INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

Big Data: Definition, Architecture & Applications

Humam Khalid Yaseen [#], Ahmed Mahdi Obaid [#]

[#] Master of computer science, Iraqi Sunni Affairs, Iraq
E-mail: humamk84@gmail.com

Abstract— **Big data is a term for massive data sets having large, more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results. The process of research into massive amounts of data to reveal hidden patterns and secret correlations named as big data analytics. These useful informations for companies or organizations with the help of gaining richer and deeper insights and getting an advantage over the competition. For this reason, big data implementations need to be analyzed and executed as accurately as possible. In this paper; Firstly, we will discuss what big data and how it is defined according to different sources; Secondly, what are the characteristics of big data and where should it be used; Thirdly, the architecture of big data is discussed along with the different models of Big data; Fourthly, what are some potential applications of big data and how will it make the job easier for the persisting machines and users; Finally, we will discuss the future of Big data.**

Keywords— **Big Data, Data set, Predictive Analysis.**

I. INTRODUCTION

Big Data [1] is a global term now and is used commonly in industrial ventures and academia. Using a word frequently in more than one context under different conditions poses a threat to the definition if the word. Due to this reason it is of immense importance for Big Data to have a standard definition. In this way it will easier for it to achieve systematic evolution as well as reduce the confusion that is related to its usage. Big Data has had revolutionary success and its use in a large number of industries it now has various meanings depending on its use in these industries. By considering the use of Big Data in academic literature as well as business literature Big Data can be divided into four themes that are:

1. Information
2. Technologies
3. Methods
4. Impact

It can be said that the above mentioned themes cover a vast majority of fields in which Big Data is used.

Looking at the reasons to why Big Data has become such a success we see that the major factor is the amount of information that has been generated up till now and has to be made available. Digitization has been a major factor in producing massive amounts of data. Digitization can be defined as a process in which analogue information is being continuously converted into discrete and digital information that can be read by a machine or is in a machine-readable format. The reason behind the popularity of digitization was the conversion of books into digital format. This led to the introduction of digital libraries that used to work on optical character recognition (OCR) [2]. Google Print Library Project 3 was the most popular attempt to of mass digitization. It started in 2004 and its aim was to digitize 15 million books that were a part of libraries around the world and some in some well-known institutes such as Stanford, Harvard and Oxford. The next step of digitization is datafication and there is a slight difference between these two terms [3]. Digitization stores the obtained information in a convenient method whereas datafication aims to organize the digitized data in such a way that the analog signals of the respective digitized signal would generate useful insights that would not have been obtained from the original signal.

According to an estimate by Cisco in between 2008 and 2009 the number of devices in this world overtook the number of living people [4]. According to a research by Gartner [5] there will be approx. 26 billion devices in this world by 2020. That makes it three devices per person. The ability of technology such as RFID tags, mobile phones, actuators and sensors to communicate with each other and reach certain common goals is known as Internet of Things (IoT) [6] [7]. Due to such massive numbers of operating devices companies get massive data sets and through these data sets they can improve their operating procedures, business models and can reduce risk and get better profit rates [8]. It is safe to say that IoT is the biggest application of Big Data and also the most promising one. There has been a lot of variety in structured data (Numeric information) and it is now being joined by unstructured data such as text,

audio files, video files etc. and also semi-structured data such as RSS feeds [9]. This diversity is a challenge that for organizations that need to give their data any sorts of value [10].

Due to the absence of any definite or formal definition of Big Data scholars have made many definitions according to the characteristics, technology or trends related to Big Data. The present definitions of Big Data present a very different picture of what actually Big Data is rather they consider it as a term that defines a social phenomenon, analytical technique, a process or a data set.

By looking at the definitions proposed by the scholars it is seen that these definitions can be divided into four groups. The first group caters for the definitions related to the attributes of data. The second group caters for definitions related to the technical aspect and needs. The third group caters for thresholds and ways to overcome them. The fourth and the last group caters for the social impact. If we look closely these are related to the four themes of Big Data that were explained earlier. Now we will discuss the definitions that fall in these four groups and come up with a formal definition that caters for all these aspects.

