# Big Data Reference Architecture for e-Learning Analytical Systems

K. Palanivel
Computer Centre, Pondicherry University,
Ponducherry 605014, India
*kpalanivel.cce@pondiuni.edu.in*

T. Chithralekha
Department of Computer Science, Pondicherry University,
Ponducherry 605014, India
*Tchithralekha.csc@pondiuni.edu.in*

**Abstract:** The recent advancements in technology have produced big data and become the necessity for researcher to analyze the data in order to make it meaningful. Massive amounts of data are collected across social media sites, mobile communications, business environments and institutions. In order to efficiently analyze this large quantity of raw data, the concept of big data was introduced. In this regard, big data analytic is needed in order to provide techniques to analyze the data. This new concept is expected to help education in the near future, by changing the way we approach the e-Learning process, by encouraging the interaction between learners and teachers, by allowing the fulfilment of the individual requirements and goals of learners. The learning environment generates massive knowledge by means of the various services provided in massive open online courses. Such knowledge is produced via learning actor interactions. Also, data analytics can be a valuable tool to help e-Learning organizations deliver better services to the public. It can provide important insights into consumer behavior and better predict demand for goods and services, thereby allowing for better resource management. This result motivates to put forward solutions for big data usage to the educational field. This research article unfolds a big data reference architecture for e-Learning analytical systems to make a unified analysis of the massive data generated by learning actors. This reference architecture makes the process of the massive data produced in big data e-learning system. Finally, the BiDRA for e-Learning analytical systems was evaluated based on the quality of maintainability, modularity, reusability, performance, and scalability.

***Keywords:*** *Big data, reference architecture, e-Learning, analytical systems, MOOCs.*

*****

## I. INTRODUCTION

The traditional teaching and learning has changed drastically in past few years and will continue to transform. Technology is no longer an academic subject, as it is rapidly being used as a tool to deliver new and better ways of engagement among learners and teachers. Digital books, MOOCS, personalized and Mobile learning are already making waves, The future of technology in education industry will broadly be influenced by virtual reality, wearables, location based services and sensor technologies.

The development of information and communication technology and services has great impact on the occurrence of large amounts of data, but also the need to access that data and perform analysis in real time. Currently, the field of education is flourishing rapidly throughout the world, due to the changes that have been occurring in this area with the implementation of massive open online courses (MOOCs) [1]. MOOCs generate big data in the form of activity traces.

Learning Analytics conceived as the process of visualization and interpretation of huge amounts of e-Learning data applied to education (Ferguson 2012) offers the possibility to different educational stakeholders to be better informed and thus make better and timely decisions. However, this will not be possible if the skills related to learning analytics continue to be restricted to highly technical profiles. Whereas learning analytics is an emerging issue in education, a systematic review of literature on the subject is considered useful to identify its key issues and complexities so that it can be incorporated in a critical discourse on the use of ICT education.

As with educational data mining [Prakash 2014], providing a visual representation of analytics is critical to generate actionable analyses; information is often represented as ―dashboards‖ that show data in an easily digestible form. A key application of learning analytics is monitoring and predicting students' learning performance and spotting potential issues early so that interventions can be provided to identify students at risk of failing a course or program of study. Several learning analytics models have been developed to identify student risk level in real time to increase the students' likelihood of success Higher education institutions have shown increased interest in learning analytics as they face calls for more transparency and greater scrutiny of their student recruitment and retention practices. Data mining of student behavior in online courses has revealed differences between successful and unsuccessful students in terms of such variables as level of participation in discussion boards, number of emails sent, and number of quizzes completed. Analytics based on these student behavior variables can be used in feedback loops to provide more fluid and flexible curricula and to support immediate course alterations based on analyses of real-time learning data.

### 1. Motivation

Recently, big data has become one of the buzzwords in Information Technology. Initially it was shaped by organizations which had to handle fast growth rates of data like Web data, data resulting from scientific or business simulations or other data sources. The business models are fundamentally based on indexing and using this large amount of data. The pressure to handle the growing data amount on the Web. Efforts were made to rebuild those technologies as open source software and they laid the foundation for technologies summarized today as 'big data'.

With this groundwork traditional information management companies IBM, Oracle, HP, Microsoft, SAS and SAP stepped in and invested to extend their software portfolios and build new solutions especially aimed at Big Data analysis. The big data portfolios are often blended with existing technologies. This is the case when big data gets integrated with existing data management solutions, but also for complex event processing solutions which are the basis to handle stream processing of big data.

Big data analysis [Abdelladim 2017] represents the combination of big data techniques with learning analysis. This combination enables to envision integrating learning analytics algorithms with learning systems based on big data. Learning analytics represent a set of algorithms useful for the analysis and pre-processing of the massive data originally generated in the MOOCs.

### 2. Problem Statement

Motivated by a current lack of clear guidance for approaching the field of big data and *how to design an effective and flexible big data architecture for e-learning analytical systems? T*he goal of this research article is to functionally structure this space by providing a big data reference architecture for e-Learning analytical systems. This reference architecture has the objective to give an overview of available technology and open source software within the space and to organize this technology by placing it according to the functional components in the reference architecture. The reference architecture shall also be suitable to serve as a basis for thinking and communicating about e-Learning i.e., big data application and for giving some decision guidelines for architecting them.

### 3. Solution to the Problem

The solution to the above problem statement is to design a big data reference architecture for e-Learning analytical systems that makes a unified analysis of the massive data generated by learning actors. The proposed reference architecture is called BiDRA for e-Learning analytical systems. This reference architecture makes the process of the massive data produced in an e-learning system.

The structure of the article is as follows: Theoretical background is provided in section2. Survey of related work in section3. The design and construction of the reference architecture is presented in section4. Also, the results are analyzed discussed. A conclusion is provided in section5.

## II. BACKGROUND TECHNOLOGY

This section presents the theoretical background required to write this research article. The current learning system generates knowledge in the form of educational big data. Such knowledge is the consequence of the interactions undertaken by pedagogical actors in MOOCs. They focus on the handling of the educational actors experiences in a system based on big data, availing themselves of the learning analytics techniques for the extraction of massive knowledge relating to the distance learning field.

### 1. Big Data Definition and Concepts

Big data is a term which refers to a large amount of data which cannot be processed and analyzed in a traditional manner, due to their complexity (Liu, 2013). For a big data description 3V model is often used. This model points out three characteristics: the amount of data (volume), processing speed (velocity) and a variety of data types (variety), trustworthiness of the data(veracity) and value. Big data spans the following dimensions:

- *Volume:* Big data comes in one size enormous! Enterprises are awash with data, easily amassing terabytes to petabytes of information.
- *Velocity:* Often time-sensitive, big data must be used as it is streaming into the enterprise to maximize its value to the business. Data ingestion comprises both batch and real-time loading methods.
- *Variety:* Big data extends beyond structured data, including unstructured data of all varieties, such as text, audio, video, click streams, and log files.
- *Veracity* refers to the messiness or trustworthiness of the data. With many forms of big data, quality and accuracy are less controllable but big data and analytics technology now allows us to work with these type of data. The volumes often make up for the lack of quality or accuracy.
- *Value:* It is all well and good having access to big data that can turn it into value it is useless. So the 'value' is the most important of Big Data. It is important that businesses make a business case for any attempt to collect and leverage big data. It is so easy to fall into the buzz trap and embark on big data initiatives without a clear understanding of costs and benefits.

