

BIG DATA ANALYTICS: – CHALLENGES, OPEN RESEARCH ISSUES FOR COMPUTER SCIENCE EDUCATION

Mr. Vatsal K. Bhuvra

B.E. (Computer Engineering)

Abstract

This paper contains the Big data definition, analysis, challenges and open issues of Big data. It also contains how we can interpret Big data and how it is useful to conclude or take correct decision regarding that data. So, I tried to write importance of Big data analysis.

Keywords: *Big Data analytics.*



Scholarly Research Journal's is licensed Based on a work at www.srjis.com

Introduction:

As word suggest the data is nothing but real fact of an entity. It mean whenever we are going to collect the real fact to process or analyze whatever the collection we made is data. Sometimes data is known as raw material because the process in future we will do over data and what we will get is known as information.

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

Now a days that world is moving through the Internet technology used to get information, do E-transactions like financial and trading. So, the word data are now transferred in to Big Data (BD). BD mean that data with large volume like data regarding health, banking, commodity etc. The concept of data processing is central to computer science and hence to computer science education. Not only manipulating data through programming, but also the efficient storage, management and retrieval of data are seen as central aspects. These topics are strongly affected by Big Data, a phenomenon that arose in recent years and which introduced several new innovations. For example, established concepts of data management, such as avoiding redundancies and inconsistencies by saving data in normalized relational database management systems (RDBMS), are dropped. In this paper we will point out major

challenges' computer science education will have to deal with, when considering the new aspects arising from the outlined developments into teaching. First, Big Data is discussed in terms of its innovations to data management. Then, we will provide the educational context by analyzing the state of research on databases and Big Data in computer science education. On this basis, we will derive and discuss the relevance of changes arising from these developments and we will describe the challenges within computer science education.

It is expected that the growth of big data is estimated to reach 25 billion by 2015. From the perspective of the information and communication technology, big data is a robust impetus to the next generation of information technology etc. Considering this advantages of big data it provides a new opportunities in the knowledge processing tasks for the upcoming researchers. However opportunities always follow some challenges. To handle the challenges we need to know various computational complexities, information security, and computational method, to analyze big data. For example, many statistical methods that perform well for small data size do not scale to voluminous data. Similarly, many computational techniques that perform well for small data face significant challenges in analyzing big data. Various challenges that the health sector face was being researched by many researchers. Here the challenges of big data analytics are classified into four broad categories namely data storage and analysis; knowledge discovery and computational complexities; scalability and visualization of data; and information security. We discuss these issues briefly in the following subsections.

A. Data Storage and Analysis

B. Knowledge Discovery and Computational Complexities

C. Scalability and Visualization of Data

D. Information Security

Institutions of higher education are operating in an increasingly complex and competitive environment. They are under increasing pressure to respond to national and global economic, political and social change such as the growing need to increase the proportion of students in certain disciplines, embedding workplace graduate attributes and ensuring that the quality of learning programmes are both nationally and globally relevant. In addition, different stakeholders are expecting higher education institutions to in a timely manner to these demands, albeit with declining government funding, declining support from business and private sectors, growing regulatory demands for transparency and accountability

(Hazelkorn, 2007), declining admissions rates due to increasing tuition and upsurge in high

schools dropout and increasing operational costs (Thornton, 2013). How can institutions of higher education respond effectively and in time to global changes affecting their environment? The decisions required for dealing with these rapid changes are complex and many are made without recourse to vast data sources that have been generated but are not available to those entrusted to make relevant and timely choices. These data can play a major part in how we understand the often contested nature of higher education governance (Clarke, Nelson & Stoodley, 2013) and so ensure that institutions are not only able to respond effectively to changes happening within and outside them, but that they also remain pertinent to their purpose in the societies that they serve.

This paper examines the role of Big Data and Analytics in addressing contemporary challenges facing higher education (see Figure 1). The paper first identifies the key global trends affecting institutions of higher education and explores the potential of Big Data and Analytics in addressing these changing trends. Secondly, the paper outlines opportunities and challenges associated with the implementation and governance of Big Data in higher education. The paper concludes by outlining future directions relating to the development and implementation of institutional project on Big Data.

