



Universitat d'Alacant
Universidad de Alicante

Arquitectura para la gestión de
datos en un campus inteligente

William Villegas Chiliquina



Tesis **Doctorales**

UNIVERSIDAD de ALICANTE

Unitat de Digitalització UA

Unidad de Digitalización UA



Universitat d'Alacant
Universidad de Alicante

INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN
INFORMÁTICA

ESCUELA POLITÉCNICA SUPERIOR

Arquitectura para la gestión de datos en un campus inteligente

William Eduardo Villegas Chiliquinga

Tesis presentada para aspirar al grado de

DOCTOR POR LA UNIVERSIDAD DE ALICANTE

DOCTORADO EN INFORMÁTICA

Dirigida por:

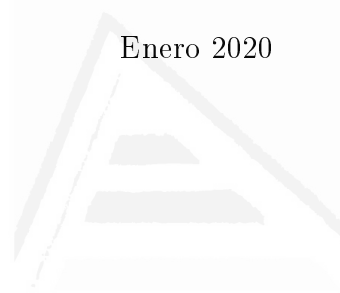
Dr. Sergio Luján Mora

Enero 2020

TESIS DOCTORAL EN FORMA DE COMPENDIO DE PUBLICACIONES

Arquitectura para la gestión de datos en un campus inteligente

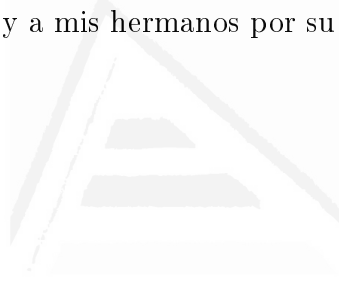
El presente documento contiene una síntesis del trabajo realizado por William Villegas Chiliquinga, bajo la dirección del Dr. Sergio Luján Mora, para optar por el grado de Doctor en Informática. Se presenta en la Universidad de Alicante y se estructura según la normativa establecida para la presentación de tesis doctorales en forma de compendio de publicaciones: una primera parte con una síntesis, una segunda parte que reproduce las publicaciones científicas realizadas y una tercera parte con las conclusiones.



Universitat d'Alicant
Universidad de Alicante

Dedicatoria

Esta tesis está dedicada a mis hijas Emma y Zoe porque son la razón y el motor de mi vida; a mi esposa María José por ser la compañera que me impulsa a ser mejor; a mis padres Segundo y María y a mis hermanos por su apoyo incondicional a todos mis proyectos.



Universitat d'Alacant
Universidad de Alicante

Agradecimientos

A través de estas líneas quiero expresar mi más sincero agradecimiento a todas las personas que con su soporte científico y humano han colaborado en la realización de este trabajo de investigación.

Quiero agradecer en primer lugar a la Universidad de Las Américas, institución en la que me desempeñé laboralmente y que ha hecho posible la realización del trabajo presentado, gracias, por la toda la disposición y ayuda brindada.

Además, quiero mencionar que la escritura de esta tesis y los artículos que la componen no hubiesen sido posible sin todas las enseñanzas y guía de mi director Sergio Luján Mora. Es por ello que el agradecimiento principal de este trabajo es para Sergio. Por la acertada orientación, por el soporte y discusión crítica que me permitió un buen aprovechamiento en el trabajo realizado y que este, mi doctorado, llegara a buen término. Muchas gracias por guiarme por el camino correcto, sus enseñanzas siempre las voy a tener presente y junto a sus consejos nunca los olvidaré y los voy a tener presente como el regalo más grande que puedo recibir de alguien, muchas gracias Sergio.

Universitat d'Alacant
Universidad de Alicante

Alicante, enero de 2020
William Eduardo Villegas

Resumen

En la actualidad, las tecnologías de la información y comunicación (TIC) se han convertido en herramientas invaluableles en el desarrollo de la sociedad. Estas tecnologías están presentes en las empresas, la medicina, la educación, etc. Prácticamente la sociedad ha llegado a un punto en que el principal asistente en cada una de las actividades son las TIC. Esto ha permitido la globalización de todas las áreas donde estas son aplicadas. Las ventajas del uso de las TIC han permitido mejorar y automatizar los procesos en todo nivel, sea en una empresa, una ciudad, una universidad, un hogar, etc. Para hacerlo, las tecnologías se ajustan a las necesidades del usuario y son capaces de interactuar con él, incluso, están en capacidad de interactuar entre sí sin la intervención de un humano. ¿Pero cómo lo hacen y para qué? Las nuevas tecnologías ahora integran varios sistemas y plataformas que están en la capacidad de adquirir información de las personas y sus entornos, analizar esta información y tomar decisiones con base en los resultados del análisis. Estas decisiones se ven plasmadas, por ejemplo, en la mejora de las ventas de una empresa o en la mejora de los procesos de manufactura. Como estos, existen muchos ejemplos que son resultado de numerosas investigaciones que tienen como objetivo mejorar la calidad de vida de las personas en ecosistemas sostenibles.