1. 1st Group: Attributes of Data

The definition proposed by Dijicks is:

"The four characteristics defining big data are Volume, Velocity, Variety and Value [11]."

The definition proposed by Bayer and Laney is:

"High volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making [12]."

The definition proposed by Intel is:

"Complex, unstructured, or large amounts of data [13]."

The definition proposed by Shroeck, Tufano, Smart, Romero-Morales and Shockley is:

"Big data is a combination of Volume, Variety, Velocity and Veracity that creates an opportunity for organizations to gain competitive advantage in today's digitized marketplace [14]."

2. 2nd Group: Technological needs

The definition proposed by Ward and Barker is:

"The storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning [15]."

The definition proposed by Microsoft Research is:

"The process of applying serious computing power, the latest in machine learning and artificial intelligence, to seriously massive and often highly complex sets of information [16]."

The definition proposed by NIST Big Data Public Working Group is:

"Extensive datasets, primarily in the characteristics of volume, velocity and/or variety that require a scalable architecture for efficient storage, manipulation, and analysis [17]."

3. 3rd Group: Thresholds

The definition proposed by Schneiderman is:

"A dataset that is too big to fit on a screen [18]"

The definition proposed by Manyika, Bughin, Chui and Brown is:

"Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze [19]."

The definition proposed by Fisher, DeLine, Czerwinski and Drucker is:

"Data that cannot be handled and processed in a straightforward manner [20]."

The definition proposed by Chen, Chiang and Storey is:

"The data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies [21]."

The definition proposed by Dumbill is:

"Data that exceeds the processing capacity of conventional database systems [22]."

4. 4th Group: Social Impact

The definition proposed by Boyd and Crawford is:

"A cultural, technological, and scholarly phenomenon that rests on the interplay of Technology, Analysis and Mythology [23]."

The definition proposed by MayerSchönberger and Cukier is:

"Phenomenon that brings three key shifts in the way we analyze information that transform how we understand and organize society: 1. More data, 2. Messier (incomplete) data, 3. Correlation overtakes causality [24]."

Having examined and read all the different types of definitions of Big Data we can say that in order to define Big Data we have to include the following characteristics of Big Data:

- 'Volume', 'Velocity' and 'Variety', to describe the characteristics of Information involved [25].
- Specific 'Technology' and 'Analytical Methods', to clarify the unique requirements strictly needed to make use of such Information [25].
- Transformation into insights and consequent creation of economic 'Value', as the principal way Big Data is impacting companies and society [25].

Therefore, the formal definition of Big Data is proposed as:

"Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value [25]."

Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD, while others view data mining as merely an essential step in the process of knowledge discovery. The knowledge discovery process as an iterative sequence of the following steps:

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined) A popular trend in the information industry is to perform data cleaning and data integration as a preprocessing step, where the resulting data are stored in a data warehouse.
3. Data selection (where data relevant to the analysis task are retrieved from the database)
4. Data transformation (where data are transformed and consolidated into forms appropriate for mining by

performing summary or aggregation operations) Sometimes data transformation and consolidation are performed before the data selection process, particularly in the case of data warehousing. Data reduction may also be performed to obtain a smaller representation of the original data without sacrificing its integrity.

5. Data mining (an essential process where intelligent methods are applied to extract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledge based on interestingness measures
7. Knowledge presentation (where visualization and knowledge representation techniques are used to present mined knowledge to users) [25][26].

II. CHARACTERISTICS

There are specific characteristics associated with Big Data that help in identifying it. These characteristics are as follows:

