Proposal to Improve the Management System of Storage in a Virtual Learning Environment

Propuesta para Mejorar el Sistema Gestor de Almacenamiento en un Entorno E-learning

Moreno, Yaneth[1*]; Valera, Junior[2]; Mejía, Gustavo[2]; Hidrobo, Francisco[3]; Díaz, Gilberto[4]; Meneses, Jathinson[4]
[1]CEMISID, Facultad de Ingeniería, Universidad de Los Andes. Mérida, Venezuela
[2] Departamento de Computación, Facultad de Ingeniería, Universidad de Los Andes. Mérida, Venezuela
[3]SUMA, Facultad de Ciencias, Universidad de Los Andes. Mérida, Venezuela
[4] Facultad de Ingeniería, Universidad Industrial de Santander. Bucaramanga, Colombia
yanethmoreno2002@gmail.com

DOI: https://doi.org/10.53766/CEI/2021.43.02.09

Abstract

This article introduces a theoretical model of a storage management system for big data-based virtual learning environments. The system uses artificial intelligence technologies such as machine learning and deep learning to explore strategies for optimizing how stored data is processed. To this end, the storage management requirements currently presented by these platforms, as well as big data techniques and tools, and their applications on those platforms were investigated. This model is designed to be distributed, scalable, fast, and fault tolerant. In addition, it is expected to work with any web application that needs to improve file downloads.

Keywords: VLE, Big Data, Machine Learning, Deep Learning.

Resumen

En este artículo se presenta un modelo teórico de un sistema de gestión de almacenamiento para entornos e-learning basado en Big Data. Este sistema contempla estrategias para optimizar el manejo de los datos almacenados utilizando técnicas de inteligencia artificial como el Machine Learning y el Deep Learning. Para ello se estudiaron las necesidades en cuanto a gestión de almacenamiento que presentan actualmente estas plataformas, así como también se analizaron técnicas y herramientas de Big Data, y su aplicación en ellas. El diseño de este modelo está orientado a ser distribuido, escalable, rápido y con tolerancia a fallos. Además, se prevee que pueda usarse en cualquier aplicación web donde sea necesario mejorar la descarga de archivos.

Palabras clave: EVA, Big Data, Máquinas de Aprendizaje, Deep Learning.

1Introduction

E-learning platforms, supported by learning management systems (LMS), are one of the many web tools that educational institutions and companies currently use to offer their courses at a distance or in a blended way.

These courses range from workshops, degrees, undergraduate and post-graduate studies, amongst many other education and training programs. Examples of this, the Distance University of Madrid (UDIMA, for its acronym in Spanish) and the University of Phoenix, which has a non-face-to-face modality for full degrees. As well as universities such as Hardware and MIT that offer some online courses. Thus, the need arises to improve the performance of this type of web applications, with the aim of enriching the quality, in terms of the user experience. That is, to look at whether the service meets the expectations and needs of the user.That is, to seek that the service satisfies the expectations and needs of the user. Because your perception of the quality of classes can be affected if during your training the platform responds slowly or unexpectedly.

This work proposes the design of a model that optimizes storage management performance, by using artificial intelligence (AI) techniques on a Big Data infrastructure.

However, storage management in e-learning platforms is one of the most important aspects, since valuable resources are administered with which the training is transmitted to the student, such as files, and depending on the performance provided by this part of the platform. To ensure that these resources are effectively managed whenever needed by the user, the teaching and learning process will be perceived as timely and faster.

It is thus necessary to improve the performance of this type of web applications, in order to enrich the quality, in terms of user experience. In other words, make sure the service meets the expectations and needs of the user. Because your perception of the quality of classes can be affected if during your training the platform responds slowly or unexpectedly.

2State of the art.

To create a Storage Management System for e-learning environments, based on Big Data, that implements strategies that allow optimizing the management of stored data, it is convenient to refer to the documentary consultation of works that motivated this research and that are related to the objective set.

Among the studies detailing some e-learning platforms and their features (Clarenc et. al, 2013). He indicated that LMS can be used to create, approve, administer, store, distribute and manage virtual learning activities. Which can even be used as a supplement to face-to-face classes, and which are generally installed on a web server (it can be installed on an intranet).

