Handling Big Data in E-Learning

Ahmed Ashraf, Hazem El-Bakry, Samir M. Abd El-razek, Yehia El-Mashad

Information Systems Dept., Faculty of Computer Science and Information Systems, Mansoura University, Mansoura, Egypt

Delta Higher Institute for Engineering & Technology, Talkha, Dakahlia, EGYPT

Abstract
The main objective of this paper is to present a new E-Learning model. There exists large amount of different types of data in E-Learning environment both structured and unstructured. So, there is a need for enhancing the performance of big data processing in educational systems. Handling big data efficiently is required to serve educational systems and learners. Our suggested model can deal with big data efficiently. Important data will be collected in a smart way through interaction between learners and learning content. The proposed model can determine all learners' behavior in E-Learning system by handling educational big data using hadoop and NoSQL.

Keywords
Big Data, Big Data Analytics, Database, Hadoop, NoSQL.

I. Introduction
Big Data, a highly innovative provides a unique forum for world-class research exploring the challenges and opportunities in collecting, analyzing, and disseminating vast amounts of data, including data science, big data infrastructure and analytics, and pervasive computing[1]. Big Data is a loosely defined term used to describe data sets so large and complex that they become awkward to work with using standard statistical software. The rise of digital and mobile communication has made the world become more connected, networked, and traceable and has typically led to the availability of such large scale data sets. Some of the keepers of Big Data sets develop interfaces for everyone to access and analyze some of the data, e.g. Google provides freely available Google Insights, while others hesitate to offer any access [2]. Big Data coverage includes: Big data industry standards, new technologies being developed specifically for big data, Data acquisition, cleaning, distribution, best practices, Data protection, privacy, and policy, business interests from research to product, The changing role of business intelligence, visualization and design principles of big data infrastructures, physical interfaces and robotics. E-learning is an education via the Internet, network, or standalone computer. E-learning is basically the network-enabled convey of skills and knowledge [3-5]. Distance education is the most renowned descriptor used when referencing distance learning. It often describes the effort of providing access to learning for those who are geographically distant [27-53]. Big data concepts and analytics can be applied to a variety of higher education administrative and instructional applications, including recruitment and admissions processing, financial planning, donor tracking, and student performance monitoring [5].

A. The Big Deal with "Big Data?"
Now, Data is available at a larger scale. Obviously, this is a lot of data. But what exactly is new about it? The short answer is that data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available. Modern data sets also have much less structure, or more complex structure, than the traditional cross-sectional, time-series, or panel data models that we teach in our econometrics classes. While business analytics are a big deal and surely have improved the efficiency of many organizations, predictive modeling lies behind many striking information products and services introduced in recent years [7].

B. Types of big data which we need
Not only in e-learning or educational system but also in any big organization needs to be cognizant of the types of data they need to deal with. The main types of data are structured data and unstructured data. Generally the term structured data (SD) is applied to databases (DB) and unstructured data (UD) applies to everything else.

C. Big data storage technologies
Big Data requires huge amounts of storage space. A typical big data storage and analysis infrastructure will be based on clustered network-attached storage (NAS). Clustered NAS infrastructure requires configuration of several NAS “pods” with each NAS “pod” comprised of several storage devices connected to an NAS device. The series of NAS devices are then interconnected to allow massive sharing and searching of data. Owners of small to medium sized businesses who are unable to afford adoption of clustered NAS technology can consider a number of cloud computing models to meet their big data needs. So we have to know that data storage using cloud computing is a viable option for small to medium sized businesses considering the use of Big Data analytic techniques [8]. Note that cloud computing is becoming a reality for many businesses, with private cloud deployments often leading the way. Cloud technology is maturing and addressing barriers to adoption with improvements in security and data integration, while IT organizations are evolving to support cloud services delivery. As a result, businesses are demonstrating growing trust in cloud delivery models.

II. Overview about big data and e-learning
We will discuss the meaning of big data and eLearning both separately.
D. Big data analytics
The big data analytics initiative should be a joint project involving both IT and business. IT should be responsible for deploying the right big data analysis tools and implementing sound data management practices. Both groups should understand that success will be measured by the value added by business improvements that are brought about by the initiative. [9].