1. Volume – This characteristic is associated with the amount of data that is being generated. This is a very important aspect of Big Data. The quantity and size of the data determine whether such data set falls in the category of Big Data or not. Looking at the name of Big Data it is obvious to say that the size of data set is an important characteristic in making it related to Big Data [26].
2. Variety – This characteristic is associated with the category of the data set in Big Data. There are various categories of data set and knowing in which category the data set falls in is an essential aspect of Big Data. This is a very important characteristics from the point of view of the data analysts as this helps them in using the data set efficiently and taking maximum advantage from it.
3. Velocity – This characteristic of Big Data deals with how quickly is the data generated and is processed. It deals with how fast data is being produced and processed to meet the needs and requirements of the consumer.
4. Variability – This characteristic is associated with the variability present in the data set of Big Data that is being analyzed. This is referring to the variety of data and the inconsistency that comes in the data set while analyzing it and this characteristic poses to be a problem for the analysts that are handling an analyzing data and hampers the process of data management and data handling [26].
5. Veracity – This characteristic is associated with the quality and accuracy of the data set. The quality of the data that is gathered can vary a lot and this affects the accuracy of the analysis. Therefore, the accuracy and correctness of the analysis depends upon the veracity of the data set and its source.
6. Complexity – This characteristic deals with data management and data analysis. Data management and data analysis is a complicated and complex process especially when we have massive data sets that are being received from multiple sources. In order to understand and analyze such massive data

sets, they have to be interlinked, connected and related to each other. This makes the data set a bit complex and hence, is termed as complexity of Big Data. [27].

III. ARCHITECTURE

The architecture of Big Data consists of methods and mechanisms for collecting and storing data, securing it, processing it and then converting it into data base structures and file systems. The analysis tools help us analyze the data that is collected and then make intelligent decisions on the basis of this collected data. Hence, the greater the amount of data collected and analyzed the better will be the decision taking ability of the machine or device. The architecture of Big Data consists of multiple layers. Firstly, the logical layers of Big Data are discussed. There are four logical layers that are as follows:

1. Big data sources layer: The data that comes to Big Data has a lot of various sources. These sources can be company servers, third-party data providers and various sensors relate to companies. Big Data has the ability to store and take in data in two modes namely real-time mode and batch mode. Some examples of the sources of data include applications and softwares such as MS Office docs, ERP, Relational Database Management System (RDMS), mobile devices, social media, sensors, data warehouses and email.
2. Data messaging and storage layer: All the data from various sources is received by this layer. If the data received is unstructured and is not in a format that could be understood by the analytic tools then this layer converts this data into a format that is readable by the analysis tools. In Big Data, unstructured data is stored in specialized file systems such as Hadoop Distributed File System (HDFS) or in a NoSQL database whereas, structured data is stored in RDBMS.
3. Analysis layer: This layer deals with the analysis of stored data. In this layer the stored data is analyzed to extract various trends and business intelligence from it. Many different sorts of tools operate in big data environment. For the analysis of structured data techniques such as sampling is used whereas, for unstructured data advanced and never specialized analytics toolsets are required.
4. Consumption layer: All the analyzed data is received by this layer. The task of this layer is to present this analyzed data as an output to the desired receiver. There are various types of outputs such as applications, business processes and human viewers [28].

There are four types of processes that operate in between these four logical layers. These cross-layer operations are: connecting to data sources, governance, systems management and quality of service (QoS). These operations are explained in detail as follows:

1. Connecting to data sources: The data received by Big Data is at a very fast rate. In order to quickly receive and analyze this data we need to have connections that can support these actions at a fast rate. For that the architecture requires adapters and connecters that can connect data to the storage system, protocols and networks [28].
2. Governing big data: The architecture of Big Data provides privacy as well as security of the data that it

receives and analyzes. The organizations using Big Data has a choice to use a security tool of their own on the analytics storage system, spend in a specialized software to keep their Hadoop environment safe and secure or they can sign an agreement with their cloud Hadoop provider that provide service level security. The policies that deal with the protection and security of data should include the security of the process starting from data ingestion till analysis and deletion or archiving of data [28].

3. Managing systems: The architecture of Big Data is a large-scale cluster which has a distributive structure that has highly scalable performance and capacity. It should regularly and continuously check the health of the system with the help of central management system consoles. If the consumer is using cloud as an environment for Big Data then they should establish and monitor Strong Service Level Agreements (SLAs) with their cloud provider.

4. Quality of service: Quality of service is an important aspect of Big Data and it is the framework that helps define the quality of data, security and compliance policies, sizes and frequency of incoming data sets and filtering data [28].