Based on this research, there are many LMSs, each with its own tools and features. LMS can be of two types: proprietary, that is, licensed to use, among the best known are Blackboard, WebCT, OS Media, Saba, eCollege, Fronter, SidWeb, e-ducativa and Catedr@, among others; free software, are those that, once obtained, can be freely used, studied, changed and redistributed, among which are ATutor, Dokeos, Claroline, dotLRN, Moodle, Ganesha, ILIAS and Sakai. On the other hand, cloud computing has allowed the development of MOOCs (Massive Online Open Courses), which are managed on platforms such as: audacity, Coursera, Udemy, ex, Ecaths, Wiziq and Edmodo, to name a few.

There are also several works on the analysis and design of architectures for e-learning systems that make it possible to study the needs in terms of infrastructure. Among them is one (Álvarez et. al, 2016) and (Bourkoukoukou et. al, 2016). The first performs a review and comparison of some models or standards of software quality, based on the identification of common criteria among them, in order to identify the most complete criteria at the level of evaluation of virtual learning platforms; while the second focuses its work on the design of a personalized e-learning system, based on the learning profile. At present, e-learning technological platforms require storage systems that support more complex processes, which in addition to organizing and storing information, which in many cases does not have any type of structure. It also requires a series of predictive analyses that enhance the management of the resources on which these platforms are supported. The relevance of this is discussed in (Ashraf et. al, 2015) and (Ashraf et. al, 2015). They provide a model to effectively process large amounts of data and intelligently collect data from interactions between learners and content. This model can determine the behavior of all students in the e-learning system through Big Data, using tools such as Hadoop and NoSQL; whereas, (Marjanovic et. al, 2014) describes a model based on Big Data, using Hadoop to improve educational processes, integrated into the Moodle platform and capable of handling structured and unstructured data from heterogeneous sources.

Online learning platforms record the interactions of different actors, not just the content. This allows to characterize the teaching activity, as is the case of (Roldán 2016), where a set of mechanisms or algorithms have been developed that enable big data analysis of the information generated in the Moodle environment of the Valladolid University virtual campus. A study that records each module that makes up an application in a file called log. As described in (Gómez O, 2012), user-information system (IS) interaction is an important source of knowledge, so these logs determine if there is an operational problem with the platform. Used to perform analysis. Throughout its history, this has been made possible by policies that manage event registers (logs). This has provided the benefits reported by AI to enhance the security, operation, and management of IS business processes.

In this study, it was important to consider issues that might exist, primarily from a memory management perspective, and solutions that others could implement. Consider an e-learning platform or other web-based system. An example of this is (Ramachandra et. al, 2014), which suggests a way to optimize the performance of database-based web applications by automatically reloading the query results. To do this, the proposed system and method will automatically insert a pre-search instruction as early as possible. This is due to procedural calls in the application source code when conditional loops and branches are present. A data flow analysis technique called look-ahead analysis extends to the predictability of queries. The benefits of pre-acquisition are limited. This is because there is an assignment statement and a conditional branch before the query execution statement. Improvements such as code movement, chaining, and write forward are aimed at increasing the benefits of prefetching.

On the other hand, works such as (Kroth B, 2015) and (Merceron et. al, 2015) were helpful. First, they evaluated and analyzed various database configurations using the cache usage and user session data currently supported by Moodle in terms of scalability, reliability, security, and performance impact. This allowed them to find many anomalies and general performance optimizations for the various systems we investigated, as well as several opportunities to reduce network costs. The second paper describes the current state of data analysis in e-learning and identifies some trends in recent studies. In addition, several directions that this field may follow have been identified, including the incorporation of multimodal data (gestures, gaze tracking, biosensors) and the diversification of learning environments (MOOC, classrooms, hands-on learning).

According to (Esposito et. al, 2015), there is a strong need for a platform for managing growing amounts of data in contexts featuring complex event handling systems and multiple heterogeneous sources. This is done to efficiently handle broadcasts and data collection and análisis in a fully distributed manner. Their work has nothing to do with e-learning, but it serves as a model to consider when analyzing e-learning platforms to find knowledge-based solutions.

3Analysis of the e-learning platform.

According to (Bernárdez, 2010), e-learning systems consist of the following areas:

1. An **Integrated Learning System (ILS)** contains a variety of tools for synchronous and asynchronous learning. B. Access to virtual classrooms, discussion forums, presentation areas, and content and exercises.
2. A **Learning Management System (LMS)** that tracks each student's learning, participation, and mentoring status.
3. A **Learning Content anagement System (LCMS)** that stores and connects content to reusable modules.

