Aide-Mémoire

Vol 2, Issue 9, 29 May 2020

# Big Data: Too Big to Ignore

Kong Rattanakthida*

Have you ever searched for a specific thing and later seen an advertisement for that particular thing or similar categories popped up on your Facebook newsfeed or Instagram homepage? In the information era, it comes as no surprise that we find ourselves in the kernel of a surge in data usage. Every day we are generating data without knowing. When we read, write or speak, we produce data. As this article is being written, for instance, the data is being collected to produce the content that you are now reading. Indeed, data is everywhere, but the question is, where are all of this unlimited amount of data stored? Does this enormous tank of data communicate with each other or stay interconnected? What does it serve when all of these data are recorded and stored? Certainly now, in 2020, the world is powered by big data, so you are in the big data environment. More and more people start to connect themselves with digital lives ranging from social networking to shopping and working purposes, and this number of internet users continue to increase rapidly. At the same time, billions of connected devices generate, collect, store and share IoT data analytics, as Internet of Things incorporates with big data, all of which leave a digital shadow – traceable data we leave whenever we use a digital service. It is likely that we have heard so much about big data, but how many people know what exactly the big data is? Many people may assume that big data is a flood of data. But is big data just about a lot of data? Is the vast amount of data that important? A bitter fact is that, at this moment, people still treat the term ‘big data’ as something they choose to ignore, while big data has significant impacts on almost every aspect of life. While the fourth industrial revolution (Industry 4.0/IR 4.0) is running toward us, big data is not just something that passes by, but a big deal. This aide memoire will discuss about the essential concepts that everyone needs to know about big data and the dawn of the data age.

## ❖ Understanding Big Data Basic Concept

Today, the globalized world is generating data at a very fast pace which is unprecedented in human history. In 2017, it was estimated that more than 3 billion people were connected to the internet, a booming increase compared to only 2.3 million users in 1990 (Jarrar, 2017). As of April 2020, the number has climbed to roughly 4.57 billion active internet users, encompassing 59 percent of the total global population (J. Clement, 2020). The 4 billion global digital population are those who produce data through everything they do at every second, leaving a digital trail in almost every aspect. Combined with the rise of computers, the internet, connectivity and all of the advanced technologies that are capable of producing and capturing data, the global data has increased in an impressive amount and size.

To better understand big data, it is crucial to understand what ‘data’ actually is. Data refers to the quantities, characters or symbols, defined as facts, figures and information stored, processed and transmitted via the computer system

***Kong Rattanakthida** is a Program Coordinator at Cambodia Development Center.

(Puntambekar, 2008). Data presentation can be in many forms, such as numbers, words, images, sounds and multimedia, etc. The term ‘big data’ is vague and has now appeared to be ubiquitous when discussing about IR 4.0 or information technology industry. Big data is defined by global research and advisory firm called Gartner, as “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” (Gartner, n.d.). In other word, big data refers to the data sets gathered from various sources and produced in huge amounts and high frequency, which surpasses the processing capacity of the conventional database infrastructure that none of the traditional data management tools can store, process and analyze at this efficiency (Upadhyaya & Kynclova, 2017).

The first concept of big data was originally characterized in 2001 by an industrial analyst, Doug Laney, as 3Vs concept: Volume, Velocity, and Variety. Later, the V concept of big data has gradually expanded to many other Vs such as Volume, Variety, Veracity, Velocity, Validity, Volatility, Variability, Visualization and Value (Firican, 2017). Even though there is no agreed-on definition of big data, the below 9Vs concept can elucidate what big data actually is.

- **1st V – Volume:** Volume of big data refers to a large size of datasets from all sources. Inputs of big data could be extracted from text, audio, image, video, social network, research studies, medical data, reports, GPS trails, banking transactions and final market data. The list goes on. Such volume of data is disorganized, and cannot be stored, processed and queried with a traditional database system.

- **2nd V – Velocity:** Velocity refers to the super-fast speed of data processing. The faster the data is generated, the higher the velocity is. Velocity highlights the importance of the speed not only for incoming data, but also streaming the fast-moving data into the storage for later processing and analysis. It is about processing data and using it at a fast rate.

- **3rd V – Variety:** Variety refers to the increasing diversity of data generation sources and data format. Data usually appears in many forms, including audio, video, text, image and so on. Such mixture has made data even more complex.

- **4th V – Veracity:** Veracity is the reliability and accuracy of data. It simply stresses on the question of how certain we are about the data or how truthful a dataset is likely to be. More specifically, the accuracy of big data does not only focus on the quality of the data but also the trustworthiness of the data sources, types, and its processing methods.

- **5th V – Validity:** Validity may sound similar to the concept of veracity, but the two are different. Validity is about the correctness and accuracy for the intended use. In some contexts, the data may be correct but invalid. In other words, the same set of data may be valid for a particular application but invalid for another application. Hence, the elements of data are required to be verified or to validated against the intended usages.

- **6th V – Volatility:** Volatility here refers to the duration of usefulness, such as how long the data is valid and how long it should be stored. In term of volatility, two main elements of big data which need to be carefully considered are volume and velocity. Key questions for this factor of consideration include how long the data needs to be kept for before it is viewed as irrelevant or not useful anymore.

- **7th V – Variability:** Variability of big data refers to the variety and rapid changes of meaning. Big data is variable due to the multitude of data dimensions resulted from multiple disparate data types and sources. The constantly changing meaning might lead to inconsistency in the data, such as the inconsistent speed at which big data is being loaded to the database.

- **8th V – Visualization:** Visualization is the visual presentation of data and information which enables users to glean insights from data in a more convenient way. It engages a presentation of data of almost any type, including the audio, behavior and webpage, in a simple graphical format to a more complex presentation, easing the interpretation and understanding of the meaning as well as the correlation of data.