The big data technology group includes relatively new type of lightweight, non-relational database that are often part of a big data solution called NoSQL databases. The purpose of NoSQL databases is to store unstructured and semi-structured data such as files, documents, email, and social media. Here, the data are not structured and that they are in different formats: text, audio, video, log files, etc. The big data are mainly unstructured data and it is necessary to apply new principles of data storing, which are different from the traditional, which use relational databases. The amount of data in big data is measured in terabytes, and that is the reason why special attention should be given to data storage. The important big data characteristic is fluctuations in the amount of data. An additional requirement is the data processing speed, since the data is often time-sensitive and require rapid transfer and analysis.

Big data technology is for analyzing very large collections of data sets on parallel-distributed commodity hardware. The big data technology has concentrated around a number of free and open-source software (FOSS) components.

### 1. Big Data Processing Technologies

Currently, individuals and enterprises focus on how to rapidly extract valuable information from large amounts of data. Due to the rapid rate of increase in data production, big data technologies [Ibrar et.al. 2016] have gained much attention from IT communities. In this context, state-of-the-art processing technologies based on *stream* and *batch* processing. Apache Hadoop allows to process large amounts of data. Many companies, such as SwiftKey(Amazon), Industry (Microsoft), redBus (Google), Nokia (Cloudera), Alacer (Alacer) are using Apache Hadoop technology in different fields (e.g., business and commerce). In order to process large amounts of data in real time the tools are available, namely Storm, S4, SQL Stream, Splunk, Apache Kafka, and SAP Hana.

Big data has provided several opportunities in data analytics. Thus, it has become very challenging due to the complexity and real-time processing demands of streaming data to design and implement new security mechanisms that can protect the data. This include data analytics, data quality, security and visualization. Extraordinary big data techniques are required to efficiently analyze large amounts of data within a limited time period. Currently, only a few techniques like statistical method, machine learning are applicable to be applied on analysis purposes (Lin, 2005). The state of the art big data analysis techniques are data mining, web mining, visualization methods, machine learning, optimization methods and social network analysis. The available algorithms, tools and also demonstrates suitable analysis techniques for specific big data applications.

### 2. Big Data Applications & Architecture

To get to know the origins of big data applications, it considered the application architecture, chronological development, and gradual evolution of major application models - the web, Internet, and big data applications. The study of the genesis of big data applications is beneficial to comprehending the conceptual foundation, vision, and trend of big data. Big data is a combination of different types of granular data. The big data applications that are the main sources of producing voluminous amounts of data, namely Internet of Things (IoT), self-quantified, multimedia, and social media data. The big data applications are evaluated by various metrics, namely, storage architecture, computing distribution, storage technology, analytics technology, and user experience.

Big data architecture must perform in line with the organization's supporting infrastructure. To date, all organizations do not use operational data (Khan 2014a). Big data architecture is the logical and/or physical layout / structure of how big data will stored, accessed and managed within a big data or IT environment. It logically defines how the big data solution will work, the core components (hardware, database, software, storage) used, flow of information, security, and more. Big data architecture primarily serves as the key design reference for big data infrastructures and solutions. It is created by big data designers/architects before physically implementing a big data solution. Creating big data architecture generally requires understanding the business/organization and its big data needs.

Behind big data architecture, the core idea is to document a right foundation of architecture, infrastructure and applications. It is created by big data designers/architects before physically implementing a solution. Creating big data architecture generally requires understanding the business/organization and its big data needs. Typically, big data architectures outline the hardware and software components that are necessary to fulfil big data solution. Big data architecture documents may also describe protocols for data sharing, application integrations and information security. The logical layers of the application architecture are shown in Fig.1 and presented as below:

- *Data source identification*: This layer knows where this data is sourced from. It involves identifying different source systems and categorizing them, based on their nature and type.
- *Data ingestion strategy and acquisition*: This layer process to input data into the system. The required data is extracted (extraction) from the above mentioned sources. This data is stored in the storage and then after is transformed for further processing on it.
- *Data storage*: This layer stores large amounts of data of any type and should be able to scale on need basis.
- *Data processing*: This layer has the tools that provide analysis over big data. The processing of the large amount of data has increased multifold. The processing methodology can be categorized into batch, real-time or hybrid.

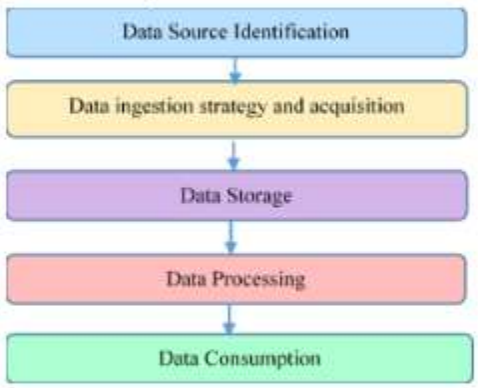


Fig. 1. Logical layers of the reference architecture

- *Data consumption/utilization*: This layer consumes output provided by the processing layer. Different users like administrator, Business users, vendor, partners etc. can consume data in different format. Output of analysis can be consumed by recommendation engine or business processes can be triggered based on the analysis.

The designer is responsible for developing, integrating, or modernizing a number of systems that all deliver similar capabilities within a domain, creating a reference architecture can provide a framework for comparing, combining, and reusing solution elements.

### 3. Big Data Analytics Architecture

In order to approach big data and analytics holistically, it is important to consider what that means. The strategy used to develop this reference architecture includes three key points to set the context:

1. Any data, any source – It views data in terms of its qualities. This includes its degree of structure, volume, method of acquisition, historical significance, quality, value, and relationship to other forms of data. These qualities will determine how it is managed, processed, used, and integrated.
2. Full range of analytics. There are many types of analysis that can be performed, by different types of users (or systems), using many different tools, and through a variety of channels. The architecture design must be universal and extensible to support a full range of analytics.
3. Integrated analytic applications. Intelligence must be integrated with the applications that knowledge workers use to perform their jobs. Likewise, applications must integrate with information and analysis components in a manner that produces consistent results. There must be consistency from one application to another, as well as consistency between applications, reports, and analysis tools.

This big data analytics architecture is designed to address key aspects of the above three points. Specifically, the architecture is organized into views that highlight universal information management, real-time analytics, and intelligent processes. They represent architecturally significant capabilities that are important to most organizations today.

Fig.2 presents concepts of the big data analytics architecture are depicted in the following description: *Data sources* are defined in two dimensions, *mobility* and *structure* of data. Mobility refers to data, which does not move. *Streaming* data refers to a data flow to be processed in real time, e.g. video stream. The structure of the data source is defined as structured, unstructured and semi structure. *Structured* data has a strict data model i.e., contents of a relational database, which is structured based on a database schema. *Unstructured* data is raw, and is not associated with a data model i.e. Web page or images. *Semi-structured* data is not raw data or strictly typed. Examples of semi-structured data include XML and JSON documents.

