International Journal of Scientific Research in Information Systems and Engineering (IJSRISE)
Volume 1, Issue 2, December-2015. ISSN 2380-8128

The Future of Big Data in Cloud

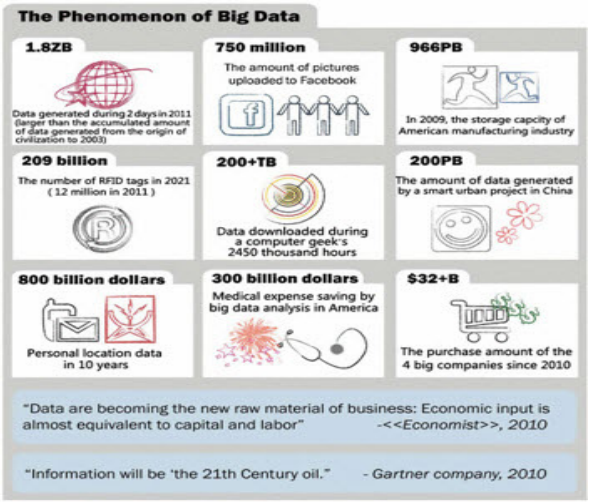
Emmanuel Oshogwe Akpeokhai
Management Information System (MBA), Girne American University, Cyprus

Abstract - Big data describes a large volume of data. This data is used on daily basis, it is not the amount of data that is most important, it is what the data is used for that matters and is most important, especially when it has to do with using the data to increase the efficiency and effectiveness of the business. Big data can be analyzed for insights that lead to better decisions and strategic business moves. The size of this data created is beyond the ability of typical database software tools to capture, store, manage, and analyze. Nowadays Big data ranges from a few dozen of terabytes to multiple petabytes (thousands of terabytes). This data can be measure in the value of data initiatives along the dimensions of the four Vs of big data. Measuring the value of data is an endless process, it is the way we make use of data that allows us to fully recognize its true value and importance to improve our decision making capabilities. These approach place high emphasis on the importance of every individual data item that goes into these systems and, as a result, highlight the importance of every single outcome linking to business impacts delivered. Big Data is now becoming a critically important driver of business success across sectors, but many executives say they don't think their companies are equipped to make the most of it.

Keywords *– Big Data, Cloud, Computing, Organizations, Technology.*

1. Big Data

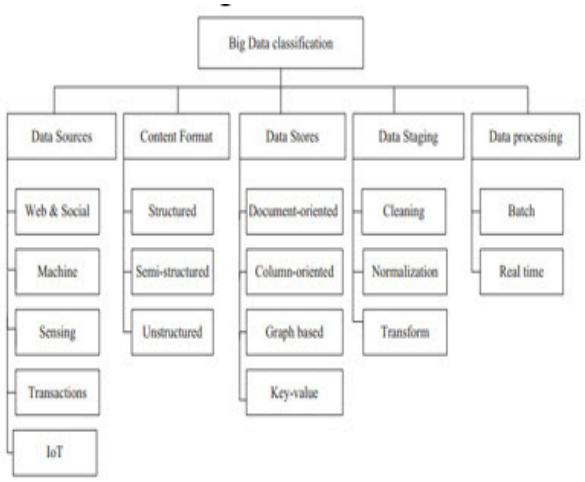
Big data describes a large volume of data, this data could be structured or unstructured, which is used by an organization on daily basis. It is not the amount of data that is most important, it is what organizations do with the data that matters and is most important especially when it has to do with using the data to increase the efficiency and effectiveness of the business. Big data can be analyzed for insights that lead to better decisions and strategic business moves [1]. The size of this data created is beyond the ability of typical database software tools to capture, store, manage, and analyze. Today Big data in many sectors range from a few dozen of terabytes to multiple petabytes (thousands of terabytes). Figure 1 shows Phenomenon of big data.



Source*: Springer* ***Figure 1****: Phenomenon of Big data*

Big Data is relatively new, the concept gained momentum in the early 2000s when an industry analyst **Doug Laney** articulated the now mainstream definition of Big Data [2]. Big data are classified into five different aspects: Data sources, content format, data stores, data staging, and data processing. Each of these categories has its own characteristics and complexities as described in the figure 2.

Corresponding Author
Emmanuel Oshogwe Akpeokhai
Management Information System (MBA), Girne American University, Cyprus



Source: *Elsevier (Big Data Research)*
Figure 2*: Big Data Classification*

2. The Value of Big Data

In most part of the world and in various organizations, the leaders wonder whether they are getting full value from the data they already have in their organizations. New technologies are collecting more data than ever before, yet many organizations are looking for a better way to obtain value from their data [3].

There are many ways that big data can be used to create value across sectors of the global economy. Many pioneering companies are already using big data to add value to their company, and others need to explore how they can do the same if they are to compete. The Governments also have a significant opportunity to boost their efficiency and the value for money. The possibilities of big data to continue to evolve rapidly, is driven by innovation in the underlying technologies, platforms, and analytic capabilities for handling data, as well as the evolution of behavior among its users as more and more individuals live digital lives by contributing to more storage of data in the cloud [4].

