E-learning and educational data mining in cloud computing: an overview

Alberto Fernández*

Department of Computer Science,
University of Jaén,
23071, Jaén, Spain
E-mail: alberto.fernandez@ujaen.es
*Corresponding author

Daniel Peralta, José Manuel Benítez and Francisco Herrera

Department of Computer Science and Artificial Intelligence,
CITIC-UGR (Research Center on Information and Communications Technology),
University of Granada,
18071 Granada, Spain
E-mail: dperalta@decsai.ugr.es
E-mail: herrera@decsai.ugr.es
E-mail: jmbs@decsai.ugr.es

Abstract: E-learning is related to virtualised distance learning by means of electronic communication mechanisms, using its functionality as a support in the process of teaching-learning. When the learning process becomes computerised, educational data mining employs the information generated from the electronic sources to enrich the learning model for academic purposes. To provide support to e-learning systems, cloud computing is set as a natural platform, as it can be dynamically adapted by presenting a scalable system for the changing necessities of the computer resources over time. It also eases the implementation of data mining techniques to work in a distributed scenario, regarding the large databases generated from e-learning. We give an overview of the current state of the structure of cloud computing, and we provide details of the most common infrastructures that have been developed for such a system. We also present some examples of e-learning approaches for cloud computing, and finally, we discuss the suitability of this environment for educational data mining, suggesting the migration of this approach to this computational scenario.

Keywords: e-learning; educational data mining; EDM; cloud computing; big data; MapReduce; Hadoop; Not Only SQL; NOSQL.

Reference to this paper should be made as follows: Fernández, A., Peralta, D., Benítez, J.M. and Herrera, F. (2014) 'E-learning and educational data mining in cloud computing: an overview', *Int. J. Learning Technology*, Vol. 9, No. 1, pp.25–52.

Biographical notes: Alberto Fernández received his MSc and PhD in Computer Science from the University of Granada, Granada, Spain, in 2005 and 2010, respectively. He is currently an Assistant Professor with the Department of Computer Science, University of Jaén, Spain. His research interests include data mining, classification in imbalanced domains, fuzzy rule learning, evolutionary algorithms, multiclassification problems with ensembles and decomposition techniques, and big data applications by means of cloud computing. He received the 'Lofti A. Zadeh Prize' of the International Fuzzy Systems Association for the 'Best Paper in 2009–2010' for his work of hierarchical fuzzy rule-based classification system with genetic rule selection for imbalanced data-sets.

Daniel Peralta received his MSc in Computer Science in 2011 from the University of Granada, Granada, Spain. He is currently a PhD student in the Department of Computer Science and Artificial Intelligence, University of Granada. His research interests include data mining, biometrics and parallel and distributed computing.

José Manuel Benítez is an Associate Professor at the Department Computer Science and Artificial Intelligence, Universidad de Granada, Spain. He holds an MS and PhD in Computer Science, both from the Universidad de Granada. He is an active researcher in computational intelligence, where his work covers the whole spectrum from foundations to applications in a number of engineering and scientific areas. His fields of interest are time series analysis and modeling, distributed computational intelligence, high performance computing, cloud computing, data mining, biometrics, and statistical learning theory. He is a member of a number of scientific associations, including IEEE, IEEE Computational Intelligence Society, and EUSFLAT.

Francisco Herrera received his MSc in Mathematics in 1988 and PhD in Mathematics in 1991, both from the University of Granada, Spain. He is currently a Professor in the Department of Computer Science and Artificial Intelligence at the University of Granada. He has been the Supervisor of 29 PhD students. He has published more than 270 papers in international journals. He currently acts as Editor-in-Chief of the international journals *Information Fusion* (Elsevier) and *Progress in Artificial Intelligence* (Springer). He acts as editorial member of a dozen journals. His current research interests include among others, soft computing, decision making data sciences and cloud computing.

This paper is a revised and expanded version of a paper entitled 'An overview of e-learning in cloud computing' presented at Workshop on Learning Technology for Education in Cloud (LTEC'12), Salamanca, Spain, 11–13 July 2012.

1 Introduction

Electronic learning, better known as e-learning, refers to the issues related to virtual distance education through electronic communication mechanisms, specifically the internet (Mayer and Clark, 2011). It is based on the use of approaches with several formats and functionalities that may support the teaching-learning process, such as e-mail, web pages, forums, various learning platforms, and so on. In this way the learning experience is managed with greater success, through the online integration of the community of students, content developers, and experts. The main advantages defined by studying through online tools include flexibility, convenience, ease of access, consistency and repeatability of the proposed tasks.

Within information technologies (ITs), there is a growing trend regarding the research and exploitation of e-learning or virtual teaching platforms. There are several related initiatives at different educative levels, from which some examples are the Khan Academy (http://www.khanacademy.org), the Virtual Learning Center of Granada University (CEVUG-UGR), the Open University of Catalonia, the MIT Open CourseWare, or the 'free online courses' of Standford University.

Virtual courses, which are clearly supported by the e-learning approach, have a clear advantage for the educative framework with respect to the classical attendance group, namely the higher number of users which can have access to their content. As an example, in the first edition of the 'machine learning' course of Stanford (http://www.ml-class.org/course/class/index) more than 160,000 students were registered worldwide. These dimensions affects different issues; on the one hand, the infrastructure provisions that are necessary to provide a concurrent service for that amount of students clearly exceed the capabilities of a conventional web server. Furthermore, the demand on teaching resources usually varies in a dynamic and rapid way, presenting high peaks of activity. To attend to requests during these periods of time without other system services being weakened, it will be necessary to prepare a quite superior infrastructure than that required for the regular working of the learning institution. An alternative would be to provide services depending on demand, and only pay for resources that are actually used. The answer to these requirements is the cloud computing environment.

Cloud computing (Mell and Grance, 2011; Buyya et al., 2011) has its origins in the concept of grid computing, which has the aim of reducing computational costs and increasing the flexibility and reliability of the systems. However, the difference between the two lies in the way the tasks are computed in each respective environment. A computational grid is more static, from the point of view of hardware resources, and is designed mostly with the aim of obtaining the best computer performance. On the other hand, cloud computing is intended to allow the user to obtain various services without being aware of the underlying architecture while offering a transparent scalability. It is therefore not so restrictive and can offer many different services, from web hosting, to word processing (Buyya et al., 2009).

A service-oriented architecture (SOA) (Papazoglou and Van Den Heuvel, 2007) is one of the base of cloud computing. This type of system is designed to allow developers to overcome many distributed organisation computing challenges including application integration, transaction management and security policies, while allowing the use of multiple platforms and protocols and leveraging numerous devices to which we can have access, and legacy systems (Alonso et al., 2004). All the services offered by a cloud platform are provided to the users without letting them know the location and other details of the computing infrastructure (Furht and Escalante, 2010).

The advantages of this new computational paradigm with respect to other competing technologies are clear. First, cloud application providers strive to provide the same or better services and performance as if the software programs were installed locally on end-user computers, so the users do not need to spend money buying complete hardware equipment for the software to be used. Second, this type of environment for data storage and computing schemes allows companies to get their applications up and running faster, with a lower requirement of maintenance from the IT department since it automatically manages the business demand by dynamically assigning IT resources (servers, storage and/or networking) depending on the computational load in real time (Velte et al., 2010).

Additionally, the e-learning platforms of the large dimensions which we mentioned above generate extensive registers of interaction among students-platform-teachers. These data bases contain significant information not defined in an explicit way. Data mining techniques must be applied to extract this information (Witten et al., 2011; Romero and Ventura, 2010). In this situation, 'educational data mining' (EDM) (http://www.educationaldatamining.org/) arises as a tool that benefits both teachers and students in order to improve the learning procedure. The object of interest of this discipline is the development of new methodologies for exploring the data generated in the activity of the educational systems, mentioned above. The final aim of the application of such methods is the achievement of a better understanding of the behaviour of the students, and how to design procedures and material that may facilitate the learning process.

In a clear connection with this process we may find the intelligent tutoring systems (http://aaai.org/AITopics/IntelligentTutoringSystems), which are computer-based systems designed to support the teaching-learning process. They are usually intelligent systems able to drive the learning process of the student, providing him/her with feedback based on their progress and the results of periodic tests. The process of EDM interacts with an intelligent tutoring system by extending and refining its knowledge base. Taking into account the dimensions and growing capacity of the computational resources (stable storage, memory and CPUs), a cloud platform is also a natural structure for the implementation of data mining techniques and their application to growing datasets (Big Data) (Agrawal et al., 2011; Madden, 2012). However, many of the data mining techniques do not have an adequate scalability. This is an aspect that grows in importance and that has attracted the interest of researchers and companies.

In order to review all these aspects, this paper is arranged as follows. In Section 2, we introduce the main concepts on cloud computing, including its infrastructure and main layers. Then, Section 3 presents the features of the e-learning approach, stressing the advantages (and also drawbacks) of the migration of such a system to a cloud computing environment and showing some examples of real applications of this kind. Section 4 is devoted to describing the subject of EDM and how can we take advantage of the cloud computing environment for this kind of task. Finally, the concluding remarks obtained throughout the development of this work are given in Section 5.