*Extraction* refers to input of data into the system. When *data* is *extracted*, it may be stored temporarily into a data store or transferred, and loaded into a *raw data store*. Streaming data may also be extracted, and stored temporarily. Efficiency may be improved by *compressing* extracted data before *transfer* and *load* operations. The purpose of the *raw data store* is to hold unprocessed data. Data from the *raw data store* may be *cleaned* or *combined*, and saved into a new *preparation data store*, which temporarily holds processed data. *Cleaning* and *combining* refer to quality improvement of the raw unprocessed data. Raw and prepared data may be *replicated* between data stores. Also, new *information* may be *extracted* from the *raw data store* for *deep analytics*. *Information extraction* refers to storing of raw data in a structured format. The *enterprise data store* is used for holding of cleaned and processed data. The *sand-box store* is used for containing data for experimental purposes of data analysis.

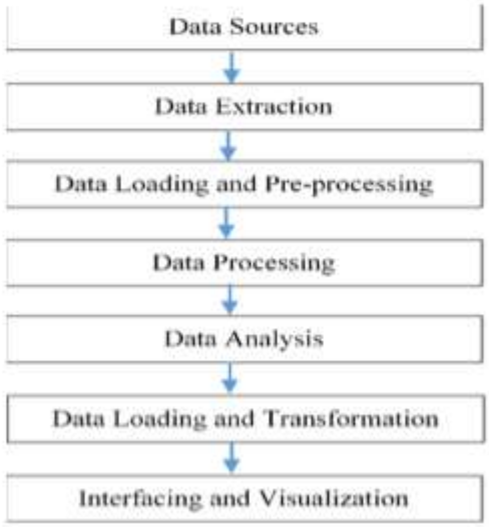


Fig 2. Concepts of big data analytics architecture

*Deep analytics* refers to execution of batch-processing jobs for data. Results of the analysis may be stored back into the original data stores, into a separate *analysis results store* or into a *publish & subscribe* store. *Publish & subscribe* store enables storage and retrieval of analysis results indirectly between subscribers and publishers in the system. Stream processing refers to processing of extracted streaming data, which may be saved temporarily before analysis. *Stream analysis* refers to analysis of *streaming* data. Results of the data analysis may also be *transformed* into a *serving data store*, which serve *interfacing and visualization applications*.

Analyzed data may be visualized in several ways. *Dashboard application* refers to a simple user interface, where typically key information is visualized without user control. *Visualization application* provides detailed visualization and control functions, and is typically realized with a Business Intelligence tool in the enterprise domain. *End user application* has a limited set of control functions, and could be realized as a mobile application for end users.

### *2. Massive Data in the Education Field*

The field of education witnessed evolution at all levels: volume, storage location, the nature and type of massive data. This progress directly influences the pool of massive knowledge produced by learning actors in MOOCs, which explains the existence of many models and approaches in the literature that deals with massive data in a novel architecture based on massive data [Gibson 2017].

It is proposed state of art drawing on massive data produced by learning systems, notably the massive data of pedagogical field's which constitute the key elements of the big data reference model and architecture. It is available some methods to the creation of communication interfaces between a learning system and the new big data architecture. Also, available some techniques and methods of learning analytics for such massive data. Hence, it is proposed an architecture to give a large coverage of the massive data produced by learning actor interactions. The reference architecture aims at making use of its services in order to put at the disposal of educational actors the optimal methods for exploiting the massive data generated by actor interactions.

On the other hand, the author in [Zhou 2017] studied big data implementation in open education. This study arouses a considerable interest among researchers due to the changes in the way data are processed given the use of novel technology. The author also laid out the challenges of the implementation of learning systems that produce massive data for open education throughout the world, focusing on the contributions of MOOCs implementation based on big data. We find in learning system such produce a massive data, three dimensions studied by researchers [Bechini 2016]. These dimensions are the key elements that have made researchers change their point of view and shift away from the traditional method to No-SQL, being mindful that the bulk of the massive data generated today is characteristic of an unstructured aspect. Therefore, we are forced to shift to the No-SQL method, which represents the mode of pre-processing the massive data generated by learning actors.

Learning analytics technique integrated into education field and involved several methods for the analysis and pre-processing of the massive data initially produced by actors in an e-learning system. These methods are useful for the creation of relevant knowledge owing to their bearing upon the initial massive data. The leaning analysis methods uses in open education.

The big data based systems revealed several methods of massive data analysis in the field of education. The field of education has taken on an approach called Educational Data mining, which represents a range of methods and techniques of pre-treatment for educational big data. It is noted the existence of methods that deal with unstructured data such as videos, audios as well. Worth noting, massive unstructured data represent a significant amount of data in learning systems that generate massive data. Some of the methods are discussed in [Olson 2017, Sliman 2016] that tackle unstructured data.

The Big Data phenomenon has obviously impacted the learning environment and the distance training. It has facilitated the creation of a mixture full of learning opportunities and allows the learners to improve their training practices and experiment with open educational resources, especially the massive and open online courses (MOOC) and the distance learning via e-learning platforms. With the emerging technologies in the Web, access to information has become easier with the ability to work and learn effectively, regardless of educational structures that have been the norm for centuries [Fournier 2011]. This put in place new structures and new working environments, enabling independent learning, but that does not mean that everyone is able to do it effectively. Two major factors are the basis for the study of learning in a massive and open online environment: learner's autonomy and the quality of the massive submitted information.

### 1. E-Learning

Electronic learning or e-Learning means to learn and educate people through electronic environment. This field becomes very vast and makes the learner to explore various things, due to the fact that data on the web. There are types of e-Learning, which are formed by the existence of technology, and are known as mobile learning, distant learning, education learning, etc.

Mobile learning is the new mode of communication and technology which has brought a new change to the abilities of the e-Learning. This has made things faster. This learning is helping students a lot but has a major hurdle during downloading of files.

Distant Learning is a technique which is being used by educational institutions. Various tools like, digital dashboards, video conferences and all other media tools, which can be used by e-Learning, are used in this type of learning. Many students have been benefitted from this as this also provides flexibility in timing for students and teachers. Many educational institutions are running on this concept and also own their own channels where they keep on repeating lectures, which helps slow movers to be more active and learn more as well. It has a direct relation with Big Data.

In educational sector, educational institutions, try to mine out the data which is categorized or can be viewed from different angles. Through big data mining, educational institutes are able to manage the user data as well as the better record keeping of the students and also letting the different departments of an organization to study how each student can benefit to any entity. The evaluations may include assessing student's knowledge, the types of students, the genuine learning problems, and making computational models and resources.

### 2. E-Learning Data

It has seen during the last decades, rapid changes in many aspects of human life. Most of these changes have been produced by a direct influence of the development and massification of Information and Communication Technologies (ICT). One of the most highlighted trends in this matter is the advent of a very interesting concept: "Internet of things", which will be part of education in the coming years and will produce huge amounts of educational data. In this context, learning activities and objects allow users to be interconnected by a constant flow of information in large quantities.