Emergence of Big Data

Many organizations are currently using data to make better decisions about their strategic and operational directions. Using data to make decisions is not new; business organizations have been storing and analyzing large volumes of data since the advent of data warehouse systems in the early 1990s. However, the nature of data available to most organizations is changing, and the changes bring with them complexity in managing the volumes and analysis of these data. Basu (2013) observed that most businesses today run on structured data (numbers and categories). However, this does not reflect the complexity on the nature of available corporate data and their untapped hidden business value. According to IBM, 80% of the data organizations currently generated are unstructured, and they come in a variety of formats such as text, video, audio, diagrams, images and combinations of any two or more formats.

Most of these unstructured data make their way to corporate data warehouse. The term 'Data warehouse' refers to a central repository of data or a centralized database. It represents an ideal vision of maintaining a central data repository and a living memory of data that can be leveraged for better decision making. Recent developments in database technologies made it possible to collect and maintain large and complex amounts of data in many forms and from multiple sources. In addition, there are analytical tools available that can turn this complex

Copyright © 2017, Scholarly Research Journal for Interdisciplinary Studies

data into meaningful patterns and value, a phenomenon referred to as Big Data. Big Data describes data that is fundamentally too big and moves too fast, thus exceeding the processing capacity of conventional database systems (Manyika et al, 2011). It also covers innovative techniques and technologies to capture, store, distribute, manage and analyze larger sized data sets with diverse structures.

With new concepts, critiques emerge. Some critics contested that the notion of Big in the term itself is misleading and that it does not reflect only data size but complexity. Yang (2013) pointed out the definition of Big Data has little to do with the data itself, as the analysis of large quantities of data is not new, but rather Big Data refers to emergent suite of technologies that can process mass volumes of data of various types at faster speeds than ever before. This conceptualisation of Big Data was echoed by Forrester defining Big Data as “technologies and techniques that make capturing value from data at an extreme scale economical.” The term economical suggests that if the costs of extracting, processing and making use of data exceed the advantages to be collected, then it is not worth indulging in the process. Generally, Big Data has come to be identified by a number of fundamental characteristics. Key among them are:

- Volume—large amount of information is often challenging to store, process, and transfer, analyze and present.
- Velocity—relating to increasing rate at which information flows within an organization—(reorganizations dealing with financial information have ability to deal with this).
- Veracity refers to the biases, noise and abnormality in data. It also looks at how data that is being stored, and meaningfully mined to the problem being analyzed. Veracity also covers questions of trust and uncertainty.
- Variety—referring to data in diverse format both structured and unstructured.
- Verification—refers to data verification and security.
- Value—most importantly, has the data been utilized to generate value of the insights, benefit and business processes, etc. within an organization?

Douglas (2001) in the Gartner’s report proposed the three of the most common properties of Big Data. The report made three fundamental observations: the increasing size of data, the growing rate at which it is produced and the cumulative range of formats and representations employed, proposed threefold definition encompassing the “three Vs” (Volume, Velocity and Variety).

There are also other properties of Big Data such as data validity, which refers to accuracy of data, and volatility, a concept associated with the longevity of data and their relevance to analysis outcomes, as well as the length required to store data in a useful form for appropriate value-added analysis. In addition to these properties, there are three stages required to unlock the value of Big Data in any organisation. These include data collection, data analysis, visualisation and application (see Figure 2).

Collection

Data collection is the first step in unlocking the value accrued from Big Data. This requires identifying data that can reveal useful and valuable information. Data must be filtered for relevance and stored in a form that is useful, as little is gained in investing in huge amounts of data and storage infrastructure if the vast majority of the data in it is not usable.