Some other important components of Big Data architecture are:

• Batch processing: In some cases of Big Data the data sets are very large and in order to provide a solution the data files using long-running batch jobs that help filter, combine and prepared data for analysis. Such jobs involves the reading of source files, processing them and then writing an output for the new file. Some of the options that can choose from consists of running U-SQL jobs in Azure Data Lake Analytics, using Hive, Pig, or custom Map/Reduce jobs in an HDInsight Hadoop cluster, or using Java, Scala, or Python programs in an HDInsight Spark cluster [29].

• Real-time message ingestion: If the solution of the Big Data consists of real-time data sources then the architecture of Big Data should consist of a process that captures and saves real-time messages for stream processing. This can be a store that has simple data in it and in that store the incoming messages are dropped for processing. Most of the solutions require a message ingestion store that acts as a buffer for the incoming messages, provides reliable delivery, and supports the scale-out process and other message queuing semantics. Some of the Options for this process include Azure Event Hubs, Azure IoT Hubs, and Kafka [29].

• Stream processing: Once the real-time messages are received the solution will process these messages by filtering them, combining them and then preparing it for analysis. Once the data is processed it is then written on an output sink [29].

• Analytical data store: There are many Big Data solutions that firstly prepare the data for analysis and then convert it in the form of a structured data that can be used by the analytical tools. Analytical data store that is used to answer such queries can be explained as a Kimball-style relational data warehouse similar to the ones that are seen in most traditional business intelligence (BI) solutions. Another way to present data is through a low-latency NoSQL

technology such as HBase, or an interactive Hive database that provides a metadata abstraction over data files in the distributed data store [29].

• Analysis and reporting: The purpose of Big Data and its solution is to give meaning to data by analyzing it and then reporting the findings. In order to give users the ability to analyze data there is a layer in the architecture of Big Data known as the Data Modeling Layer such as tabular data model in Azure Analysis Services. Data scientists and Data analysts can interact with the data and explore it with the help of analysis and reporting. For large-scale data exploration Microsoft R Servers can be used [29].

• Orchestration: Many Big Data solutions comprise of data processing operations that are repeated and they transform the source data and move it between different sources. Then they either load the processed data into analytical data storage or push forward the results into a dashboard or report [29].
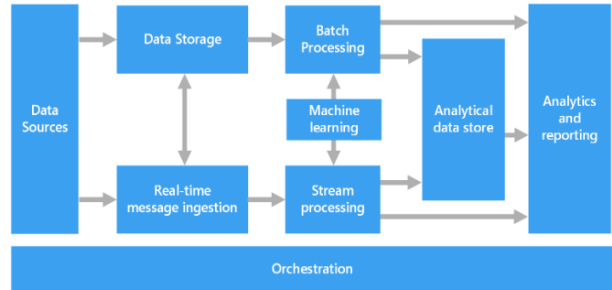

Fig. 1 Logical components of Big Data architecture

IV. APPLICATION

The amount of data produced is growing at a rapid pace and it is anticipated that by the end of 2020, 1.7 Megabyte of information will be created every second for every human being on this planet [30]. By then the total amount of data produced will be 44 trillion gigabyte [31]. By 2020, there will be 6.1 billion smartphone users in the world [32] and there will be approx. 50 billion smart devices that are connected to each other and they will be developed to collect, analyze and share data [33]. Hadoop which is a software that is used for distributive computing is expected to grow annually at the rate of 58% surpassing $1 billion mark by 2020 [34]. 73% of organizations either plan or already have invested in Big Data and data analysis [35]. One interesting fact about analyzed data is that only 0.5% of the existing data is analyzed which means that Big Data has a big future [36]. This proves that there are numerous applications of Big Data [37]. Some of the applications of Big Data analytics are as follows:

1. *Health care*

It is anticipated that the data related to health care will grow dramatically in the coming years [38]. The field of healthcare is headed in the direction of technologies and procedures such as pay for performance and meaningful use of resources. Using the data obtained efficiently and intelligently will help save millions of dollars in revenue and profits. Healthcare organizations are should use the available tools and technology to handle data efficiently [39]. The existing data in the field of healthcare can be analyzed with the help of the analysis techniques. By understanding the

48

result of these analysis techniques we can devise and propose better solutions in the field of healthcare. The way in which this data can be utilized is that the analysis of the data will inform the physician or doctor and their patient about the best possible treatment according to the data of vitals of the patient. By using Big Data and digitizing the data obtained from patients, health related organizations can get significant benefits [40]. Some of the potential benefits are:

1. Detecting and treating diseases at an early stage.
2. Managing specific individual and population health.
3. Detecting health care fraud more quickly and efficiently.