2 Basic concepts of cloud computing

Cloud computing is a paradigm in which resources and/or services are provided over the internet. This leads us to the concept of SOA (Papazoglou and Van Den Heuvel, 2007), which is an integration platform based on the combination of a logical and technological architecture oriented to support and integrate all kinds of services. In general, a 'service' in the framework of cloud computing is a task that has been encapsulated in such a way that it can be automated and supplied to the clients through a consistent and constant procedure. Any component can be considered to be a service, from the components which are closest to hardware, such as the storage space or the computational time, to software components aimed at authenticating a user or managing the mail, the management of a database or the monitoring of the use of the system resources.

In this section, we will give a brief introduction to the cloud computing environment, first describing its main features (Section 2.1). We will then present the layers form which this platform is built (Section 2.2). Then, we will stress the goodness of the use of cloud computing with respect to other similar technologies such as grid computing (Section 2.3). Finally, we will point out several technological difficulties that remain to be addressed in order to improve the quality of this paradigm (Section 2.4).

2.1 Introduction to cloud computing

The philosophy of cloud computing mainly implies a change in the way the problems are solved by using computers. The design of the applications is based upon the use and combination of services. In contrast to what happens in more traditional approaches, i.e. grid computing, the provision of the functionality relies on this use and the combination of services rather than on the concept of process or on algorithms.

Clearly, this brings advantages in different aspects, for example the scalability, reliability, and so on. For example, in the presence of a peak of resource demand, because of an increase of users or an increase of the data size, additional instances of a determinate service could be raised so that the application response time remains acceptable for users. In the corresponding case of a fall in demand, resources should be released. All of these actions are carried out in a way that is transparent to the user.

The main features of this architecture are its loose coupling, high interoperativity and the presence of interfaces that isolate the service from the implementation and the platform. In an SOA, the services tend to be organised in a general way in layers or levels (not necessarily with strict divisions) where, normally, some modules use the services that are provided by the lower levels in order to offer other services to the superior levels. Furthermore, those levels may have different organisational structures, a different architecture, etc. Depending on the type of services offered, we can find three main types of layer which, together, compose what is known as a 'stack' (Youseff, 2008):

1 a storage cloud, which provides storage services based on blocks (chunks) or files
2 a data cloud, which provides data management services, based on registers, columns or objects
3 a computation cloud, which provides high performance execution services.

The cloud computing paradigm benefits large projects. As is well known, many applications in science or business domains face a high computational load. This may result from the requirement of handling large amounts of data stored in static databases, which implies a high data storage capacity; or may be the result of a dynamic data stream, which requires a high throughput network connection.

The current benefits of cloud computing for application providers and end users can be summarised as follows (Han et al., 2009):

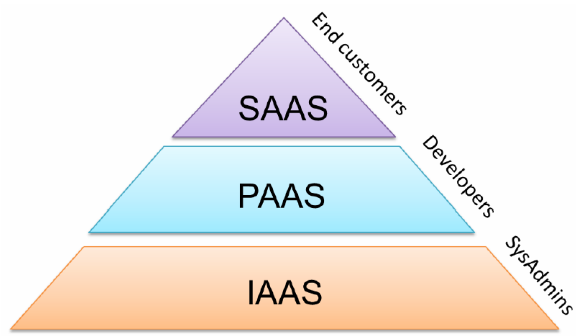
- *Scalability in resource usage:* a cloud appears as an illusion of infinite computing resources accessible on demand, and allows the users safe investment in infrastructures. Therefore, users may start with a minimal scale of resources and increase their resource occupancy in accordance with their needs.

- *Flexibility in the pricing model:* the scalability in resource occupancy stated above directly prompts flexible pricing models of clouds.
- *Low administration effort:* the administration and maintenance is on the server side, liberating the application user from this responsibility.
- *More mobility:* the cloud environment is accessible from any device as long as it has an internet connection, since all the workload is carried out on the cloud (data centre) instead of local facilities.

2.2 Cloud computer layers

There are different categories in which the service-oriented systems can be clustered. One of the most widely used criteria to group these systems is the abstraction level that they offer to the system user. In this way, three different levels are often distinguished, as we can observe in Figure 1. In the remainder of this section, we will first describe each of these three levels, providing the features that defines each of them, and some examples of the most well known systems of each type. Then, we will present some technological challenges that must be taken into account for the development of a cloud computing system.

Figure 1 Illustration of the layers for the SOA (see online version for colours)



2.2.1 Infrastructure as a service

Infrastructure as a service (IaaS) is the supply of hardware as a service, that is, servers, net technology, storage or computation, as well as basic characteristics such as operating systems and the virtualisation of hardware resources (Hurwitz et al., 2010). Making an analogy with a monocomputer system, the IaaS will correspond to the hardware of such a computer together with the operating system that takes care of the management of the hardware resources and facilitates access to them.

The IaaS client rents the computational resources instead of buying and installing its own data centre. The service is usually billed based on its actual usage, so it has the advantage that the client pays for what is used, and uses what he/she needs at any given moment. Also, due to the dynamical scaling associated with cloud computing, in the case of low loads of work he/she uses (and pays for) less resources. In the presence of a higher resource demand, IaaS can provide them in order to attend to the needs of that client. It is also common that the contract of the service includes a maximum point that the user cannot exceed.

Among the more typical IaaS clients are scientific researchers and practitioners, who thanks to the IaaS and the wide volume of infrastructure that it offers as a service, can develop tests and analyse the data at a level that could not be possible without access to this big scale computational infrastructure. Nowadays, one of the main IaaS providers is Amazon with their Elastic Computer Cloud (EC2). We may also note the Google Compute Engine, Windows Azure, and RackSpace as other significant IaaS providers.

2.2.2 Platform as a service

At the platform as a service (PaaS) level, the provider supplies more than just infrastructure, i.e., an integrated set of software with all the stuff that a developer needs to build applications, both for the developing and for the execution stages. In this way, a PaaS provider does not provide the infrastructure directly but, by making use of the services of an IaaS it presents the tools that a developer needs to have an indirect access to the IaaS services and, consequently, to the infrastructure (Hurwitz et al., 2010).

The PaaS can be viewed as a kind of 'software layer' that enables components to be developed for applications, as well as applications themselves. It will be an integrated developer environment, or a set of independent tools, that allows an engineering software problem to be developed in all its stages, from the analysis and model of the problem, the design of the solution, its implementation and the necessary tests, before carrying out the stage of deployment and exploitation. In the same way, a programming language that counts on compilers and libraries for the different operating systems will enable the same application to be deployed in different systems without the necessity of rewriting any piece of code. Some examples of PaaS are Google App Engine, Microsoft Azure, Force.com, Amazon Web Services, Heroku, and OpenShift (from Red Hat), among others.

2.2.3 Software as a service

At the last level we may find the SaaS, which offers software as a service. This was one of the first implementations of the cloud services, along with the gaining in importance of internet usage. It has its origins in the host operations carried out by the application service providers, from which some companies offered to others the applications known as customer relationship managements (Duer, 2003).

Over time, this offer has evolved to incorporate a wide range of possibilities, both for companies and for individual users. Regarding net support, although these services are performed through the internet, as it provides the geographical mobility and the flexibility needed, a simple exchange of data in this way will not assure its privacy. For this reason, virtual private networks are often employed to this end, as they allow the transmission of data through the internet in an encrypted way, maintaining the privacy and security of the information exchange between the client application of the user and the SaaS application store in the cloud. Some systems that can be found within this layer are well known to the common user such as Facebook, eBay, Skype, Dropbox, GoogleApps (Gmail, Google Docs, ...) Quickbooks online (Salesforce.com) or Evernote.

2.3 Cloud computing versus other computational paradigms

Traditionally, in order to process a high amount of data in a short period of time, a grid computing environment was the most suitable solution in order to reduce computational costs and to increase the flexibility of the system. The cloud computing platform contains characteristics of both clusters and grids, since both of them are composed of loosely coupled, heterogeneous nodes, and grids are also geographically dispersed. However, there are significant differences among them, which mainly refer to the target of their application, and the implementation technologies used (Foster et al., 2009). Cluster and grid computing were initially designed and developed to facilitate high performance computing applications, while cloud computing also aims at supporting many diverse services for the user without investing in the underlying architecture, hence making management costs much lower than those of clusters and grids and management less cumbersome (Buyya et al., 2009). Additionally, the use of virtual machines supports failover and content replication, making the cloud environment more reliable. The distributed nature of clouds gives, however, the cloud user less control over the location of data and computation (Mateescu et al., 2011).

Additionally, the advantages of this new computational paradigm with respect to other competing technologies are clear. First, cloud applications strive to provide the same or better service and performance as if the software programs were installed locally on end-user computers, so the users do not need to spend money buying complete hardware equipment for the software to be used, i.e., a simple PDA device is sufficient to run the programs on the cloud.

Second, this type of environment for the data storage and the computing schemes allows companies to get their applications up and running faster. It also requires less maintenance from the IT department, since it automatically manages the business demand by dynamically assigning resources (servers, storage and/or networking) in real time depending on the computational load (Velte et al., 2010). This inherent elasticity is accordingly reflected in the billing for the use of the infrastructure.

In brief, grid computing mainly focuses on high performance computing whereas cloud computing offers additional services with much more flexibility over a new business model.