According to the above, educational institutions would be part of that flow both as producers and consumers of information. In that sense, it is not a matter of only providing educational institutions with strong Wi-Fi networks and computers or mobile devices through which students and teachers can connect to the Internet, but also of achieving the School, as a whole, produces information potentially useful for educational stakeholders (Kortuem 2010).

Because things will be interconnected, school tables, chairs, walls, windows and corridors as well as Learning

Management Systems and other support tools should provide educationally useful data that reveal what happens inside and outside the classrooms and thereby allow to make better-informed decisions and actions. Moreover, if students develop the skills related with Learning Analytics it would be possible that they, wherever they are (inside or outside the school) may be connected to the library, to learning support systems, to academic networks, to their teachers and their peers, and mainly to information generated by themselves, enabling them to improve their learning experiences.

Big data with e-Learning supports with various benefits which an individual may use to learn. To learn not only means to educate himself, but to also learn clusters. Predictive modeling is used by all major educational organizations to understand how they can enhance teaching and learning. Big data when related to e-learning, the learners share their knowledge on websites and various discussion forums. This can also be regarded as memory based learning, although in this it can save the document on any handheld device.

The results of the data analysis for the online search scenario will allow to delineate the context of the future system and to better understand the design of a methodology based on the tool that will integrate the mixing of structured and unstructured data in one layer to facilitate access in addition to an optimal search relevance with adequate and consistent results according to the needs of the learner.

## 3. Big Data e-Learning System

It is important to identify a sophisticated strategy to combine different types of data in a way that they provide the best result to the learner, the user of the e-learning platform. In this context, it is proposed to design Big Data e-Learning system (BiDeL) that integrates the mix of structured and unstructured data in one data layer to facilitate access in addition to an optimal relevance of search with adequate and consistent results according to the expectations of the learner. The adopted method will consist initially in a quantitative and a qualitative study of the variety of data and their typology, followed by a detailed analysis of the structure and harmonization of the data to finally find a fictional model for treatment of such data. Fig. 3 shows the functional architecture of the BiDeL.

The data server will handle the capture and collection of massive data and then, the application server can carry out the treatment, structuring, formatting and classification required for these raw data and thus make them consumable by the analytical systems at the end and that will be accessible via the web browser by the user of the BiDeL.

The BiDeL performs the capture of all types of data (text, image, video, audio, etc.) related to the subject of the theme and group them in its raw data store or mart. It then includes data of any type, such as posts, pictures, videos, audio tracks, etc. The BiDeL thereafter proceed with the treatment of the raw data to make it consumable by the user, via classification, structuring and formatting of these data at the presentation layer, thus allowing an organized display and ergonomics at the user interface of the system. The data related to the video type classified in the same category even though they come from different sources like YouTube, Vimeo, Dailymotion, etc., also, for the type of post data like Facebook, Twitter, Google +, etc. audio data type and finally the type of the image data.

## 4. Learning Analytical System

Learning analytics (LA) is the process of discovering learning patterns among learners. The analytic usually involves variety of techniques including machine learning techniques to mine the data. It involves the process of select and capture data, aggregate and report, predict, use, refine and share [Lias 2011]. In the early phase of learning analytic, select and capture relevant data from e-learning database need to be done [Larusson 2014]. Next, data pre-processing and processing is conducted that include data aggregate and producing report. The tools or methods need to be selected wisely to fit the data type. From the process, come out the application of several techniques on the result which is useful for prediction.

The analytic result is shared with learning centre for strategic decision making and to inform students of their progress. Those processes are enhanced with big data engine to process large data efficiently. From other perspective, the process involved are to make the raw data to reach wisdom level as depicted in knowledge continuum.

Most of the data that used in learning analytics applications comes from Learning Management System either from the login information, rates of participation in specific activities, assessment results or grades and time students' interaction with online resources. The knowledge gained from e-Learning data can help in improving the adaptive learning environment [Tirthali 2014]. Learning analytics can assist by providing a more personalized learning experience through the use of historical and current data to respond to students' needs.

## *5. New e-Learning Architecture*

Many campus IT organizations role is simply to administer the Learning Management System (LMS) and maintain the status quo. The status quo represents a barrier to connected learning and often forces those faculty who want to innovate to go around the central IT organization. The current institutional IT infrastructure in higher education institutions explore new applications and services, let alone encourage personalization.

With the advent of connected learning, the rate of change in the appearance and therefore evaluation for potential use of new technologies is already far exceeding that in the traditional e-learning infrastructure. It is more important for educational institution to come up with scalable approaches for academic technology support as connected learning accelerates.

Hence, it also needs an open architecture that can provide a much more efficient foundation for adding and evolving innovative technologies. It encourage open source to lead the way in promoting interoperability and connected learning. Specifically, this new open architecture for

 ISSN: 2321-8169
55 – 67

connected learning needs to support and enable unprecedented agility, flexibility, and personalization.

The new architecture [Rob 2013] needs to enable an instructional environment in which technology "gets out of the way" and becomes highly supportive of teachers' and students' needs. The architecture based on open standards can reduce the support costs and provide consistent mechanisms for a variety of tools. The mobile app movement and the mash-up capabilities of the web exemplify this type of "innovation environment." The following are several specific examples of what the open standards and services must enable to make this new architecture for learning a reality:

- Digital content and applications must be easily, quickly, and seamlessly integrated into any platform that supports a set of vendor-neutral open standards and, importantly, are not trapped inside a single platform.
- User, course, and context information must be synchronized among selected applications so that making set-up and use of new software much easier for all concerned.
- Data that describes usage, activities, and outcomes must flow from learning content apps to the enterprise system of record, learning platforms, and analytics platforms.
- Systems, services, and tools must be virtualized and must increasingly move toward the elastic computing model that enables sharing scenarios across systems or other federations of users.

The new IT architecture also requires that tools be able to "work together" to provide better information (analytics) to both faculty and students on their progress and to administrators on their usage.

## 3. *Reference Architecture*

Software architecture is a flexible and hierarchical reusable architecture based on domain-specific and software product lines. Under this architecture, It we encapsulate business logic into software component firstly, and then weave the unchanged or uneasy changing parts of the e-learning domain logic into respective domain frameworks , and margin the easy changing parts as pluggable user interfaces in order to insert or delete or replace the components according different requirements in the future.

Before defining the term reference architecture, it must first establish an understanding of the term 'architecture'. IEEE Standard defined as *Architecture is the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution* [IEEE 2000u]. It is also defined as *the software architecture of a system is the set of structures needed to reason about the system, which comprise software elements, relations among them, and properties of both* [Bass 2012].

The above definitions describe architecture to be about structure and that this structure is formed by components or elements and the relations or connectors between them. An architecture is abstract in terms of the system it describes, but it is concrete in the sense of it describing a concrete system. It is designed for a specific problem context and describes system components, their interaction, functionality and properties with concrete business goals and stakeholder requirements in mind.

