

Real-Time Based Big Data and E-Learning: A Survey and Open Research Issues

Wael W. Hadeed

wael.hadeed@uomosul.edu.iq

Dhuha B. Abdullah

prof.dhuha_basheer@uomosul.edu.iq

*Department of Computer Science,
College of Computer Science and Mathematics,
University of Mosul, Mosul, Iraq*

Received on: 02/10/2021

Accepted on: 22/11/2021

ABSTRACT

Big data is considered a remarkable aspect of development communication and information systems. In terms of delivery, retrieval, and storage, a vast volume of complex data exceeds the capability of conventional software and device capabilities. As a result, advanced alternative solutions that allow their control flow is becoming more prevalent. One of the most difficult fields of information management and technology is the real-time management, analysis, and processing of big data. These issues can be seen in a large amount of everyday produced data in a variety of places, including online social networks and the method of logging cell phone data. This survey looks at several studies that use a range of methods to handle, interpret, and process big data in real-time. Hence, the objective of this survey is to provide a comprehensive overview of the integration between real-time and big data fields of study with the field of E-learning. Finally, this survey also presents the colorful aspects of big data and their relationship to E-learning domains such that e-learning platforms, big data frameworks, and datasets used.

Keywords: Real-Time processing, Big data, Data analytics, E-Learning, Big Data Platform.

I. Introduction

Big data is an evolving term that describes a very large amount of structured and unstructured data that can be analyzed mathematically to obtain methods, relationships, and linkages, especially those related to human behavior and interaction, to the effects of what we do in the information age [1]. Through the Internet, computers, or our mobile phones, which describes this vast amount of information [2].

A. Overview

Another type of data source is user behavior, which includes things like searching the Internet for a service, product, or other types of information, as well as the number of times a page is visited on the web [3][4], like those emerging from executing a program, Regardless of what this program does, such as bank records, electronic health care, insurance records, hospital visits, electronic education, etc... Commercial sources are another data source, for instance, transactions over the Internet by mobile devices and credit card transactions [2][5]. Networks too depend on some sources like tracking devices and sensors, for instance, climate sensors, tracking data from cell phones, road sensors, satellite imaging, the Global Positioning System, and others [6].

One of the most important domains that use the term big data is e-learning, especially after the spread of the Covid-19 epidemic a global trend has been made to use

e-learning platforms [7][8]. To increase the data scale, e-learning platforms and interaction-based systems are increasingly being used in the education area. In learning environments, we obtain detailed information and also information regarding social relationships, as well as comprehensive data on learning events, such as texts, media, and videos, and these data differ in quality and depth to varying degrees [9] [10] [11]. We find data sets about learners' academic expertise and learners due to the disparity between the types of data sets that can be obtained from learning environments [12]. These types of big data research can be utilized in education to provide a number of options and possibilities for students to learn more effectively, such as competency-based learning or adaptive learning, in which faster learning leads to better learning [13].

B. E-Learning

E-Learning is a reactive educational method that allows students to use information and communication technologies to learn [14]. It is built on an integrated digital electronic environment that displays courses through e-networks, provides tools for organizing assessments, guiding, and directing, as well as assessing processes and resources and managing them [15]. Traditional learning refers to an educational system based on the classroom in its traditional form, in which the classroom is a physical thing as well as a collection of classrooms that include students and learners from diverse backgrounds, and learning is imparted to them through the instructor, who is the focal point of the educational process and the primary source of information. The distinction between traditional education and e-learning derives from the tools utilized in the educational process, and because the two techniques are so dissimilar, there are several differences between them. Some of the differences between e-learning and traditional learning can be summarized in Table 1. E-Learning has three types:

- Synchronous e-learning: is education that takes place over the air or directly, and allows a large group of students to engage in dialogue and interaction at the same time, both among themselves and with the instructor [16].
- Asynchronous e-learning: This form of education is known as indirect education because it does not require the presence of learners at the same time. Instead, the learner will study at his own pace and with the commitment he desires [16].

Table 1. E-Learning vs. Traditional Learning

Des	E-Learning	Traditional Learning
Culture	It introduces a new sort of culture, digital culture, that focuses on knowledge processing and helps the student, not the teacher, to be at the center of the educational process.	It is based on traditional culture, which focuses on the production of knowledge, and the teacher is the basis of the process.
Cost	To equip the infrastructure of computers, manufacture software, train teachers and students on how to deal with this technology, and design scientific material electronically, it requires a considerable expense, especially at the beginning of its use. Assistants are also required to establish an interactive atmosphere for instructors and assistants.	In terms of infrastructure and teacher and student training to obtain technical competences, traditional education does not require the same cost as e-learning. He also does not require helpers since, in a traditional learning environment without the use of modern technological media or teacher assistants, it is the instructor who imparts knowledge to the minds of students.

Actual attendance	E-learning does not have to give education at the same time or in the same place; instead, the student is free to choose a specific location and time for the learning process..	All students receive traditional education in the same place and at the same time.
Teacher Status	The teacher is a guide and guide to information, as well as an advisor and assistant to the student by providing the necessary advice.	The educator is the transmitter and the primary source of information.
Feedback	Attention to immediate feedback on development work.	Feedback does not have a clear and satisfactory role.
Class activity	Because scientific knowledge is delivered in the form of written texts, still and moving images, video clips, graphics, graphs, and simulations, as well as an electronic course or a visual e-book, students are more excited and motivated to learn.	The educational content is presented to the student in the form of a printed book containing editorial texts, even if some pictures are more than that and the technical accuracy is not available.

- Co-education: It gives the learner more freedom and achieves a form of socialism in education by becoming asynchronous at times and synchronizing at other times, depending on the activities suggested by the lecturer. [7] [18] [17] [16].

