Integrating big data in higher education

by Lavanya Madamshetty and Krishna Suresh Babu Naidu, Durban University of Technology

The dynamic working atmosphere of today's competitive world puts all organisations in need of innovative solutions for better understanding of their performance. This is expected not only for companies but also for schools and universities. Big data is one such innovation. This article gives overviews on the concept and challenges related to the analytical processing of big data and how these features impact higher education institutions.

Globalisation processes and development of information and communication technologies (ICT) have led to the accumulation of large volumes of data. Nevertheless, the availability of data alone does not add value for organisations unless that data is processed and transformed into valuable information to support decision making [1]. The term "big data" tends to refer to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of dataset [2].

From business leaders to academics, big data is the buzzword that invites attention and, to some extent, fright. Surprisingly, the unexpected growth of big data has left many unrehearsed [3]. The wild leap of big data related technologies and reception of concepts by public and private sectors has left little time for the discourse to develop and mature in the academic domain. Doug Laney, an analyst with the META Group (now part of Gartner), described big data with a collection of "v" words, referring to 1) the increasing size of data (volume); 2) the increasing rate at which it is produced and analysed (velocity); and 3) its increasing range of sources, formats, and representations (variety) [5].

With the advent of new technologies, higher educational institutions are quickly engaging blended learning (BL), the combination of face-to-face and technology mediated instructions [6]. Moreover, research in education articulates that in present learning environments, users (students) frequently use online communities like discussion forums, online chats, instant messaging clients and various learning management systems. Recent learning methods greatly depend on online activities, which result in the accumulation of an enormous amount of unused data that is wasted as traditional learning analytics are not capable of processing it. This has lead to the penetration of big data technologies and tools into education. This article analyses the challenges related introduction of "big data" at higher education institutions and introduces readers to recent developments.

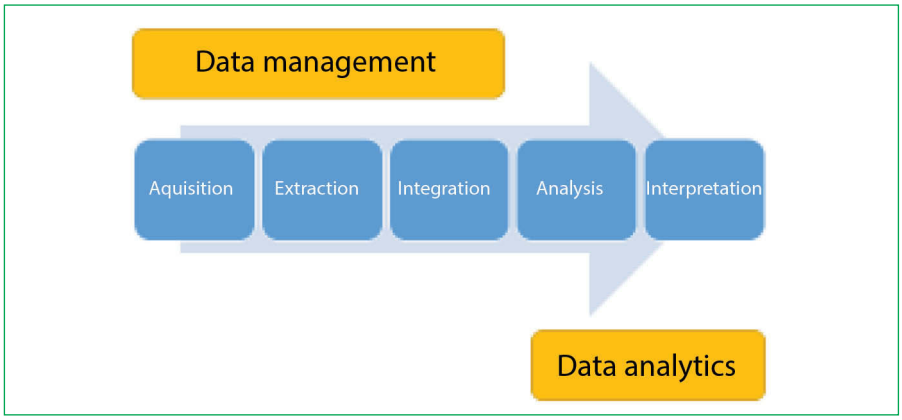

Fig. 1: Different phases in big data management.

What is big data?

Big data is a term that defines the accumulation of large volumes of data (both structured and unstructured) that submerges a business on a day-to-day basis [9]. Higher educational institutes are flooded with data about learners, educators, achievements and performance. To date, this data has rarely been extracted intelligently with the goal of refining learning and informing teaching practice, although evidence from other sectors such as marketing, sports, retail, health and technology suggests that the effective use of big data can offer the education sector the potential to enhance its systems and outcomes [10]. Nevertheless, big data comprises gathering datasets that are huge and complex to process by any data-processing software applications. Traditional software applications are inadequate as they lack the practical, technical and financial capacity to effectively gather, manage and mine big data [11].