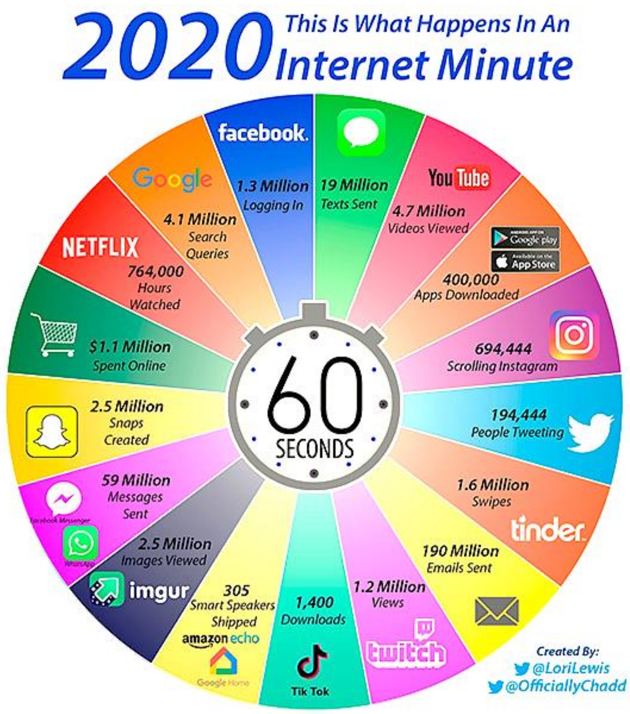
- **9th V – Value:** Value refers to the resourceful and useful information of the big data, and it is the most special of all. Unlike the other Vs of the big data, value is the most important asset for big data processing and analytics. The other characteristics of big data would be meaningless if the desired outcomes were not extracted from data.

### ❖ How Big is Big Data?

Data neither sleeps nor hibernates, and is continuously growing rapidly, making people's lives overwhelming with all kinds of data. How much data is produced each day? How much data is it in the world? What happens on the internet within a minute? People would probably be familiar with some terms of the data scales in terms of units of measurement, such as terabytes, gigabytes, megabytes, kilobytes, bytes, or even bits which are used to quantify the amount of data that people may run into. Despite the intriguing questions towards the exact amount of data, there are no definite answers to either count or measure the data.

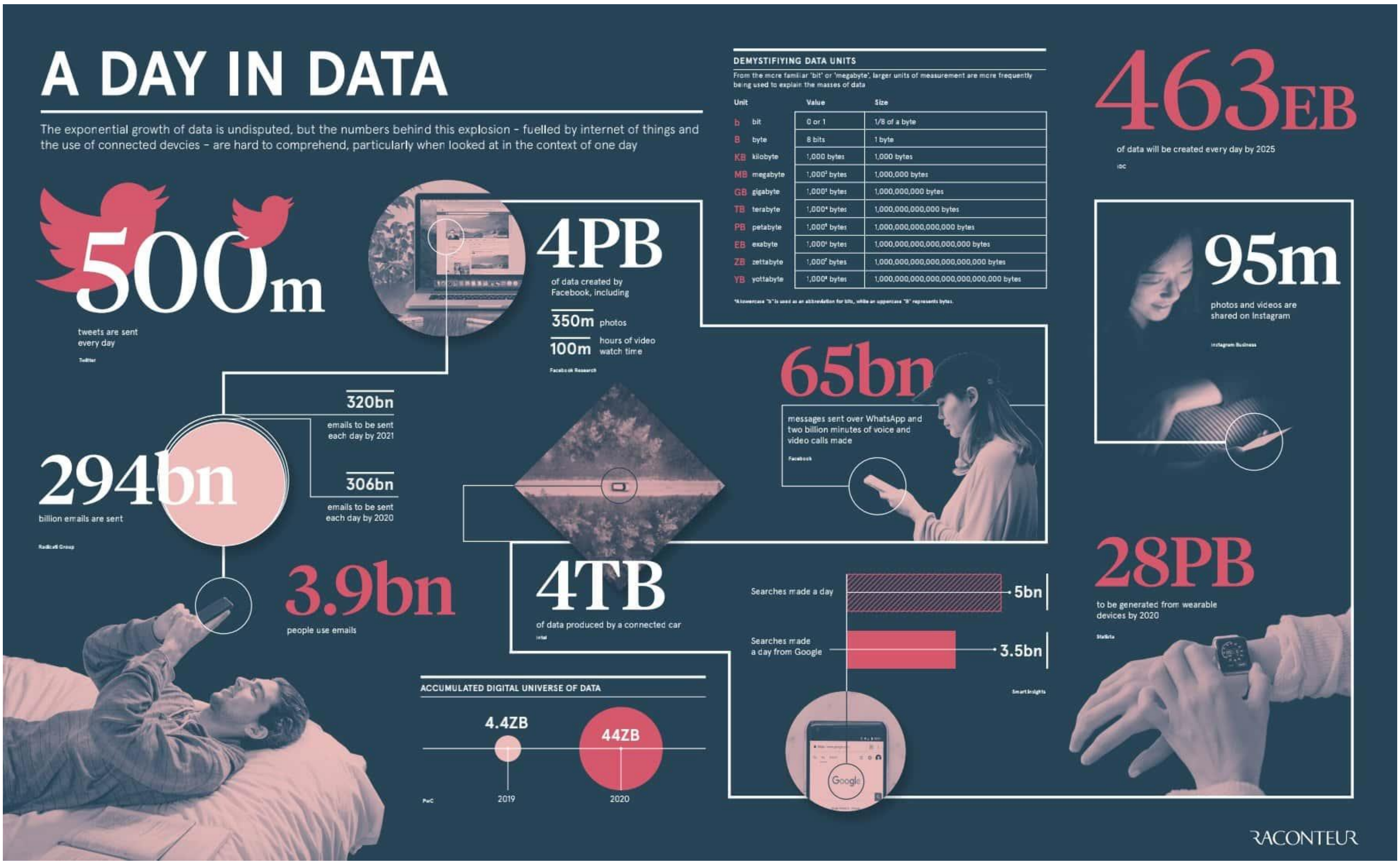
However, there are a number of facts and statistics which highlight the consumption and data actions. It is estimated that the digital universe would reach 44 zettabytes of data by 2020 and is expected that roughly 463 exabytes of data will be created every 24 hours globally at the dawn of 2025 (Desjardins, 2019). Furthermore, there are anticipated 50 billion Internet-of-Things (IoT) connected devices in use worldwide by 2030, comparing to 22 billion available IoT devices in 2018 (Statista Research Department, 2020). The Raconteur's infographic below indicates the key statistics of data on a daily basis. At least approximately 500 million tweets and 294 billion emails are sent; 65 billion messages are sent on WhatsApp; 4 petabytes of data are generated on Facebook; 4 terabytes of data are created on the connected car; and 5 billion searches are created (Desjardins, 2019). This is only in a single day.