The shift from Traditional to Online learning

The computer is being utilized to teach not only typical students, but also kids with hearing and vision impairments, as well as pupils with learning problems. Distance education, libraries, and electronic publication all benefited from the computer. There are accredited and non-accredited computer programs on the Internet to teach listening, speaking, reading, writing, spelling, grammar, vocabulary, pronunciation, dictionaries, encyclopedias, and programs to correct grammatical and spelling errors in the field of teaching and learning English for non-native speakers.

Therefore, the transition from traditional education to e-learning based on technology, whether wholly or partially, has become a necessity. [46][47][48]

C. Big Data

There is no consensus to give an explicit definition of the term “big data”. In contrast, there are a number of attributes that describe big data, the most important of which are:

- Volume [19] [20]

The weight and capabilities of the data to be classified within big data are determined by the amount of data collected from a source.

- Variety [20]

Refers to the variety of extracted data such as images, audio and video clips, SMS, call logs and GPS data, and it requires stress and time to get ready it in a form suitable for analysis and processing

- Velocity [20]

Refers to the speed at which data can be extracted and generated. where speed is a key factor in making a decision based on this data, and it refers to the time it takes from the time the data is received to the time the decision is taken.

D. Big Data Analytics

Analyzing data from information that is collected in raw form or taken from a data warehouse, is complex and requires several steps that can be summarized through the following steps and shown in Figure 1. [19]

1. Data Sources: Collecting the information from different data sources.
2. Data Integration: Combine data from different sources in worthy information.
3. Data Management: Applying some data transformation rules such as transformation, correction, enrichment, and manipulation.
4. Data analysis: Make the right decisions at right time based on suitable data analysis techniques.
5. Data Repository: Storing and accessing this information into/from the big data repository.
6. Real-Time Data Stream Processing: This way of continuous computation happens almost instantaneously which data streaming from one device to another through the system.

E. Big data framework Types

It provides a structure for organizations that want to work with big data [1] [21]. The big data framework includes all the organizational aspects that must be taken into consideration in a big data organization [22]. The frameworks are many and they fall into many areas, depending on the method of processing big data [23] [24]. Figure 2 illustrates the classification of some frameworks.

Studies in this field have dealt with different types of big data frameworks, and each according to the requirements, this can be shown in Figure 2

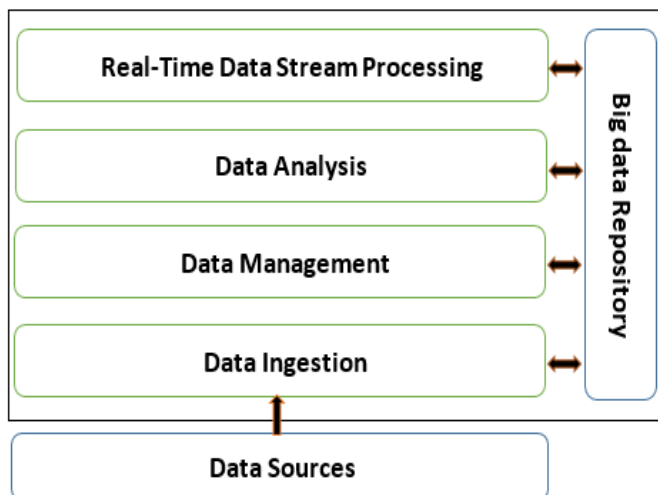


Figure 1. Big Data Analysis

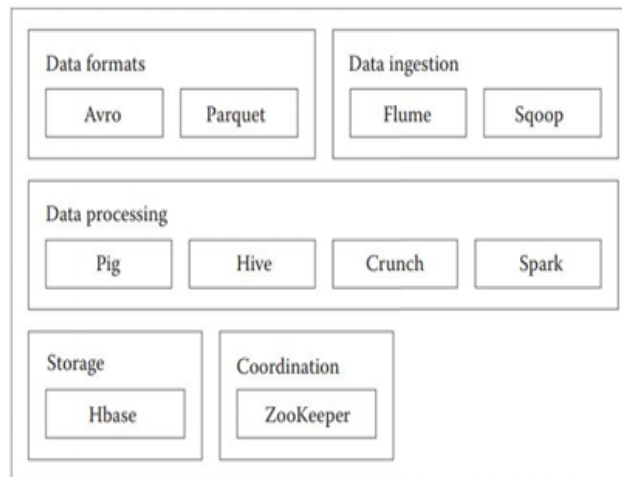


Figure 2. Big Data Analysis

Big data and E-Learning integration benefits

Big data can provide new possibilities in eLearning. the competency lies in the ability to manage the big data effectively to streamline eLearning instructional strategy. Online teaching is aided by Big Data. Learning analytics can provide you with a better idea of how the student interacted with the material. This will allow us to track:

- Their learning styles and preferences in order to create personalized learning experiences.
- Whether the eLearning program solved the workforce's performance problems or not.
- The efficiency of the online training program while determining its cost-effectiveness.
- The areas where learners get stuck.

II. Big Data and E-Learning

In the past few years, the central position for big data analysis has provided authors and scientists with opportunities to develop a range of projects, methods, and analytical techniques related to challenges and their applications in many fields. In this survey, we sampled previous studies that give a clear understanding of the use of big data with e-learning.

Dahdouh et al. [1], by incorporating emerging technologies in the ecosystem of big data and cloud services in the e-learning area, several solutions for cloud computing and big data were presented. The aim was to improve e-learning platform efficiency in terms of Computing Environments for Human Learning (CEHL), relying on cloud-based services over the Internet. To build dealing with big data over the Internet, the author used the LMS Moodle platform. The author used Spark as a framework within the big data pool to analyze the learning dataset obtained from the EsTenLigne platform. And both Orozova and Popchev [25] presented a study on the techniques for storing, searching, analyzing, and managing big data in the field of virtual e-learning space and the problems encountered. They implemented the Evidence Exploration System for students from the Burgas Free University by using the Orange platform. The Hadoop Big Data Framework was used to organize the data of the students of the Burgas Free University which forms the data set that the authors used during the study. However, Ashraf et al. [16] added NoSQL to Hadoop and found that there is a need to improve the performance of big data processing in educational systems. There are many different types of data in the e-learning environment. The main goal is to introduce a

new e-learning model that manages big data effectively to support learners and educational systems. The proposed model collects important data using intelligent methods through interaction between learners and learning content, in addition to determining the behavior of all learners in the electronic learning system through the processing of large educational data with the use of Hadoop as a framework of the NoSQL.