2.4 Technological challenges in cloud computing

Cloud computing has been shown to be a very effective paradigm due to its features such as on-demand self-service. Some of its advantages are listed below:

- the provision of computing capabilities to the customers without requiring any human interaction
- broad network access from heterogeneous client platforms
- resource pooling to serve multiple consumers
- rapid elasticity as the capabilities appear to be unlimited from the consumer’s point of view
- a measured service allowing for a pay-per-use business model.

However, in order to offer such an advantageous platform, there are some weak points that should be taken into account. We present some of these issues as follows:

- *Security, privacy and confidence:* Since the data can be distributed on different servers, and 'out of the control' of the customer, there is a necessity of managing hardware for computation with encoding data by using robust and efficient methods. Also, in order to increase the confidence of the user, several audits and certifications of the security must be performed.
- *Availability, fault tolerance and recovery:* to guarantee a permanent service (24 × 7) with the use of redundant systems and to avoid net traffic overflow.
- *Scalability:* In order to adapt the necessary resources to the changing demands of the user by providing an intelligent resource management, an effective monitoring must be used to identify a priori the usage patterns and to predict the load in order to optimise the scheduling.
- *Energy efficiency:* It is also important to reduce the electric charge by using microprocessors with a lower energy consumption and adaptable in their use.

3 Cloud computing for e-learning tasks

As we stated in the introduction of this work, with the huge growth in the number of students, educational content, services that can be offered and resources made available, e-learning system dimensions grow at an exponential rate. The challenges with regard to optimising resource computation, storage and communication requirements, and dealing with dynamic concurrency requests highlight the necessity of the use of a platform that meets scalable demands and cost control. This environment is cloud computing.

Defining the potential of SaaS solutions for efficient and sustainable online learning platforms in contrast to the former 'classic' online learning platforms, may lead us to understand the benefits of cloud computing on both a technological and educational level. We need to introduce the 'path' for promoting a migration to such a model in order to obtain a valuable system for web tools and collaborative services, such as lesson plans, videos, curriculum resources, student interactions, and so on. By referring to a wide number of current developments, we will measure the real impact of this novel computational scheme and we will describe some examples on how to start using a system with these characteristics.

Throughout this section we will introduce the main advantages that e-learning systems offer and, more importantly, which drawbacks establish the basis for acknowledging the necessity of a new technology (subsection 3.1). Then, in accordance with the points made in this section, the significance of selecting cloud computing for this kind of tools will be emphasised (subsection 3.2). The organisation and infrastructure necessary for the virtual platform is then described (subsection 3.3). Then, we will review some of the e-learning applications that have already been developed within the cloud computing platform (subsection 3.4). Finally, we will point out some challenges that remain to be addressed regarding the adaptation of e-learning to a cloud environment (subsection 3.5).

3.1 Drawbacks of current e-learning systems

Among learning technologies, web-based learning offers several advantages over conventional classroom-based learning. Its most notable advantages are reduced costs, as a physical environment is no longer required and it can therefore be used at any time and place at the convenience of the student. Moreover the number of students that can follow the class is not limited by the site of a physical classroom. Additionally, the learning material is easy to keep updated and the teacher may also incorporate multimedia content to provide a friendly framework and to facilitate the understanding of the concepts. Finally, it can be viewed as a learner-centred approach which may address the differences among teachers, so that all of them may compare their material in order to evaluate and re-utilise common areas of knowledge (Jolliffe et al., 2001).

However, there are some disadvantages that must be addressed prior to the full integration of e-learning into the academic framework. Currently, e-learning systems are weak on scalability at the infrastructure level. Several resources can only be deployed and assigned for specific tasks so that when receiving high workloads, the system needs to add and configure new resources of the same type, making resource acquisition and management very expensive.

This key issue is also related to the efficient utilisation of these resources. For example, in a typical university scenario, PC laboratory resources (software, hardware) that can accommodate all the students in one department, may sometimes remain under-utilised due to low demand in other departments. Another hitch is dynamic rates of use: there is a high demand for use of these resources towards the end of a semester, whereas it is small at the beginning of the semester and practically non-existent during breaks. Therefore, the physical machines are held even when they are idle, wasting their full potential.

The computing infrastructure does not only include computers (hardware) but also civil infrastructure, i.e., buildings, climatisation, security devices, and so on. Additionally, there is also an increased cost regarding the software licensing with corresponding costs of installation, maintenance and technical support (Kwan et al., 2008).

3.2 On the suitability of cloud computing for e-learning

The use of cloud computing in education has great potential, as can be observed from the number of educational institutions that have implemented these systems already (Boyatt and Sinclair, 2012; Sultan, 2009). For example, in the USA 27% of schools make use of cloud facilities (CDW-G, 2011). Other programmes such as that in the UK (JISC, 2012) are underway to provide a private education cloud with appropriate data management, storage and tools.

E-learning in the cloud can be viewed as education SaaS. Its deployment can be performed very quickly since the hardware requirements of the user are very low. Furthermore, as we stated previously, it lessens the burden of maintenance and support from the educational institution to the vendor, allowing them to focus on their core business, also obtaining the latest updates in the system without charge, and sharing key resources using Web 2.0 technology. From a technical perspective (Bensch and Rager, 2012), mechanisms of the e-learning system architecture and the cloud computing infrastructure were examined with respect to stability, balance, efficient usage of resources and the sustainability of the e-learning ecosystem (Dong et al., 2009b).

In what follows, we summarise the consequences and implications of the development of e-learning services within the cloud computing environment, as pointed out by Masud and Huang (2011):

- *Access via the web:* implies an ease of access since anyone, anywhere, and at any time can access the application, meeting the greater demand for web development skills.
- *No client-side software needed:* therefore, it has reduced costs for the subscriber, as there are no installation, software maintenance, deployment or server administration costs. There is also a lower total cost of ownership, reduced time-to-value, and fewer IT staff are required by the institution.
- *Pay by subscription based on usage:* this is suitable for the software model education market, and can provide access to more sophistified applications.
- *SaaS server may support many educational institutions:* Since the application is running on a server farm, scalability is inherent to the system. As student usage grows, the software performance will not degrade.
- *All subscriber data held on SaaS server:* a very high level of security is needed by the SaaS provider in order to gain the trust of subscribers and sophisticated multi-tenanted software architecture. The subscriber data is distributed between many providers and it must be integrated in order to gain an overview of business, a higher demand for the system and data integrators.

From a technical perspective, some authors have already explored reasons why the possibilities provided by a cloud approach are likely to be beneficial in education. The most common issue is cost, but further factors include those noted for cloud use in general (Ouf and Nasr, 2011; Sultan, 2009; Katzan, 2010; Stein et al., 2013):

- No need to backing everything up to a thumb drive and transferring it from one device to another. It also means that students can create a repository of information that stays with them and keeps growing as long as they want.
- Crash recovery is almost completely unnecessary. If the client computer crashes, there are almost no data lost because everything is stored in the cloud (Pocatilu et al., 2010).
- Allows students to work from multiple places (home, work, library, and so on), find their files and edit them through the cloud, and browser-based applications can also be accessed through various devices (mobile, laptop and desk top computers, provided internet access is available) (Al-Zoube et al., 2010).
- Allows students to create content through the browser, instead of only searching through the browser.
- Provides a low cost solution to academic institutions for their researchers, faculty and students.

- Flexibility: Scale infrastructure to maximise investments. Cloud computing allows the user to dynamically scale as demands fluctuate (Ercan, 2010).
- Improved improbability: it is almost impossible for any interested person (thief) to determine where the machine that stores some desired data (tests, exam questions, results) is located, or to find out which is the physical component he needs to steal in order to get a digital asset (Pocatilu et al., 2010).
- Virtualisation: makes possible the rapid replacement of a compromised cloud located server without major costs or damages. It is very easy to create a clone of a virtual machine so the cloud downtime is expected to be reduced substantially.
- Extending machine life cycle: this is closely related to the former. Since the computational power needed to run applications is provided remotely, schools can save costs through the use of older and less powerful computers.
- Centralised data storage: losing a cloud client is not a major incident while the main part of the applications and data is stored in the cloud so that a new client can be connected very fast. Imagine what happens today if a laptop storing examination questions is stolen.
- Monitoring of data access becomes easier in view of the fact that only one place should be supervised, not thousands of computers scattered over an extensive geographical area, for example. Also, the security changes can be easily tested and implemented since the cloud represents a unique entry point for all the clients (Wheeler and Waggener, 2009).

Finally, although little is yet to be found concerning cloud-related pedagogies or evaluating learning (Stein et al., 2013), from an academic point of view, one of the positive features of the cloud is its accessibility (Geith, 2008), since it is specifically designed to enable users to work from anywhere at any time. Additionally, it has the potential to offer more material to a wider range of learners in a greater variety of contexts, as it has the ability to reach more students beyond the traditional classroom (Boyatt and Sinclair, 2012; Wolsey, 2008) and meeting their expectations (Katz, 2008).