**Visualizing the Internet in a Minute**




***Source:*** *Cloudnine*

## A Day in Data




***Source:*** *Raconteur*

### ❖ Knowing Data Types of Big Data

It is essential to recognize the classification of data when talking about big data. Data is classified into three major categories, which are structured data, semi-structured data and unstructured data (Gyamfi, Albert, Williams, & Idongesit, 2019). First, according to Bernard Marr, the strategic business and technology advisor, 'structured data' is the pre-defined type of data that is straightforward and easy to search and organize because its element is contained in the row-column database and can be mapped into fixed fields (Marr, 2019). Data stored in the excel spreadsheet is an example of this structured type as it can be grouped to form a relation that is convenient to store and analyze. Structured data can be generated by human and machines and is seen as the easiest form compared to the other two types for business operation. However, in 2017, it had been estimated that structured data only accounts for less than 20 percent of all the data in the world (Muse, 2017). Second, a much larger percentage of data is 'unstructured data'. The unorganized nature makes the unstructured data more challenging to search, manage, process and analyze, and it cannot be contained in the traditional data system like a row and column. Such attributes have made unstructured data forsaken by businesses until the recent proliferation of advanced technologies such as artificial intelligence and machine learning algorithms, which help smoothen the process (Marr, 2019). Storage-wise, unlike the structured data which can be stored in spreadsheets or relational databases, unstructured data is usually stored in data lakes, NoSQL databases, applications and data warehouse (Taylor, 2018). Lastly, the 'semi-structured data,' which has defining characteristics yet does not conform to a rigid structure as required by the relational database, stays between the structured and unstructured data. For instant, email message is unstructured in its actual content; however, it comprises of structured data, such as names and email addresses of senders and recipients as well as the time that it was sent. In short, the structured data conforms to a rigid format easier to access, process and analyze while the unstructured data is complex and often appears as qualitative information which cannot be organized. The semi-structured, on the other hand, contains both elements.

### ❖ How does data become Big Data?

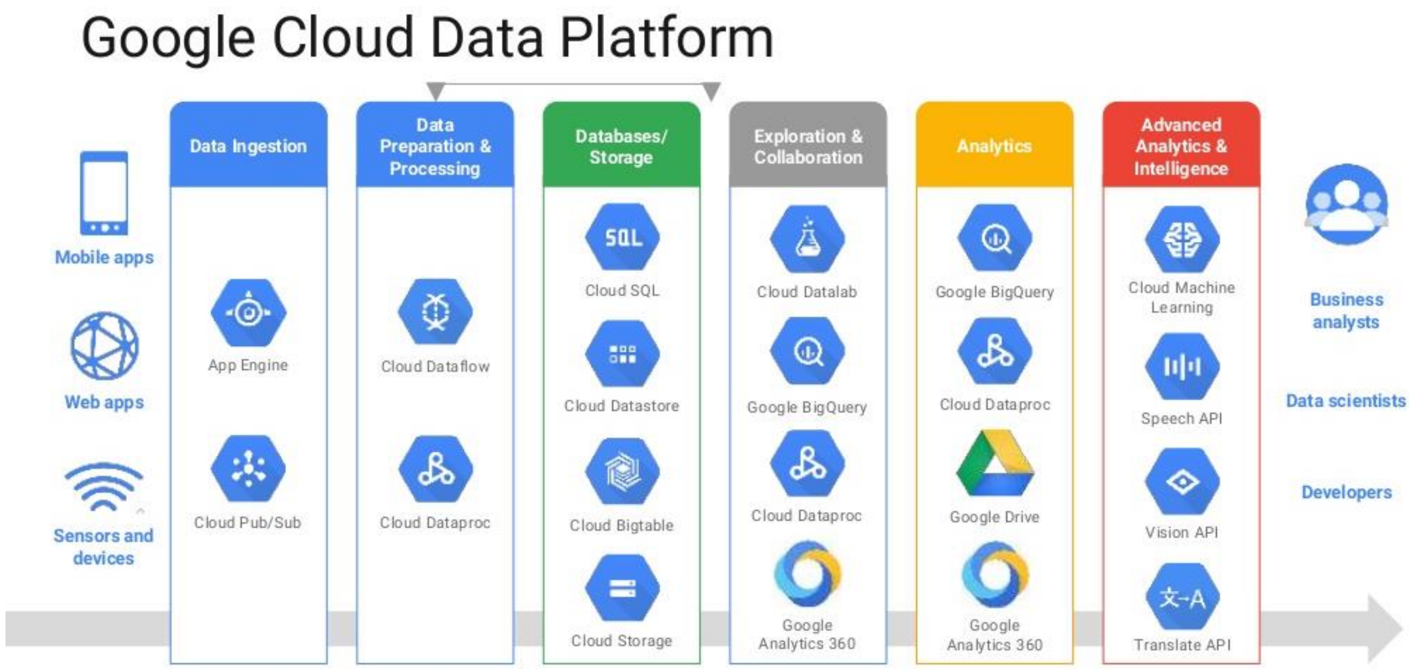
Data collected in its raw form has no value and disorganized. The input of big data is unstructured, which thus requires different storages and processes than those found in the traditional methods. Former Chief Marketing Officer and Vice President Business Development of IBM Watson, Stephen Gold, see big data this way:

> "Big Data is the fuel. It is like oil. If you leave it on the ground, it does not have a lot of value. But when we find ways to ingest, curate and analyze the data in new ways, such as in Watson, Big Data becomes very interesting."

Data can be generated from various formats, including text files, picture files, audio files, video files, webpages, social media sites and presentations, etc. The advancement of IoTs and ICTs – through the buildout of connected devices such as smartphones, sensors, machine-to-machine and artificial intelligence – have contributed to the generation of big data. The process of big data generation is intricate. It has to go through various stages of data life cycles. The Director of the Data Science Institute and Professor of Computer Science at Columbia University, Jeannette Wing, described data life cycle as the sequence of phases that data goes through from its initial generation to its transformation to be useful information. This cycle begins with the generation of data. Everything people do ranging from every search they query, every link they click, every message they send, every book and article they read and every locations they go, all of which contribute to the digital footprint of the people on the internet (Wing, 2019). Following the generation of data is the collection phase. Data collection refers to the process of obtaining and gathering the raw data of different forms and natures for conversion purposes in the next phase (Arass & Souissi, 2018). After the collection, the data arrives at the processing phase, which consists of data cleaning, data wrangling and data formatting to data compression that are necessary for data encryption as well as efficient and secure storage. The next phase is the storage, which primarily supports input/output operations on storage with a large number of data files and objects. Big data storage infrastructure is designed specifically to store, manage and retrieve a massive amount of data, allowing data to be sorted in a way that is easily accessed, used and processed. Then, the data comes to a management stage – a process which includes acquiring, validating, storing, protecting and processing required data to ensure the accessibility, reliability and timelines of the data for the expected outcome. After the management stage, it comes to the analysis, whose function is to cover all the computational and statistical techniques to analyze data for the intended purpose. The algorithms and methods for the analysis underscore the process conducted by the artificial intelligence, data mining, machine learning, and statistical inference. In addition to the analysis, data visualization helps put data in a clear-cut organization, in which people can readily contemplate and visualize the data. Beyond the presentation of data, one last critical part is the interpretation, which is a guide to the explanation of data that get people to understand the picture's context, point of view and implication. At the end of the cycle, all phases will have transformed raw bits to valuable data for the end users for their intended usage and application. These elements will then make certain that the data are kept up-to-date, useable and safely protected (Wing, 2019).