We can predict certain outcomes with the help of the analysis performed by Big Data. The data gathered is for analysis include the following information:

1. Length Of Stay (LOS)
2. Patients that choose elective surgery
3. Patients that are not likely to get benefit from surgery
4. Complications
5. Patients that are at the risk of medical complications
6. Patients at the risk of sepsis, C. difficile, MRSA or other hospital-acquired illness
7. Progression of illness
8. Possible comorbid conditions (EMC Consulting).

According to McKinsey approx. $300 billion can be saved per year in U.S healthcare. Clinical operations and Research and Development are the two biggest department that offer savings of up to $165 billion an $108 billion. This can result in saving 8% in national healthcare expenditures [41]. According to McKinsey Big Data can help remove inefficiency in the healthcare field in the following three areas:

1. Clinical operations: Conducting research to figure out methods that are cost-effective and help diagnose and treat patients efficiently. This will result in efficient use of resources.
2. Research & development: Following changes will help reduce ineffectiveness:
   • Produce a well-defined and targeted Research and Development pipeline for drugs and devices by the help of predictive modeling and analysis techniques offered by Big Data.
   • Statistical tools and algorithms to improve clinical trial design and patient recruitment to better match treatments to individual patients, thus reducing trial failures and speeding new treatments to market
   • Processing the information related to the patients such as their clinical trials and medical history to analyze the adverse effects of medical products before they are introduced in the market.
3. Public health: Following changes will help reduce ineffectiveness:
   • Collect, process and analyze information related to the outbreaks of diseases and patterns in diseases transmissions that will help in the betterment of public health surveillance and quick response to disease outbreaks.

• Development of vaccines for a specific disease in a much more efficient and quick manner.
• Converting massive amounts of data in actionable information which can be utilized to identify needs, predict crisis, provide services and prevent crises that will result in the benefit of masses [41].

Some of the other ways in which Big Data Analytics can contribute in the field of healthcare are:

• Evidence-based medicine: Big Data can help predict the medicines that are a best match to certain patients. It does by analyzing the unstructured data-EMRs, genomic data, financial data, operational data, and clinical data.
• Genomic analytics: Such type of analysis are done on the basis of the data related to the genes of the specific patient. Executing gene sequencing in a cost effective manner can result in using genetic analysis a regular medical care procedure which will help in a better analysis of the patient [42].
• Pre-adjudication fraud analysis: Big Data can analyze a large number of requests and claims related to potential fraud, waste and abuse which will result in saving costs of expenses in the field of healthcare. Rapidly analyze large numbers of claim requests to reduce fraud, waste and abuse.
• Device/remote monitoring: With the help of remote connected devices, the data of patients can be analyzed continuously. Safety monitoring of the patients can be done with the help of remote devices in real-time and any hazardous event can be predicted with the received data and can be stopped.
• Patient profile analytics: By analyzing the analytics profiles of patients it will be easier to identify what sort of lifestyle will suit a specific patient. For example, patients that suffer from diabetes will be prescribed to keep a proactive lifestyle [43].

2. Manufacturing

Manufacturing sector has always been related to data and this sector has always been using data to transform itself and its ways pf working to achieve better quality of work, effectiveness and designing better products. This sector is the backbone of economy of many developed countries. The expansion of a certain business in the manufacturing sector results in new challenges that the organization should overcome in order to maintain themselves. Due to the ever changing technological needs and globalizations some of the countries have specialized themselves in certain stages involved in the production process. In order to maintain their growth and sustainability such countries have to analyze large data sets that can help them achieve higher levels of efficiency and produce high-quality products. Manufacturing sector stores more data than any other sector. This sector store approx. 2 Exabyte of new data in 2010. The sources of such massive data sets include process control, instrumented production machinery, monitoring systems for performance of sold products, supply chain management, etc. In 2011 the number of RFID tags sold was 12 million and this number is expected to rise to 209 billion in 2021. It systems help extend the memory in which this generated data is stored. In order to use the data more effectively the

manufactures combine data received from various systems such as computer-aided manufacturing, computer-aided design, collaborative product development management and computer-aided engineering [44]. According to McKinsey Global Institute [45] analysis big data levels across the manufacturing value chain are as follows:

