BIG DATA AND E-LEARNING: THE IMPACT ON THE FUTURE OF LEARNING INDUSTRY

Valentin PAU[1]
Marius Iulian MIHAILESCU[2]

Rezumat. *În prezent, unul dintre cele mai interesante aspecte ale e-Learning-ului este acela de evoluție continuă, arhitectura big data reprezentând o componentă importantă, asupra căreia comunitățile e-Learning au început să se oprească din ce în ce mai mult. Lucrarea de față își propune analizarea beneficiilor tehnologice ale conceputului de tip big data și impactul asupra viitorului e-Learning-ului dar și menționarea aspectelor critice ce țin de integritatea datelor.*

Abstract. *In nowadays, one of the most interesting aspects of e-Learning is that he is continuously evolving, where, the big data architecture represents an important component over which the e-Learning communities has stopped more and more. In our work paper we will analyze the technological benefits of the big data concept and the impact on the future of e-Learning but also we will mention the critical aspects regarding the integrity of the data.*

Keywords: big data, e-Learning, platform, integrity

1. Introduction

When we are talking about *big data*, especially when it comes up about e-Learning industry, the definition of big data is suffering a small modification, which can be stated as "the data created by learners during their online training and courses in which are enrolled".

There are many definitions for *big data* and most of them are adapted according to the needs of the market, infrastructure and user requirements. Below, we have presented some of the latest definitions about *big data*, which from this year, 2015, are more flexible and easy to understand with the environment itself. Analyst, *Doug Laney* [6], has characterized *big data* as "*data that's an order of magnitude greater than data you're accustomed to";* IBM's chief executive, *Virgina Rometty*, preconize that "*there will be 5.200 gigabytes of data for every human on the planet by 2020'';* one of the Forbes contributor, *Raj Sabhlok* [7], state that "*he right data and the right time not only reduce stress but also boosts your productivity. And that goes for your company too.";* the director from Social Innovation Fund [8], *Michael D. Smith*, believe in a possitive way as being" ...*a*

---
[1]Prof., PhD, PhD (ABD), Eng., *Titu Maiorescu* University, full/member of the Academy of Romania Scientist (e-mail: v_pau@utm.ro).
[2] Lect. Ph.D, *Titu Maiorescu* University.

great opportunity to figure out how you leverage the interest graph, how you leverage the social graph, how you leverage everything that we know about people on social networks. Don't think about it as scary—think about it as an opportunity."; the member of team for Big Spaceship [9], *Tony Clement* [10], state that" *Giving your data a meaningful voice is about choosing to invest in the right human skills. Skills that increase an organization's analytical literacy, eye for information design and ability to build data driven stories.";* also one of the well-known in his field, *Dorie Clark,* branding expert and former professor, says that "*Where the owner knows you and knows your family and is absolutely aware that with this new product, maybe your child would like it and they suggest it to you. That's something that as our country and the world has grown and become globalized, it became impossible to do at such a large scale. But Big Data is enabling us to go back to those roots and give people the personalized experience that they craved in the past, and is part of human nature.*"

Our work will present four main sections which will cover some of our research activity regarding the subject, *first research* consist in a study in which we will cover the *big dat*a benefits which are offered to the e-learning professionals (see Section 2); *second research* consists in explaining how the *big data* will impact the future of e-learning industry (see Section 3); *third research* will go through several cryptographic mechanisms which can used in order to assure the integrity of the collected data by different tools (Learning Management System (LMS), e-learning authoring tool, multimedia, social networks) – see Section 4; *forth research* will go to some ideas regarding the security of *big data* and data that are begin collected from different tools, and we will provide also some of ours contributions in this field in order to assure integrity of the data analytics in *big data*. As our contribution, we propose a classification of the main vulnerabilities and challenges and an examination of them will be provided.

Let's consider the following scenario as an example: *We have an employee who interacts with a specialized module with the goal of training, mapped on the company policies, progress, the results from the assessment, sharing on social platforms and any other types of data that are produced from the process of the e-Learning course".* According to the small modification that we have presented above, the example covers the most concentrate and pure meanings of the technological terms.