A reference architecture abstracts away from a concrete system, describes a class of systems and can be used to design concrete architectures within this class. Put differently a reference architecture is an 'abstraction of concrete software architectures in a certain domain' and shows the essence of system architectures within this domain [Samuil 2012]. A reference architecture shows which functionality is generally needed in a certain domain or the solve a certain class of problems, how this functionality is divided and how information flows between the pieces (called the reference model). It then maps this functionality onto software elements and the data flows between them [Bass 2003]. Within this approach reference architectures incorporate knowledge about a certain domain, requirements, necessary functionalities and their interaction for that domain together with architectural knowledge how to design software systems, their structures, components and internal as well as external interactions for this domain which fulfil the requirements and provide the functionalities [Samuil 2012].

The goal of bundling this kind of knowledge into a reference architecture is to facilitate and guide future design of concrete system architectures in the respective domain. As a reference architecture is abstract and designed with generality in mind, it is applicable in different contexts, where the concrete requirements of each context guide the adoption into a concrete architecture [Samuil 2012]. The level of abstraction can however differ between reference architectures and with it the concreteness of guidance a reference architecture can offer.

### 1. *Building Reference Architecture*

Building the reference architecture will consist of the following four steps.

i. Step 1: To conduct a qualitative literature study to define and describe the space of 'big data' and related work (sections 3) and to gather typical requirements for analytical big data e-Learning.
ii. Step 2: To design the reference architecture. It will develop and describe a methodology from literature about designing software architectures, especially reference architectures. Based on the gathered requirements, the described methodology and design principles for big data e-Learning systems.
iii. Step 3: To gather an overview of existing technologies available for handling and processing large volumes of heterogeneous data in reasonable time.
iv. Step 4: To verify and refine the resulting reference architecture by applying it to case studies and mapping it against existing big data architectures from academic and industrial literature..

While developing a reference architecture, it is important to keep some of the points in mind. The result should be relevant to a specific domain i.e. big data e-Learning, that is incorporate domain knowledge and fulfil domain requirements, while still being general enough to be applicable in different contexts.

Following a design method for reference architectures helps accomplishing that and the basis for the reference architecture to be well-grounded and valid as well as to provide rigour and relevance. However, the research about reference architectures and respective methodology is significantly more rare than that about concrete architectures. Therefore it is decided to loosely follow the proposed development process, which consists of the following 6 steps.

i. Decide on the reference architecture type
ii. Select the design strategy
iii. Empirical acquisition of data
iv. Construction of the reference architecture
v. Enabling reference architecture with variability
vi. Evaluation of the reference architecture

### *2. The Big Data Reference Architecture*

The big data reference architecture (BiDRA) was created after investigating the literature, the experts, using grounded theory to perform quantitative data analysis, and determining the representations of the various elements. The reference architecture is a *guideline*, not a *prescription*. Each element in the model is optional in the solution architecture that is ultimately created. The reference architecture [Geerdink 201x] consists of categories - *components & interfaces, architectural patterns, architecture principles and architectural best practices*, according to the elements that were defined in the "what" question (dimension D1) of Angelov's model.

*Components & Interfaces* category contains all components (business, software, and hardware) that are part of the reference architecture, as well as the interfaces between them. The components of the reference architecture are divided into the three common layers business, application, and technology. The conceptual elements of the reference architecture are all structures. The structural character of the model was chosen on purpose, to give architects a clear guidance while still presenting abstract components. It allows for an easy translation of the abstract reference architecture to real physical components in the big data solutions that are implementations of the model.

Architectural patterns are important to the working of the reference architecture. They define the way the reference architecture is constructed and the way the architects should work with the reference architecture. The BiDRA contains two patterns and they are Pipes and Filters and Layers. The BiDRA contains two architectural principles loose coupling and interoperability. The principles were derived from analysis of the literature review.

The BiDRA contains "*use free and open-source software (FOSS)"* best practice. Most architectures that are addressed in the literature are based on FOSS components. The FOSS products are simply better than the commercial ones in terms of usability, modifiability, performance, reliability, and costs. Vendor lock-in is avoided, and the architecture principles loose coupling and interoperability can be applied more easily with FOSS. Preferring FOSS components over proprietary software can have some impact on organization, especially if this best practice is not implemented yet. With FOSS, organizations cannot rely on support contracts and have to build up knowledge of the components in-house.

## III. LITERATURE REVIEW AND RELATED WORKS

This section presents various big data architecture studied and reviewed with respect to e-Learning systems.

Big Data are becoming a new technology focus both in science and in industry and motivate technology shift to data centric architecture and operational models. There is a vital need to define the basic information/semantic models, architecture components and operational models that together comprise a so-called big data ecosystem. They are explained as below:

- Extended relational reference architecture is more about relational reference architecture but components like data warehouse, data mart, and OLAP cube components cannot handle big data challenges.
- Non-relational reference architecture is more about Non-relational reference architecture but still components like NoSQL databases, distributed file systems, Map Reduce and search engine cannot handle big data challenges completely.
- Data discovery architecture is more about Hadoop based big data architecture which can be handle few core components of big data challenges but not all (like Search Engine, etc.)
- Data analytics architecture adopted by Facebook. Here Facebook collects data from two different sources - user data and web servers generate event based log data. Data from the web servers is collected and then executed. The data from other source is dumped, compressed and transferred into the production. Facebook uses different clusters for data analysis. Including business intelligence tools for dimensional analysis.
- Data analytics architecture adopted by LinkedIn. Here, the data is collected from two sources: database snapshots and activity data from users of LinkedIn. The activity data comprises streaming events, which is collected based on usage of LinkedIn's services. Results of the analysis in the production environment are transferred into an offline debugging database or to an online database.
- Data analytics architecture adopted by Twitter: In the Twitter's infrastructure for real-time services, a brokers all requests coming to Twitter. Requests include searching for tweets or user accounts service. Tweets are input service to an ingestion pipeline for tokenization and annotation. Subsequently, the processed tweets enter to servers for filtering, personalization, and inverted indexing. The results of analysis are persisted and serves users of Twitter.

The existing literature was searched for big data architectures. Both scientific and non-scientific sources were used to get an overview of work that has been done on architectures considering big data and predictive analytics. An evaluation of the literature identified the usable elements for the big data reference architectures.

A framework for design and analysis of software reference architectures [Angelov 2012] contains a multi-dimensional classification space, and five types of reference architectures. It is claimed that architecture design based on the classified reference architectures should lead to better success. Also, empirically-grounded design of software reference architectures [Galster 2011] approach is based on expected empirical material gathered with interviews, questionnaires, and document analysis. The procedure is a step-wise process, which consists of deciding a type for the reference architecture, selection of design strategy, empirical acquisition of data, construction of reference architecture, enabling of variability, and evaluation.

Service-oriented reference architecture has been defined for enterprise domain [Zimmer 2013]. However, in the big data context, there exist only few architecture proposals. Demchenko et al. presented a big data architecture framework, which consists of high-level description of big data lifecycle and infrastructure [Demchenko 2013].