The study carried out by Stein et al. (2013) comprises the benefits of cloud computing in education with respect to an educational point of view. We highlight some of their main suggestions below:

- *Software license cost savings:* a virtual cloud does not include any license requirements. Instead, software has a negotiated number of ‘seat’ licenses loaded on the virtual system which only restricts maximum limits for concurrent use.
- *Diversity of software choices:* The flexibility of choosing diverse software applications is a critical component of e-learning systems. In a classical scenario, different software applications often conflict with other programs on the same machine, creating time-consuming work for IT staff to get all the desired software running smoothly. As a result, IT personnel can inadvertently restrict pedagogical practices and approaches that may be beneficial, and contribute to the types of bottlenecks that too often occur when tools and information are provided to students in public schools. One of the main advantages offered by cloud computing to make the cost of providing advanced learning applications, both in amplifying the use of

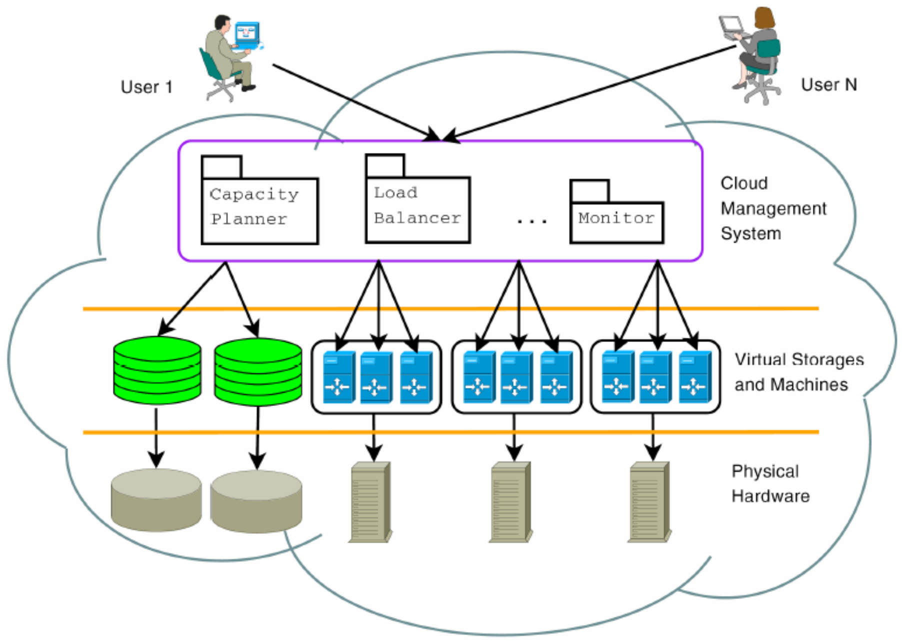
software licenses as discussed above, and in maximising the efficiency of IT personnel, affordable. An additional advantage that the virtual cloud brings to education can be seen in the use of multiple versions of the same software.

- *Affordability of class time delivery:* any software can be pre-loaded onto the cloud servers to meet the need for guaranteed on-time delivery.

3.3 Organisation of the cloud computing environment

Above we have seen the potential goodness of the implementation of a cloud computing environment in the educational sector. A further consideration is the type of cloud that such institutions need, i.e., the option of a private, public or hybrid cloud. Under these circumstances, a private cloud is mostly preferred for three reasons (Shyamala and Mukherjee, 2012): First, educational centres have the option of utilising their existing infrastructure which allows them to have a better control over the resources. Regarding this previous point, this implies a low cost process for setting up the cloud. Finally, the security issue is a major factor in the preference for this type of cloud. Certain activities such as the admission process for new students and the policies thereof, examinations/tests conducted for existing students and all research activities require confidentiality and breaches in security cannot be tolerated.

Figure 2 Overview of a cloud architecture for e-learning (see online version for colours)



Source: Taken from Sulistio et al. (2009)

Motivated by all the aforementioned benefits of integrating cloud computing in educational institutions, and especially those related to private clouds, throughout this section we will describe the characteristics of building a private cloud inside an educational institution, from both the architectural and functional point of view. The features for implementing a system with these characteristics are shared for other similar cloud applications. The architecture of a cloud computing platform as depicted in Figure 2 is usually common to most e-learning approaches on the cloud, which includes three main layers: the physical hardware at the bottom, a virtualisation environment at the next level above, and the cloud management system and service layer at the top. A description of each layer is provided below.

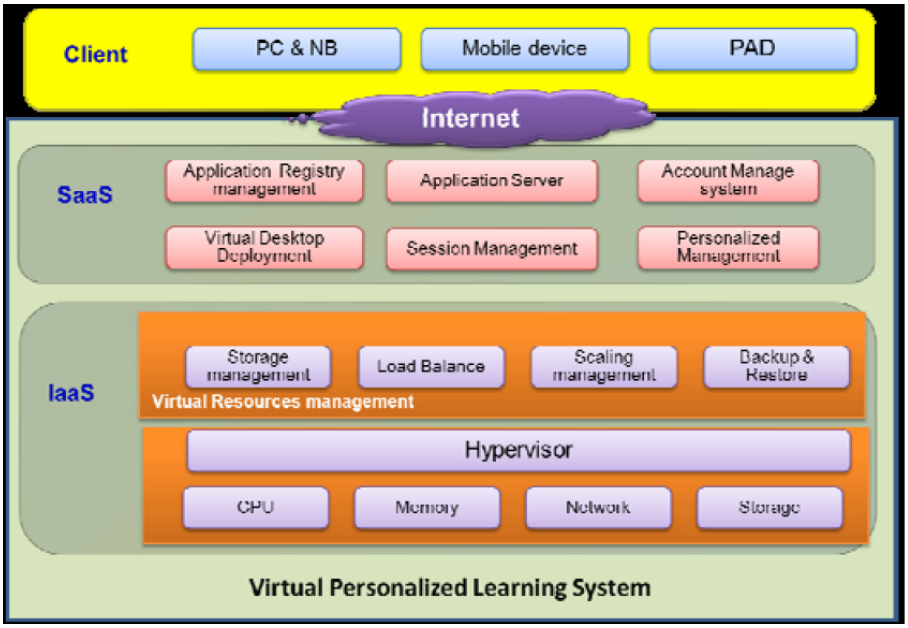
- *The physical layer:* The private cloud architecture is built on top of an existing hardware infrastructure that consists of two computer pools used for teaching purposes: a PC pool with a thin client [such as Thinstation (http://www.thinstation.org/)] and a server pool with the hypervisor (vSphere), both of them managed using VMware's Vsphere Vcenter which has the important feature of ensuring that the most important applications get the computing, network and storage resources they need by prioritising workloads. In this way, all hosts and services can be visualised and managed through a web browser, displaying real-time data about the virtual infrastructure including performance, configuration, storage, alarms and access permissions.
- *The virtualisation layer:* In order to allow multiple operating systems to share a single hardware host, a hypervisor is needed. A hypervisor is a program that controls the host processor and resources by allocating what is needed to each operating system in turn and making sure that the virtual machines cannot disrupt each other. A native hypervisor, which runs directly on the physical hardware is a better choice in this case, such as the VMware's Vsphere ESXi (http://www.vmware.com/products/vsphere/overview.html).
- *The service layer:* This layer is the interface with the cloud environment, and provides software for supporting the PaaS and the SaaS the cloud users need. In this case, virtual machines are created by choosing base images and post installing software packages selected by the course coordinators. In this way, predefined development tools are defined for carrying out different course projects, so that the students simply access the required VM using the thin client.

Additionally, in Liang and Yang (2011) the authors describe the functions used in the cloud IaaS and Saas which are required in the development of such a system. These features can be observed in Figure 3 and are enumerated below:

- from the IaaS perspective:
  1. storage management for the learning system and the users
  2. load balance for all learning systems
  3. scaling management for virtual machines
  4. backup and restore for the learning applications
- from the SaaS perspective:
  1. application registry management for the commercial provider, in order to register their applications
  2. application server for managing and deploying the subscribed learning contents to the users
  3. account management system for the authorised users

4 virtual desktop deployment to provide the personalised desktop including the subscribed learning contents
5 session management to provide the virtual desktop used by the authorised user
6 personalised management for the subscription of favourite learning contents.

Figure 3 The architecture of the virtual personalised learning environment (see online version for colours)



Source: Taken from Liang and Yang (2011)

3.4 Studies and applications of cloud computing for e-learning

Nowadays, the combination of cloud technologies and e-learning has been insufficiently explored. The pedagogical possibilities of the collaborative aspect of the cloud are studied in Stevenson and Hedberg (2011). In the aforementioned work, the authors refer to the lack of research that might provide a theoretical framework with which a pedagogy could be developed. However, the mobility inherent in the cloud approach could be stressed as a major advantage in developing conceptual frameworks and establishing effective teaching and learning strategies (Park, 2011; Taylor et al., 2006).

The drawback in this area is that there are few reports which provide a specific treatment of the matter. Instead, in the literature, the general attributes of the cloud are related to social interaction and collaborative working (Boyatt and Sinclair, 2012). For example, Blau and Caspi (2009) investigate students' perceptions of quality and ownership with respect to different modes of collaboration within Google Docs. Kittle and Hicks (2009) discuss practical activities which utilise technology to change and improve the collaborative experience of students writing a joint paper. Furthermore, we can find some cloud-related works for performing a comparison of the efficiency of online models versus traditional models (Hu and Zhang, 2010). The most representative work in this area is developed in Vaquero (2011), in which the authors focus on the impact of supporting technologies or the perceived ease of use and acceleration of the learning process. Furthermore, they analyse the appropriate level of abstraction (i.e., IaaS or PaaS) that should be delivered to students to enable them to focus on the course topics.