Uno de estos ecosistemas que ha adquirido gran importancia recientemente son las ciudades inteligentes. El valor de las ciudades inteligentes se basa en satisfacer las necesidades de los miembros de su comunidad en armonía con la naturaleza. Esto involucra una mejor administración de los servicios como el transporte, la generación y consumo energético, la seguridad, la gobernabilidad, etc. Sin embargo, transformar una ciudad común en una ciudad inteligente requiere de muchos esfuerzos y recursos, tanto económicos como humanos. Ante este problema, es necesario contar con escenarios similares que incluso sirvan como un banco de pruebas para la implementación de tecnologías y que su implementación en entornos más grandes sea efectiva y con los recursos adecuados. Las universidades, como generadoras de conocimiento, son las llamadas a realizar los procesos de implementación, pruebas y generación de nuevas tecnologías. Su ambiente, administración y organigrama estructural, sumada a extensas áreas que conforman sus campus, permite compararlas con pequeñas ciudades. Esto permite establecer una línea base donde se apliquen todos los componentes necesarios para transformarlos en campus inteligentes (*smart campus*).

Los campus inteligentes buscan mejorar la calidad de la educación a través de la

Resumen

convergencia de nuevas tecnologías. Es importante establecer que un campus universitario pone a disposición de los estudiantes y los miembros de la comunidad todas las condiciones para garantizar la calidad de la educación. Los campus inteligentes, al igual que las ciudades inteligentes, basan sus entornos en satisfacer las necesidades de sus miembros; para esto, es necesario crear procesos o sistemas que adquieran información sobre ellos. Es por esto, que el Internet de las cosas (IoT, acrónimo en inglés de *Internet of Things*) se convierte en uno de los componentes necesarios para la transformación de un campus tradicional. La información recolectada necesariamente debe convertirse en conocimiento para ejecutar acciones con base en este conocimiento. Estas acciones responden a una toma de decisiones efectiva y eficiente que satisfaga las necesidades de las personas. Para realizar el análisis de datos es necesario contar con una arquitectura que gestione un gran volumen de datos independientemente de su formato. La tecnología que ofrece estas capacidades es el *big data*, su integración al campus inteligente genera una estructura lo suficientemente robusta para soportar toda la carga del IoT y el análisis de datos requerido por los usuarios.

Estas tecnologías, en compañía de la computación en la nube (*cloud computing*), permiten a los miembros del campus inteligente desarrollar sus actividades en total armonía con los recursos y la naturaleza. Este trabajo de investigación está enfocado en proponer una arquitectura para la gestión de datos en un campus inteligente. Este enfoque trata todas las variables que influyen en la educación universitaria. Descubrir estas variables, tratarlas y establecer sus relaciones entre sí, requiere de la integración de las tecnologías mencionadas incluso con modelos de inteligencia artificial que permitan tomar acciones sobre los resultados del análisis de datos.

Universitat d'Alacant
Universidad de Alicante

Abstract

At present, information and communication technologies (ICT) have become invaluable tools in the development of society. These technologies are present in companies, medicine, education, etc. Virtually society has reached a point where the main assistant in each of the activities in ICT. This has allowed the globalization of all areas where they are applied. The advantages of the use of ICT have allowed to improve and automate processes at all levels, whether in a company, a city, a university, a home, etc. To do so, the technologies adjust to the user's needs and are able to interact with him, even, they are able to interact with each other without the intervention of a human. But how do they do it and for what? New technologies now integrate several systems and platforms that are capable of acquiring information from people and their environments, analyzing this information and making decisions based on the results of the analysis. These decisions are reflected, for example, in the improvement of a company's sales, in the improvement of manufacturing processes. As there are many examples that are the result of research that aims to improve the quality of life of people in sustainable ecosystems.