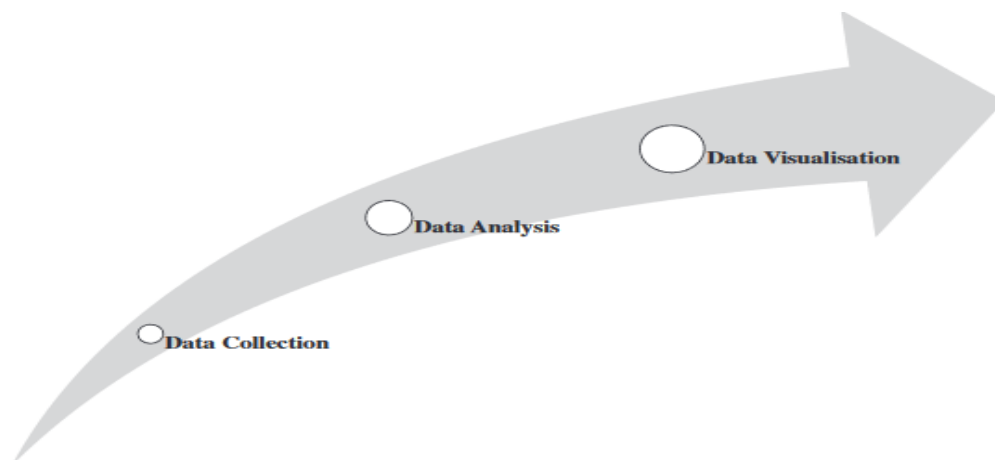
Analysis

Once data have been rendered into a usable form, it has to be analyzed to generate actionable information. However, with the growing diversity in the nature of data, managing and analyzing diverse data set is becoming a very complex process. Analysis needs to include linking, connecting correlating different data sets to be able to grasp the information that is supposed to be conveyed by these data. This situation is, therefore, termed as the 'complexity' of Big Data.

Visualization and application

This is the last stage where the analyzed data is made available to users in a form that is interpretable and integrated into existing processes, and ultimately used to guide decision making.

The Value of Big Data in Higher Education



Big Data and analytics in higher education Big Data is a knowledge system that is already changing the objects of knowledge and social the oryzin many fields while also having the potential to transform management decision-making theory (Boyd & Crawford, 2012). Big Data incorporates the emergent research field of learning analytics (Long & Siemen, 2011), which is already a growing area in education. However, research in learning analytics has largely been limited to examining indicators of individual student and class performance. Big Data brings new opportunities and challenges for institutions of higher education. Long and Siemen(2011) indicated that Big Data presents the most dramatic framework in efficiently utilizing the vast array of data and ultimately shaping the future of higher education. The application of Big Data in higher education was also echoed by Wagner and Ice(2012), who noted that technological development shave certainly served a catalysts for the move towards the growth of analytics in higher education. In the context of higher education, Big Data connotes the interpretation of a wide range of administrative and operational data gathered processes aimed at assessing institutional performance and progressin order to predict future performance and identify potential issues related to academic programming, research, teaching and learning (Hrabowski, Suess & Fritz, 2011a, 2011b; Picciano, 2012). Others indicated that to meet the demands of improved productivity, higher education has to bring the tool of analytics into the system. As an emerging field within education, a number of scholars have contended that Big Data framework is well positioned to address some of the key challenges currently facing higher education (see, eg, Siemens, 2011). At this early stage much of the work on analytics within higher education is coming from interdisciplinary research, spanning the fields of Educational Technology, Statistics, Mathematics, Computer Science and Information Science. Acores element of the current work on analytics in education is centered on data mining. Big Data in higher education also covers database systems that store large quantities of longitudinal data on students' right down to very specific transactions and activities on learning and teaching. When students interact with learning technologies, they leave behind data trails that can reveal their sentiments, social connections, intentions and goals. Researchers can use such data to examine patterns of student performance over time—from one semester to another or from 1 year to another. On a higher level, it could be argued that the added value of Big Data is the ability to identify useful data and turn it into usable information by identifying patterns and deviations from patterns. Long and Siemen (2011) indicated that Big Data is now well positioned to start addressing some of the key challenges currently facing higher education. An OECD (2013)

report suggested that it may be the foundation on which higher education can reinvent both its business model and bring together the evidence to help make decisions about educational outcomes. From an organization all earning perspective, it is well understood that institutional effectiveness and adaptation to change relies on the analysis of appropriate data (Rowley, 1998) and that today's technologies enable institutions to gain insights from data with previously unachievable levels of sophistication, speed and accuracy (Jacqueline, 2012). As technologies continue to penetrate all facets of higher education, valuable information is being generated by students, computer applications and systems (Hrabowski & Suess, 2010). Furthermore, Big Data Analytics could be applied to examine student entry on a course assessment, discussion board entries, blog entries or wiki activity, which could generate thousands of transactions per student per course. These data would be collected in real or near real time as it is transacted and then analyzed to suggest courses of action. As Siemens (2011) indicated that "[learning] analytics are a foundational tool for informed change in education and provide evidence on which to form understanding and make informed (rather than instinctive) decisions.