Regarding the use of cloud computing for educational purposes, we can distinguish between the public cloud and the private cloud solutions. We will review these two different approaches in the following sections.

3.4.1 Public cloud solutions in education

The interest of academic organisations in cloud computing is growing fast. This is shown by the increasing number of platforms, applications and services being developed to this end. Below, we enumerate some examples of enterprises which are publicising their services in this framework (Alshuwaier et al., 2012):

- *Amazon Education Cloud Computing:* this is aimed at assisting educators in providing cloud computing instruction by using the Amazon web services (AWS). AWS provides research and teaching grants for academic researchers using AWS in their work, and also tutorial and project grants for the student organisations using AWS for self-directed learning.
- *Microsoft Live@edu:* this gives students and researchers the ability to make full use of the same Microsoft technologies in the educational institution (Microsoft-Education, 2013). It is usable through popular web browsers for various types of operating systems.
- *Google App Education:* this enables its users such as the faculty, researchers, students and so on, to operate web applications within the Google infrastructure. It is available at no cost to institutions, universities and education communities. Specifically, Google Apps Education Edition (http://www.google.com/apps/intl/en/edu/index.html) lets technical administrators provide a collection of web-based messaging tools such as Google Mail, Google Talk, Google Sites , Google Video and Google Calendar to the faculty, students and staff for free in addition to productivity and collaboration tools such as Google Docs Package.
- *IBM SmartCloud for Education:* this a set of cloud services designed to help education systems leverage predictive analysis to get real-time insights into educators and institutional performance, enhance researcher effectiveness, and alleviate constrained lab resources for learning purposes. Educators can also benefit from the self-service reservation of, and seamless access to virtual computer resources both on campus and on the IBM public cloud (IBM-Cloud-Academy, 2013).

3.4.2 Private cloud solutions in education

As stated in Section 3.3, the development of a private cloud in an educational centre is a straightforward solution that provides many benefits from a technological perspective. There are several examples of private clouds infrastructures that have already been developed such as the ‘Cloud Infrastructure and Application’ (CloudIA) (Doelitzscher et al., 2011) at the Hochschule Furtwangen University (Germany), which is mainly dedicated to HFU staff and students running e-Learning applications, and external users for collaborative purposes. It is a framework which provides on-demand creation and configuration of VM images so that students are able to have their own Java servlet environment for experimentation, containing MySQL, Tomcat, PHP, and Apache web server. With this approach, students can focus more on developing, deploying and testing their applications in a servlet container.

Additionally, the 'Virtual Computing Laboratory' (VCL) (Vouk et al., 2008), developed by North Carolina State University (USA) enables students to reserve and access virtual machines (VMs) with a basic image or specific applications environments, such as Matlab and Autodesk along with physical computing hardware management.

Another example of an application that can be found in the specialised literature is the BlueSky cloud framework (Dong et al., 2009a) developed by Xi'an Jiaotong University (China). Its architecture has several components aimed at the efficient provision and management of e-learning services, and is able to preschedule resources for the hot contents and applications before they are needed, to safeguard the performance in concurrent accesses, although no details have been found with regard to how this is achieved.

In Liang and Yang (2011), the authors present a new service model that enhances efficiency within a virtual personalised learning environment. This system is intended to subscribe the selected learning resources as well as to create a personalised virtual classroom, and allows the learning content providers to register their applications in the server and the learners to integrate other internet learning resources to their learning application pools.

The Seattle cloud computing platform (Cappos et al., 2009) has been developed for educational networking. It is a free, portable, and lightweight platform using donated cloud computing. Seattle allows students to learn the concepts of networking and distributed systems on computers spread throughout the internet. It can also emulate cloud computing, peer-to-peer computing, and classify computing within a simple area.

Finally, Noor et al. (2010) proposed an architecture of cloud computing for education based on the availability of widespread resources for the entirety Bangladesh. This architecture may provide an effective and flexible way of matching resources with the current economic condition through the utilisation of unused resources and the abstraction of third party involvements, as well as providing a more flexible environment so that the client can now configure his own security policy.

3.5 Future challenges for e-learning systems in the cloud

Above, we have enumerated a wide number of positive features related to the development of a cloud platform in educational centres that might improve the current e-learning systems. However, some drawbacks remain to be overcome in order to exploit the full potential of this paradigm (Masud and Huang, 2012). From a technological point of view, in Section 2.4 we have already pointed out several challenges that may affect the decision to adopt cloud computing. In what follows, we identify those problems for the e-learning scenario in order to define future work on the topic:

- Security and data privacy are two major concerns in the migration to a cloud computing environment. The multi-tenancy, reuse of hardware and software profiles, and resiliency due to the redundant nature of cloud implies a greater risk of incomplete or unlocked deletion or denial of service attacks on institutions. Furthermore, the data requires encryption technology for storage and transmission, and a robust system for preventing attacks. It is necessary to find a way of implementing security with the least effect on performance, as this is one of the most important aspects of cloud in which academia is interested. Finally, in the case of public clouds, priority access to the data does not pertain the owner, but the cloud computing service provider. Therefore, we cannot rule out the possibility that e-learning confidential data may be leaked.

- The availability of cloud services is a further concern. Since all e-learning cloud computing services are internet-based, a lot of network traffic is generated, so it will work quite slowly for connections with a low bandwidth. Obviously, it also requires constant connection to the internet, thus limiting the guarantee of a permanent service. Additionally, cloud-based software services offer less features than standard applications.

- Users need to add tenant-specific data extensions without affecting the overall data schema. Best practice designs include using a pre-allocated field in each table to hold custom XML data extensions. In any case, if data portability is not supported, this causes a dependency on a particular cloud service provider for service preparation.

- The introduction of cloud computing to e-learning is the introduction of a market mechanism, so how to charge becomes a particularly important problem. Charges can be made for educational institutions or for personnel, but it is very complex. To set up a market-oriented charging mechanism, and combine two types of fees: institutional fees and individual fees, with institutional charging for general resources and individual charging for special resources is essential for the successful implementation of cloud-based solutions.

- Scalability issues are critical for adapting to the changing demands of the resources. Effective monitoring must be included in order to optimise the scheduling (Brzozowska et al., 2012). Fortunately, for standard e-learning purposes, it is straightforward to deliver virtual machines to students' labs, i.e., the working hours of the required virtual machines are known and are strictly provided by university staff. Therefore, these machines are easy to predict and have typical configurations. In most cases, low performance is required but accessibility is critical. However, in other cases, machine requirements have load peaks that must be managed properly.

- Teachers should play leading roles and participate in developing and making use of e-learning cloud. Establishing a suit of comprehensive management rules including teaching content management, course management, examination management, performance management, student management, teacher workload management and so on should be undertaken when implementing a cloud-based system. The primarily disadvantage is the period of adaption required by educators to a new educational context in which new technologies are coupled with innovative methodologies for creating and publishing didactic resources (la Prieta et al., 2012).

4 EDM and cloud computing

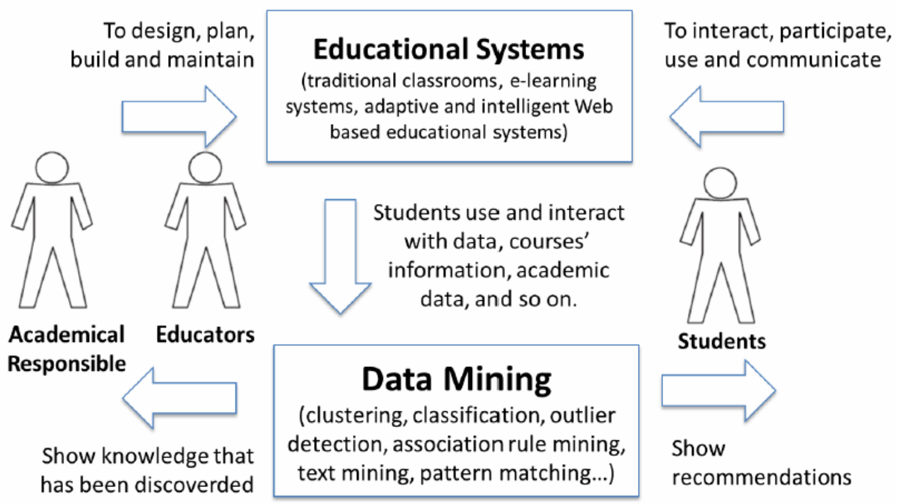
EDM is an emerging inter-disciplinary research area that deals with the development of methods to explore data originating from an educational context (Romero and Ventura, 2007, 2010). In this field, different types of educational data are addressed with machine learning algorithms with the objective of analysing hidden information that will enable the resolution of educational research issues. In accordance with this, its aim is to develop specific methodologies that can deal with the features of the educational settings and, by the application of these models, better understand the behaviour of the students and adapt to the context of learning.