An interesting question that is raised, *who will collect the data?* The answer consist in a set of tools (see Fig. 1), such as Learning Management System (LMS), e-learning authoring tool, multimedia, social networks etc. The tools are looking very professional when they are mentioned, but there is a word in the answer, *collect,* which means that the collected data is stored and transferred through internet, and this transfer process can be very vulnerable and sensitive. In

Section 3, we will go through some cryptographic techniques which can be applied in the big data environment, in order to assure the integrity of the data that are collected.
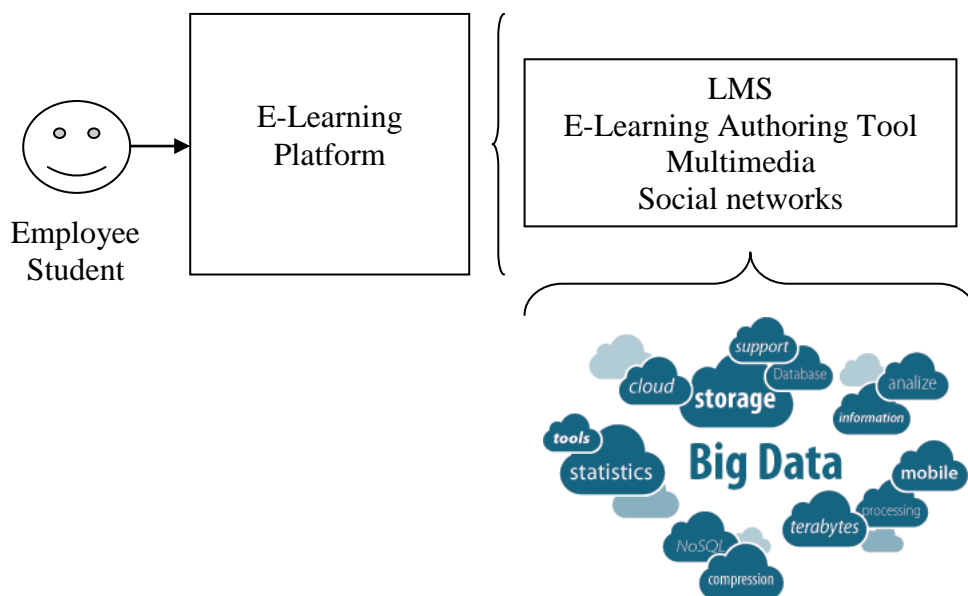

Fig. 1. Big Data Tools and E-Learning components.

The term *big data* sometimes can be very confused and it does not apply to the volume of data itself, but also on the independent pieces of data that are collected. These components are analyzed in order to give to the companies or e-learning professionals the chance to determine the way how the learner is acquiring the information, at what rhythm, and to pin in the problems that could exist within the e-Learning generalship itself.

2. *Big Data* Benefits Offered to E-Learning Industry and Professionals

In this section, we will present several benefits that are offer to e-learning professionals and not only. These benefits are chosen by their impact on the future of e-learning and revolutionize the method through which we analyze and assess the experience regarding e-learning. Below, we have presented a list with the most important benefits that go hand-in-hand with big data.

    a. Giving the possibility to the e-learning industry and professionals to figure out how the students are digesting the information and which learning is necessary to recourse the most to them. For example, the *big data* enable e-learning professionals to determine if a scenario that is based on reality, is more effective than a text-based which has a goal the activity of problem solving.

b. Another benefit consist in giving the possibility to professionals to pin out different zones on which is necessary to fine-tuned within the digital course or module. Let's consider the following example, if we have multiple learners which are time consuming for finishing a specific module, this probably means that the module will need some special improvements in order to make it more customizable for the learners.

c. An important benefit is that of providing analysis of which e-learning modules more and more visited, and in the case of social learning, which digital modules or web links are shared with other learners. Let's take for example the following case, where we have the possibility to determine what link was shared the most through Facebook.

d. The data are received instantly, a good benefit when we don't want to wait for long period of time to receive different tests, homework's etc. In this way, the professionals which are using e-learning, can start to implement the changes or using the data to improve their e-learning strategy immediately.

e. One of the last benefit that need to be taken into consideration, is based on patterns where e-learning professionals have the possibility to predict in which cases the learners are struggling or exceling. In this way, there is the possibility to develop different e-learning modules in such way that the learners get the chance to gather the best possible results.

3. The impact of the *big data* on the future of e-learning

On this part, we will focus on several reasons for which *big data* will revolutionize e-learning industry. The impact is based on an important aspect, which is a positive one, more precise, it will allow eLearning professionals and academicians to customize their courses and modules. *Big data* has an important potential as an impact on the future of e-learning by:

a. **Having precious feedback.** They will have the opportunity to receive an important feedback. The feedback can be used later to pin out where the learner, and the digital course itself, need to be improved. Let's take for an example the following case, in which a learner has the possibility to look at an analysis of where the student (electronic learner or *elearner*) felt short during the progress on the course, and where there is the possibility to figure out the possibility to correct the issue by moving forward. In the same time, if the administrator of the platform or e-learning system will observe that the majority of the learners will face difficult with particular modules or assignments, then they will have the possibility to make the proper adjustments in order to improve the performance for the students.