- Research & Development and product design: Using Big Data helps accelerate the process of research as well as product development. It helps the producer produce products that are according to the needs of the consumer and does all that in a cost efficient manner. Following topics fall under the category of such lever that is:
  a. Product lifecycle management
     In the past the manufacturing companies were able to produce large amounts of data and store them in the implemented IT systems but they were not able to utilize or analyze these data sets that as they were trapped in these IT systems. With the help of Big Data manufacturers have introduced Product Lifecycle Management (PLM) platform that allows the already existing data set to collaborate.
  b. Design to value
     With the help of data analysis techniques offered by Big Data, manufacturers can study the market and extract vitals findings related to the market and demands of customers. These useful insights can help the manufacturers improve their products and make them according to the needs of the customers.
  c. Open innovation
     Manufactures are always looking for innovative ideas and they sometimes get these ideas from outside innovative sources. Web 2.0 can help introduce a platform for these manufacturers meet innovative individuals and perfume business with them on web-based platforms. Organizations such as Procter & Gamble and L'Oréal invite people to submit their creative ideas for innovations.

- Supply Chain: The type of manufacturers that work in fast-moving goods require a predictive analysis that can help them know what will be the demands of the product in the market in future as they have to be prepared for the production of these products. There are many trends that affect the sales and production of goods such as change in prices of the products, increase in marketing of the product, fluctuating rates of products in the market, etc. Manufacturers can improve their demand forecasting with the help of Big Data as it will provide them with an analysis of the market and its predictive analysis is made for exactly such circumstances. Manufacturers can use the data obtained from sources such as promotion data, inventory data and launch data, and then integrate them. This will help the manufacturer in getting smooth order patterns.

- Production: Big Data helps provide more efficiency to the production process. With the implementation of Internet of Things (IoT) it is becoming easier for manufacturers to use real-time data to monitor the operations of machinery:
  a. Digital factory
     With the help of the inputs obtained from the product development and historical production data

the manufacturers can create a digital model of the whole manufacturing process.
  b. Sensor-driven operations
     With the introduction of Internet of Things (IoT) manufacturers get real-time data and this data can be used to optimize to reduce waste of materials, save costs and maximize output. They also enable the manufacturers to use Nano-manufacturing which was not possible before.

3. Education

Big Data can become an alternative approach that can offer innovative ways of teaching for higher level institutes. Students generate data via smartphones, laptops, social networks, Learning Management Systems (LMS) and various other sources. Universities also have data related to students that is gathered from sources such as academic records, financial records, library information systems, etc. Integrating all such type of data will result in a profile of the student.
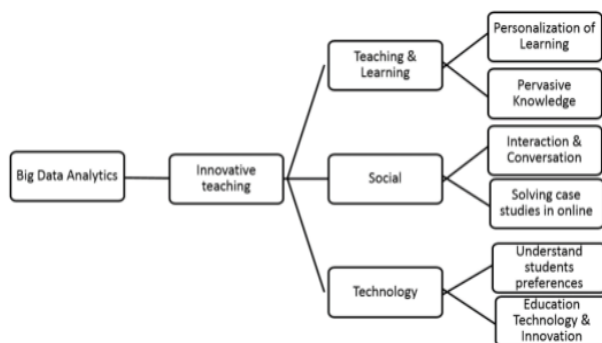

Fig. 1 Big Data and innovative teaching process

The profile of the student is based on the data collected from the above mentioned sources. With the help of Big Data analysis the student can get predictions about learning content, subject preferences, library visiting habits, etc. With the help of this analysis the teachers can provide them personalized services. If the student is diverting from the devised pattern than the system will send a notification to the student let them know. There are three innovative ways that can be implemented with Big Data in the field of education.

