

BIG DATA ANALYTICS AND E LEARNING IN HIGHER EDUCATION

Tulasi.B¹ and Suchithra.R²

¹Department of Computer Science, Christ University, Bangalore, India

²Department of Computer Science, Jain University, Bangalore, India

ABSTRACT

With the advent of internet and communication technology the penetration of e-learning has increased. The digital data being created by the higher educational institutions is also on ascent. The need for using "Big Data" platforms to handle, analyse these large amount of data is prime. Many educational institutions are using analytics to improve their process. Big Data analytics when applied onto teaching learning process might help in improvising as well as developing new paradigms. Usage of Big Data supported databases and parallel programming models like MapReduce may facilitate the analysis of the exploding educational data. This paper focuses on the possible application of Big Data Techniques on educational data.

KEYWORDS

Big Data, e-learning, Higher Education, MapReduce, MongoDB, Association rule mining

1. INTRODUCTION

Operational processes of higher educational institutions are increasing becoming complex. The quality of learning, accountability to the stake holders, the requirements of the digital-era Millennials are few of the issues which are currently associated with higher education. The digital revolution has given rise to newer approaches to learning in form of Massive Open Online Courses and Learning Management Systems where the interaction of the learner with the teacher moves beyond the physical walls of the class room leading to design of flexible class room sessions [1]. With the augmentation of internet and communication technology, demand for online learning has seen greater growth. The amount of data being stored by higher educational institutions has always had been huge. But the digital penetration into the processes of teaching-learning and evaluation has paved way to availability of huge amount of data which can be termed as "Big Data". Big Data is loosely defined as data sets which are of very large scale and complex in nature [2]. With the penetration of digital and mobile communication into the field of education data being stored has exponentially grown.

The digital learning or e-learning is being seen as a futuristic approach of learning by the current generation of learners [3]. The ease of availability and usage of handheld devices like Tablets has provided the required leap for e-learning. Open source LMS like MOODLE provide avenues for the students to interact not only with the teacher but also with their peers which would result in richer learning experience. The huge amount of data being generated by these educational modules is on ascent. This data when analysed can provide greater insights on learning patterns of students, gaps in the teaching learning processes and many more. The storage requirements of this humongous data are quite different from the traditional storage. The analytics to be applied on this data also would require newer technologies and platforms. Parallel computation of these large

data sets would be required to derive actionable knowledge. The horizontal scalability feature of “NoSQL” data stores makes it a potential alternative over the traditional storage. The trait of NoSQL “shared nothing, replicating and partitioning of data over many servers” allows it to support large number of read/write operations per second [4].

This paper explores the possible role of big data analytics, e-learning in higher education. Big Data analytics can be applied at various levels in higher education like teaching learning, resource allocation, student retention, course advisor. The paper tries to identify the penetration of e-learning into higher education and the impact big data analytics could provide to the various processes of higher education.

2. BACKGROUND

Digital data being created as a result of direct, indirect usage of technology and communication when analysed would provide profound data on the teaching- learning trends. This would facilitate in providing recommendations to both the students and the teachers regarding the learning curve. Tools like NodeXL, SNAPP, Gephi are being utilized by researchers and educational institutions to obtain information regarding learning advancement of a student [5]. “Signals” at Purdue University has helped students, teachers, administrators through the intervention model [6]. With digitalization of most of the activities of a higher educational institution the digital trail left by the student plays an important role in analysing his learning [7].

3. “BIG DATA” IN HIGHER EDUCATION

The environment of higher education is transforming to cater the changing learning needs and diversity of students. Global changes like advancement, availability and ease in use of technology has steered the change in higher education. Though there have been developments in the educational field, the role of data has been overlooked [1]. E-learning technologies have provided a new platform to enhance the teaching learning. The data trails left by them provide higher education institutions actionable data to adapt to the required changes. The data being stored by the higher education institutions along with the digital data being created by the use of technology has resulted in huge chunks of data which can be termed as “Big Data”. Big Data is associated with five V’s Velocity, Volume, Variety, Value and Veracity. The same is depicted in the figure 1. The exploration of Big Data in association with higher education would be beneficial in understanding the social, cognitive and emotional aspects of students and teachers.