When one (or more) e-learning platforms are combined with a virtual workspace, teachers, developers, content writers, and programmers collaborate to create and install online courses with training, which is what it is. You can get "e-performance". The latter is shown in Fig. 1. Here you can see that the learning platform is an integrated set of interactive online services. It's about between tutors and students, and between students. You can also perform evaluations, transfer files, forums, chat, and use a variety of other tools.

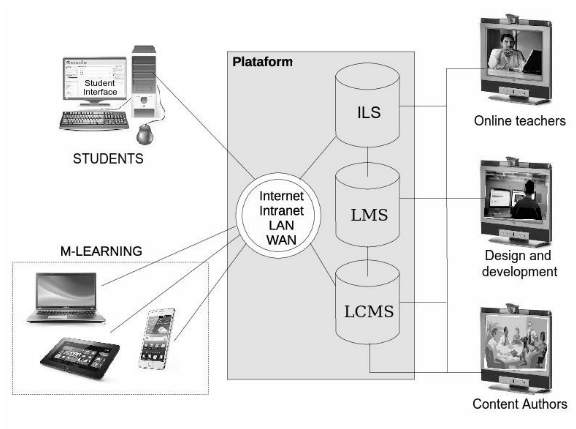

Fig. 1. E-learning System. Modified from (Bernárdez, 2010)

In general, these learning platforms are designed and implemented to support the e-learning process itself. This does not provide the advantage that some computing resources (especially storage networks) can be generated when managing the data generated by the e-learning process or considering the following characteristics:

Stability: Maintains current functionality despite potential changes or conversions.

Scalability: Deploy clusters that provide reasonable performance by increasing the number of available resources (scale up), automatically reducing them when your application needs them (scale down), or by autoscaling. This capacity is provided by cloud services that automatically increase or decrease, guaranteeing capacity during periods of high demand without wasting resources.

Standardization: The ability to extract and interpret information from different sources by importing or exporting between different versions or distributions of the same software. Standardization: The ability to extract and interpret information from different sources by importing or exporting between different versions or distributions of the same software.

Analyzing this informacion from the big data paradigm, it is a technology that is the trend of information management today. According to (Dumbill. 2012), what is defined as data that exceeds the computational power of traditional database systems because the data is too large, too fast to generate, or does not correspond to the structure of the database architecture. It provides useful tips for choosing an alternative. The way they manage. Management of stored data with specific technologies to find relevant information, suggest inferences, and support decision making. For example, using this data to generate information and knowledge to improve e-learning infrastructure, especially storage mechanisms, performs data analysis tasks, but storage resource performance in this type of e-learning infrastructure. Implement AI technology to improve.

3.1. Storage system optimization model architecture.

E-learning platforms, like most web systems with databases, usually suffer from the latter performance issues, no matter how well designed. In addition, information production is growing exponentially and may need to be maintained indefinitely. In addition, it can be used in real time from anywhere. Performance is impacted, especially in the event of similarities, bottlenecks, interruptions, and lack of response.

4Storage System Optimization.

It has been suggested to deal with file downloads to improve performance from a memory management perspective. We propose a system design that can predict the files requested by users from the e-learning platform. To later store them in a component that contains some storage, avoid transactions with the database and the underlying file system and speed up this process.

Initially, we were looking to implement the appropriate transaction scheme built into the platform so that the server could quickly access the files requested by the user.

A second component is needed to tell this element which files to extract from the database. It needs to be intelligent and able to learn about user behaviors and patterns. Therefore, the idea of applying artificial intelligence technology in combination with big data technology to predict which files will be required produces transaction component recommendations and reduces response times.

4.1 Storage system optimization model architecture.

A prototype model was developed based on a traditional e-learning system. Its architecture is shown in Fig. 2, which combines four components:

1. User Experience: This is where user interactions with the platform are recorded and metadata is generated. The e-learning platform behaves like a typical web system with a database as shown in Fig. 3. There are also different types of actors, depending on the role they play, as explained in the analysis presented in (Moreno et. al, 2019). For this particular task, in addition to sharing resources when enrolling in the same course, we considered using student records because we usually have simultaneous access to the platform.

2. E-Learning Platform: Consists of modules that contain SQL databases that correspond to the applications that users of the e-learning system interact with.

3. Main Recommended Components for Optimizing Storage Systems: It aims to predict specific file downloads, taking into account various factors through techniques and strategies aimed at artificial intelligence. Not just using big data tools. The generated forecasts are delivered to the transaction component responsible for streamlining the file transfer process. These systems include a Data Transformer module, real-time data collection, HDFS-formatted data warehouse, real-time queuing system, streaming and learning engine, recommender system API, and deep learning server.