The challenge of creating a new learning platform is the most important issue that has been addressed by Karim and Muhammed [27], They were given a brief overview of the possibilities for disseminating big data in the educational sector, as well as the benefit that students and businesses gain from highlighting obstacles, choosing resources, and extracting useful data from complex educational data sets. According to the authors, educational institutions produce vast volumes of data to conceal important information; future studies should focus on publishing other platforms such as Super Doop for e-learning and unified data protocols that ensure protection and privacy.

There may be non-contact learning, that is, the so-called passive or offline learning, the paper presented by Nair et al. [28] deals with online learning acceleration through the use of offline data sets. They proposed a new algorithm that combines effective dynamic programming with maximum likelihood policies update to accelerate data analysis and processing. This gives a simple and effective framework for dealing with large amounts of offline data, thus obtaining a higher speed between offline data sets and online learning. The proposed algorithm "Advantage Weighted Actor-Critic (AWAC)" can quickly learn successful policies in difficult tasks with high action dimensions and binary sparse rewards, and is much better than previous methods in this field.

Abdelouarit et al. [20] suggested developing a tool called Big-Learn that works in the environment of big data and e-learning to speed up the process of searching in big data. This tool requires a combination of data diversity, improved performance, resilience, durability, and scalability. Indeed, the tool is based on the technique of merging and mixing structured and unstructured data into a single data layer, thus facilitating access to more research relations with consistent and appropriate results for the learner. The authors relied on a qualitative and quantitative study of an assortment of data, and their classification, followed up by a detailed analysis of the data structured to find a fictional model for processing at the end.

There must be some kind of adaptation in the field of e-learning, Birjali et al. [29] proposed a model for adaptive e-learning to supply the most appropriate educational content for every learner, in this context this model relies on two layers of adaptive e-learning starting from which the relevant future educational goals are determined by appropriate e-learning for the learner using genetic algorithms based on the MapReduce tool To create an adaptive learning path for every learner by using the "Ant colony Optimization" algorithm, also a proposal for analyzing social networks based on the MapReduce to determine the learner's motivation to set a specific learning cadence for every learner, the initial result of using this algorithm on the big data collecting better than those That uses traditional businesses.

Providing an advanced unified electronic educational platform suitable for a large group of universities, the research presented by authors Logica and Magdalena [9] discussed aspects related to developing big data, how to apply it to e-learning, and its impact on the universal environment. They have created three-level device architectures for a community of universities to analyze massive data sets and access them in a cloud environment, depending on existing software solutions. This was accomplished by

using the Gephi graphical tools to explore unstructured data. The aim was to create a cohesive education management system based on open-source software that would allow universities to take advantage of this trend.

The data generated as a data warehouse needs to select an appropriate tool for generating big data in its practical sense, the author Yulia [22] reviews the implementation of big data technologies while exploring the modern data warehouse that can replace the traditional data warehouse that cannot work with big data in the educational system. The tools used with data warehouses cannot handle the load and data analytics processes, so big data technologies must be used to expand data warehouse solutions. The author suggests using Hadoop technology as a tool to analyze the big data generated by universities and this data is growing exponentially. Also, it can support the decision-making process. The description of the approaches is summarized in Table 2.

III. Big Data with other Domains

Big data provides a competitive advantage in several areas, so if institutions effectively use and analyze it, they will be able to provide excellent service to their customers and meet their needs, as well as make more informed decisions within the organization based on information extracted from customer databases, resulting in increased efficiency and profit.

Cloud computing with the real-time requirement of big data greatly influences the accurate and fast rendering of results, Zheng et al. [23] presented a report on the challenges of big data and summarized these challenges into six topics, with a recommendation of a four-layer architecture model and hierarchical computing to increase the efficiency of real-time big data processing (RTDP) based on cloud computing technology. To meet real-time requirements and bypass heterogeneity in the RTDP framework, the model stores a multi-level and publishing process that relies on the LMS educational system, and through the use of a set of techniques for analyzing big data such as MapReduce, the data was organized and then uploaded to the cloud server.

Author Chen et al. [30] introduced a big data PRF (Parallel Random Forest) algorithm with Apache Spark as a framework, in addition to improving the aforementioned algorithm to look like a hybrid view that combines parallel data optimization with parallel tasks, moreover, the use of Apache Spark with big data gave good working accuracy in terms of algorithm optimization. Apache Spark helps reduce dimensions in the training and prediction process, thus obtaining accurate classification, performance, and scalability.

Frameworks can be used in several fields, including environmental, banking, and health systems. Authors Munshi and Mohamed [31] presented an ecosystem for big data with a smart grid and use LMBADA as a framework. This system can implement real-time operations and parallel batch on distributed data. This system relies on Hadoop big data lake to store various types of smart grid big data, such as Smart Meter, photos, and video data, with the use of cloud computing. The system gave the response and the ability to perform many analyzes on big data.

IV. Techniques for Big Data and Online Systems

Many technologies serve as a link between big data and online systems. This field is broad and developed, especially after the global epidemic, Covid-19. E-learning has become mandatory in a large group of countries, so this space took the researchers'

attention to develop and give opportunities for continuing education with the suspension of schools, universities, and others.