Educational data is not just humungous data but also heterogeneous data [9]. Traditional data mining approaches are applied to obtain patterns and move towards predication. But mining of educational data tends to focus on development of new tools to determine patterns and apply techniques to analyse large data sets [13]. Due to the sheer volume and variety associated with Big Data the traditional approaches might fall short for deliverance. The big storage requirement is also one of the major concerns to be handled. Big Data is usually associated with no-schema, high scalability. NoSQL data stores like Cassandra, MongoDB, CouchDB try to meet the requirements of “Big Data”

3.1. Big Data Supportive databases

NoSQL databases are inherently schema less and highly scalable [9]. These databases support parallel processing of the large amount of data. MapReduce is a data processing paradigm for squeezing the large volumes of data into useful and comprehensive results [10]. MapReduce provides platform to access data in distributed file systems with intermediate data being stored on

local disks. MongoDB, Cassandra, Accumulo, MonetDB, Apache Hadoop, Hive are few platforms which have emerged to store large chunks of data [11]. This section deals with “MongoDB” and how it is suitable for storing educational data.

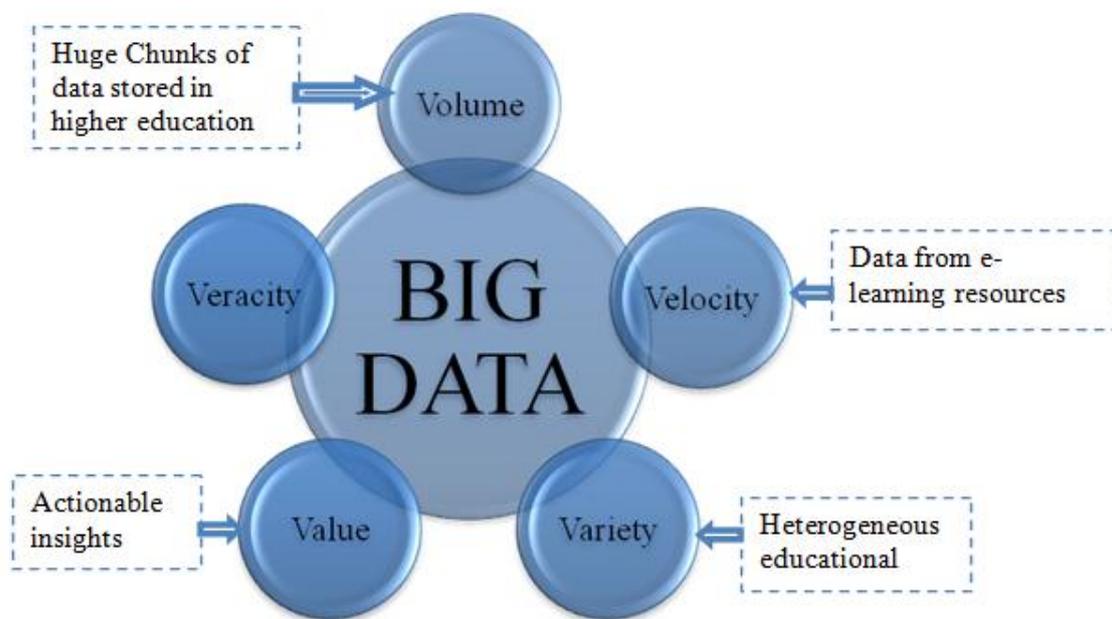


Figure 1: Big Data – Five V’s

MongoDB stores data in form of documents which are JSON like field and value pairs [12]. MongoDB documents are BSON, which is binary representation of JSON. All documents are stored in collections. Due to non enforcement of document structure to collections, there is a greater flexibility in mapping a document to an object. To tackle large datasets and the issues of scalability MongoDB uses the concept of “sharding”. It achieves sharding through configuration of sharded clusters [10]. MongoDB provides rich set of aggregation and map-reduce operations to perform aggregation. The two stages of map-reduce: map phase which processes each document and emits one or more objects. Reduce phase that combines the output of map phase. The flexibility and scalability features of MongoDB makes it suitable to store educational data. Few salient features are mentioned in the Table 1.

Table 1: Features of MongoDB

Sno	Feature
1	Schema less document based database
2	MapReduce and Aggregation tools
3	Use of secondary and geospatial indexes
4	Designed for High performance
5	Easy to scale
6	Supports Sharding, Replication and high availability

CRUD operations in MongoDB are simpler and flexible when compared to the traditional SQL statements. The following example depicts the flexibility in storing documents with dynamic schemas which stores student information into the collection "Studentdetails".

```
Student1= {
  "Student_id": "1"
  "Deptid": "D1"
  "FirstName": "Ravi"
  "LastName": "Kumar"
  "Age" :20
  "Email": ravi.kumar@gmail.com
  "CoreCourses": ["Java Programming",
  "Computer Networks"],
  "InternalAssesmmnt" :[78,56] }

Student2={
  "Student_id": "2"
  "Deptid": "D1"
  "FirstName": "Geeta"
  "LastName": "Rao"
  "CoreCourses": [" Java Programming",
  "Computer Networks"],
  "Certifications": ["Oracle", "RedHat" ] }
```

3.2. Analytics in higher education

In recent years, there has been increasing focus on the use of data in the processes of educational institutions. Data-driven decisions would help the teaching learning process to evolve and also indulge in creation of new pedagogy. Learning Analytics would pave way to identify the learning patterns and behaviours of students [16]. Few popular methods which being used on educational data are clustering, classification, discovery with models, Association rule mining.