2.1. How to measure the value of big data

Organizations can measure the value of data initiatives along the dimensions of the four Vs of big data. Data itself is quite often inconsequential in its own right. Measuring the value of data is an endless process. Data is as valuable as the business outcomes it makes possible. It is the way we make use of data that allows us to fully recognize its true value and importance to improve our decision making capabilities. There are multiple approaches in improving a business's decision making process and to determine the ultimate value of data, including data warehouses, business intelligence systems, and analytics sandboxes and solutions. These approach place high emphasis on the importance of every individual data item that goes into these systems and, as a result, highlight the importance of every single outcome linking to business impacts delivered.

Big data characteristics are defined popularly through the four Vs: volume, velocity, variety and veracity. Adapting these four characteristics provides multiple dimensions to the value of data at hand. Essentially, there is an assumption that the data has great potential, but no one has explored where that might be. Unlike a business intelligence system, where analysts know what information they are seeking, the possibilities of exploring big data are all linked to identifying connections between things we don't know. It is all about designing the system to decipher this information

The possible approach here is to take the four Vs into consideration and determine what kind of value they deliver while solving a particular business problem [5].

Volume-based value

Most organizations have the ability to store as much data as possible in a cost effective manner, they have the capabilities to do broader analysis across different data dimensions and also deeper analysis going back to multiple years of historical context behind data. The scenario applies heavily into developing true customer centric profiles. The more data businesses have on the customers, both recent and historical, the greater the insights. This will in turn lead to generating better decisions around acquiring, retaining, increasing and managing those customer relationships [5].

Velocity-based value

This has to do with the speed, now speed is more important than ever. The faster businesses can inject data into their data and analytics platform, the more time they will have to ask the right questions and seek answers. Rapid analysis capabilities provide businesses with the right decision in time to achieve their customer relationship management objectives [5].

Variety-based value

In the digital era, capability to acquire and analyze varied data is extremely valuable, as the more diverse customer data businesses have, the more multi-faceted view they develop about their customers. This in turn provides deep insights into successfully developing and personalizing customer journey maps, and provides a platform for businesses to be more engaged and aware of customer needs and expectations [5].

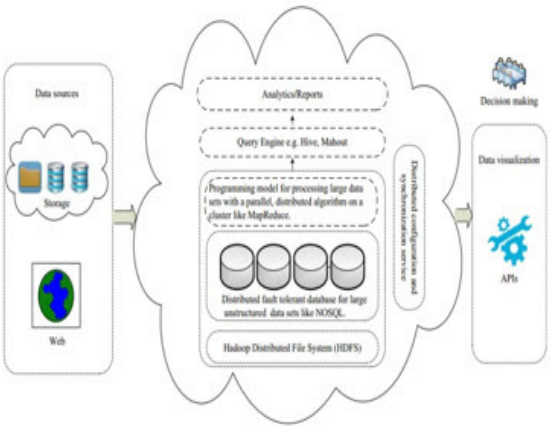
Veracity-based value

While many question the quality and accuracy of data in the big data context, but for innovative business offerings the accuracy of data is not that critical – at least in the early stages of concept design and validations. Thus the more business hypotheses that can be churned out from this vast amount of data, the greater the potential for business differentiation edge [5].

3. Relationship between big data and cloud computing

Big Data and Cloud Computing are conjoined. Big data provides users the ability to use commodity computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. Cloud computing provides the underlying engine through the use of Hadoop, a class of distributed data processing platforms. The use of cloud computing in big data is shown in Fig 3.

Large data sources from the cloud and Web are stored in a distributed fault tolerant database and processed through a programming model for large datasets with a parallel distributed algorithm in a cluster. The main purpose of data visualization, as shown in Fig. 3, is to view analytical results presented visually through different graphs for decision making. Big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast growing cloud based applications developed using [6].



Source: *Elsevier* ***Fig. 3****: Cloud computing usage in big data.*

4.Analysis of Big Data

Big Data is quickly becoming a critically important driver of business success across sectors, but many executives say they don't think their companies are equipped to make the most of it. **Bain & Company** surveyed executives at more than 400 companies around the world, most with revenues of more than $1 billion. We asked them about their data and analytics capabilities and about their decision-making speed and effectiveness. Nowadays, the data that need to be analyzed are not just large, but they are composed of various data types, and even including streaming data. Since big data has the unique features of "massive, high dimensional, heterogeneous, complex, unstructured, incomplete, noisy, and erroneous," which may change the statistical and data analysis approaches. Although it seems that big data makes it possible for us to collect more data to find more useful information, the truth is that more data do not necessarily mean more useful information. It may contain more ambiguous or abnormal data. For instance, a user may have multiple accounts, or an account may be used by multiple users, which may degrade the accuracy of the mining results. Therefore, several new issues for data analytics come up, such as privacy, security, storage, fault tolerance, and quality of data.