Author Li Du and Qian Sun [49] gave an in-depth study on current online learning in the year 2020. The popularity of online learning is very strong, and the importance of teaching has been stressed in various ways, according to the poll. And, because the technology is still in its early stages of implementation, there are still some technical flaws, such as a lack of data security, the teaching platform's performance being unsteady, student platform management not being scientific, and so on. To arrive at suitable proportions, the author used a sample and a field survey. By interviewing 60 teachers and 500 students, he performed this survey in the form of a questionnaire and a field interview. A total of 560 questionnaires were distributed, with 551 of them being returned. The recovery rate was 98.3 percent. Some of the terminology used in the survey (use the Internet to teach, they are satisfied with the effect of teaching, improve learning efficiency). The author concluded that e-learning should be used in an organized and well-planned manner, as it is a process that needs to be developed, and all parties involved in this field must build a strong electronic educational system with ways to include stakeholders in educational courses to achieve the best possible result.

Since today's learning environments are digital, learning analytics applications, like virtual mentors, come to the fore in today's education systems. Author Dulger [50] believes that students feel lonelier and less motivated during e-learning activities, making managing the learning process more difficult. To present a clear image of learning analytics and how it might adapt to educational institutions, the author employed the LAID program (Learning Analytics Implementation Design). Later and by taking a sample of Sakarya University students, he used remote electronic technology to teach a mathematics course. The author employed the notion of coordination and comparison in the LAID program to acquire time and manipulate the data through a series of attempts, and this provided the engagement that the student desired.

Choosing the educational system is one of the most important points through which the degree of success of the work can be determined. The author Fitria [51] in 2021 relied on an e-learning project developed by an Indonesian institution to create an e-learning platform, by using LMS as an educational system. A system based on learning standards helps manage the learning process. The administrator, instructor, and student each have their user access permissions in this program. Everyone has a unique level of access. He can manage the lecturer and student data as well as the application system as a whole. During the present COVID-19 pandemic, this work was made as an alternative to traditional learning. By offering materials, assignments, discussions, and messages with the students, the project was able to compensate for the lack of presence instruction. Students can only get information, materials, tasks, conversations, and communications from the lecturer in the meantime.

A closer examination of the paper by Vlasenko et al. [52] reveals that the author's educational system is PLE (Personal e-Learning Environment) based on evidence of teaching activities carried out by mathematics teachers to perform educational, planning, design, cognitive, administrative, and social matters. The findings of the study led to the conclusion that it is dependent on the teacher's desire to construct a PLE (Personal e-Learning Environment) model, as well as other aspects such as the layout of web resources and the teacher's willingness for professional growth. The methodology was put to the test by assessing the level of awareness of teaching staff in Ukrainian

universities regarding the usage of web tools in the classroom. It is preferable to consider the views of the widest possible set of stakeholders in this type of application.

V. Big Data Frameworks Usage

As mentioned previously, it is very important to choose the appropriate framework for dealing with data to obtain the most detailed form of large data. The following is an overview of researchers' use of framework compositions that are not specified, Author Kumar et al. [32] gives a basic concept of big data with the benefit of work in using this concept in addition to the method of using apache Hadoop. The author gives the way of the apache Hadoop works by listing the components of the Hadoop (HDFS, MapReduce). In addition, the author touches on a tender presented on NewSQL and NoSQL with the characteristics and analysis of the big data.

The author Huafeng Yu [53] preferred to use the Spark framework over the Hadoop framework to extract useful information from big data, this choice was based on discussing the relationship between the various influencing factors to evaluate moral education, and through the results that appeared, it was shown that the Spark framework gives better results than the Hadoop framework in parallel implementation to the proposed algorithm, so care must be taken when choosing the framework by taking enough time to understand the big data and what is required of it. Figure 3 shows the workflow and data hierarchy in the Spark framework

Author Wang et al. [55] also proposed an efficient model that could detect duplicate profiles inside a single platform dating back to the same individual by successfully integrating user profiles across platforms using Apache Spark for distributed execution.

When dealing with structured big data, NoSQL database systems should be utilized to execute retrieval and contiguous operations on a vast amount of data, rather than relational models, which are generally useless. They were primarily employed for real-time applications and statistical analysis of ever-increasing data sets.

Author Kaur [56] attempted to highlight the two databases, Cassandra and MongoDB, that have the most common security issues. The most well-known problem with the two databases is the lack of encryption in the information documents, as well as poor validation between the client and servers and among some server personnel. Simple consent without the use of role-based access control and denial of service are just two of the issues that hackers can exploit. Figure 4 shows that the NOsql database performs a data classification process.

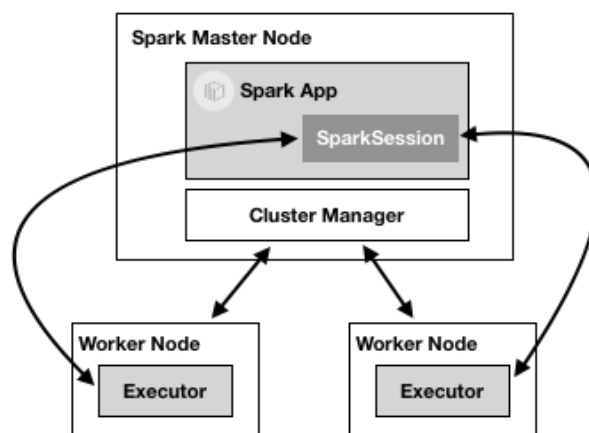


Figure 3. Spark Framework Architecture [54]

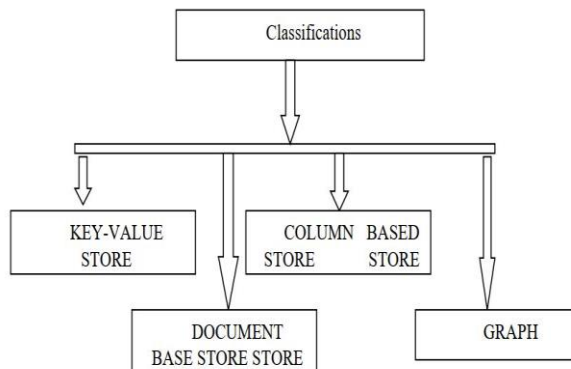


Figure 4. NoSQL database system [56]