Nevertheless, there are also challenges to data life cycles, which are characterized into three main categories, namely data challenges, process challenges and management challenges (Vassakis, Petrakis, & Kopanakis, 2018). Data challenges are concerned with volume, velocity, variety and veracity while the process challenges refer to the techniques needed for big data acquisition, integration, transformation and analysis. Last but not least, the data management challenges are challenges related to data security, privacy, governance, and cost/ operational expenditures.

## Data Life Cycle in Google Cloud




***Source:*** *House of Bots*

### ❖ Why Big Data Matters?

The most essential question lies at the question of why the world is attentively focusing on big data? Big data is now the big deal for businesses and industries considering its incredible ability to encompass the world of plain figures into useful and resourceful information. The onslaught of the super-fast and most advanced technologies of the fourth industries revolution and hyper-connectivity has permitted massive quantities of data that an organization can collect, manage, process, and analyze. The utilization of the big data unlocks the potentials and significant insights for almost every industry and in almost every sector from large to small, those of which include retails, manufacture, education, finance and banking, healthcare, small and medium enterprises, social services, government, insurances and transportation, etc. More importantly, big data allows us to observe, predict and understand the pattern of human behaviors as well as unexpected downturns, where such resources are beneficial to optimize the production process. Big data is portrayed as an important and a dispensable asset from the new science breakthrough that brings benefits to many forms of life, including law enforcement, governing and administration, business, social network, community building, education, healthcare, agriculture, communication and transportation, etc. The following discussion brings an insight and some examples of how big data can be helpful in the sectors.

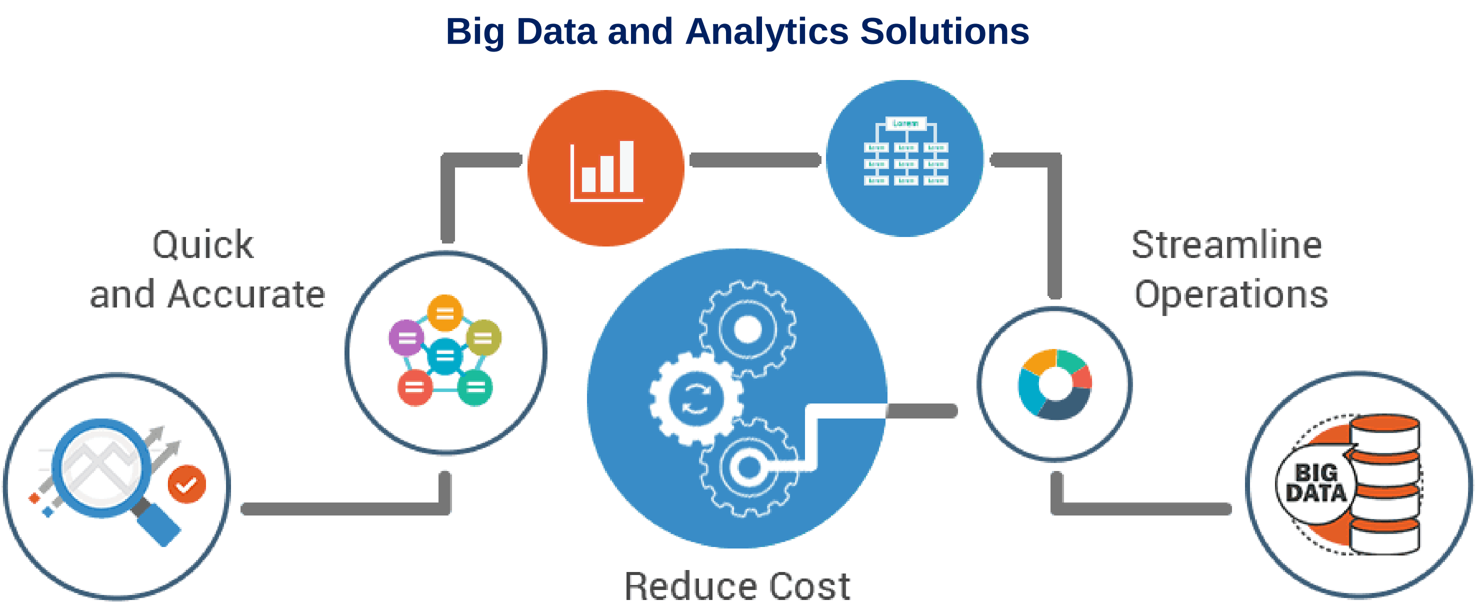
### ➢ Business

There have been remarkable trends that have contributed to a considerable proliferation of data generation. The first trend – the growth in traditional transactional databases – takes place when organizations begin to gather granular data in a broader frequency for developing a more comprehensive business analysis regarding marketplace and customer's behavior prediction due to growing market competition, turbulent business environment and rising customer's expectation (Wielki, 2013). Later on, the growth of multimedia content comprising of text, audio, graphic image, video and other interactive content emerges as the second trend, which correlates with a rapid increase in the use of multimedia and its all-important roles in various aspects of life and industries of the economy today. The third trend is marked by the proliferation of the Internet of Things (IoT), which enables a fruitful growth of data generation. The Internet of Things is a giant network of connected physical devices that operates on the internet to collect and share data (Clark, 2016). As the Internet of Things plays out, sensory technology generates more data than ever before throughout the history of human and history of data itself.