Generalized software architecture was proposed for predictive analytics of historical and real-time temporally structured data [Westerlund].

Meier conducted design of reference architecture covering functionality in realized big data use cases [Meier 2013]. The author initially defined requirements for reference architecture, conducted architecture design, and validated the presented architecture against published implementation architectures of Facebook, LinkedIn, and Oracle. The design was conducted in the empirically-grounded design framework for reference architectures [Angelov 2012, Galster 2011].

e-Learning data [Andres 2017] sets grow rapidly - in part because they are increasingly gathered by low cost and numerous information-sensing mobile devices, aerial devices (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks. Big data spans the volume, velocity and variety dimensions.

Martin 2015 presented that *technologies* can be positioned in the areas of this reference architecture where most effective. Also *data modelling* techniques can be positioned in the appropriate areas. Zooming into the core of this Architecture on data modelling reveals interesting complementarity and similarity.

In total, ten literature sources were investigated. Most articles mention several aspects of big data models and architecture concerned with business processes, infrastructure, patterns, principles, and best practices. Big data architectures in literature points contain hardware and software components, architecture principles, and best practices. The important observations were made from the literature review.

i. The literature review clearly defines the core of a big data architecture that all sources contain a parallel batch-processing engine (e.g. Hadoop MapReduce), a distributed file system (e.g. HDFS), and a NoSQL database (e.g. HBase) components.

ii. The big data architectures have many components including data sources, data mining processes, coordination and configuration engines, databases, monitoring, etc. All these components play a role in the literature, some more than others. Several other components are typical for a big data architecture, simply because they surface often in the literature.

iii. Several architecture principles exist in the articles and websites on big data. Loose coupling, cloud computing, and scalability are popular principles in literature. There are several principles about whom the literature sources disagree.

iv. The best practice indicates that a big data architecture is like a pipeline through which data flows. Several literature sources point to another best practice that is "data exploration and discovery" method. This best practice is actually a type of big data analytics where the data is not retrieved or imported, but remains at its source and is approachable directly for analytical purposes.

v. There seems to be more consensus about the hardware and software component than about the principles and best practices. This indicates that people agree about big data technology, but have yet to reach a common understanding about the approaches and patterns in big data architecture.

The components, architecture principles, and best practices found in literature were put forward in the expert interviews to confirm their place in the final model. In this way, the literature review served to create a provisional model of the final big data reference architecture.

## IV. BIG DATA REFERENCE ARCHITECTURE

This section presents the proposed reference architecture called "Big data reference architecture (BiDRA) for e-Learning analytical systems". The BiDRA for e-Learning analytical systems might look highly complex. The more important it is not to face the end user with this high complexity.

### 1. Requirements

A BiDRA for e-Learning analytical systems supports both functional and non-functional requirements. The functional key requirements are given below:

- Various e-learning, big data and analytical application types, including batch and real-time analytics
- Industry-standard interfaces, so that existing e-learning, big data and analytical application applications can work for real-time streaming and processing of data
- Various data types and databases and interfaces
- Large volumes of e-Learning data in different formats

Customers require their big data solution to deliver business value without significant overhead. The non-functional requirements are simplified, reliable, fast and scalable, and secure.

## 2. Functional Components

The main functional components of the BiDRA for e-Learning system represent the different technical roles within a big data e-Learning system. The functional components are system orchestrator, data provider, big data application provider, big data framework provider and data consumer.

The system orchestrator includes defining and integrating the required data application activities into an operational vertical system. Typically, the system orchestrator involves a collection of more specific roles, performed by one or more actors, which manage and orchestrate the operation of the big data e-Learning system.

The data provider role introduces new data or information feeds into the big data e-Learning system for discovery, access, and transformation by the big data e-Learning system. The data provider actors can be anything from a sensor, to a human inputting data manually, to another big data system. It is the ability to import and use data from a variety of data sources. Data sources can be internal or public records, tapes, images, audio, videos, sensor data, web logs, system and audit logs, cookies, and other sources.

The big data application provider role executes a specific set of operations along the data life cycle to meet the requirements established by the system orchestrator, as well as meeting security and privacy requirements. The Big data application provider is the architecture component that encapsulates the business logic and functionality to be executed by the architecture. The big data application provider activities include collection, preparation, analytics, visualization and access.

The big data framework provider consists of one or more hierarchically organized instances of the components in the BiDRA for e-learning analytical system as shown in Fig. 2. Most big data implementations are hybrids that combine multiple technology approaches in order to provide flexibility or meet the complete range of requirements, which are driven from the big data application provider.

The data consumer role can be an actual end user or another system. In many ways, this role is the mirror image of the Data Provider, with the entire Big Data framework appearing like a data provider to the data consumer. The data consumer uses the interfaces or services provided by the big data application provider to get access to the information of interest. These interfaces can include data reporting, data retrieval, and data rendering.

## 3. Features of BiDRALS

BiDRA for e-Learning analytical systems has features of flexible technology integration and efficient approach to analysis. The BiDRA for e-Learning analytical systems BiDRA verified the combination of products by vendors and open source software and thus, it can support selecting the combination of products by different vendors. The combination of products with high frequency of use is provided as a set; besides, it is possible to select products fitting with existing IT systems. The e-Learning data analysis work tends to be individualistic and a wide range of industries and business seek data analysis. The data analysis and systematizes the analytics model analysis purpose by categorizing and summarizing purposes, procedures, and techniques of data analysis with the template of those scenarios. Thus, it can get results about requests from any industries.

## 4. BiDRA for e-Learning Analytical Systems - Logical View

A big data application architecture should be able to consume in a fast and inexpensive manner. Fig. 4 illustrates the logical view of major components of the big data analytics architecture. It can choose either open source framework or packaged licensed products to take full advantages of the various components in the proposed architecture stack. This logical view contains e-Learning data source, data management, services, analytical applications and visualization systems.

The *e-Learning data source* (infrastructure) includes the hardware and platforms on which the big data analytics components run. This includes infrastructure to support traditional databases, specialized big data management systems, and infrastructure that has been optimized for analytics.

The *e-Learning data store and management* includes all information management components, i.e. data stores, as well as components to capture, move, integrate, process, and virtualize data. At the bottom are data stores that have been commissioned for specific purposes, such as individual operational data stores, content management systems, etc. The *e-Learning services* includes components that provide or perform commonly used services. Presentation services and information services are types of services. They can be defined, cataloged, used, and shared across solutions. Business activity monitoring, business rules, and event handling provide common services for the processing layer(s) above.

 ISSN: 2321-8169
55 – 67

Fig 4: Logical view of big data e-Learning analytical system

The *e-Learning analytical applications* (processing) represents components that perform higher level processing activities. For the purpose of big data analytics, this layer calls out several types of applications that support analytical, intelligence gathering, and performance management processes.