In the same context, a new open-source platform was proposed that facilitates the process of storing and managing data in one system, and Apache Hadoop is the core of the big data operating system. There is another tool that gives a clear concept of dealing with very large data that the authors Konda and Rohini [33] discussed where they first dealt with the architectural framework of Hadoop and the method of overcoming the tolerance of errors in (HDFS) which includes checkpoint and recovery and copying data. The authors demonstrated the efficiency of Hadoop by storing large quantities of data. And to stream those data sets at high bandwidth to user applications. In addition to the use of Zookeeper and Flume, which deal with the fundamentals of synchronization and coordinates of distributed systems, thus dealing with large and varied data with the ability to efficiently transfer data from many different sources to a data central warehouse.

Artificial intelligence techniques can be used in the analysis, organization, and processing of big data. The interaction between artificial intelligence and big data gives positive results in the process of organizing the data. Thus, analyzing and manipulating big data is key to solving AI problems. Authors Gupta et al. [34] proposed a new framework that combines the computational distributive capabilities of SPARK with multilayer perceptron MLP. The researchers say the results were good and satisfactory, which proves improved methods for analyzing big data.

Databases are classified as structured data, so an exemplary framework can be proposed for dealing with databases. This can be observed through the presentation of the researchers Moniruzzaman and Hossain [35], as well as the student Zaki [36], and Venkatraman et al. [37] in dealing with organized or structured databases, where the NoSQL framework can be used that It gives excellent results for handling this type of data.

From the above, it can be said that the method of choosing a framework is considered one of the most important foundations in the success of the work, where a set of factors must be taken into consideration, the most important of which are: the environment (where the raw data is located), the presence of the data on a network (Structured) and other influences. Table 3 illustrates researchers' use of big data frameworks.

A. Facts and Recommendations

The importance of dealing with big data and e-learning can be seen from the above studies, as authors have flourished by listing several applications and proposals in this field, using big data framework such as HADOOP or SPARK or NoSQL, etc., with some intelligence technology which given us a good data organization and thus faster

access in terms of storage and retrieval, especially by dealing with data-sensitive such as e-learning.

B. E-Learning Systems and Platforms

Many types of e-learning systems depend on their use in the learning environment, infrastructure, and education system, among the types of e-learning systems learning management Systems (LMS), virtual learning environments (VLE), content management systems (CMS), etc.

Authors have several options in choosing the appropriate educational system and e-learning platform, Table 4 shows the options of authors in this field.

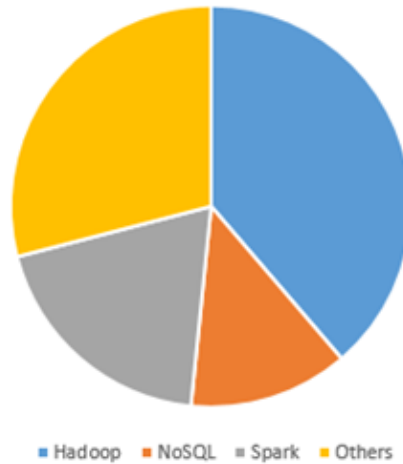


Figure 5. Big Data Framework Usage



Figure 6. Educational Systems Usage

Table 2. Researches Advantage and Data Sets Usage

<i>Author</i>	<i>Year</i>	<i>Dataset</i>	<i>Advantage</i>	<i>Limitation</i>
Karim Dahdouh et al. [1]	2020	Data collected	<ul style="list-style-type: none"> • Enhance the online learning environment for all students. • Integrate big data technology into e-learning platforms. • Moodle LMS was used to build online learning platforms. • Used parallel FP-growth. 	In the storage process, a better mechanism must be created, and the problem of available resources must be addressed, all while speeding up implementation and lowering processing costs.
Orozova and Popchev [4]	2019	Burgas Free University	<ul style="list-style-type: none"> • Implemented the Evidence Exploration System for students. • Use data mining tools for analysis and prognosis. 	Experimenting with several learning platforms to give better results.
Ahmed Ashraf And et al. [16]	2015	-----	<ul style="list-style-type: none"> • Improving big data analysis efficiency in educational systems • Introduce a new e-learning model that manages big data effectively. 	Data mining techniques can be used to speed up project operation so that this technology can work better.
Karim Dahdouh and et al. [26]	2019	Proposed and Generated	<ul style="list-style-type: none"> • Proposes a method for integrating big data and online learning systems with cloud computing to improve distance learning performance. • In the e-learning world, big data technology is used to increase the quality of distance learning. 	Realistically implementing the project requires real evidence, then some results may change with the use of more frameworks to reach the best.
Karim and Muhammed [27]	2020	-----	<ul style="list-style-type: none"> • Enhanced big data analysis • Changed learning environment • Analyzed career counseling based on student data • Reduce the number of students dropping out of classes • Assist in decision-making and administrative tasks. 	There is a challenge in choosing the tool and extracting the value from the complex set of educational data, in addition to security restrictions and data privacy. Therefore, there must be accuracy in dealing with education data.
Nair et al. [5]	2020	-----	<ul style="list-style-type: none"> • Online learning acceleration • Proposed a new algorithm AWAC that used effective dynamic programming and big data analysis • AWAC is much better than previous methods in this field. 	Using raw data that is not connected to the Internet, the algorithm should preferably handle all types of data, including online data.
Abdelouarit et al. [20]	2015	-----	<ul style="list-style-type: none"> • Suggested a tool called Big-Learn that works in the environment of big data and e-learning • The tool is based on the technique of merging and mixing structured and unstructured data 	The development of this tool requires creating and configuring a questionnaire for a small sample of online learners
Birjali et al. [29]	2018	•CACM collection •Medline collection •Cranfield collection	<ul style="list-style-type: none"> • Proposed a model for adaptive e-learning to provide the most appropriate educational content for each learner • Use the "Ant colony Optimization" algorithm • Analyzing social networks based on the MapReduce 	Other intelligent technologies can be tried to obtain better results by finding an adaptive model between e-learning and big data and giving the best picture in this field.
Logica and Magdalena [7]	2015	General	<ul style="list-style-type: none"> • Designed a three-level system architecture for a group of universities. • Exploring unstructured data using the Gephi graphical tool. 	Analyzing the performance of many analytical tools and testing them against workloads with interpretation and query of e-learning data. This gives unclear and ambiguous results.