Big data may be created by handheld device, social network, internet of things, multimedia, and many other new applications that all have the characteristics of volume, velocity, and variety. As a result, the whole data analytics has to be re-examined from the foregoing perspectives[7].

There are some big data analysis frame-

works and platforms which can be divided into Processing/Compute: Hadoop, Nvidia CUDA or Twitter Storm , Storage: Titan or HDFS, and Analytics: MLPACK or Mahout Although there exist commercial products for data analysis most of the studies on the traditional data analysis are focused on the design and development of efficient and/or effective "ways" to find useful things from the data.

4.1. OSINT and Big Data From The Deep Web

OSINT is employed by government intelligence agencies. OSINT stands for "Open-Source INTelligence", which refers to any unclassified information and includes anything freely available on the web. OSINT is the opposite of close-source intelligence or classified information [8]. Common OSINT sources include social networks, forums, business websites, blogs, videos, and news sources etc OSINT is unclassified and available, but it is not always easily found. Much of it is available on the Deep Web or "hidden internet" as most people often call it. The information is publicly available, but link-crawling search engines like Google do not always access it. **Big Data from the Deep Web is OSINT**. Your organization can benefit from exploiting that information. There is information publicly available online right now that you are missing by searching with Google, or not searching for at all because you aren't even aware it exists. If you want to find counterfeiters of your product online? You can use OSINT to track them down. If you want to monitor real-time activities for your current financial investments or possible future investments? That information is publicly available on the Deep Web and can be harvested and delivered to you in real-time. If want to know all the available grants for your area of research, who is offering them, and who has won them in the past? Again, that is OSINT and ready to be harvested. OSINT, Big Data, unstructured data and the Deep Web can all be nearly impossible to navigate without the proper tools. Even if you find the data, harvesting the data at scale, analysis, and visualizations often require additional licenses, subscriptions, and a bigger budget [8].

5.Conclusion

Big data will change how we live in every area of our life [9]. Gradually, we are moving from text based communications to richer data which include images, videos, and interactive maps as well as associated metadata such as geo-location information and time and date stamps. Twenty years ago, ISDN (Integrated Service Digital Network) lines couldn't handle much more than basic graphics [10], but today's high-speed communication networks enable the transmission of storage-intensive data types. For instance, Smartphone users can take high-quality photographs and videos and upload them directly to social networking sites via Wi-Fi and 3G or 4G LTE networks. Advances in data storage and mining technologies make it possible to preserve increasing amounts of data generated directly or indirectly by users and analyze it to yield valuable new insights. For example, companies can study consumer purchasing trends to better target marketing. In addition, near-real-time data from mobile phones could provide detailed characteristics about shoppers that help reveal their complex decision-making processes as they walk through malls.

While big data can yield extremely useful information, it also presents new challenges with respect to how much data to store, how much this will cost, whether the data will be secure, and how long it must be maintained. For example, both companies and law enforcement agencies increasingly rely on video data for surveillance and criminal investigation. Closed-circuit television (CCTV) is ubiquitous in many commercial buildings and public spaces. Corporations are using big data to learn more about their workforce, increase productivity, and introduce revolutionary business processes. Emerging online apps will not only enable users to upload video via mobile social networking but will soon incorporate wearable devices in the form of a digital watch or glasses to allow for continuous audiovisual capture. People will essentially become a camera. This publicly available data will dwarf that generated by today's CCTV cameras.

Investments in big-data computing will have extraordinary near-term and long-term benefits. The technology has already been proven in some industry sectors; the challenge is to extend the technology and to apply it more widely.

Reference

[1] SAS academy of data Science (2015) http://www.sas.com/en_us/insights/big-data/what-is-big-data.html#dmhistory
[2] Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.
[3] LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N.(2013). Big data, analytics and the path from insights to value. MIT sloan management review, 21.
[4] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A. (May 2011) Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
[5] Rossi B. (21 August 2015). How to measure the value of big data. Insight and analysis for IT leaders. http://www.information-age.com/it-management/strategy-and-innovation/123460041/how-measure-value-big-data.
[6] Ibrahim I., Targio A., Hashem a., Mokhtar S., AGani a, Samee U. The rise of "big data" on cloud computing: Review and open research Information Systems 47 (2015) 98–115.
[7] Tsai, C-Feng Lai, H Chao, Athanasios Vasilakos Big data analytics: a surveyJournal of Big Data 2015, 2:21 (1 October 2015) DOI 10.1186/s40537-015-0030-3C.
[8] What is osint and how can your organization use it. Bright Planet Deep Web Inteligence. http://www.brightplanet.com/2013/04/what-is-osint-and-how-can-your-organization-use-it/
[9] Michael, K., & Miller, K. (2013). Big data: New opportunities and new challenges [guest editors' introduction]. Computer, 46(6), 22-24. DOI:10.1109/MC.2013.196.
[10]Stallings, W. (1992). ISDN and broadband ISDN. Macmillan Publishing Co., Inc..