   Data Transformer: Converts SQL data to the format specified by the Hadoop Distributed File System (HDFS). Hadoop is a framework designed for big data management and analysis, and the technologies used, such as Hive, Flume, and Kafka, work with that framework and require conversion to a file system format.

   Real-time data collection: This is where each user's requests to the e-learning system are intercepted. This should be done by an agent that is constantly monitoring database changes. The agent sends a message to the real-time queuing system module.

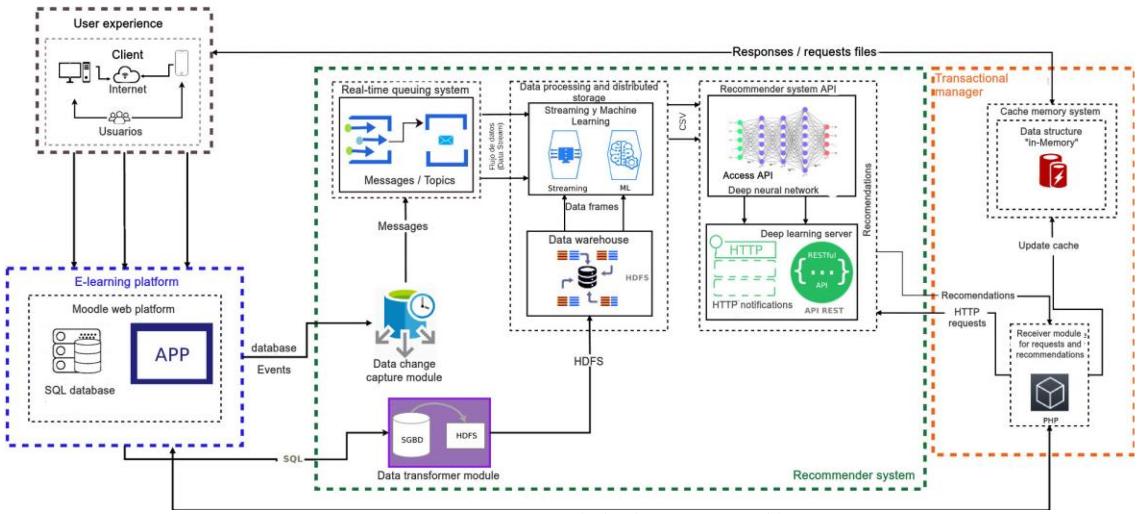

Fig. 2. Storage Management Optimization System Architecture

HDFS format data warehouse: This module has already been cleaned up, filtered, converted and stored in mysql with data extracted from the database history and converted to HDFS format. The result is a new data warehouse with the required information in the required format.

Real-time queuing system: This system uses messages that arrive synchronously or asynchronously and are stored later to respond to requests (data of requests made to the e-learning system in real time). The purpose is to prevent user requests from being lost or ignored via a data stream from the Streaming and Machine Learning module. In addition, it reduces overhead and enables operations in busy transactional environments.

Streaming and machine learning: Allows you to receive real-time data streams sent by the queuing system and new data stored in the data warehouse. These are processed together using machine learning. Machine Learning uses a collaboration algorithm to generate an initial list of recommended files, the number of records of which can reach thousands or millions. This first list serves as an entry for the recommender system API.

System API Recommendations: This is where the final prediction system will run. Neural network models can be used because the system is non-linear and has many variables, which complicates the problem. Therefore, you need to use a technique that allows you to train and test your model in an abstract way. This API can be implemented within a cluster or over the web, so it must be instantiated as a service so that the prediction results are passed to the Deep Learning Server Module.

Deep Learning Server: Responds to HTTP requests from external modules through a list of replicas. The idea is to create this service as a REST API model that sends a list of recommendations from the transaction manager to the Receiver of Recommendations and Requests module.

4. Transaction Manager: This is a scheme for transferring data at the backend level developed within the e-learning platform. This component contains two modules: Recommendations and Request Recipients and Caching System.

Recommendations and Request Recipients: This is an element integrated as an extension (plug-in) into the PHP code of the e-learning system. This consumes a list of database recommendations and requests and allows you to update your caching system. On the other hand, it compares the actual requested file with the recommended file and provides the recommended system with a list of visits to help adapt and evaluate it.

Cache memory system: Represents the cache system that stores the files most likely to be requested and the actual request made. This allows the user to get a faster response when the user requests a file. This memory should be dynamic. That is, it adapts to the size of the memory the system is running on.