Zheng et al. [23]	2015	-----	<ul style="list-style-type: none"> • Improve the performance of real-time big data processing (RTDP). • Proposal of a four-layer architecture model and hierarchical computing. 	There is a lack of support for real-time massive processing with big data frame and real-time platform processing
Yulia [22]	2017	General	<ul style="list-style-type: none"> • Suggests using Hadoop technology as a tool to analyze the big data generated by universities. • Take control of data that is growing exponentially, rich and growing. • The information generated can support the decision-making process. 	Modern data warehouse deals with big data, which gives methods that may be unclear to system users in this area

Table 3. Big Data Framework Usage

Reference	Big data frameworks			
	Hadoop	NoSQL	Spark	Others
[1]	√		√	
[2] [3] [22] [27] [29] [7] [38]	√			
[3]	√	√		
[4] [13]				Proposed
[39]				Hive & Cassandra
[40]				Kafka & Storm
[24]				Yarn
[41]	√			Mahout
[42] [31]			√	Lambda
[30] [34] [43]			√	
[18]				SAMOA
[44]				general
[32]	√	√		
[33]	√			Zookeeper & Flume
[35] [36]		√		

Table 4. Educational System and E-Learning Platform Usage

<i>REF.</i>	<i>PLATFORM</i>	<i>LMS</i>	<i>CMS</i>	<i>VLE</i>	<i>OTHER</i>
[1]	ESTENLIGUE	√			
[27]	SUPERDOOP				√
[29]	GENERAL				√
[11]	GENERAL			√	
[25]	ORANGE				√
[16]	GENERAL	√			
[45]	MOODLE				√
[12]	GENERAL	√			
[13]	MOOC	√			
[14]	GENERAL	√	√		

VI. Conclusion

Big data field has an important, effective, and crucial role in the education sector. As it provides a new vision for teachers towards learners and directing them in modern ways. E-learning is a future path that provides both sides (learner and teacher) an electronic style that gives an understanding of realistic education. This paper surveyed the literature for variety of studies that deal with the processing and analysis of big data in real-time with e-learning. This study gave an understanding of the latest developments in this field. Moreover, we suggest that time be given great importance to the interaction between big data and e-learning while using platforms and frameworks that use real-time algorithms to obtain best response time. E-learning is one of the areas in which the data is the center of the process due to its importance. Finally, this survey can be considered a comprehensive guide for researchers since it provides the most recent techniques that integrate different areas to achieve one goal, which is having efficient education process.

Acknowledgment

The authors would like to thank the Computer Science Department/ College of Computer Science and Mathematics/ University of Mosul/ IRAQ for all the support in achieving this research.

REFERENCES

- [1] K. Dahdouh, A. Dakkak, L. Oughdir, and A. Ibriz, "Improving Online Education Using Big Data Technologies," in *The Role of Technology in Education*, 2020.
- [2] N. Mohamed and J. Al-Jaroodi, "Real-time big data analytics: Applications and challenges," *Proc. 2014 Int. Conf. High Perform. Comput. Simulation, HPCS 2014*, no. July 2014, pp. 305–310, 2014, doi: 10.1109/HPCSim.2014.6903700.
- [3] A. A.-2018 I. 3rd I. C. on B. Data and undefined 2018, "Real-time big data warehousing and analysis framework," *ieeexplore.ieee.org*, Accessed: Feb. 20, 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8367649/>.
- [4] O. Soumaya, T. M. Amine, A. Soufiane, D. Abderrahmane, and A. Mohamed, "Real-time Data Stream Processing - Challenges and Perspectives," *Int. J. Comput. Sci. Issues*, vol. 14, no. 5, pp. 6–12, 2017, doi: 10.20943/01201705.612.
- [5] M. Banane and A. Belangour, "Towards a new scalable big data system semantic web applied on mobile learning," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 1, pp. 126–140, 2020, doi: 10.3991/ijim.v14i01.10922.
- [6] M. Dayarathna, P. Fremantle, S. Perera, and S. Suhothayan, "Role of real-time big data processing in the internet of things," in *Big Data Management and Processing*, 2017, pp. 239–262.
- [7] B. Logica and R. Magdalena, "Using Big Data in the Academic Environment," *Procedia Econ. Financ.*, vol. 33, no. 15, pp. 277–286, 2015, doi: 10.1016/s2212-5671(15)01712-8.
- [8] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics," *IEEE Access*, vol. 6, pp. 39117–39138, 2018, doi: 10.1109/ACCESS.2018.2851790.
- [9] S.-H. Ju and K.-S. Noh, "A Study on Policy for Data Convergence infrastructure of e-Learning Industry," *J. Digit. Converg.*, vol. 13, no. 1, pp. 77–83, 2015, doi: 10.14400/jdc.2015.13.1.77.
- [10] S. Kusuma and D. Kasi Viswanath, "IOT and Big data analytics in e-Learning: A technological perspective and review," *Int. J. Eng. Technol.*, vol. 7, no. 1, pp. 164–167, 2018, doi: 10.14419/ijet.v7i1.8.11540.
- [11] S. T. Tom and B. Tulasi, "Analytics in e-Learning," vol. 11, no. 2, pp. 319–331, 2018.
- [12] B. C. Czerkawski, "When Learning Analytics Meets E - Learning," *Online J. Distance Learn. Adm.*, vol. 18, no. 2, pp. 5–9, 2015, [Online]. Available: <http://www.westga.edu/~distance/ojdla/summer182/czerkawski182.html>.
- [13] K. Palanivel and T. Chithralekha, "Big Data Reference Architecture for e-Learning Analytical Systems," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 1, 2018, [Online]. Available:

- http://www.ijritcc.org/download/browse/Volume_6_Issues/January_18_Volume_6_Issue_1/1516261400_18-01-2018.pdf.
- [14] K. Kim, S. Trimi, H. Park, and S. Rhee, "The Impact of CMS Quality on the Outcomes of E-learning Systems in Higher Education: An Empirical Study Subject Areas: CMS Benefits, Course Management System (CMS), E-learning, E-learning Success, Information Quality, Instructional Quality, User Satisfact," Oct. 2012. doi: 10.1111/j.1540-4609.2012.00360.x.
- [15] B. Rienties, S. Cross, V. Marsh, and T. Ullmann, "Making sense of learner and learning Big Data: reviewing five years of Data Wrangling at the Open University UK," *Open Learn.*, vol. 32, no. 3, pp. 279–293, 2017, doi: 10.1080/02680513.2017.1348291.
- [16] A. Ashraf, H. El-Bakry, S. M. A. El-razek, and Y. El-Mashad, "Handling big data in e-learning," *Int. J. Adv. Res. Comput. Sci. Technol.*, vol. 3, no. 1, pp. 47–51, 2015.
- [17] S. M. J. Jalali, E. Mahdizadeh, M. R. Mahmoudi, and S. Moro, "Analytical assessment process of e-learning domain research between 1980 and 2014," *Int. J. Manag. Educ.*, vol. 12, no. 1, pp. 43–56, 2018, doi: 10.1504/IJMIE.2018.088371.
- [18] R. Pecori, "A virtual learning architecture enhanced by fog computing and big data streams," *Futur. Internet*, vol. 10, no. 1, pp. 1–30, 2018, doi: 10.3390/fi10010004.
- [19] D. Aggarwal, "Application of Big Data Analytics for the development of eLearning," *Ijrece*, vol. 6, no. 3, pp. 1188–1192, 2018.
- [20] K. Aoulad, B. Sbihi, and N. Aknin, "Big-Learn: Towards a Tool Based on Big Data to Improve Research in an E-Learning Environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 10, pp. 59–63, 2015, doi: 10.14569/ijacsa.2015.061008.
- [21] P. K. Udupi, P. Malali, and H. Noronha, "Big data integration for transition from e-learning to smart learning framework," 2016 3rd MEC Int. Conf. Big Data Smart City, ICBDS 2016, pp. 268–271, 2016, doi: 10.1109/ICBDSC.2016.7460379.
- [22] L. W. Santoso and Yulia, "Data Warehouse with Big Data Technology for Higher Education," *Procedia Comput. Sci.*, vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.
- [23] Z. Zheng, P. Wang, J. Liu, and S. Sun, "Real-time big data processing framework: Challenges and solutions," *Appl. Math. Inf. Sci.*, vol. 9, no. 6, pp. 3169–3190, 2015, doi: 10.12785/amis/090646.
- [24] U. Demirbaga, A. Noor, Z. Wen, P. James, K. Mitra, and R. Ranjan, "SmartMonit: Real-time big data monitoring system," *Proc. IEEE Symp. Reliab. Distrib. Syst.*, no. October, pp. 357–359, 2019, doi: 10.1109/SRDS47363.2019.00049.
- [25] I. P. Popchev and D. A. Orozova, "Towards Big Data Analytics in the e-Learning Space," *Cybern. Inf. Technol.*, vol. 19, no. 3, pp. 16–24, 2019, doi: 10.2478/cait-2019-0023.

- [26] K. Dahdouh, A. Dakkak, and L. Oughdir, “Big data: a distributed storage and processing for online learning systems,” *Int. J. Comput. Intell. Stud.*, vol. 8, no. 3, p. 192, 2019, doi: 10.1504/ijcistudies.2019.102536.
- [27] K. Moharm and M. Eltahan, “The Role of Big Data in Improving E-Learning Transition,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 885, no. 1, 2020, doi: 10.1088/1757-899X/885/1/012003.
- [28] A. Nair, A. Gupta, M. Dalal, and S. Levine, “Accelerating Online Reinforcement Learning with Offline Datasets,” arXiv, 2020.
- [29] M. Birjali, A. Beni-Hssane, and M. Erritali, “A novel adaptive e-learning model based on Big Data by using competence-based knowledge and social learner activities,” *Appl. Soft Comput. J.*, vol. 69, pp. 14–32, 2018, doi: 10.1016/j.asoc.2018.04.030.
- [30] J. Chen et al., “A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 4, pp. 919–933, 2017, doi: 10.1109/TPDS.2016.2603511.
- [31] A. A. Munshi and Y. A. R. I. Mohamed, “Data Lake Lambda Architecture for Smart Grids Big Data Analytics,” *IEEE Access*, vol. 6, pp. 40463–40471, 2018, doi: 10.1109/ACCESS.2018.2858256.
- [32] R. Kumar, B. B. Parashar, S. Gupta, Y. Sharma, and N. Gupta, “Apache Hadoop , NoSQL and NewSQL Solutions of Big Data,” *Int. J. Adv. Found. Res. Sci. Eng.*, vol. 1, no. 6, pp. 28–36, 2014.
- [33] S. Konda and R. More, “Balancing & Coordination of Big Data in HDFS with Zookeeper and Flume,” pp. 869–874, 2015.
- [34] A. Gupta, H. K. Thakur, R. Shrivastava, P. Kumar, and S. Nag, “A Big Data Analysis Framework Using Apache Spark and Deep Learning,” *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2017-November, no. 1, pp. 9–16, 2017, doi: 10.1109/ICDMW.2017.9.
- [35] A. B. M. M. and S. A. Hossain, “No TitleNoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison,” *Int. J. Database Theory Appl.*, vol. 6, no. 4, 2013.
- [36] . A. K. Z., “Nosql Databases: New Millennium Database for Big Data, Big Users, Cloud Computing and Its Security Challenges,” *Int. J. Res. Eng. Technol.*, vol. 03, no. 15, pp. 403–409, 2014, doi: 10.15623/ijret.2014.0315080.
- [37] S. Venkatraman, K. F. S. Kaspi, and R. Venkatraman, “SQL Versus NoSQL Movement with Big Data Analytics,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 8, no. 12, pp. 59–66, 2016, doi: 10.5815/ijitcs.2016.12.07.
- [38] Y. MADANI, J. BENGOURRAM, M. ERRITALI, B. HSSINA, and M. Birjali, “Adaptive e-learning using Genetic Algorithm and Sentiments Analysis in a Big Data System,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 8, 2017, doi: 10.14569/ijacsa.2017.080851.
- [39] F. V. Lima, C. Costa, and M. Y. Santos, “Real-Time Big Data Warehousing,” pp. 28–57, 2019, doi: 10.4018/978-1-5225-5516-2.ch002.