Big data and analytics have become a major part of the higher education system. Big data in the education context is researched under the name Educational Data Mining (EDM), and is also termed as Knowledge Discovery in Database (KDD). Big data in the education sector offers exceptional opportunities through automated analytics that provide feedback not only on learner study behaviours, but also progress and outcomes. This will not only improve academic performance but also develop student self–regulated learning (SRL) skills, and SRL proficiency in unique ways [12]. Big data analytics enable educators to understand students' educational experience which will help identify, evaluate and develop custom study plans based on the student's learning behaviour, or embrace new programs based on the additional requirements of the students.

The main motive of any educational system and big data with respect to education is to improve student results. Better the results,

better the students; and it is better for society, organisations and educational institutions. Student actions can be monitored, answering questions such us:

- How much time do they seek to answer a question?
- What resources do they depend on?
- What kinds of questions do they like to answer?
- What questions do they skip?
- Where are their strengths?
- What information do they avoid?

EDM software applications trace whether a student was able to achieve the learning goals and if not, it helps them achieve them. In order to do this, the system has to be intelligent enough to implement various processes, which requires effective algorithms and models.

EDM looks around the new data patterns and designs new algorithms and statistical models that can be applied in digital learning. It is used to analyse data to check on learning theories and to refine the design of the learning systems themselves. It generally specifies learning lesser components that can be explicated and then utilised by software to adjust to the student.

Two prominent EDM researchers, Ryan Baker at Teacher's College Columbia University in the US and Kalina Yacef of the University of Sydney [13], characterise the goals of EDM research as:

- Forecasting students' future learning behaviour by generating student models that integrate comprehensive data, for instance students' knowledge, motivation, metacognition, and attitudes.
- Learning domain models that distinguish the content to be learned and optimum instructional classifications.
- Reviewing the effects of diverse types of pedagogical backing that can be provided by learning software.
- Evolving scientific knowledge about information and pupils through superstructure of computational models that include models of the student, the domain, and the software's pedagogy.

Big data: Acquisition, extraction, integration and analysis

These days, large volumes of data are collected and stored by organisations with the hope of it being useful in the future. Capture, report, predict, act and refine are the different phases identified to be involved in big data analysis [14]. Bearing in mind the nature of big data, the majority of phases involved in analysing big data can be acquisition, extraction, integration, analysis and interpretation as shown in Fig. 1.

Big data acquisition includes data collection, data transmission, and data pre-processing [15]. Data, in real time, may be attained from social networking sites (like Facebook, Twitter, Blogs) course management systems (CMS), learning management system (LMS), and physical world data like library usage. The collected datasets often result to petabytes and may comprise much redundant or unusable data which needlessly surges storage space and impacts on the succeeding data analysis. Therefore, identifying data of interest is very important, and describing filters which abandon unwanted data and preserve only useful information is very advantageous. Furthermore, applying data compression techniques can be useful to reduce the redundancy. Therefore, data pre-processing operations are crucial to ensure efficient data storage and exploitation. Upon the completion of raw data collection, data will be conveyed to a data storage infrastructure (data centre) for processing and analysis. Data transmission from the data centre is via two phases: inter-dynamic circuit network transmissions and intra-dynamic circuit network transmissions [15].

Data integration is a critical step in big data management and essentially involves combining data from several diverse sources into a cohesive view. There is no universal approach to big data integration (BDI) [16]. Today, the progress of the internet enables access to data from different information systems and domains. Examples include web texts, web logs, large scale e-commerce, social networks, sensor networks, genomics, biomedical records, surveillance, etc. [17]. Data integration is a three-step process as shown in Fig. 2 [18].