Data of data is increasing at a fast-moving pace, which come as unstructured, high-volume and fast-paced, posing challenges when applying traditional approaches based on the relational database model. Such nature requires a growing demand for a new class of technologies and analytical methods for processing and managing the data. As argued, the essence of big data is not wholly dependent on the large percentage of data, yet it depends on the intention of what people are going to do with it. In the business-related field, big data analytics is profitable to the organization as it enables lower cost of production and time reduction. Meanwhile, it is competent to introduce new product development, increase productivity and advance business intelligence, all of which bring out a more competitive potential for businesses. Therefore, when big data is entwined with high-power analytics, it is able to help accomplish tasks and generate answers to essential questions in just seconds rather than the traditional ways that took days or months. In the decision-making context, big data paves the way for business intelligence and analytics to guide decisions, such as predicting customer-oriented innovations, creating a futuristic business model and even designing or testing a new state-of-the-art product (Wielki, 2013).

Big data is indeed behind the success of numerous corporations such as Facebook and Google. Though these two businesses have become giant data producers, they continue to exploit big data to further expand their businesses. Facebook is globally famous for making everyone stay connected more conveniently. It relies on algorithms to identify the connections that people shared, and uses such interaction to sort out what kind of posts, preferences or topics that people want to see, and then make them visible to their newsfeed (Montcheuil, n.d.). Facebook uses the data from each of their users, such as their search history, pages or content they like or dislike, the user's interaction with their friends and the locations they have been to, etc., as a guide to understanding each individual's interests. The guiding information gives a reliable insight to Facebook so that it has the capacity to customize specific content for identified users and publish ads with the preferences of individual users.

It is undeniable that big data ecosystem lies at the heart of Google's operation, where a mountain of data is stored and processed and, at the same time, continues to rise. For instance, Google search engine can bring uncountable things to you counting from weather information, flight schedule, books, breaking news, music, images, location, history and future, and the list goes on and on. Furthermore, Google knows a lot more than you think it does and even your own data. At times, you may forget about what you did and where you visited, but Google has it all stored. It is true that people now are living with Google, making communication through Google Mail (G-mail), storing the data in Google Cloud, traveling with the assistance from Google Map, resolving the language barrier through Google Translate, learning in Google Classroom and sometimes talking to Google Voice. How are all these possible? It is because big data and Google go hand in hand. When people search for some random topics on the search bar, Google notes every search, runs complex algorithms to sort through various websites with all the available data, identifies what people are looking for and then fetches the answer to their inquiries and even preferences (Marr, 2017). For more complicated operations such as translation, Google operates on its built-in algorithms based on big data, which studies millions of translated texts and speeches to identify the most accurate interpretation for its services (Marr, 2017).




***Source:*** *Signity*

### ➢ Education

The integration of big data into the education sector is revolutionizing and digitalizing the conventional way of education. By taking full advantage of big data for education and mainstreaming it into various practices and activities in schools, students will have more opportunities to learn and explore something new. It is inevitable that most modern teaching and learning, such as digital learning and customized curricular for online classroom system rely heavily on technologies. In the digital classroom, teachers are able to exploit big data to collect information related to students, which thus allows them to better develop customized lesson plans and methodologies to fit the student's needs and learning purposes accordingly. It is distinctive from the traditional classroom, where its curriculum consists of static learning plans that are applied consistently for all students irrespective of their academic performance. More importantly, big data will provide performance-enhancing functions and certain level of innovation to improve the quality of education. In contemporary modern classrooms, teachers use big data technology to better understand student's behaviors and find out student's problems, rather than just relying on the standardized tests and traditional assessments. By further integrating adaptive learning technology, slow-learning students can spend extra-time working and studying against the subject matter while, at the same time, continuing to keep pace with other students in the entire class.

Last but not least, big data system is also dedicated to the new learning pathway. Although age is a sole indicator to determine student's learning progress and the likelihood of advancement in traditional learning paths, the contemporary learning path encourages exceptionally talented students to continue advancing their learning progress even though they remain in the same class with their less gifted peers since their teachers can gauge the performance continuously and customize the curriculum for the students accordingly. As education and big data continue to evolve over time, the proliferation of big data technology in classroom will play an increasingly significant role to bridge the loopholes and missing dots in education that contributes consistently to students' learning outcomes and their future career.

- **Healthcare**

Big data analytics in the healthcare industry offers immense potentials and opportunities, notably in eHealth (electronic health) and mHealth (mobile health). eHealth refers to healthcare services provided with the support of the information and communication technology for healthcare system, including computers, mobile phones and satellite communication. Meanwhile, mHealth deals with the use of smart portable devices for health service and information, which varies from electronic prescribing and medical recording to constant alerts of key health indicators (Moss, Süle, & Kohl, 2017). Big data analytics in medical service is impacting the way patients and doctors handle and respond to healthcare management. It transforms a current healthcare system into a preventive, pro-active, and value-based system through seamless exchanging and sharing of health data (Zillner & Neururer, 2016). With big data technology, health records are digitalized so that patients' diagnosis can be identified more quickly, in which the patients, as a result, can receive treatment more effectively. Several big data applications are applied to align with the needs for improved healthcare quality and to importance of early diagnosis before disease becomes critical. In addition, the integration of big data analytics into the healthcare industry allows doctors to track the vitals and statistics of different patients, such as heart rate, sleep, fitness, pulse, blood pressure and glucose levels. This brings out a more proactive and forward-thinking system for the healthcare industry. Equally crucial, the digitalization of hospital and patient records makes it more convenient for high-risk patients to be better assisted, closely monitored and identified using the data gathered from their previous health records. Furthermore, one remarkable thing that big data can do is to limit human errors such as wrong medication. For instance, big data can help reduce the error by analyzing the patient's records with all medications prescribed and identifying anything that appears in a disorderly form. In a more advanced healthcare industry, through the access of big data and the storage of all information in the cloud, patients can experience new treatment models, which has the capability to carefully follow up their health as well as their wellbeing in a new way that they can pay more attention to any changes regarding their health. Last but not least, the versatile of big data technology and the growth of mHealth can potentially boost and leverage public health initiatives for healthcare delivery and concerns. Resourceful data collected from big data analytics and mHealth will provide insights for the government for well-informed decisions needed for efficient policy formulation and implementation, effective use of resources and timely response to specific public health challenges (Madanian, Parry, Airehrour, & Cherrington, 2019).