E. The differences between distance learning, e-Learning, and online learning.
Distance education is the most renowned descriptor used when referencing distance learning. It often describes the effort of providing access to learning for those who are geographically distant. Online learning is direct relationships between previously described modes and online learning by stating that one uses the technology used in the other and described by most authors as access to learning experiences via the use of some technology. E-Learning education via the Internet, network, or standalone computer. E-Learning is essentially the network-enabled transfer of skills and knowledge. E-learning refers to using electronic applications and processes to learn. E-learning applications and processes include Web-based learning, computer-based learning, virtual classrooms and digital collaboration. The content of e-learning is delivered via the Internet, intranet/extranet, audio or video tape, satellite TV, and CD-ROM [10].

F. E-Learning objectives
E-Learning represents an innovative shift in the field of learning, providing rapid access to specific knowledge and information [11]. E-Learning enables organizations to overcome distance and other organizational gaps by providing a cohesive virtual learning environment. Companies must educate and train vendors, employees, partners, and clients to stay competitive and E-Learning can provide such just-in-time training in a cost-effective way [12].

G. Student motivation in e-learning
Students enrolled in higher education courses come from a variety of backgrounds and have different reasons for studying. While it is generally accepted that online learning designers should use intrinsic motivation strategies, extrinsic motivation may also be used. A university student may be extrinsically motivated in only doing what is required in order to pass units without a significantly deep interest for the subject. Students studying in distance mode need to feel that they are part of a group of learners and are able to obtain assistance with the unit’s requirements and technical difficulties. For students who are intrinsically motivated to study due to a desire to develop a deeper understanding of the subject matter content which fosters deeper understanding of the subject and relates to real-life and employment situations should also be included [13].

III. Big Data in eLearning: The Future of eLearning Industry
Big data, in terms of the eLearning industry, is the data that is created by learners while they are taking an eLearning course or training module.

A. Benefits That Big Data Offer To eLearning Professionals
There are a variety of benefits that big data offer to eLearning professionals, all of which have the power to impact the future of eLearning and revolutionize the way we analyze and assess the eLearning experience [14]. Here are just a few of the most significant advantages associated with big data: Allows eLearning professionals to understand how the learners are digesting the information and which learning needs appeal the most to them, Enables eLearning professionals to pinpoint areas that may need to be fine-tuned within the eLearning course or module, Provides an analysis of which eLearning modules are visited the most and in the case of social learning which eLearning modules or links are shared with other learners and Data is received almost immediately, rather than having to wait for long periods of time to receive assessments.

B. Big data will impact the future of eLearning
Offering invaluable feedback, allowing, allowing eLearning professionals to design more personalized eLearning courses, Targeting effective eLearning strategies and eLearning goals, and tracking learner patterns and expanding our understanding of the eLearning process.

C. Big Data with e-Learning development
In the e-Learning world, when learners interact with content in your course, they produce data—or Big Data. We’re now able to collect and track this data through learning management systems (LMSs), social networks and other media that tracks how learners interact with aspects of the e-Learning course. The good news is that we’ve come a long way for the origin of the term Big Data. Today, Big Data be processed and analyzed, which is especially helpful for the e-Learning industry [15].

3.3.1 Data driven personalization in eLearning
Data driven personalization is a lesson you can apply to your e-Learning content with the help of big data and new technology like Tin Can API, you have the opportunity to know more about your learners and their behavior patterns than ever before. You can use that knowledge to come up with e-Learning courses that truly engage your learners, through scenarios that speak to their job situations and needs, characters they will relate to and more [16].

D. How to manage data in eLearning?
Big Data is a Data Management & Analytics market opportunity driven by new market requirements. In-Database Analytics – Data Mining there are used Big Data Connectors to combine Hadoop and DBMS data for deep analytics. Also there is the need to re-use SQL skills to apply deeper data mining techniques or re-use skills for statistical analysis. Everything is all about “Big Data” instead of RAM-scale data. This is how the predictive learning of relationships between knowledge concepts and business events is done [17].

IV. The Potential of Learning Analytics and Big data
The understanding of Big Data is mainly very important. In order to determine the best strategy for a company it is essential that the data that you are counting on must be properly analyzed. Also the time span of this analysis is important because some of them need to be performed very frequent in order to determine fast any change in the business environment [18]. The first international conference on Learning Analytics and Knowledge was held in 2011 in Alberta [19] defining Learning Analytics “as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”.
A. Improving learning between students and their teachers

While reflection on one’s learning is critical, how do we learn to reflect and make effective decisions about improving our learning? Personalized feedback is extremely useful but for the teacher it is time-consuming and cannot always be available in a timely manner. Personal analytics that potentially identify one’s common errors in understanding or less-than-appropriate techniques used by the students can provide them with a context in which to reflect on their own development and to address improvements in a timely manner. How might we best present such feedback and information to students? There are several examples of systems and Web sites that attempt that. There are web site with open resources for learning mathematics and science through a series of online video and interactive activities. Detailed bar graphs show students’ progress, histograms presenting the problems students encountered with particular activities and a tree structure represent the knowledge. Can the use of various feedback cases and contextual knowledge be useful in supporting such feedback? How can teachers leverage such knowledge to provide the appropriate and effective feedback efficiently? How can the data help students help each other in developing further understanding and help them improve their learning experience? [20].