The *e-Learning analytical interaction systems* (interaction) is comprised of components used to support interaction with end users. Common artifacts for this layer include dashboards, reports, charts, graphs, and spreadsheets. In addition, this layer includes the tools used by analysts to perform analysis and discovery activities. The results of analysis can be delivered via many different channels. The architecture calls out common IP network based channels such as desktops and laptops, common mobile network channels such as mobile phones and tablets, and other channels such as email, SMS, and hardcopy.

The architecture is supported by a number of components that affect all layers of the architecture. These include information and analysis modeling, monitoring, management, security, and governance.

## 5. Big Data Reference Model for e-Learning Analytical Systems

The big data analytics reference model provides high level view of the functional components in a large scale analytics platform within a big data solution ecosystem. In order to better understand the several roles and responsibilities of the several functional components of the platform it isolate the layers of responsibilities. This way we can also achieve a common understanding of the big data analytics domain.

The Fig. 5 provides the high level overview of the big data reference model for e-Learning analytical system and specific functional layers of the platform. All layers provide external/internal APIs that serve both other layer functions and external third party applications, in respective levels of data relevance and data density.

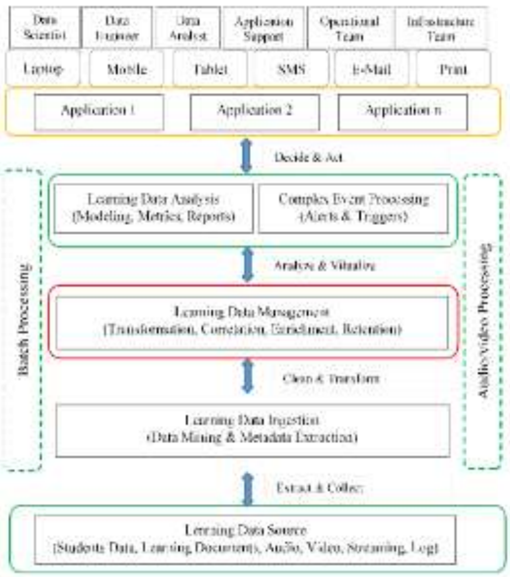


Fig. 5: Big data reference model for e-Learning analytical systems

## 6. BiDRA for e-Learning Systems

The big data reference architecture (BiDRA) defines a multi-tier architecture template that can be used to describe many types of technology solutions. The conceptual view for the reference architecture, shown in Fig. 6, uses capabilities to provide a high-level description of the big data and analytics solution for e-Learning systems. It illustrates the proposed reference architecture called BiDRA for e-Learning Systems. This reference architecture layered architecture which contains e-Learning data source, data management, services, analytical applications and visualization layers.

The *Learning data source layer* contains functions that gather various data generated and stored in various data sources such as web media, sensors and databases, changing them into a form that can be easily analyzed. It implements integration of different types of data by ETL, and deals with the improved reliability, availability, and accessibility by messaging/replication and shared information between different resources such as software and hardware in this layer.

The *Learning data* store layer contains database functions for flexibly storing and processing massive amounts of data. For example, distributed data store which realizes the processing of massive amounts of data, an in-memory database which realizes processing at high-speed, and NoSQL which realizes high scalability and flexibility, are contained in this layer.

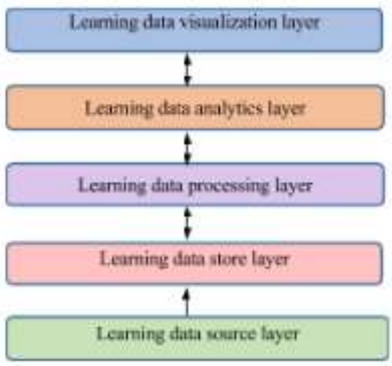


Fig.6 BiDRA for e-Learning analytical systems

The *Learning data processing layer* contains a function for high-speed processing of massive amounts of data collected and a pre-processing function for analysis. For example, the core functions of big data solution such as distributed parallel processing which realizes massive data processing technology and complex event processing technology that realizes processing at high-speed, are in this layer.

The *Learning data analytics layer* contains functions for analyzing stored and collected data such as correlation analysis, natural language analysis and machine learning. For example, text mining and data mining are contained in this layer. Moreover, this layer optimizes various analytical methods and utilizes them in multiple ways.

The *Learning data visualization layer* contains functions for decision support with the results of analysis. Data visualization, OLAP, and business process management are contained in this layer.

The proposed architecture shown in Fig.6 illustrates how major components of the big data and analytics architecture fit within the universal BiDRA high-level structure.

### 7. Analysis of BiDRA for e-Learning Analytical Systems

A detailed and high level view of BiDRA for e-Leering systems was designed inductively based on published material of the big data use cases. The BiDRA for e-Learning systems is comprised of semi-detailed functional components and data stores, and data flows between them. The presented design contained description of the architectural elements. The reference architecture was created inductively based on the published big data functional architectures. Thus, only the observed functional components and data stores are present in the architecture. If the model is extended in the future based on other big data use cases, the model may need to be updated with new functional components or data stores.

Big data integration technologies fall into data, bulk data movement and data replication categories. Bulk data movement includes technologies such as ETL to extract data from one or more data sources, transform the data, and load the data into a target database. Data replication technologies like change data capture can capture big data, such as utility smart meter readings, in near real time with minimal impact to system performance. Data virtualization (data federation) allows an application to issue SQL queries against a virtual view of data in heterogeneous sources such as in relational databases, XML documents, and on the mainframe. Big data quality [Sunilsoares 2012] includes the methods to measure and improve the quality and integrity of an e-Learning's data. It address data quality natively within Hadoop. It leverage unstructured content to improve the quality of sparse data.

## V. CONCLUSION

This paper described a research article for a reference architecture of big data solutions. The reference architecture was designed using qualitative data analysis and grounded theory, and evaluated using a questionnaire that investigated several quality criteria. The BiDRA for e-Learning analytical systems was justified and evaluated based on the quality of maintainability, modularity, reusability, performance, and scalability. Since the BiDRA for e-Learning systems is a 'good' reference architecture for its purpose.

An obvious future step is to create a solution architecture with guidance of the reference architecture. By conducting one or more case studies with the model, its practical use and quality can be investigated. An example of a case study is to create a solution architecture for a Digital Library with the goal of combining open data sources (e.g. learning data) with enterprise data (e.g. library data) to produce a forecast (e.g. buying/subscribing e-books, e-journals, on-line courses, etc.). Another example is to create an architecture in an educational organization that helps to predict the amount of students in the upcoming years by combining course data, internet data, traffic data, weather data, and others.

Finally, a future research project could investigate the way in which organizations move from traditional business intelligence systems to modern big data analytics. Also, the student behavior patterns identified by the analytic component will be integrated with real-time data, such as the student's performance and interactions with LMS. A future research project could repeat the evaluation with a larger sample, to gain more accurate insight into the quality of the reference architecture.

## REFERENCES

[1] Abdelladim 2017, Abdelladim Hadioui, Nour-eddine El Faddouli, Yassine Benjelloun Touimi, and Samir Bennani, Machine Learning Based On Big Data, Extraction of Massive Educational Knowledge, IJET – Vol. 12, No. 11, 2017.