- **Transportation**

The digital transformation and the utilization of big data technologies have assisted public transportation and various relevant business, including supply chain management, express delivery, e-commerce, reverse logistics, freight logistics, maritime, air and land logistics (Borgi, Zoghlami, & Abed, 2017). Big data technologies inform both the transportation companies and government about how the routes are being used while predicting passenger volumes as promptly as possible. For instance, particular events such as change in weather, road infrastructure, traffic congestion and customer feedback from transportation operation can be stored, analyzed and processed in real time. Moreover, big data enables smart transportation that can produce better results in operational efficiency, improved end-to-end customer experience, fuel reduction and increased flexibility (Rijmenam, 2019). Given delivery sequences and recipient's need, big data can speed up deliveries accordingly using real-time route optimization, satellite navigation and sensors so that high customer satisfaction will be achieved.

- **Government**

The potential of big data in assisting government work is pervasive, ranging from public service delivery to smarter policy-making and deeper citizen engagement. All of these are what the societies and general public are expecting to see how the government acts in bringing digital technologies to better serve the public interests. Moving from the first generation of e-government (web 1.0) to the second generation of e-government (web 2.0), citizen-citizen and citizen-government interaction have rapidly changed in a manner never seen before, to a closer and active engagement (Jarrar, 2017).

Big data analytics can be used by government bodies to improve the existing public services and coordinated activities. A vast adoption of big data technology has made people more convenient in various ways, making services more responsive and efficient, as evident in the healthcare, employment, transportation, education and business sector that has been discussed earlier. Beyond this, it also contributes to the policy-making process in a number of ways. Policy-makers are using satellite, sensors, mobile phone data and other super advanced technologies to produce alternative indicators for both economic and social aspects in real-time update. The data gathered is reliable with profound insights of human mobility and economic behaviors, which the government can use to better understand the population's tendency, allowing the leaders and agents of the government to better formulate policy decisions as well as implementing such decisions. Despite the traditional instruments, such as censuses, surveys and administrative data – which is time-consuming and labor-intensive – the big data can generate and deliver information in a more frequent, disaggregated and cost-effective manner. It comprehensively analyzes the dynamic changes in various areas, including transportation and urban planning, demographic statistics, food security, employment services, environmental policy and so on. Lastly, the citizen engagement can be encouraged and enhanced through robust utilization of big data technology. Using big data analytics, citizens can monitor and continuously offer constructive feedback on public services and policies to the government, which contributes to the decision-making process. In term of government-citizen interaction, big data analytics enables a more informed and engaged electorate, in which voters are informed about the politicians' performance; at the same time, they are able to keep a check and feedback on public services and policies on a real-time basis. Therefore, voters receive sufficient data and information to make their decisions.

## ❖ Big Data Is Not Perfect, at Least for Now

When implementing big data analytics, the process itself is not solely focused on the hardware equipment or software tool. First and foremost, it requires specification of what and how much information a person is looking for and in what timeframe. It is inevitable that different flow of information will require a different set of tools and a very different approach to that of machine data. Despite value gained, just like every other IT-related initiative, big data has its own set of pros and cons, encompassing data privacy and security, data governance, data sharing, cost operational expenditures and data ownership. The Economic Intelligent Unit research suggests some impediments to the effective utilization of big data for decision-making. Rather than pooling the benefits of the entire organization, data are connected with particular functions (Wielki, 2013). It cannot be ignored that the need for skilled people, like data scientists, to handle the data analysis is also prominent to analyze and act on the data in real time. Additionally, there are also numerous challenges concerning the legal aspects and the use of data such as copyright, database right, confidentiality, trademarks, contract law, competition law and more. These require transparency in data collection practices and cautious utilization of big data to ensure that data used are not associated with any risks. There is somehow still exclusion or gaps in data signals – a method of how data or information is transferred or sent from one device to another network. Commonly, it is transferred in binary code in signals which produce biases in data collection and prevent the big data from providing the whole picture for a particular situation. At the same time, the challenges and threats posed by the utilization of big data, such as the security of data and information collected, are still at high concern. Such concerns are related to the problem of safeguarding highly sensitive data from cybercriminals and hackers. This problem is connected with the broadly defined security of IT infrastructure of the organizations and protection against attacks and cyberwarfare.

Regardless of the problems above, big data is continuously implemented by many institutions and companies who see the potentials of big data that outweigh those of its shortcomings. The rapid growth of the big data and opportunities related to the practical utilization are increasingly changing the process of decision-making at all levels. More importantly, big data analytics is highly necessary for companies or organizations to design new products with latest, smart and exciting functions to engage and improve customer experiences with a safe option. With the power of big data, information about customer's behaviors and their preferences will be used to develop new techniques and methods which are specifically tailored for customer's needs.

However, becoming a data-driven company is more than just having and using analytical techniques. In order to utilize the big data to an even higher degree, it requires analytical models, tools and skilled people with systematic thinking and organizational capabilities. In the growing data-oriented environment, organizations have to be well-equipped and adaptive to the essential skills needed to utilize big data; otherwise, better data-driven performance would be difficult and unmanageable. At present day, data users should be ready to be tech-savvier, up-to-date and flexible with the emerging technologies, or it would be very hard to catch up to the real-time technologies which keep transforming the society every hour. In short, big data is the certain way to the biggest game-changer when one knows how to entail and utilize all the information and technologies that they have in hand. For these reasons, when deciding to put big data and its analytics into use, institutions and companies should always consider all factors and conduct critical assessment on both the strong and weak points so that they can utilize the benefits to a greatest extent that big data could offer and avoid any unexpected setback from the technology.

## References


(2008). *Data and International Organisations: Navigating Cross-Sectoral Data Challenges.* Geneva: Geneva Internet Platform. Retrieved from https://www.giplatform.org/sites/default/files/DataIntOrganisations.pdf

(2017). *Big Data In Action for Government.* World Bank Group. Retrieved from http://documents.worldbank.org/curated/en/176511491287380986/pdf/114011-BRI-3-4-2017-11-49-44-WGSBigDataGovernmentFinal.pdf

*5th International Conference on Big Data for Official Statistics*. (2019, May 3). Retrieved from BigData UN Working Group: https://unstats.un.org/unsd/bigdata/conferences/2019/default.asp

*6 Ways Big Data Is Changing Healthcare Industry*. (2019, October 29). Retrieved from Analytics Insight: https://www. analyticsinsight. net/6- ways-big-data-is-changing-the-healthcare-industry/