b. **Giving the possibility to the e-learning professional to design more customized e-learning courses.** If the students have the opportunity to understand how their e-learning administrator will acquire the information and what is best for them, in terms of delivering and content, the results will be more customized and the digital course are becoming more engaged. The courses need to be personalized with the needs of the student, in this case, we will obtain a high level of quality and experience regarding the e-learning experience.

c. **Having different strategies and goals regarding the e-learning.** Using *big data* is very nice thing to do, especially when it comes about strategies that are working and figuring out those who are and not necessarily in achieving e-learning goal. With other words, considering the following example in which we are able to determine what digital courses are representing as an contribution to skill development and which e-learning modules, elements, or components, may be irrelevant.

d. **Applying patterns and tracking students.** Is very rare the possibility of tracking a student through the hole process of studying and following the digital courses. With other words, we can see how they answer to tests, how fast they have finished a module. This will help the administrator to set some patterns that will not only enable the opportunity to learn more about the behaviors during the learning process of the students, but also as a students' group as an entire group.

e. **Understanding the e-learning process must be expanded.** Being e-learning student or professional, is very important to learn as much as we can about how learners will gather and digest the knowledge. *Big data* will allow gaining a deep understanding of the e-learning process and also will give us a status about how the learners will respond to the e-learning courses that we, as administrators, will deliver to them. Many statistics are possible to create, such as the moment from a day when the learners are more productive, or which methods for delivering allow them to retain the information in a most efficiently way. Those information's will be used later to move further with our strategies.

4. Confidentiality, Integrity, and Authenticity. A major problem for *big data*

Securing analytics is an important aspect that needs to be secured in *big data*. Security in *big data* sometimes could be a real and taught challenge, cryptographic techniques being related to the differences according to the cost of performance. Our survey presented in this section will cover three cryptographic techniques – homomorphic encryption, verifiable computation, and multi-party computation. Another survey regarding the cryptographic techniques, it can be read here [1].

First technique, homomorphic encryption, gives the possibility for the functions to be computed on data that has been encrypted without decrypting first.

Let's consider $E_{key}(message)$ is the encryption of the *message* using a key. A homomorphic encryption scheme is according to a function f if we have a corresponding function f' in such way that

$$ [\![D]\!]_{key} (f^{T'} (E_{\downarrow} key \ (message) \ ) = f(message) \ , $$

where D_{key} represent the algorithm with key.

In order to assure the integrity of the analytics using *first technique*, is very recommended to use the encryption allowing the cloud to do it and to take action over the encrypted data, without sharing the encryption key (see Fig. 2.

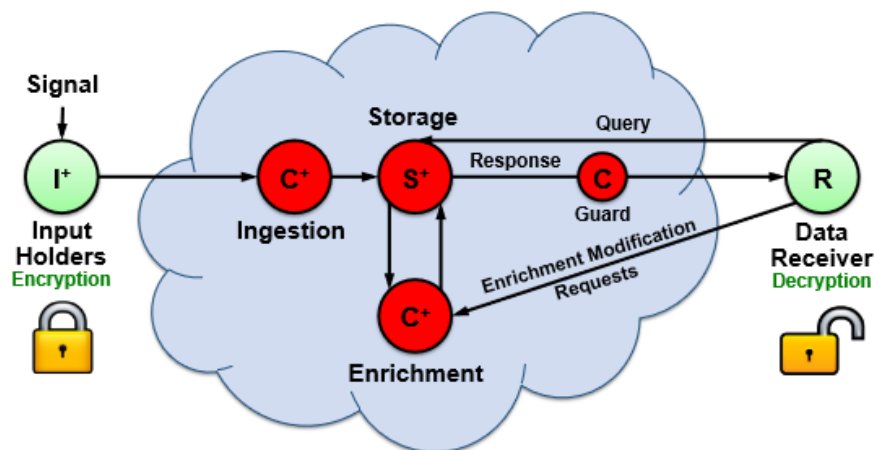This is a very interesting approach, which have been described with many details in [1].

Fig. 2. Using homomorphic encryption [1].

Second technique will consist in using a *verifiable computation,* which will check the integrity of the computation.

The computation is done by the owner.

This is done using the following scheme.

The data owner will provide the data together with the data by using a specification for the computation that wish to be used, to some component which will be named *prover* (see Fig. 3).