1. Innovative Social
   With the advancement in technology students can now study online and can find all the required material of their courses online. Students can now work online, make and distribute surveys online, share documents, interact in discussions online, study case studies and solve them, watched tutorials and online videos, and published articles on the web.
   Social media analytics include the analysis of data that is obtained from sources such as blogs, Facebook, twitter and forums.

2. Innovative teaching
   The innovation in teaching means that the students these days do not use the conventional methods of taking notes in the classes rather they search the web for content related to the topic of their study. However, the data available online is in a massive amount and in order to get the desired data we need the help of Big Data. Big Data enables us to get massive amount of data refined at a fast pace.

50

3. Innovative technology

Big Data can help predict the future trends such as education technology trends. For example the data generated from social media apps and instant messaging universities can understand the expectations of their students and stakeholders. Data analytics provide the behavioral patterns of students which is based on the online activity of the students [46]. Data analytics can also help determine the personalities and preferences of students on the basis of collected data. Big Data can also help determine the future problems that could possibly arrive in the education system [47].

V. CONCLUSION

Throughout the history of human society, the demands and willingness of human beings are always the source powers to promote scientific and technological progress. Big data may provide reference answers for human beings to make decisions through mining and analytical processing, but it could not replace human thinking. It is human thinking that promotes the widespread utilizations of big data. Big data is more like an extendable and expandable human brain other than a substitute of the human brain. With the emergence of IoT, development of mobile sensing technology, and progress of data acquisition technology, people are not only the users and consumers of big data, but also its producers and participants. Social relation sensing, crowdsourcing, analysis of big data in SNS, and other applications closely related to human activities based on big data will be increasingly concerned and will certainly cause enormous transformations of social activities in the future society. In our paper we focused on what is the big data and the layer of it and its application in order to enhance the imagination of the researchers about big data and its utilization for making the maximum utilization of its application.

REFERENCES

[1] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss and M. McKeon, " IEEE Trans. Vis. Comput.," p. Graph 1121, 2007.

[2] K. Coyle and J. Acad, " Librariansh.," p. 641, 2006.

[3] V. Mayer-Schönberger and K. Cukier, "Big Data: A Revolution That Will Transform How We Live," 2013.

[4] D. Evans, "The Internet of Things - How the Next Evolution of the Internet Is Changing Everything," 2011.

[5] Gartner, 2014. [Online]. Available: http://www.gartner.com/newsroom/id/2684616.

[6] D. Estrin, D. Culler, K. Pister and G. Sukhatme, "IEEE Pervasive Comput. 1,," IEEE, p. 59, 2002.

[7] L. Atzori, A. Iera and G. Morabito, " Comput. Networks 54," p. 2787.

[8] M. Chui, M. Löffler and R. Roberts, " McKinsey Q. 291," p. 10, 2010.

[9] P. Russom, " TDWI Best Pract. Report," Fourth Quarter, 2011.

[10] J. Manyika, M. Chui, B. Brown and J. Bughin, " Big Data: The next Frontier for Innovation, Competition,," 2011.

[11] J. Dijcks, " Big Data for the Enterprise," Oracle Report, 2012.

[12] M. A. Beyer and D. Laney, "The Importance of "Big Data": A Definition," Gartner report, 2012.

[13] Intel, "Big Data Analytics. Intel's IT Manager Survey on How Organizations Are Using Big Data," Intel, 2012.

[14] M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales and P. Tufano, " Analytics: The Real-World Use of Big Data," IBM, 2012.

[15] J. Ward and A. Barker, " preprint arXiv:1309.5821," 2013.

[16] M. Research, Microsoft, 2013. [Online]. Available: https://www.microsoft.com/en-us/news/features/2013/feb13/02-.

[17] N. B. D. P. W. Group, " Big Data Interoperability Framework: Definitions," 2014.

[18] B. Shneiderman, "Proc. 2008 ACM SIGMOD Int. Conf. Manag. Data," 2008.