In this section, we will first present a brief introduction to EDM in order to provide some preliminary concepts regarding this topic. Taking into account the features described in the former subsection, we then carry out a short discussion on how cloud computing can be used to take advantage of EDM processes by enhancing its productivity and reducing its cost.

4.1 Introduction to EDM

The EDM process follows the same steps as general data mining (see Figure 4). First, it starts with the preprocessing of the data in order to prepare it for the learning stage, i.e., using cleaning techniques, reducing the number of instances and/or input variables to facilitate the working of the algorithms, and so on. Then, the learning algorithms are applied to extract useful information from the data, whose type will depend on the final aim of the user, such as classification, clustering, association rule mining, sequential mining or text mining. Finally, a postprocessing step can be carried out to enhance the obtained results or to provide a more interpretable representation of the system.

Figure 4 The cycle of applying data mining in educational systems (see online version for colours)



Source: Taken from Romero and Ventura (2007)

However, it is worth pointing out that EDM has a broader purpose than just applying traditional data mining techniques. It is used in a wider sense by including other approaches, such as regression, correlation, or visualisation. All these approaches are necessary to carry out a better treatment of the educational data for all kind of purposes and to generate summarisation reports that facilitate the understanding of the extracted knowledge.

From a general perspective, EDM allows us to discover new knowledge based on students' usage data. Its aims include the validation and evaluation of educational systems, to potentially improve some aspects of the quality of education and to lay the groundwork for a more effective learning process. The specific objectives that are addressed in the framework of EDM are applied research to improve the learning process and guide students' learning and pure research to achieve a deeper understanding of educational phenomena.

The great advantage of the EDM methodology is its use in research and in building models in various areas that can influence e-learning systems. One of the main areas is user modelling, covering what the student knows, his/her behaviour and motivation, how the user experience is, and how different user requirements are satisfied by e-learning (Bienkowski et al., 2012).

The data that is managed in this scenario have special properties such as intrinsic semantic information, for example taking pedagogical aspects of the learner and the system into account. Additionally, they often present relationships with other data. An example of the former is to represent the data in a Q-matrix, which shows relationships between items/questions of a test/quiz system and the concepts evaluated by the test. Also, observing the domain model we may discover a relationship among the concepts of a specific subject in a graph or hierarchy (multiple hierarchy levels), i.e., a course is composed of chapters, which are divided into several lessons, and each lesson includes diverse concepts. According to these facts, educational problems have special characteristics which imply that standard methods have to be adapted to the specific educational problem at hand, or rather that it is necessary to implement specific data mining techniques to properly solve the problem.

To conclude this section, we highlight the main topics that can be developed in the field of EDM below:

- *analysis and visualisation of data:* to highlight useful information and support decision making
- *providing feedback for supporting instructors:* this feedback is also used to support course authors/teachers/administrators in decision making
- *recommendations for students:* according to the student's profile, the system can give him/her some personalised activities, links to visits, new tasks or problems, and so on
- *predicting student's performance:* this is used to estimate unknown values that describe the student
- *student modelling:* to develop cognitive models of human users/students, including a modelling of their skills and declarative knowledge
- *detecting undesirable student behaviours:* some patterns of unusual behaviour or some type of problem may be detected and, in this manner, erroneous actions, low

motivation, academic failure and the dropping out of students, may be identified as potential problems and thus avoided

- *grouping students:* according to their customised features, personal characteristics, etc.
- *social network analysis:* studying relationships between individuals
- *developing concept maps:* constructing concept maps is aimed at helping instructors/educators in the automatic process of developing/constructing concept maps
- *constructing courseware:* to help instructors and developers to carry out the construction/development process of courseware and to learn contents automatically
- *planning and scheduling*.

4.2 Taking advantage of cloud computing for EDM

We have noted an increase in the instrumental educational software in recent years. This has made the gathering of more and more information for teaching purposes possible. As we have commented above, instructors may use different frameworks that apply data mining algorithms to the representation and extraction of dynamic learning processes and learning patterns to support students' deep learning, efficient tutoring and collaboration in web-based learning environments (Romero and Ventura, 2008; Guo and Zhang, 2009).

Among the different techniques that can extract knowledge from these databases, we may find data mining tools that apply association rule mining and collaborative filtering (García et al., 2011), the use of genetic algorithms (Minaei-Bidgoli and Punch, 2003) which are known to be effective techniques for achieving good results in large spaces of data, or ensemble techniques for online learning processes (Kotsiantis et al., 2010).

The active use of e-learning systems implies the creation of large repositories of data. These large amounts of information about teaching-learning interaction are endlessly generated and ubiquitously available (Castro et al., 2007), and reflect how students learn (Koedinger et al., 2008). However, the use of the aforementioned data mining techniques is limited in regard to these huge data-sets, because the quantity of computational resources needed to support an intelligent system with these characteristics is very large.

The scenario we are currently addressing is known as the Big Data problem (Madden, 2012). In this framework, the application of a cloud computing environment is mandatory. Due to the possibilities that cloud computing offers with respect to high and easily scalable computing and storage resources, the development of EDM over cloud computing makes sense. Standard software and computer systems may lose many useful pedagogical features as they are unable to manage all of this information in a reasonable time. Please recall that the final aim of EDM is to reinforce teaching and learning by enabling a detailed feedback from the students' results.

We must stress that there have been relatively few scientific journal papers published since 2008 which cover the topic of cloud computing and, to the best of our knowledge, even fewer for those related to e-learning and EDM tasks. Studying the specialised literature we have observed conceptual similarities between EDM and e-commerce systems (Burke et al., 2011; Sever, 2011), which are aimed at determining clients' interests so as to be able to increase online sales. Accordingly, we can migrate the basic structure that has already been proposed for e-commerce and business intelligence solutions (Abelló and Romero, 2012; Jun and Jun, 2011) to the field of EDM. The most suitable management organisation architecture is based on a four layer architecture:

- A file system for the storage of the Big Data, i.e., a large amount of archives of high capacity. This layer is implemented within the IaaS level as it defines the basic architecture organisation for the remaining tiers.

  Hadoop Distributed File System (HDFS) (White, 2009) is the open source project of Apache that implements the Google File System, the initial solution conceived to deal with this problem (Ghemawat et al., 2003). Hadoop installation is composed of multiples nodes, which are divided into two classes: a Master node (namenode) and a large number of fragments storages or datanodes. Files are divided into fixed size chunks of 64 megabytes, in a similar way to the clusters or sectors of the traditional hard disk drives. Datanodes store these fragments, which are assigned a unique id label of 64 bits in the name node when it is generated. The HDFS features parallel access and fault-tolerantcy.

- A database management system for organising the data and which is able to access them in an efficient way. It can be viewed as occupying a space between IaaS and PaaS, since it shares common characteristics with both schemes. It is used by developers to access the data, but its implementation lies in the hardware level.

  A new structure called 'Not Only SQL (NOSQL)' databases emerges (Han et al., 2011) to facilitate data organisation in this context. The main difference with standard relational databases is to allow horizontal scalability of the data and to modify the storage and retrieval of key/value pairs (as opposed to the foreign-key/primary-key relationships-based model).

- An execution tool to distribute the computational load among the computers of the cloud. This layer is clearly related to PaaS, as it is a kind of 'software API' for the codification of the applications.

  The most well known example of a cloud computing execution environment is probably Google MapReduce (Dean and Ghemawat, 2008) and its open source implementation, Hadoop (The-Apache-Software-Foundation, 2012). These environments aim at providing elasticity by enabling the adjustment of resources according to the application, handling errors transparently and ensuring the scalability of the system.

- A query language for extracting the knowledge and information required by the user of the system. This will be the interface to the user and will provide transparency to the other tiers of the architecture. In this case, we may assume that the tools related with the query language are also at the PaaS level, following the same considerations as in the previous case. Then, the SaaS application will be built on top of all these layers at the higher level of abstraction.

  MapReduce programming model is quite restrictive with respect to its format. It has a simple input and just two stages, in order to facilitate the programming, so that format queries can be translated into functional data flow programs. Several data processing and query languages have been designed for NOSQL-type databases. The idea behind this is to achieve a trade-off between the declarative style of SQL and the low-level/procedural style of MapReduce.

It will be noted that this system has significant similarities to that introduced for e-learning purposes in Section 3.3, specifically at the lower layers of the infrastructure. In this way, all the guidelines introduced in the former section remain valid for that which we are currently describing. Both systems may coexist, but it is also possible to include a cloud computing framework for EDM to process data from a standard e-learning approach.

Additionally, the 'pros and cons' of the development of a cloud computing system for EDM are practically identical as in the case of e-learning tasks. As stated above, the advantages of being able to process more information in a shorter time are critical in allowing a quick response to improve learning quality in accordance with the students' progress with the current pedagogical tools. On the other hand, for a major deployment of the cloud environment, the personal and technological drawbacks that this new paradigm implies (Section 3.5), remain necessary to be overcome.

5 Concluding remarks

In this work we have identified the main components of e-learning, focusing on the flexibility, convenience, ease of access, consistency and repeatability of this kind of system. An e-learning system faces the challenge of optimising large-scale resource management and provisioning, in accordance to the huge growth of users, services, education contents and media resources. We have established the goodness of a cloud computing solution in this area of education.