[2] Andres 2017, Andres Chiappe, Lilian Patricia Rodríguez, Learning Analytics in 21st century education: a review, Ensaio: aval. pol. públ. Educ., Rio de Janeiro, v.25, n. 97, p. 971-991, out/dez. 2017

[3] Angelov 2012, Angelov A, P. Grefen, D. Greefhorst, A Framework for Analysis and Design of Software Reference Architectures, Inf. Softw. Technol. 54 (2012) 417–431.

[4] Bass 2003, Len Bass, Paul Clements, and Rick Kazman. *Software Architecture in Practice*. SEI Series in Software Engineering. Addison-Wesley, 2nd edition edition, 2003.

[5] Bass 2012, Len Bass, Paul Clements, and Rick Kazman. *Software Architecture in Practice*. SEI Series in Software Engineering. Addison-Wesley, 3rd edition, 2012.

[6] Bechini 2016, Bechini, A., Marcelloni, F., Segatori, A., 2016. A MapReduce solution for associative classification of big data. Information Sciences 332, 33–55.

[7] Dan 2017, Dan Kangas, Weixu Yang, Ajay Dholakia, James Sun, Deployment considerations for high-performance, cost-effective and scalable solutions, Lenovo Big Data Reference Architecture for the MapR Converged Data Platform, 30 March 2017.

[8] Demchenko 2013, [9] Y. Demchenko, C. Ngo, P. Membrey, Architecture framework and components for the Big Data Ecosystem, SNE Technical Report, University of Amsterdam, September 12, 2013.

[9] Ferguson 2012, Ferguson R. Learning analytics: drivers, developments and challenges, *International Journal of Technology Enhanced Learning*, v. 4, n. 5–6, p. 304-17, 2012.

[10] Fournier 2011, Fournier, H., & Kop, R. Factors affecting the design and development of a Personal Learning Environment: Research on super-users, in the International Journal of Virtual and Personal Learning Environments, 2 (4), 12-22., 2011.

[11] Fujitsu 2016, White paper, Solution Approaches for Big Data, FUJITSU Technology Solutions GmbH, Germany, 2016.

[12] Galster 2011, M. Galster, P. Avgeriou, Empirically-grounded reference architectures: a proposal, in: Joint

ACM SIGSOFT Conference on Quality of Software Architectures and ACM SIGSOFT Conference on Quality of Software Architectures and ACM SIGSOFT Symposium on Architecting Critical Systems, Boulder, Colorado, USA, June 20–24, 2011.

[13] Geerdink 201x, Bas Geerdink, A Reference Architecture for Big Data Solutions, Introducing a model to perform predictive analytics using big data technology, Faculty of Science and Engineering Utrecht University of Applied Sciences, Utrecht, the Netherlands

[14] Geerdink 201x, Bas Geerdink, A Reference Architecture for Big Data Solutions, Introducing a model to perform predictive analytics using big data technology, Faculty of Science and Engineering Utrecht University of Applied Sciences, Utrecht, the Netherlands

[15] Germin 2014, Germin Ong, A Smarter Library: Using Data Analytics to Improve Resource Management and Services at NLB, Civil Service College, Singapore, 2014.

[16] Gibson 2017, Gibson D.C., Ifenthaler D. (2017) Preparing the Next Generation of Education Researchers for Big Data in Higher Education. In: Kei Daniel B. (eds) Big Data and Learning Analytics in Higher Education. Springer, Cham.

[17] HPE 2016, HP Enterprises, Technical white paper HPE Reference Architecture for Hadoop on HPE Elastic Platform for Big Data Analytics (EPA) Deploying Hortonworks HDP with HPE Workload and Density Optimized (WDO) system, 2016.

[18] Ibrar et.al. 2016, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Abdullah Gani, Salimah Mokhtar, Ejaz Ahmed, Nor Badrul Anuar, Athanasios V. Vasilakos, Big data: From beginning to future, Elsevier International Journal of Information Management 36 (2016) 1231–1247, 2016

[19] IEEE 2000u, IEEE Recommended Practice for Architectural Description of Software-Intensive Systems. Standard Specification IEEE 1471-2000, IEEE Computer Society, 2000u.

[20] Khan 2014, Nawsher Khan, et.al., Big Data: Survey, Technologies, Opportunities, and Challenges, The Scientific World Journal Volume 2014 (2014).

[21] Larusson 2014, Larusson, J. A. and White, B. (2014). *Learning Analytics: from research to practice*. Springer.

[22] Lias 2011, Lias, T. E., & Elias, T. (2011). Learning Analytics: The Definitions, the Processes, and the Potential.

[23] Martin 2015, Martijn Imrich, A Big Data Reference Architecture, the Data Lab practice.

[24] Meier 2013, M. Meier, Towards a big data reference architecture, Master's thesis, Eindhoven University of Technology, October 2013.

[25] NTT 2015, NTT DATA Big Data Reference Architecture Ver. 1.0, NTT DATA Corporation, 2015.

[26] Olson, D.L., Wu, D., (2017). Predictive Models and Big Data, in: Predictive Data Mining Models. Springer, pp. 95–97.

[27] Oracle 2016, An Enterprise Architect's Guide to Big Data, Reference Architecture Overview, Oracle Enterprise Architecture White Pape, March 2016.

[28] Rob 2013, Abel, Malcolm Brown and John Suess, A New Architecture for Learning, IMS Global Learning Consortium University of Maryland, Baltimore.

[29] Samuil 2012, Samuil Angelov, Paul Grefen, and Danny Greefhorst. A framework for analysis and design of software reference architectures. *Information and Software Technology*, 54(4):417–431, April 2012.

[30] Slimani, H., El Faddoul, N., Bennani, S., Amrous, N., 2016. Models of Digital Educational Resources Indexing and Dynamic User Profile Evolution. International Journal of Emerging Technologies in Learning, Vol. 11.

[31] Tirthali 2014, Hollands, F. M., & Tirthali, D. (2014). Why do Institutions Offer MOOCs? *Online Learning*, *18*(3).

[32] Westerlund [89] M. Westerlund, et al., A generalized scalable software architecture for analyzing temporally s

[33] Zhou 2017, Zhou, L., Pan, S., Wang, J., Vasilakos, A.V., (2017). Machine Learning on Big Data: Opportunities and Challenges. Neurocomputing.

[34] Zimmer 2013, [18] A. Zimmermann, et al., Towards and integrated service-oriented reference enterprise architecture, in: International Workshop on Ecosystem Architectures, Saint Petersburg, Russia, August, 2013.

[35] Prakash 2014, B R Prakash, Dr.M. Hanumanthappa, Vasantha Kavitha, Big Data in Educational Data Mining and Learning Analytics International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 12, December 2014.

Author(s)

K. Palanivel has completed her Masters in Computer Science & Engineering from Bharathiar Pondicherry University, Pondicherry, India. Prior to this, he received her Bachelor's in Computer Science & Engineering University, Tamil Nadu, India. His research coverage also includes software architecture, e-Learning and Web Services Technology. He is responsible for driving the unit's web and network monitoring and presence.