Association Rule Mining (ARM) is one of the most useful methods and has been successfully applied on educational data [18]. Association rule mining is a procedure which is used to find frequent patterns, associations from data sets. Though the main applications of it have been basket data analysis, cross marketing, it would be applicable to the field of education. Association rule consists of two parts: the antecedent which an element available in the data and the consequent which is the element obtained along with the antecedent. These kinds of if/then rules can be applied on educational data to frame a hypothesis which can be investigated further. A sample rule can be "Student spends daily a minimum of a hour in library" \Rightarrow "Good Learning behaviour". The if/then patterns identified by association rules by analysing the data help in identification of important relationships. In the large data sets of educational data this would provide information regarding the most significant relationship.

4. CONCLUSIONS

Educational institutions are generating huge volumes of data through the admission process, evaluation and teaching learning. The field of education is gaining insight and is obtaining actionable data from large chunks of varied data known as Big Data. With the advent of e-learning provide in many universities, the amount of data available to all the stake holders of the educational system is enormous. The Big Data paradigms are needed in current world to add value to the processes of educational institutions.

REFERENCES

- [1] Ben Daniel ,”Big Data and analytics in higher education: opportunities and challenges” , British journal of educational technology 2014.
- [2] Ahmed Ashraf,Hazem M, Yehia El-Mashad, Nikos MAstorakis ,”Enhancing Big Data processing in Educational Systems” , Advances in Computer and Technology for Education,2014
- [3] Anthony G.Picciano, “Big Data and Learning Analytics in Blended Learning Environment; Benefits and Concerns”, International Journal of Artificial Intelligence and Interactive Multimedia, Vol2, No7,2014
- [4] Cattell R (2011) ,”Scalable sql and nosql data stores”. ACM SIGMOD Record 39(4):12–27
- [5] <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf> <accessed on 2nd October 2015>.
- [6] Arnold, K. E. (2010). “Signals: Applying Academic Analytics”, EDUCAUSE Quarterly 33(1).
- [7] Phil Long and Gorge Siemens ,”Penetrating the fog: Analytics in learning and education”, EDUCAUSE Review vol 46,no 5 September/October 2011
- [8] A survey of use of weblogs in education, Current developments in technology-assisted education ,1,260-264
- [9] Jyotsna Talreja Wassan ,”Discovering big Data Modelling for Educational World” , Procedia- Social and Behavioral Sciences 176(2015) 642-649
- [10] <https://docs.mongodb.org/manual/core/map-reduce/> <accessed on 5th October 2015>.
- [11] <http://nosql-database.org/> <accessed on 5th October 2015>
- [12] <https://docs.mongodb.org/manual/core/crud-introduction/> <accessed on 7th October 2015>
- [13] Rmoero.C.R and Ventura.S ,”Educational Data Mining: A Review of the State of the Art” , IEEE Transactions on Systems, Man and Cybernetcs, Part C:Applications and Reviews, 40,6,610-618
- [14] Pradeep soni, narendra singh yadav, “Quantitative analysis of document stored databases” International journal of computer applications Volume 118 – No.20, May 2015
- [15] Ciprian-Octavian Truică, Alexandru Boicea ,”Operations in MongoDB “, International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013),
- [16] Mr.Suhas G. Kulkarni , Mr.Ganesh C. Rampure , Mr.Bhagwat Yadav ,”Understanding Educational Data Mining (EDM)”, International Journal of Electronics and Computer Science Engineering, volume 2, Number 2, 2013.
- [17] Rodrigo, M.M.T., Baker, R.S.J.d. (2011) “Comparing Learners' Affect While Using an Intelligent Tutor and an Educational Game”, Research and Practice in Technology Enhanced Learning, 6 (1), 43-66
- [18] ROMERO, C., ROMERO, J.R., LUNA, J.M., VENTURA, S.,” MINING RARE ASSOCIATION RULES FROM E-LEARNING DATA”, PROC. OF EDUCATIONAL DATA MINING CONFERENCE, PP. 171–480 (2010)