The features of the cloud computing platform are appropriate for the migration of this learning system, so that we can fully exploit the possibilities offered by the creation of an efficient learning environment that offers personalised contents and easy adaptation to the current education model. The benefits of integrating an e-learning system into the cloud are, in particular, good flexibility and scalability for the resources, including storage, computational requirements and network access. We must also stress the saving of both software and hardware. On the one hand, it provides a greater diversity of educational programs with a reduced license cost. On the other hand, since machine life is extended, the replacement rate for students' computers is reduced. Those savings are amplified by the reduction in IT personnel costs related to maintaining computer labs and updating software.

We have enumerated several approaches that have already been proposed to address e-learning on cloud computing, describing these models and how they take advantage of this environment to enhance the features of the educational system. The significance of the application of a cloud computing platform for EDM has also been stressed. We have emphasised the goodness of EDM for enhancing the quality of current education from a pedagogical point of view, but it has the disadvantage of requiring the management of a large amount of information, usually generated within e-learning courses. In order to extract useful knowledge from it by running data mining algorithms, and to efficiently present the reports obtained from them to the instructors, the computational advantages of this new paradigm of cloud computing are evident.

In spite of all of the above, the actual use of cloud computing in educational centres is at an early stage. Human users are always reluctant to adopt any technological change, from teachers that must learn the use of a new software paradigm, to IT personnel that must change their policies in order to make the implementation of this new technology possible. Other related challenges are security issues regarding online access, and the confidentiality concerns of users with regard to the storage of their data in a public cloud environment.

As we have mentioned above, these are just the initial steps towards an open line for the research and exploitation of online learning systems. For the full establishment of this paradigm, further research should include the simplification of the deployment of a private cloud, considering the hardware infrastructure and management issues of the system, as well as stressing the impact cloud delivery may have on greater advancements in pedagogical effectiveness.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science and Technology under projects TIN2011-28488, TIN2009-14575 and the Andalusian Research Plans P10-TIC-6858 and P11-TIC-7765.

References

Abelló, A. and Romero, O. (2012) 'Service-oriented business intelligence', *eBISS 2011*, Vol. 96 of LNBIP, pp.156–185.

Agrawal, D., Das, S. and Abbadi, A.E. (2011) 'Big data and cloud computing: current state and future opportunities', in Ailamaki, A., Amer-Yahia, S., Patel, J.M., Risch, T., Senellart, P. and Stoyanovich, J. (Eds.): *EDBT*, pp.530–533, ACM, Uppsala, Sweden.

Alonso, G., Casati, F., Kuno, H. and Machiraju, V. (2004) *Web Services: Concepts, Architectures and Applications*, Springer, Heidelberg.

Alshuwaier, F.A., Alshwaier, A.A. and Areshey, A.M. (2012) 'Effective use of cloud computing services in education', *Journal of Next Generation Information Technology (JNIT)*, Vol. 3, No. 4, pp.62–77.

Al-Zoube, M., El-Seoud, S.A. and Wyne, M.F. (2010) 'Cloud computing based e-learning system', *Intl. Arab Journal of e-Technology*, Vol. 8, No. 2, pp.58–71.

Bensch, S. and Rager, M. (2012) 'Cloud-based online learning platforms', in Abramowicz, W., Domingue, J. and Wecel, K. (Eds.): *BIS (Workshops)*, Vol. 127 of Lecture Notes in Business Information Processing, pp.165–176, Springer.

Bienkowski, M., Feng, M. and Means, B. (2012) *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief Technical Report*, US Department of Education, Office of Educational Technology, Washington, DC, USA.

Blau, I. and Caspi, A. (2009) 'What type of collaboration helps?', *Chais Conference on Instructional Technologies*.

Boyatt, R. and Sinclair, J. (2012) 'Navigating the educational cloud', *Workshop on LTEC 2012*, Vol. 173 of AISC, pp.179–191.

Brzozowska, A., Greblicki, J. and Kotowski, J. (2012) 'Cloud computing in educational applications methods of virtual desktops deployment', in Moreno-Díaz, R., Pichler, F. and Quesada-Arencibia, A. (Eds.): *EUROCAST 2011*, Part I, Vol. 6927 of LNCS, pp.568–575.

Burke, R., Felfernig, A. and Göker, M.H. (2011) 'Recommender systems: an overview', *AI Magazine*, Vol. 32, No. 3, pp.13–18.

Buyya, R., Broberg, J. and Goscinski, A. (2011) *Cloud Computing: Principles and Paradigms*, Wiley, Chichester.

Buyya, R., Yeo, C., Venugopal, S., Broberg, J. and Brandic, I. (2009) 'Cloud computing and emerging it platforms: vision, hype, and reality for delivering computing as the 5th utility', *Future Generation Comp. Syst.*, Vol. 25, No. 6, pp.599–616.

Cappos, J., Beschastnikh, I., Krishnamurthy, A. and Anderson, T. (2009) 'Seattle: a platform for educational cloud computing', in Fitzgerald, S., Guzdial, M., Lewandowski, G. and Wolfman, S.A. (Eds.): *SIGCSE*, pp.111–115, ACM, Chattanooga, TN, USA.

Castro, F., Vellido, A., Nebot, A. and Mugica, F. (2007) 'Applying data mining techniques to e-learning problems', in Jain, L.C., Tedman, R. and Tedman, D. (Eds.): *Evolution of Teaching and Learning Paradigms in Intelligent Environment*, Vol. 62 of Studies in Computational Intelligence, pp.183–221, Springer-Verlag, New York.

CDW-G (2011) *The 2011 Cloud Computing Tracking Poll* [online] http://newsroom.cdwg.com/features/feature-05-26-11.html (accessed October 2012).

Dean, J. and Ghemawat, S. (2008) 'MapReduce: simplified data processing on large clusters', *Commun. ACM*, Vol. 51, No. 1, pp.107–113.

Doelitzscher, F., Sulistio, A., Reich, C., Kuijs, H. and Wolf, D. (2011) 'Private cloud for collaboration and e-learning services: from IaaS to SaaS', *Computing*, Vol. 91, No. 1, pp.23–42.

Dong, B., Zheng, Q., Qiao, M., Shu, J. and Yang, J. (2009a) 'Bluesky cloud framework: an e-learning framework embracing cloud computing', in Jaatun, M.G., Zhao, G. and Rong, C. (Eds.): *CloudCom*, Vol. 5931 of Lecture Notes in Computer Science, pp.577–582, Springer.

Dong, B., Zheng, Q., Yang, J., Li, H. and Qiao, M. (2009b) 'An e-learning ecosystem based on cloud computing infrastructure', *ICALT*, pp.125–127, IEEE, Riga, Letonia.

Duer, W. (2003) *CRM, Customer Relationship Management*, MP editions, Madrid, España.

Ercan, T. (2010) 'Effective use of cloud computing in educational institutions', *Procedia – Social and Behavioral Sciences*, Vol. 2, No. 2, pp.938–942.

Foster, I.T., Zhao, Y., Raicu, I. and Lu, S. (2009) 'Cloud computing and grid computing 360-degree compared', *Grid Computing Environments Workshop, 2008 (GCE'08)*, pp.1–10.

Furht, B. and Escalante, A. (Eds.) (2010) *Handbook of Cloud Computing*, Springer US, Boston, MA.

García, E., Romero, C., Ventura, S. and de Castro, C. (2011) 'A collaborative educational association rule mining tool', *Internet and Higher Education*, Vol. 14, No. 2, pp.77–88.

Geith, C. (2008) 'Teaching and learning unleashed with Web 2.0 and open educational resources', *Physical and Virtual Learning Spaces in Higher Education: Concepts for the Modern Learning Environment,* pp.219–226, Information Science Reference; IGI publishers, Hershey, PA, USA.

Ghemawat, S., Gobioff, H. and Leung, S-T. (2003) 'The Google file system', *19th Symposium on Operating Systems Principles*, pp.29–43.

Guo, Q. and Zhang, M. (2009) 'Implement web learning environment based on data mining', *Knowl. Based Syst.*, Vol. 22, No. 6, pp.439–442.

Han, J., Haihong, E., Le, G. and Du, J. (2011) 'Survey on NOSQL database', *Proceedings of the 2011 6th International Conference on Pervasive Computing and Applications (ICPCA 2011)*, pp.363–366.

Han, Y., Brezany, P. and Janciak, I. (2009) 'Cloud-enabled scalable decision tree construction', *Proceedings of the 2009 Fifth International Conference on Semantics, Knowledge and Grid, SKG '09*, pp.128–135, IEEE Computer Society, Washington, DC, USA.

Hu, Z. and Zhang, S. (2010) 'Blended/hybrid course design in active learning cloud at South Dakota State University', *2nd ICETC*, Vol. 1, PP.V1-63–V1-67.

Hurwitz, J., Bloor, R., Kaufman, M. and Halper, F. (2010) *Cloud Computing for Dummies*, John Wiley and Sons, Hoboken, NJ, USA.

IBM-Cloud-Academy (2013) *Education for a Smarter Planet: Cloud Computing in Education* [online] http://www.cetpa-k12.org/ (accessed November 2013).