The goal of the *prover* is to attest that the output (result) of the chosen computation, together with the *real argument or demonstration*, is correct.
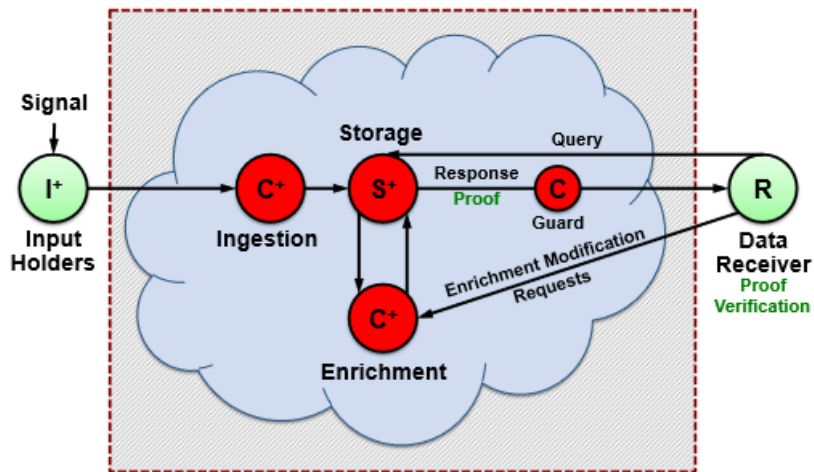
Fig. 3. Using verifiable computation [1].

Third technique, multi-party computation, is used if we want to have an advantage of the semi-trusted cloud configuration (see Fig. 4). Multi-party computation will leverage the presence of trusted components, ignoring the fact that we don't need to know which parties are honest, in order to get accomplish the goal of confidentiality and integrity. As a state-of-the-art, secure computation was firstly introduced by Yao [2] and Goldreich, Micali, and Wigderson [3] and extended a specific case by Chaum, Crepeau, and Damgard [4] and also Ben-OR, Goldwasser, and Wigderson [5].
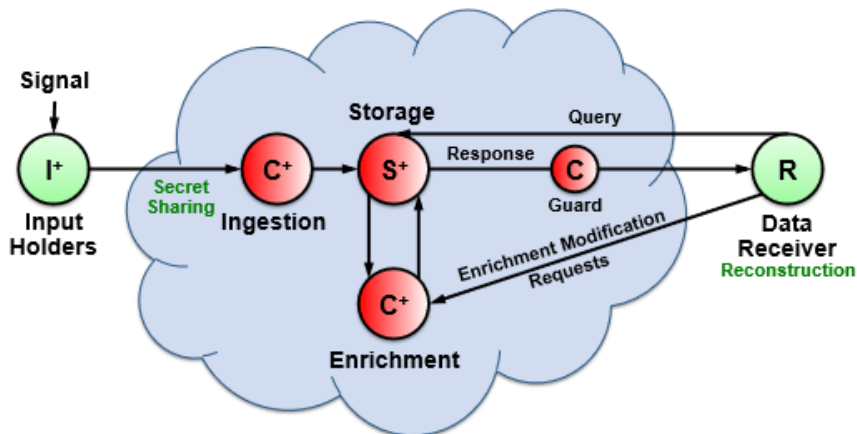

Fig. 4. Using Secure multi-party computation [1].

4.1. *Big data security challenges classification*

In this section we have provided a classification the most important challenges about security in Big Data.

Challenges regarding the infrastructure:

- Protocols used in communication;
- Management of the keys.

Challenges about encryption:

- Encryption for access-based on the policy;
- Searching and filtering through encrypted data;
- Securing the outsourcing of the computation.

Challenges regarding the privacy:

- Dissemination of the security;
- Collection and aggregation securing data;
- Collaboration security.

Challenges about data administration:

- Assuring the integrity of the data;
- Data storage and proofing.

The most vital and vulnerable points from above and on which we have to focus, are: encryption for access-based on the policy, searching and filtering through encrypted data, securing the outsourcing of the computation, dissemination of the security, data storage and proofing.

When we are talking about proving the data storage, we imagine that we have upload a huge quantity of data into the cloud and the question is, from where we know every time if the data is available on the cloud?

We have two answers, a bad one and a good one.

The bad answer consists in two aspects, *first*, asking to reply the data back to you, which means that we will have big costs; *second*, to ask to send the hash of the data, which means that we have the big chance to be replayed.

The good answer consists in proving the data storage.

This can be easily being done with RSA algorithm. For details about RSA, read [11].

We will propose a scheme for proving the data storage using RSA algorithm.

The scheme consists in two components, the user and the volume of big data, e.g. files, documents.