Aher, B. (2018, April 20). *How Big Data Impacts Smart Cities*. Retrieved from Dzone: https://dzone.com/articles/how-big-data-has-the-biggest-impact-in-smart-citie

Arass, M. E., & Souissi, N. (2018). Data Lifecycle: From Big Data to Smart Data. *5th Edition International IEEE Congress on Information Science and Technology (CiSt'18).* Retrieved from https://www.researchgate. net/publication/328769944_Data_Lifecycle_From_Big_Data_to_Smart_Data

Austin, D. (2020). *2020 This Is What Happens In An Internet Minute* . Retrieved from Cloudnine: https://cloudnine.com/ediscoverydaily/electronic-discovery/no-fooling-its-time-for-the-2020-internet-minute-infographic-ediscovery-trends/?utm_source=feedblitz&utm_medium=FeedBlitzRss&utm_campaign=ediscoverydaily

Ayers, R. (2018, September 7). *How Big Data And Education Can Work Together To Help Students Thrive*. Retrieved from Smart Data Collective: https://www.smartdatacollective.com/how-big-data-and-education-can-work-together-help-students-thrive/

*Big Data Storage*. (n.d.). Retrieved from Technopedia: https://www.techopedia.com/definition/29473/big-data-storage

*Big Data: What it is and Why it matters?* (n.d.). Retrieved from SAS: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Borgi, T., Zoghlami, N., & Abed, M. (2017). *Big Data for Transport and Logistics :A Review.* 2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET). Retrieved from https://www.researchgate.net/publication/318692922_Big_data_for_transport_and_logistics_A_review

Clark, J. (2016, November 17). What is the Internet of Things? Retrieved from IMB: https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/

Cook, K. (2019, February 12). *Use Machine Learning Applications of Google*. Retrieved from House of Bots: https://www.houseofbots.com/news-detail/11399-1-use-machine-learning-applications-of-google

Desai, P., Potia, A., & Salsberg, B. (n.d.). *Retail 4.0: The Future of Retail Grocery in Digital World .* McKinsey&Company . Retrieved from https://www.sipotra.it/wp-content/uploads/2017/06/The-future-of-retail-grocery-in-a-digital-world.pdf

Desjardins, J. (2014, July 24). *Why Big Data Keeps Getting Bigger*. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2019/07/why-big-data-keeps-getting-bigger/

Desjardins, J. (2019, April 17). How much data is generated each day?. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Dhawan, V., & Dhawan, V. (2014). *Big data and social media analytics.* Retrieved from https://www.cambridgeassessment.org.uk/Images/465808-big-data-and-social-media-analytics.pdf

Dooley, B. J. (2018). *Data Fabrics for Big Data*. Retrieved from TDWI: https://tdwi.org/articles/2018/06/20/ta-all-data-fabrics-for-big-data.aspx?m=1

Firican, G. (2017, February 8). *The 10 Vs of Big Data*. Retrieved from tdwi: https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

Gallagher, S. (2012, March 15). *How IBM's Deep Thunder delivers "hyper-local" forecasts 3-1/2 days out*. Retrieved from Ars Technica: https://arstechnica.com/information-technology/2012/03/how-ibms-deep-thunder-delivers-hyper-local-forecasts-3-12-days-out/

Gartner. (n.d.). Gartner. Retrieved from Gartner Glossary - Big Data: https://www.gartner.com/en/information-technology/glossary/big-dataEN9D3-ik_&sig=ACfU3U34jhcYZrUiSfjaGiH7q2ReoGZD-w&hl=

Gyamfi, Albert, Williams, & Idongesit. (2019). Big Data and Knowledge Sharing in Virtual Organizations. Retrieved from https://books.google.com.kh/books?id=CvGCDwAAQBAJ&pg=PA19&lpg=PA19&dq=by+2020,+the+amount+of+digital+information+available+will+have+grown+from+around+5+zettabytes+today+to+50+zettabytes.&source=bl&ots=HEN9D3-ik_&sig=ACfU3U34jhcYZrUiSfjaGiH7q2ReoGZD-w&hl=

IBM. (2016, June 15). *The Weather Company Launches 'Deep Thunder' - the World's Most Advanced Hyper-Local Weather Forecasting Model for Businesses*. Retrieved from https://www-03.ibm.com/press/us/en/pressrelease/49954.wss

*International Symposium on the use of Big Data for official statistics*. (2019, October 18). Retrieved from BigData UN Global Working Group: https://unstats.un.org/bigdata/

J. Clement. (2020, 24 24). Worldwide digital population as of April 2020. Retrieved from Statista: https://www.statista.com/sta

James, M., Chui, M., Brown, B., Bughin, J., Richard Dobbs, C. R., & Byers, A. H. (2011). *Big Data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx

Jarrar, Y. (2017, February 13). *What is the role of government in the digital age?* Retrieved from World Economic Forum : https://www.weforum.org/agenda/2017/02/role-of-government-digital-age-data/

Kapil, G., Agrawal, A., & Khan, R. A. (2016). A Study of Big Data Characteristics. *2016 International Conference on Communication and Electronics Systems (ICCES).* Research Gate. Retrieved from 315867458_A_study_of_big_data_characteristics

Khan, M. A.-u.-d., Uddin, M. F., & Gupta, N. (2014). *Seven V's of Big Data Understanding Big Data to extract Value.*

Zone 1 Conference of the American Society for Engineering Education.

Kortmann, C. (2016). Big Data Analytic Tooling and Cyber-Risk Management. In R. Taplin, *Managing Cyber Risk in the Financial Sector: Lessons from Asia, Europe and the USA* (pp. 63-76). New York: Routledge. Retrieved from https://books.google.com.kh/books?id=CIBwCwAAQBAJ&pg=PA63&lpg=PA63&dq=extremely+large+data+sets+that+may+be+analysed+computationally+to+reveal+patterns,+trends,+and+associations,+especially+relating+to+human+behaviour+and+interactions.&source=bl&ots=_B-0I

Madanian, S., Parry, D. T., Airehrour, D., & Cherrington, M. (2019). *mHealth and big-data integration: promises for healthcare system in India.* BMJ Health & Care Informatics. Retrieved from https://informatics.bmj.com/content/26/1/e100071#article-bottom

Marr, B. (2017, August 14). *Want To Use Big Data? Why Not Start Via Google, Facebook, Amazon, (Etc.)*. Retrieved from Forbes: https://www.forbes.com/sites/bernardmarr/2017/08/14/want-to-use-big-data-why-not-start-via-google-facebook-amazon-etc/#55365863d5db