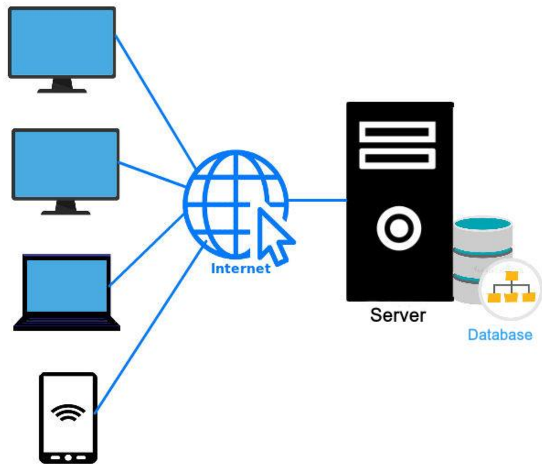

Fig. 3. Web system with database

4.2 Expected performance of the model.

The model needs to be implemented in real time because the dialogue takes place between the user and the e-learning platform. The goal is to reduce response time by saving user-downloaded files before requesting them to the intermediate element between the e-learning platform and the database.

At the beginning of the implementation, recommender systems receive a history database (SQL) converted to HDFS format. Next, a new data warehouse is created that contains only information related to user behavior and file downloads. For example, some data stored in the log (user ID, platform entry time and time, activity performed, etc.), file metadata (name, type, size, etc.). This data warehouse is processed to find out the specific classifications associated with the file and the metrics that show the user's behavioral patterns. Note that this happens only once.

In parallel, the e-learning platform receives requests from users that need to be sent to the database stored on disk. However, they are intercepted by the transaction manager on the one hand and sent to the data change capture module on the other. The first responds to the user's request by sending the file from the cache system, and if the file does not exist, it requests the file from the database on disk. The second, on the other hand, takes these requests and sends them as messages to the real-time queuing system so they aren't lost.

The streaming and machine learning components then receive both the metrics from the processing of the data warehouse (data frame) and the data flow of the queuing system. This is intended to perform an initial evaluation of the received data in order to make an initial prediction. It consists of thousands of recommendations for files that need to be transferred from the database on the hard disk to the cache system. These recommendations are sent in CSV format to the recommender system API, where a second list of predictions based on deep neural networks is generated. The latter should be significantly smaller than the former.

Lists generated by recommender systems are sometimes sent to the transaction manager as sending all requests is a waste of resources.

5Conclusion.

This paper describes the design of a theoretical model of the system to improve the data management of the e-learning platform. It is based on the process of downloading files and defines some strategies built into big data technology. On the one hand, a series of studies that combine several areas are established, and on the other hand, general designs are proposed using artificial intelligence techniques and strategies such as machine learning and deep learning. You can learn from user behavior on the e-learning platform and generate recommendations for optimizing your storage system. This will reduce response time.

The most important thing about this model is that it is distributed, scalable, fast, and fault tolerant. There is a desire to improve the process of downloading files from the e-learning platform and reduce the response time when users send requests to the platform, but be aware that it can be implemented in any web system provided by the database. please give me. This has nothing to do with the manager you use.

References.

Álvarez A, Alarcón A, Callejas M, 2016, Comparación de modelos y estándares de evaluación de calidad para una plataforma de aprendizaje virtual. Actas de Ingeniería. Vol 2. pp. 254-262. Available: http://fundacioniai.org/actas. Fecha de consulta: 12 Noviembre 2020.

Ashraf A, Hazem M, El-Bakry H, Abd El-razek S, El-Mashad Y, Mastorakis N, 2015, Enhancing big data processing in educationalsystems. Advances in Computers and Technology for Education, pp. 176-181, ISBN: 978-1-61804-2-9.