JISC (2012) *Cloud Services for Education and Research* [online] http://www.jisc.ac.uk/news/stories/2011/06/cloudservices.aspx (accessed October 2012).

Jolliffe, A., Ritter, J. and Stevens, D. (2001) *The Online Learning Handbook: Developing and Using Web-based Learning*, Kogan Page, London.

Jun, L. and Jun, W. (2011) 'Cloud computing based solution to decision making', *Procedia Engineering*, Vol. 15, No. 1, pp.1822–1826.

Katz, R. (2008) 'The gathering cloud: is this the end of the middle?', in *Physical and Virtual Learning Spaces in Higher Education: Concepts for the Modern Learning Environment*, pp.2–42, Information Science Reference; IGI publishers, Hershey, PA, USA.

Katzan, H. (2010) 'The education value of cloud computing', *Contemporary Issues in Education Research*, Vol. 3, No. 7, pp.37–42.

Kittle, P. and Hicks (2009) 'Transforming the group paper with collaborative online writing', *Pedagogy*, Vol. 9, No. 3, pp.525–538.

Koedinger, K.R., Cunningham, K., Skogsholm, A. and Leber, B. (2008) 'An open repository and analysis tools for fine-grained, longitudinal learner data', in de Baker, R.S.J., Barnes, T. and Beck, J.E. (Eds.): *1st Int. Conf. Educational Data Mining*, pp.157–166 [online] http://www.educationaldatamining.org (accessed October 2012).

Kotsiantis, S.B., Patriarcheas, K. and Xenos, M.N. (2010) 'A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education', *Knowl. Based Syst.*, Vol. 23, No. 6, pp.529–535.

Kwan, R., Fox., R., Chan, F. and Tsang, P. (2008) *Enhancing Learning Through Technology: Research on Emerging Technologies and Pedagogies*, World Scientific, Hackensack, NJ, USA.

la Prieta, F.D., de Paz-Santana, J.F., Gil, A.B. and Boureux, M. (2012) 'Retrieving learning resources over the cloud', *Workshop on LTEC 2012*, Vol. 173 of AISC, pp.47–56.

Liang, P-H. and Yang, J-M. (2011) 'Virtual personalized learning environment (VPLE) on the cloud', *Proceedings of the 2011 International Conference on Web Information Systems and Mining, WISM'11*, Springer-Verlag, Berlin, Heidelberg, Part II, pp.403–411.

Madden, S. (2012) 'From databases to big data', *IEEE Internet Comput.*, Vol. 16, No. 3, pp.4–6.

Masud, A.H. and Huang, X. (2011) 'Esaas a new education software model in e-learning systems', in Zhu, M. (Ed.): *ICCIC 2011*, Vol. 235 of CCIS, pp.468–475.

Masud, A.H. and Huang, X. (2012) 'A novel approach for adopting cloud-based e-learning system', in Miao, H., Lee, R.Y., Zeng, H. and Baik, J. (Eds.): *ACIS-ICIS*, pp.37–42, IEEE.

Mateescu, G., Gentzsch, W. and Ribbens, C.J. (2011) 'Hybrid computing – where HPC meets grid and cloud computing', *Future Generation Comp. Syst.*, Vol. 27, No. 5, pp.440–453.

Mayer, R. and Clark, R. (2011) *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, 3rd ed., Wiley and Sons, Hoboken, NJ, USA.

Mell, P. and Grance, T. (2011) *The NIST Definition of Cloud Computing*, Vol. 145, No. 6, (Draft) Recommendations of the National Institute of Standards and Technology, NIST Special Publication.

Microsoft-Education (2013) *Cloud Computing from Microsoft – Empowering Education Through Choice* [online] https://partner.microsoft.com/NZ/40142863/ (accessed November 2013).

Minaei-Bidgoli, B. and Punch, W.F. (2003) 'Using genetic algorithms for data mining optimization in an educational web-based system', in Cantú-Paz, E., Foster, J.A., Deb, K., Davis, L., Roy, R., O'Reilly, U-M., Beyer, H-G., Standish, R.K., Kendall, G., Wilson, S.W., Harman, M., Wegener, J., Dasgupta, D., Potter, M.A., Schultz, A.C., Dowsland, K.A., Jonoska, N. and Miller, J.F. (Eds.): *GECCO*, Vol. 2724 of Lecture Notes in Computer Science, pp.2252–2263, Springer.

Noor, S.A., Mustafa, G., Chowdhury, S., Hossain, Z. and Jaigirdar, F. (2010) 'A proposed architecture of cloud computing for education system in Bangladesh and the impact on current education system', *International Journal of Computer Science and Network Security*, Vol. 10, No. 10, pp.7–13.

Ouf, S. and Nasr, M. (2011) 'Business intelligence in the cloud', *IEEE 3rd International Conference on Communication Software and Networks (ICCSN2011)*, pp.650–655.

Papazoglou, M. and Van Den Heuvel, W-J. (2007) 'Service oriented architectures: approaches, technologies and research issues', *VLDB J.*, Vol. 16, No. 3, pp.389–415.

Park, Y. (2011) 'A pedagogical framework for mobile learning', *The International Review of Research in Open and Distance Learning*, Vol. 12, No. 2, pp.78–102.

Pocatilu, P., Alecu, F. and Vetrici, M. (2010) 'Measuring the efficiency of cloud computing for e-learning systems', *W. Trans. on Comp.*, Vol. 9, No. 1, pp.42–51.

Romero, C. and Ventura, S. (2007) 'Educational data mining: a survey from 1995 to 2005', *Expert Syst. Appl.*, Vol. 33, No. 1, pp.135–146.

Romero, C. and Ventura, S. (2008) 'Data mining in course management systems: Moodle case study and tutorial', *Computers and Education*, Vol. 51, No. 1, pp.368–384.

Romero, C. and Ventura, S. (2010) 'Educational data mining: a review of the state of the art', *IEEE Trans. Syst., Man, Cybern. C*, Vol. 40, No. 6, pp.601–618.

Sever, A. (2011) 'A problem of data mining in e-commerce', *Applied Mathematics and Computation*, Vol. 217, No. 24, pp.9966–9970.

Shyamala, L. and Mukherjee, S. (2012) 'Educloud: an institutional cloud with optimal scheduling policies', *ObCom 2011*, Part I, Vol. 269 of CCIS, pp.114–123.

Stein, S., Ware, J., Laboy, J. and Schaffer, H.E. (2013) 'Improving k-12 pedagogy via a cloud designed for education', *International Journal of Information Management*, Vol. 33, No. 1, pp.235–241.

Stevenson, M. and Hedberg, J. (2011) 'Head in the clouds: a review of current and future potential for cloud-enabled pedagogies', *Educational Media International*, Vol. 48, No. 4, pp.321–333.

Sulistio, A., Reich, C. and Doelitzscher, F. (2009) 'Cloud infrastructure and applications – cloudia', in Jaatun, M.G., Zhao, G. and Rong, C. (Eds.): *CloudCom*, Vol. 5931 of Lecture Notes in Computer Science, pp.583–588, Springer.

Sultan, N. (2009) 'Cloud computing for education: a new dawn?', *International Journal of Information Management*, Vol. 30, No. 2, pp.109–116.

Taylor, J., Sharples, M., O'Malley, C., Vavoula, G. and Waycott, J. (2006) 'Towards a task model for mobile learning: a dialectical approach', *Int. Journal of Learning Technology*, Vol. 2, No. 2, pp.138–158.

The-Apache-Software-Foundation (2012) *Hadoop, An Open Source Implementing of Mapreduce and GFS* [online] http://hadoop.apache.org (accessed October 2012).

Vaquero, L.M. (2011) 'Educloud: PaaS versus IaaS cloud usage for an advanced computer science course', *IEEE Transactions on Education*, Vol. 54, No. 4, pp.590–598.

Velte, A.T., Velte, T.J. and Elsenpeter, R. (Eds.) (2010) *Cloud Computing: A Practical Approach*, McGraw Hill, Columbus, OH, USA.

Vouk, M., Averitt, S., Bugaev, M., Kurth, A., Peeler, A., Shaffer, H., Sills, E., Stein, S. and Thompson, J. (2008) 'Powered by VCL – using virtual computing laboratory (VCL) technology to power cloud computing', *2nd Intl. Conference on the Virtual Computing Initiative (ICVCI)*, Research Triangle Park, North Carolina, USA.

Wheeler, B. and Waggener, S. (2009) 'Above-campus services: shaping the promise of cloud computing for higher education', *EDUCAUSE Review*, Vol. 44, No. 6, pp.52–67.

White, T. (2009) *Hadoop: The Definitive Guide*, 1st ed., O'Reilly Media, Sebastopol, CA, USA.

Witten, I., Frank, E. and Hall, M. (2011) *Data Mining. Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufman, Burlington, Massachusetts, USA.

Wolsey, K. (2008) 'Where is the new learning?', in *Physical and Virtual Learning Spaces in Higher Education: Concepts for the Modern Learning Environment*, pp.212–218, Information Science Reference; IGI publishers, Hershey, PA, USA.

Youseff, L. (2008) 'Toward a unified ontology of cloud computing', in Butrico, M. and Da Silva, D. (Eds.): *Grid Computing Environments Workshop, 2008 (GCE'08)*.