B. Using Learning Analytics to Predict (and Improve) Student Success

Learning analytics provides one of many methods to not only document student performance but also to provide tools that encourage the types of continuous improvement that accrediting bodies are seeking. On a more national level, institutions of higher education are experiencing greater demands to retain students. For example, the American Graduation Initiative (Brandon, 2009) urges five million additional higher education graduates by 2020. Learning analytics can assist with this goal by providing a more personalized learning experience through the use of data to respond to students’ needs. This kind of personalization will likely lead to greater success in the classroom. Learning analytics can be used at various levels, including the course, curriculum, institutional, and national level. There is value in being able to leverage data analytics at all of these various levels [21].

V. Big data get out of the bottleneck Oracle, Hadoop, and NoSQL

The Oracle Big Data Appliance is an engineered system of hardware and software designed to help enterprises derive maximum value from their big data strategies. It combines optimized hardware with a comprehensive software stack featuring specialized solutions developed by Oracle to deliver a complete, easy-to-deploy offering for acquiring, organizing, and analyzing big data, with enterprise-class performance, availability, supportability, and security [22]. The Oracle Big Data Appliance offers the following benefits: Rapid provisioning of a highly-available and scalable system for managing massive amounts of data, A high-performance platform for acquiring, organizing, and analyzing big data in Hadoop and Control of IT costs by pre-integrating all hardware and software components into a single big data solution that complements enterprise data warehouses. Apache Hadoop is a powerful open source software platform that addresses above mentioned problems of big data. This is the platform used for cloud computing by some of the pioneers like Yahoo!, Amazon, Facebook, eBay, etc. Two major components of Hadoop are: Hadoop Distributed File System (HDFS) for distributed storage and Map reduce for parallel processing. Hadoop includes a fault tolerant storage system called the Hadoop Distributed File System. HDFS is able to store huge amounts of information, scale up incrementally and survive the failure of significant parts of the storage infrastructure without losing data. Hadoop creates clusters of machines and coordinates work among them. Clusters can be built with inexpensive computers. If one fails, Hadoop continues to operate without losing data or interrupting work, by shifting computation to the remaining machines in the cluster [23].

NoSql architecture basically tries to achieve the shared nothing architecture, the term shared nothing means that when the data repository is being sharded the clusters should not be dependent on each other for any kind of updates any change inside any cluster should not affect the others [24].

Big Data is creating new opportunities for organizations to serve customers and markets — while also creating and extracting value — in new ways. There are several technologies to handle big data such as MapReduce, Hadoop, NoSQL, PIG and Hive. In this paper we are focusing on the various NOSQL tools. NoSQL database provides the foundation for many of these systems, not only as a real-time, operational data store but in offline capacities as well. Traditionally, the work of capturing and analyzing data has required different technologies, different infrastructure and redundant costs. It is often used for storing Big Data. This is a new type of database which is becoming more and more popular among web companies today. Proponents of NoSQL solutions state that they provide simpler scalability and improved performance relative to traditional relational databases. These products excel at storing “unstructured data,” and the category includes open source products such as MongoDB, and Redis. NoSQL encompasses a wide variety of different database technologies and were developed in response to a rise in the volume of data stored about users, objects and products, the frequency in which this data is accessed, and performance and processing needs [26].

VI. Framework handling educational big data using hadoop

As shown in fig.1 a clear map which discuss hadoop environment with big data starting with gathering structured and unstructured
data to store and manage this. After that looking on products (data) to help developer write hadoop application or otherwise analyze this data.

VII. Conclusion
This paper has clarified the manipulation of big data in educational environment. This has been done by using hadoop and nosql. Furthermore, a propose framework for handling big data in E-Learning environment has been presented. In addition, the operation of the proposed model will be accelerated by using soft computing intelligent techniques. Moreover, its performance will be enhanced through data mining approaches. We expect that the proposed technique will develop E-Learning industry.

References
[1] Editor-in-Chief: Vasant Dhar, Executive Editor: Eugene Kolker, ISSN: 2167-6461 • Published Quarterly Online, ISSN: 2167-647X, Current Volume: 2.