[19] J. Manyika, M. Chui, B. Brown and J. Bughin, "Big Data: The next Frontier for Innovation, Competition,," 2011.

[20] D. Fisher, R. DeLine, M. Czerwinski and S. Drucker, "Interactions," 2012.

[21] H. Chen, R. Chiang and V. Storey, " MIS Q. 36, 1165," 2012.

[22] E. Dumbill, "Big Data 1," 2013.

[23] D. Boyd and K. Crawford, "Information, Commun. Soc. 15," p. 662, 2012.

[24] V. Mayer-Schönberger and K. Cukier, "Big Data: A Revolution That Will Transform How We Live," 2013.

[25] A. D. Mauro, M. Greco and M. Grimaldi, "A formal definition of Big Data based on its essential features," 2016.

[26] "Online IT Guru," undated. [Online]. Available: https://onlineitguru.com/tutorial/big-data-characteristics.

[27] "Guru99," [Online]. Available: https://www.guru99.com/what-is-big-data.html.

[28] C. Taylor, "Datamation," 2017. [Online]. Available: https://www.datamation.com/big-data/big-data-architecture.html.

[29] "Microsoft," 2018. [Online]. Available: https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/.

[30] EMC, 2012. [Online]. Available: https://www.whizpr.be/upload/medialab/21/company/Media_Present ation_2012_DigiUniverseFINAL1.pdf.

[31] EMC, "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," 2014.

[32] I. Lunden, "Techcrunch," 2014. [Online]. Available: https://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/#.brp6v1:RPIH.

[33] Cisco, "Cisco," [Online]. Available: https://www.c isco.com/c/en/us/solutions/internet-of-things/overview.html.

[34] M. Analysis, "Hadoop Market Analysis: The End of Big Data Bluff but the Market is Still Here," 2019.

[35] "Gartner," [Online]. Available: https://www.gartner.com/en/newsroom.

[36] A. Regalado, "MIT Technology Review," 2013. [Online]. Available: www.technologyreview.com/s/514346/the-data-made-me-do-it/.

[37] B. Marr, "Forbes," 2015. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#7cfc0ef517b1.

[38] "IHTT: Transforming Health Care through Big Data Strategies for leveraging," 2013. [Online]. Available: http://ihealthtran.com/wordpress/2013/03/iht%C2%B2-releases-big-data-research-reportdownload-today/.

[39] S. LaValle, E. Lesser, R. Shockley, M. Hopkins and N. Kruschwitz, "Big data,analytics and the path from insights to value.," MIT Sloan Manag Rev 2011, no. 52, pp. 20-32, 2011.

[40] C. Burghard, "Big Data and Analytics Key to Accountable Care Success," IDC Health Insights, 2012.

[41] J. Manyika, M. Chui, B. Brown, J. Buhin, R. Dobbs, C. Roxburgh and A. Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," USA: McKinsey Global Institute, 2011.

[42] IBM, "IBM: Large Gene interaction Analytics at University at Buffalo," 2012. [Online]. Available: https://public.dhe.ibm.com/common/ssi/ecm/en/imc14675usen/.

[43] IBM, "IBM big data platform for healthcare," 2012. [Online]. Available: http://public.dhe.ibm.com/common/ssi/ecm/en/ims14398usen/IMS14398USEN.PDF.

[44] B. NEDELCU, "About Big Data and its Challenges and Benefits in Manufacturing," Database Systems Journal, vol. 4, pp. 10-19, 2012.

[45] McKinsey, "The Internet of Things," McKinsey, 2010.

[46] M. Anshari, Y. Alas and L. S. Guan, " Developing online learning resources: Big data, social networks,and cloud computing to support pervasive knowledge," Education and Information Technologies , pp. 1-15, 2015.

[47] M. Minelli, C. M. and A. Dhiraj, Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses, John Wiley & Sons, Inc, 2013.

[48] S. Yen, " NoSQL is a horseless carriage.," 2009. [Online]. Available: http://dl.getdropbox.com/u/2075876/nosql-steve-yen.pdf.

[49] J. O. Chan, "An Architecture for Big Data Analytics," Communications of the IIMA, vol. 13, no. 2, 2013