Initially, corresponding attributes that are used to describe the information items in the source need to be identified. The result of this step is a schema mapping, which is used to transform the data present in the sources into a common representation (renaming, restructuring). Next, the different objects that are described in the data sources need to be identified and aligned. Using duplicate detection techniques finds multiple and possibly inconsistent representations of the same real-world objects. Lastly, the duplicate

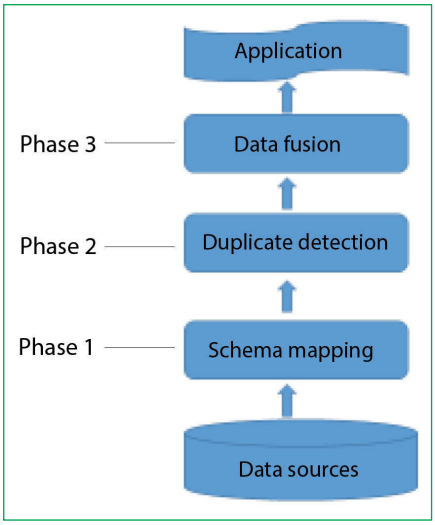

Fig. 2: Data integration process.

representations are combined and fused into a single representation while inconsistencies in the data are resolved. Capturing and fusing data from different sources into a single structure is a highly domain-dependent task. Consequently, one approach may work extremely well in one domain and fail in another.

Application of big data in higher education

Big data delivers an opportunity to higher education institutes to strategically use their advanced information technology resources, to increase educational quality and guide students to higher rates of completion [20].

Analysing and managing such huge data is vital as it can bring accountability and transparency in the management of the education sector, assist both students and organisations to recognise their achievements and areas of weaknesses and associate with other such organisations. Also, faculty and students can track behavioural progress through big data analytics [22].

Big data techniques can be used in a variety of ways in learning analytics, with some listed below [7]:

- *Performance valuation:* Student's performance can be predicted by analysing their interaction in a learning environment with other students and teachers
- *Risk management:* By analysing the student's behaviour, risk of students dropping out from courses can be perceived and suitable measures can be implemented in the beginning of the course to retain students.

- *Data conception:* Reports on educational data become increasingly complex as educational data volumes grow. Data can be envisaged using data visualisation techniques to easily classify the trends and relations in the data just by looking on the reports.
- *Smart feedback:* Learning systems can deliver intelligent and instant feedback to students in response to their inputs which will progress student communication and performance.
- *Course approval:* Innovative courses can be suggested to students based on the interests of the students by analysing their activities. This will ensure that students are not ill-advised in choosing specialisations in which they have no interest.
- *Behaviour recognition:* Recognition of student behaviours in community-based activities or games which will help in developing a student model
- *Constructing courseware:* For web-based personalised distance learning programs, data analytics will help to adapt the teaching in accordance to individual student's ability in the distance learning environment.

Challenges in implementing big data

In education, the value of analytics and big data can be found in: 1) their role in guiding reform activities in higher education; and 2) their role in assisting educators to improve teaching and learning. Some of the difficulties influencing the execution of big data in higher education are discussed below.

Acceptance of big data

It involves motivating clients to acknowledge big data as a channel for adopting new procedures and change management [12]. This is because developing a better understanding about the system will be time-consuming.

Difficulties in accessing data

This comprises difficulties in getting the required data from the database, as low quality and inaccurately organised data from less accessible database systems can likewise cause noteworthy issues towards client.

Requirement of expert knowledge

Using traditional data analysis techniques will be inadequate to manage with the dawn of big data which principally include huge data, both structured and unstructured [24]. In order to ensure the system is working properly, expert knowledge is needed in higher education. However, there is lack of skills and knowledge on big data [20].

Organisational barriers

Organisational barriers in this context are present in terms of cooperation between departments in implementing the big data and cost associated with collecting, storing, analysing and visualising the data [25].

Conclusion and further study

Higher education institutes currently work in dynamic and competitive environments. Traditional universities accumulate large volumes of academic data. These datasets have a wealth of information which increases day by day. This warrants the use of advanced tools from university's top-level management to process this information. The resultant data is expected to support the decision-making process by top-level management. The big data technology approach to data warehousing will help reduce difficulties associated with traditional data analysis. Moreover, this has the potential to elevate the education system with novel learning methods, and enable more effective and efficient decision-making at institutional level.