- [40] K. M. M. Thein, T. T. S. Nyunt, and K. N. Aye, "Security of real-time Big Data analytics pipeline," no. September, 2017, [Online]. Available: <http://onlinerresource.ucsy.edu.mm/handle/123456789/914>.
- [41] K. Alhamazani et al., "Real-time QoS monitoring for cloud-based big data analytics applications in mobile environments," Proc. - IEEE Int. Conf. Mob. Data Manag., vol. 1, no. i, pp. 337–340, 2014, doi: 10.1109/MDM.2014.74.
- [42] R. Fu, F. Gao, R. Zeng, J. Hu, ... Y. L.-2017 C. I., and undefined 2017, "Big data and cloud computing platform for energy Internet," ieeexplore.ieee.org, Accessed: Feb. 20, 2021. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8388531/>.
- [43] A. G. Shoro and T. R. Soomro, "Big Data Analysis : Apache Spark Perspective Big Data Analysis : Ap Spark Perspective BigDataAnalysisApSparkPerspective," vol. 15, no. JANUARY, 2015.
- [44] L. Caviglione and M. Coccoli, "Smart e-learning systems with big data," Int. J. Electron. Telecommun., vol. 64, no. 4, pp. 445–450, 2018, doi: 10.24425/123544.
- [45] H. A. Abu Alsaad and R. R. K. Al Taie, "Using Big Data Technology For Prediction Of Quiz Difficulty Level In E-learning Systems," *المجلة العراقية لتكنولوجيا المعلومات*, p. 176, 2018, doi: 10.34279/0923-008-004-014.
- [46] M. Pust, "Challenges of Immediate Transition to e-Learning Annmarie Gorenc Zoran Abstract :," no. 2, pp. 84–107, 2021, doi: 10.37886/ruo.2021.030.
- [47] A. Mukhametshin, N. Asratyan, A. Safina, and A. Gaifutdinov, "Students ' attitude to e-learning," vol. 01042, pp. 1–7, 2021.
- [48] R. P. Díaz Redondo, M. Caeiro Rodríguez, J. J. López Escobar, and A. Fernández Vilas, "Integrating micro-learning content in traditional e-learning platforms," *Multimed. Tools Appl.*, vol. 80, no. 2, pp. 3121–3151, 2021, doi: 10.1007/s11042-020-09523-z.
- [49] L. Du and Q. Sun, "Application of Computer Big Data in Internet Learning," *J. Phys. Conf. Ser.*, vol. 1744, no. 4, 2021, doi: 10.1088/1742-6596/1744/4/042178.
- [50] E. DÜLGER, "Günümüz Eğitim Sistemlerinde Büyük Veri Teknolojisi: Öğrenme Analitiği," *Eur. J. Sci. Technol.*, no. November, pp. 353–361, 2020, doi: 10.31590/ejosat.824182.
- [51] T. Nur Fitria, "Implementation of Institution's E-Learning Platform in Teaching Online at ITB AAS Indonesia," *EDUTECH J. Educ. Technol.*, vol. 4, no. 3, pp. 493–503, 2021, doi: 10.29062/edu.v4i3.157.
- [52] K. Vlasenko, O. Chumak, V. Achkan, I. Lovianova, and O. Kondratyeva, "Personal e-learning environment of a mathematics teacher," *Univers. J. Educ. Res.*, vol. 8, no. 8, pp. 3527–3535, 2020, doi: 10.13189/ujer.2020.080828.
- [53] H. Y.-M. and Microsystems and undefined 2021, "Apriori algorithm optimization based on Spark platform under big data," Elsevier, Accessed: Jun.

- 23, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0141933120306785>.
- [54] “Using Spark - DNAnexus Documentation.” <https://documentation.dnanexus.com/user/spark> (accessed Jun. 23, 2021).
- [55] M. Wang, W. Wang, W. Chen, L. Z.-W. W. Web, and undefined 2021, “EEUPL: Towards effective and efficient user profile linkage across multiple social platforms,” Springer, Accessed: Jun. 23, 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s11280-021-00882-7>.
- [56] H. Kaur, “Analysis of Nosql Database State-of-The-Art Techniques and their Security Issues,” *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 467–471, 2021, doi: 10.17762/turcomat.v12i2.852.