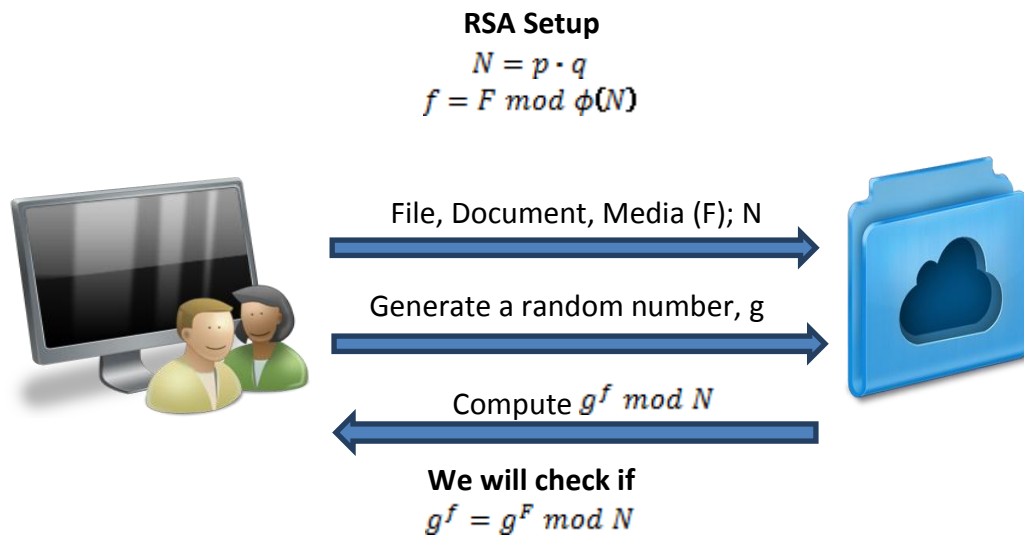Let's consider the following scenario from Fig. 5.

RSA Setup

$$N = p \cdot q$$
$$f = F \bmod \phi(N)$$

File, Document, Media (F); N

Generate a random number, g

Compute $g^f \bmod N$

We will check if

$$g^f = g^F \bmod N$$

Fig. 5. Scheme for proving the data storage using RSA algorithm.

Conclusions

The present work paper represents a short introduction into *big data* as a concept applied to learning industry. We have accomplished four main goals, from which we will mention: big data benefits, the impact on the future of e-learning and assuring integrity of the data collected with the help of tools mentioned at the beginning of the work-paper and analytics.

Security in a *big data* environment is very difficult to implement, thus there are many schemes developed and a part of them also implemented, with sufficient knowledge is very easy to break them.

E-learning industry represents the future in education, and the evolution of technologies will have an important impact on it, if the technologies will not embraced in a quickly and fast manner.

R E F E R E N C E S

[1]   Yakoubov Sophia, Vijay Gadepally, Nabil Schear, Emily Shen, Arkady Yerukhimovich, *A Survey of Cryptographic Approaches to Securing Big-Data Analytics in the Cloud.* High Performance Extreme Computing Conference (HPEC), ISBN – 978-1-4799-6232-7, pp. 1-6, IEEE, 2014, DOI: 10.1109/HPEC.2014.7040943.

[2]   A. C.-C. Yao, *Protocols for secure computations (extended abstract)* in FOCS, 1982.

[3]   O. Goldreich, S. Micali, and A. Wigderson, *How to play any mental game,* In Proceedings of the Nineteenth Annual ACM Symposium on Theory pf Computing, 1987.

[4]   D. Chaum, C. Crepeau, and I. Damgard, "Multiparty unconditionally secure protocols (abstract)," in CRYPTO, 1987.

[5]   M. Ben-Or, S. Goldwasser, and A. Wigderson, *Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract)*" in STOC, 1988.

[6]   Doug Laney, *Deja VVVu: Others Claiming Gartner's Construct for Big Data,* http://blogs.gartner.com/doug-laney/.

[7]   Raj Sabhlok, http://www.forbes.com/sites/rajsabhlok/.

[8]   Social Innovation Fund, http://www.nationalservice.gov/programs/social-innovation-fund.

[9]   Big Spaceship, http://www.bigspaceship.com/.

[10] Tony Clement, *Why Big Data is a Scary Investment, http://www.bigspaceship.com/big-data-scary-investment/.*

[11] RSA Cryptosystem, http://en.wikipedia.org/wiki/RSA_(cryptosystem).

[12] V. Pau, I. Mihăilescu, O. Stănescu – Avoiding security and privacy issens by using specific patterns in biometric and RFID systems, International Conference " *Promovarea parteneriatului în învăţământul superior pentru o societate bazată pe cunoaştere*", 2011.