ARIMA and RBFNN for short-term forecasting," *Int. J. Adv. Intell. Informatics*, 2015.

[64] Haviluddin and I. Tahyudin, "Time Series Prediction Using Radial Basis Function Neural Network," *Int. J. Electr. Comput. Eng.*, vol. 5, no. 4, pp. 765–771, 2015.

[65] S. Hardwinarto and M. Aipassa, "Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan-Indonesia," *Procedia Comput. Sci.*, vol. 59, pp. 142–151, 2015.

[66] Mislan, A. F. O. Gaffar, Haviluddin, and N. Puspitasari, "Water Level Prediction of Lake Cascade Mahakam Using Adaptive Neural Network Backpropagation (ANNBP)," in *1st International Conference on Tropical Studies and Its Application (ICTROPS)*, 2018.

[67] F. Sia, R. Alfred, and K. O. Chin, "Learning Relational Data Based on Multiple Instances of Summarized Data Using DARA," in *Communications in Computer and Information Science*, 2013.

[68] Suprihatin, I. T. R. Yanto, N. Irsalinda, T. Purwaningsih, Haviluddin, and A. P. Wibawa, "A performance of modified fuzzy C-means (FCM) and chicken swarm optimization (CSO)," in *Proceeding - 2017 3rd*

International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era, ICSITech 2017, 2018.

[69] M. Janssen and J. v. d. Hoven, "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?," *Gov. Inf. Q.*, vol. 32, no. 2015, pp. 363–368, 2015.

[70] Q. Liang, J. Ren, J. Liang, B. Zhang, Y. Pi, and C. Zhao, "Security in big data," *Secur. Commun. Networks*, vol. 8, no. 2015, pp. 2383–2385, 2015.

[71] N. Chaudhari and S. Srivastava, "Big data security issues and challenges," in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, 2017.

[72] IDC, "Big Data: Global Overview: Market Analysis: Worldwide Big Data Technology and Services 2012-2016 Forecast," 2012.

[73] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big Data in Smart Farming – A review," *Agricultural Systems*. 2017.

[74] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*. 2018.

[75] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review. Neural Networks," *Neural Networks*, vol. 61, no. 2015, pp. 32–48, 2015.

[76] M. Xu and S. Y. Rhee, "Becoming data-savvy in a big-data world," *Trends in Plant Science*. 2014.