Big Data can also address the challenges associated with finding information the right time when data are dispersed across several unlinked different data systems in institutions. By identifying ways of aggregating data across systems, Big Data can help improve decision-making capability.

Challenges of implementation There are a number of anticipated challenges associated with the implementation of analytic techniques for Big Data in higher education. Some of these include challenges associated with getting users to accept Big Data as a conduit for adopting new processes and change management. There is also a tremendous cost associated with collecting, storing and developing algorithms to mine data, a process that tend to be time consuming and complex. Furthermore, most of institutional data systems are not interoperable, so aggregating administrative data, classroom and online data can pose additional challenges (Daniel & Butson, 2013). Furthermore, data integration challenges are eminent, especially where data come in both structured and unstructured formats and needed to be integrated from disparate sources most of which are stored in systems managed by different departments. Additionally, data clean sing when performing integration of structured and unstructured data is likely to result to loss of data. There are also challenges associated with quality of data collected and reported. Lack of standardised measures and indicators make inter(national) comparison difficult, as the quality of information generated

Copyright © 2017, Scholarly Research Journal for Interdisciplinary Studies

from Big Data is totally dependent on the quality of data collected and the robustness of the measures or indicators used. A recent U.S. Department of Education (2012) report suggested that the successful implementation of Big Data in higher institution would depend on collaborative initiatives between various departments in a given institution. For instance, the involvement of IT services departments in planning for data collection and use is deemed critical. This is consistent with views that the value of Big Data will be based on the ability to co-create governing structures and delivery of more progressive and better policies and strategies currently used (Schleicher, 2013). Wagner and Ice (2012) also pointed out that by increasing collaborative ventures on Big Data initiatives help all groups take ownership of the challenge involving student performance and persistence. Dringus (2012) suggested that the practice of learning analytics should be transparent and flexible to make it accessible to educators (Dringus, 2012; Dyckhoff et al, 2012). However, there is still a divide between those who know how to extract data and what data are available, and those who know what data are required and how it would best be used, all which make collaboration difficult. Furthermore, as Romero and Ventura (2010) note, analytics has

Traditionally been difficult for non-specialist to generate (and generate in meaningful context), to visualize in compelling ways, or to understand, limiting their observability and decreasing their impact (Macfadyen & Dawson, 2012). The importance of communicating these ideas is also acknowledged by Macfadyen and Dawson (2012), who pointed out that analytics have a negative or neutral impact on educational planning. They advocate delving into the socio-technical sphere to ensure analytics are presented to those involved in strategic positions in ways that have the power to motivate organisational adoption and cultural change. Becker (2013) suggested three interactive components to be studied when collecting data for analytics: location, population and timing. Location is defined by where and how students are accessing the learning space, whereas population refers to the characteristics of the group of learners participating in the learning space. Timing can be defined by any unit, from second or minute to semester or year. Big Data utilization also raises issues around ethics of data collection in regard to quality of data, privacy, security and ownership. It also raises the question of an institutions responsibility for taking action based on the information available (Jones, 2012). Security and privacy issues pose additional challenge to implementation of Big Data in higher education. Currently, techniques such as disaster recovery plans, strong password policy, firewalls, encryption and anti-virus software that reduce the risks of losing or manipulating Big Data are still being investigated. Furthermore,

Copyright © 2017, Scholarly Research Journal for Interdisciplinary Studies

risks and security procedures for data protection and privacy are still lacking in many institutions of higher education. For instance, SladeandPrinsloo(2013)pointed out that while most higher education institutions seem to have policies to regulate and govern intellectual property, safe guard data privacy and regulate access to data, these policies might not be adequate to address contemporary challenges associated with Big Data in higher education.

References:

British Journal of Educational Technology

T. K. Das and P. M. Kumar, *Big data analytics: A framework for unstructured data analysis*,
International Journal of Engineering and Technology.

(IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 2, 2016