Marr, B. (2019, October 18). *What's The Difference Between Structured, Semi-Structured And Unstructured Data?* Retrieved from Forbes: https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/#34d053a62b4d

Mauro, A. D., Greco, M., & Grimaldi, M. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics. *Research Gate*. Retrieved from https://www.researchgate.net/publication/265775800_What_is_Big_Data_A_Consensual_Definition_and_a_Review_of_Key_Research_Topics

Mihajlovic, I. (2019, April 18). *What is Big Data? Let's answer this question!* Retrieved from Medium : https://towardsdatascience.com/what-is-big-data-lets-answer-this-question-933b94709caf

Montcheuil, Y. D. (n.d.). *Facebook: A Decade of Big Data*. Retrieved from Wired: https://www.wired.com/insights/2014/03/facebook-decade-big-data/

Moss, R. J., Süle, A., & Kohl, S. (2017). eHealth and mHealth. *European Journal of Hospital Pharmacy, 26*(1). Retrieved from https://ejhp.bmj.com/content/ejhpharm/26/1/57.full.pdf

Muhaxheri, E. (2018, September 4). *Is it really the end of Privacy? The Controversy of Fingerprint, Facial Recognition, and Big Data*. Retrieved from PECD Insight: https://insights.pecb.com/end-privacy-fingerprint-facial-recognition-big-data/

Muse, D. (2017, June 20). *Structured Data*. Retrieved from Datamation: https://www.datamation.com/big-data/structured-data.html

Niño, M., & Illarramendi, A. (2015). *Understanding Big Data: Antecedents, Origin and Later Development.* Dyna Technologies.

Normandeau, K. (2013, September 12). *Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity*. Retrieved from Inside Big Data: https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/

Panimalar.S, A., Shree.S, V., & Kathrine.A, V. (2017, September). The 17 V's Of Big Data. *International Research Journal of Engineering and Technology (IRJET), 4*(9).

Retrieved from https://www.irjet.net/archives/V4/i9/IRJET-V4I957.pdf

Polyzotis, N., Roy, S., & Whang, S. E. (2018). *Data Lifecycle Challenges in Production Machine.* SIGMOD Record. Retrieved from https://sigmodrecord.org/publications/sigmodRecord/1806/pdfs/04_Surveys_Polyzotis.pdf

Puntambekar, A. (2008). Data Structures and Algorithms. Retrieved from https://books.google.com.kh/books?id=WR-antpy_sIC&pg=RA1-PA2&dq=what+is+data?&hl=en&sa=X&ved=0ahUKEwiC8fHMktbpAhUVheYKHTdXBbYQ6AEIeTAJ#v=onepage&q=what%20is%20data%3F&f=false

Rijmenam, M. v. (2019, August 14). *4 Ways Big Data Creates A Smart Transportation Industry*. Retrieved from Medium: https://medium.com/@markvanrijmenam/4-ways-big-data-creates-a-smart-transportation-industry-247418d3ec87

Rouse, M. (n.d.). *Big Data*. Retrieved from Search Data Management: https://searchdatamanagement.techtarget.com/definition/big-data

*Signal*. (n.d.). Retrieved from Technopedia: https://www.techopedia.com/definition/14154/signal-electronics

Signity. (n.d.). *Big Data and Analytics Solutions*. Retrieved from Signity: https://www.signitysolutions.com/analytics-big-data

Singh, N., Garg, N., & Mittal, V. (2014). Big Data – insights, motivation and challenges. *International Journal of Scientific & Engineering Research, 4*(12). Retrieved from https://www.ijser.org/researchpaper/Big-Data-insights-motivation-and-challenges.pdf

Statista Research Department. (2020, February 19). Number of internet of things (IoT) connected devices worldwide in 2018, 2025 and 2030. Retrieved from Statista: https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/

Subramaniam, A. (2019, July 4). *Big Data Analytics – Turning Insights Into Action*. Retrieved from Edureka: https://www.edureka.co/blog/big-data-analytics/#Why%20Big%20Data%20Analytics?

Taylor, C. (2018, March 28). Structured vs. Unstructured Data. Retrieved from Datamation: https://www.datamation.com/big-data/structured-vs-unstructured-data.html

Upadhyaya, S., & Kynclova, P. (2017). Big Data – Its relevance and impact on industrial statistics. Vienna: United Nations Industrial Development Organization. Retrieved from https://www.unido.org/api/opentext/documents/download/9921461/unido-file-9921461

Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). *Big Data Analytics: Applications, Prospects and Challenges.* Retrieved from https://www.researchgate.net/publication/320771893_Big_Data_Analytics_Applications_Prospects_and_Challenges

Vuleta, B. (2020, January 30). *How Much Data Is Created Every Day? [27 Powerful Stats]*. Retrieved from Seed Scientific: https://seedscientific.com/how-much-data-is-created-every-day/

*What is Big Data?* (n.d.). Retrieved from NTNU: https://www.ntnu.edu/ie/bigdata/what-is

Wielki, J. (2013). *Implementation of the Big Data concept in organizations - Possibilities, impediments and challenges.* Federated Conference onComputer Science and Information Systems. Retrieved from https://www.researchgate.net/publication/257341462_Implementation_of_the_Big_Data_concept_in_organizations_-_Possibilities_impediments_and_challenges

Wing, J. M. (2019, June 23). *The Data Life Cycle*. Retrieved from HDSR: https://hdsr.mitpress.mit.edu/pub/577rq08d

World Health Organization . (n.d.). *eHealth at WHO*. Retrieved from World Health Organization : https://www.who.int/ehealth/about/en/

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., . . . Dickinson, R. (2013, September 15). The role of satellite remote sensing in climate. *Nature Climate Change*. Retrieved from https://www.eenews.net/assets/2014/02/11/document_cw_01.pdf

Zillner, S., & Neururer, S. (2016). Chapter 10: Big Data in the Health Sector. Retrieved from https://www.researchgate.net/publication/299617192_Big_Data_in_the_Health_Sector

Cambodia Development Center

Cambodia Development Center (@cd.centerkh)

Cambodia Development Center

Cambodia Development Center

Building E, University of Puthisastra, #55, Street 184, Sangkat Boeung Raing, Khan Daun Penh

info@cd-center.com | (+855) 10 950 456

មជ្ឈមណ្ឌលអភិវឌ្ឍន៍កម្ពុជា
Cambodia Development Center