References

[1] A Gandomi and M Haider: "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management, Vol. 35 No. 2, pp. 137-144, 2015.

[2] Wikipedia: "Big data". [cited 2018 25-09-2018]; Available from: https://en.wikipedia.org/wiki/Big_data.

[3] H Chen, RH Chiang, and V.C. Storey: "Business intelligence and analytics: From big data to big impact." MIS Quarterly, pp. 1165-1188, 2012.

[4] D Poddar: "Big data analysis: Industries using it!" [cited 2018 28/10]; Available from: https://upxacademy.com/big-data-use-cases-industries/, 2016.

[5] D Laney: "3D data management: Controlling data volume, velocity and variety", Application Delivery Strategies (META group), 2001.

[6] WW Porter et al.: "Blended learning in higher education: Institutional adoption and implementation." Computers & Education, Vol. 75, pp. 185-195, 2014.

[7] K Sin and L Muthu: "Application of big data in education data mining and learning analytics – A literature review." ICTACT Journal on soft computing, Vol. 5 No. 4, 2015.

[8] J O'Flaherty and C Phillips: "The use of flipped classrooms in higher education: A scoping review." The Internet and Higher Education, Vol. 25: pp. 85-95, 2015.

[9] P Zikopoulos and C Eaton: "Understanding big data: Analytics for enterprise class hadoop and streaming data." McGraw-Hill Osborne Media, 2011.

[10] LP Macfadyen et al.: "Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge." Research & Practice in Assessment, Vol. 9, pp. 17-28, 2014.

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

[12] B Daniel: "Big data and analytics in higher education: Opportunities and challenges." British Journal of Educational Technology, Vol. 46 No. 5, pp. 904-920, 2015.

[13] V Aleven, J Kay, and J Mostow: "Intelligent tutoring systems" 10th International Conference, ITS. Pittsburgh, PA, USA, June 14-18, 2010, Proceedings. Vol. 6095. Springer Science & Business Media, 2010.

[14] CJ Dede, P Mitros, and AD Ho: "Big data analysis in higher education: promises and pitfalls." 2016.

[15] M Chen et al.: "Big data generation and acquisition." pp.19-32, 2014.

[16] XL Dong and D Srivastava: "Big data integration." in Data Engineering (ICDE), 2013 IEEE 29th International Conference on. 2013.

[17] J Bleiholder and F Naumann: "Data fusion." ACM Computing Surveys (CSUR). Vol. 41 No.1, pp. 1.

[18] F Naumann et al. "Data fusion in three steps: Resolving schema, tuple, and value inconsistencies." IEEE Data Eng. Bull, Vol. 29 No. 2 pp. 21-31, 2006.

[19] H Jagadish et al.: "Big data and its technical challenges." Communications of the ACM, Vol. 57 No. 7, pp. 86-94, 2014.

[20] ZM Hanapiyah, WNW Hanafi, and S. Daud: "Issues, challenges and opportunities of big data management in higher education institutions in Malaysia." Indian Journal of Science and Technology, Vol. 11 No. 4, 2018.

[21] V Kellen, A Recktenwald, and S. Burr: "Applying big data in higher education: A case study." Cutter Consortium white paper, Vol. 13 No. 8, 2013.

[22] SF Wamba et al.: "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study." International Journal of Production Economics, Vol. 165: pp. 234-246, 2015.

[23] G Siemens and P Long: "Penetrating the fog: Analytics in learning and education." Educause Review, Vol. 46 No. 5, p. 30, 2011.

[24] MA Waller and SE Fawcett: "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management." Journal of Business Logistics, Vol. 34 No. 2, p. 77-84, 2013.

[25] E Hoover and MK Harder: "What lies beneath the surface? The hidden complexities of organizational change for sustainability in higher education." Journal of Cleaner Production, Vol. 106, pp

Contact Dr Suresh Naidu, Durban University of Technologuy, sureshk@dut.ac.za