Ashraf A, Hazem M, El-Bakry H, Abd El-razek S, El-

Mashad Y, Mastorakis N, 2015, Handling Big Data in E-Learning. International Journal of Advanced Research in Computer Science Technology Vol. 3, Issue. 1, pp. 47-51, ISSN : 2347-8446 (Online); ISSN : 2347-9817 (Print).
Bernárdez M, 2010, Organización y Mercado. Conferencia Online. Available: http://www.expert2business.com/itson/Organizacion
Bourkoukou O, Essaid E. B, El Bachari E, El Adnani M, 2016, A Personalized E-Learning Based on Recommender System. International Journal of Learning and Teaching Vol. 2, Issue. 2, pp. 99?103. Diciembre 2016, ISSN: 2377-2891(Print); 2377-2905(Online).
Castro, López, Moreno, Tosco, 2013, Analizamos 19 plataformas de e- Learning: Investigación colaborativa sobre LMS. Available: www.congresoelearning.org. Fecha de consulta: 10 octubre 2020.
Dumbill E, 2012. Big Data Now. ISBN: 978-1-449-35671-2 Published by: O'Reilly.
Esposito C, Palmieri F, Castiglione A, Ficco M, 2015, A knowledge-based platform for big data analytics based on publish/sub-scribe services and stream processing. Knowledge Based Systems Vol 79, pp 3-17. Available: http://www.sciencedirect.com/science/article/pii/S0950705114001816, ISSN: 0950-7051, doi:http://dx.doi.org/10.1016/j.
Gómez O, Estrada V, Bauta R, García I, 2012, Modelo de gestión de log para la auditoría de información de apoyo a la toma de decisiones en las organizaciones, Revista Cubana de Información en Ciencias de la Salud (ACIMED) Vol. 23, Issue. 2, pp. 187-200. Available: http://new.medigraphic.com/cgi-bin/contenido.cgi?IDPUBLICACION=3813.
Kroth B, 2015. Adventures in moodle performance, Cs764 project report, University of Wisconsin-Madison, Available: https://vdocuments.site/cs764-project-report-adventures-in-moodle-performance-pagescswiscedubpkrothcs764bpkrothcs764project.html.
Marjanovic D, Milovanovic M, Radenkovic B, 2014, Hadoop Infrastructure for Education, in XIV International Symposium on New Business Models and Sustainable Competitiveness. University of Belgrade, pp. 365?370.
Merceron A, Blikstein P, Siemens G, 2015, Learning analytics: From big data to meaningful data. Journal of Learning analytics Vol 2, Issue. 3, pp. 4-8. https://doi.org/10.18608/jla.2015.23.2.
Moreno Y, Barrios G, Hidrobo F, 2019, Entendiendo el funcionamiento de Moodle: un enfoque basado en un marco de modelado, Revista Ibérica de Sistemas e Tecnologias de Informação, pp. 327-337, ISSN: 1646-9895.
Ramachandra K, Sudarshan S, 2014, Method for optimizing performance of database/web-service backed applications by automatically prefetching query results. United States Patent Application Publication, Pub. No.: US 2014/0195512 A1. Available: https://patentimages.storage.googleapis.com/.
Roldan A, 2016, Mecanismos de análisis bigdata para la caracterización de la actividad docente en un campus virtual moodle. Repositorios Documentales Universidad de Valladolid. Available: http://uvadoc.uva.es/handle/10324/17475 (Tesis de Maestria).
Subramanian Y, Zainuddin N, Alatawi S, Javabdeh T, 2014, Che Hussin A, A Study of Comparison between Moodle and Blackboard based on Case Studies for Better LMS, Journal of Information Systems Research and Innovation (JISRI), pp. 26-33, ISSN: 2289-1358.

Recibido: *23 de noviembre de 2021*

Aceptado: *21 de febrero de 2022*

Moreno, Yaneth: *System Engineer, Master degree in Computing, Ph.D. Candidate in Applied Science. Professor at the de Los Andes University. Mérida - Venezuela.*
https://orcid.org/0000-0002-5299-5439.

Valera, Junior: *Systems Engineering student, computing area at the Los Andes University. Mérida - Venezuela.*
E-mail: juniorvaleraa@gmail.com.
https://orcid.org/0000-0003-1328-2935.

Mejía, Gustavo: *Systems Engineering, Los Andes University. Mérida - Venezuela.*
E-mail: bfmvtm@gmail.com.
https://orcid.org/0000-0002-2069-4447.

Hidrobo, Francisco: *System Engineer, degree in Computing, Ph.D in Computer Architecture and Technology. Professor at the Los Andes University. E-mail: fhidrobo@gmail.com.*
https://orcid.org/0000-0002-1495-6666.

Díaz, Gilberto: *System Engineer and Master degree in Computing. Currently is TIC Consultant Universidad Industrial de Santander in Bucaramanga - Colombia. E-mail: gjdiazt@uis.edu.co.*
https://orcid.org/0000-0001-8188-5784.

Meneses, Jathinson: *System Engineer and Master degree in Management, Application and Software*

Development. He is currently a professor at the Universidad Industrial de Santander in Bucaramanga - Colombia.
E-mail: jameneme@correo.uis.edu.co.
https://orcid.org/0000-0002-1544-497X.