One of these ecosystems that have acquired great importance recently is smart cities. The value of smart cities is based on meeting the needs of the members of their community in harmony with nature. This involves better administration of services such as transportation, energy generation and consumption, security, governance, etc. However, transforming a common city into a smart city requires many efforts and resources, both economic and human. Faced with this problem, it is necessary to have similar scenarios that even serve as a test bench for the implementation of technologies and that their implementation in larger environments is effective and with adequate resources. Universities, as knowledge generators, are the calls to carry out the processes of implementation, testing and generation of new technologies. Its environment, administration and structural organization chart, added to large areas that make up its campuses, allows comparing them with small cities. This allows establishing a baseline where all the necessary components are applied to transform them into intelligent campuses.

Smart campuses seek to improve the quality of education through the convergence of new technologies. In important to establish that a university campus makes available to students and community members all the conditions to guarantee the quality

Abstract

of education. The campuses, like smart cities, base their environments on meeting the needs of their members, for this it is necessary to create processes or systems that acquire information about them. This is why the Internet of Things (IoT) becomes one of the necessary components for the transformation to a smart campus. The information collected must necessarily become knowledge to execute actions based on this knowledge. These actions respond to effective and efficient decision making that meets the needs of people. To perform data analysis it is necessary to have an architecture that manages a large volume of data regardless of its format. The technology that offers these capabilities is big data, its integration to the smart campus generates a structure robust enough to support the entire IoT load and data analysis required by users.

These technologies, in the company of cloud computing, allow smart campus residents to develop their activities in total harmony with resources and nature. This research work is focused on proposing architecture for data management in a university campus. This approach addresses all the variables that influence university education. Discovering these variables, treating them and establishing their relationships with each other, requires the integration of the mentioned technologies even with artificial intelligence models that allow actions to be taken on the results of the data analysis.



Universitat d'Alacant
Universidad de Alicante

Índice general

Dedicatoria	I
Agradecimientos	III
Resumen	V
Abstract	VII
Índice de figuras	XIII
Índice de cuadros	XV
I SÍNTESIS	1
1 Introducción	3
1.1 Motivación	3
1.1.1 Definición del problema	6
1.1.2 Identificación de variables del aprendizaje	7
1.1.3 La inteligencia de negocios como arquitectura para la gestión de datos	9
1.1.4 Una arquitectura para la gestión de datos en un campus inteligente	11
1.2 Objetivos	17
1.3 Método de trabajo	18
1.4 Resultados	18
1.5 Estructura de la tesis	21
1.6 Convenciones de escritura	23

2	Publicaciones y visibilidad	25
2.1	Publicaciones	25
2.1.1	Revistas	25
2.1.2	Congresos	26
2.2	Visibilidad	28
II	TRABAJOS PUBLICADOS	31
3	Compendio	33
4	Big Data, the Next Step in the Evolution of Educational Data Analysis	35
5	Comprehensive Learning System Based on the Analysis of Data and the Recommendation of Activities in a Distance Education Environment	47
6	Management of Educative Data in University Students with the Use of Big Data Techniques	61
7	Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus	75
8	Application of a Big Data Framework for Data Monitoring on a Smart Campus	105
III	CONCLUSIONES	123
9	Conclusiones	125
10	Trabajos futuros	129
	APÉNDICE	133
A	Otros artículos	133
B	Analysis of Data Mining Techniques Applied to LMS for Personalized Education	135

C	Systematic Review of Evidence On Data Mining Applied to LMS Platforms for Improving E-Learning	139
D	Data Mining Toolkit for Extraction of Knowledge from LMS	143
E	Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining	147
F	Application of Data Mining for the Detection of Variables that Cause University Desertion	151
G	Artificial Intelligence as a Support Technique for University Learning	155
	Referencias	159



Universitat d'Alacant
Universidad de Alicante

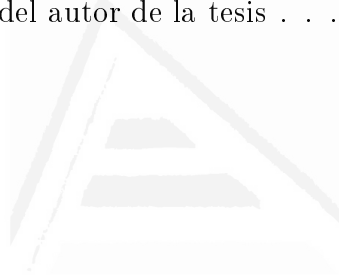
Índice de figuras

1.1	Síndrome de la deserción universitaria	7
1.2	Arquitectura de un campus inteligente basado en modelo de inteligencia de negocios	10
1.3	Componentes de un campus inteligente	12
1.4	Arquitectura de un campus inteligente basado en el análisis de datos . .	17
1.5	Diagrama de flujo de un proceso de transporte interno mediante <i>big data</i>	22
2.1	Detalles del autor en Scopus	29
3.1	Línea de tiempo del compendio de publicaciones	34
10.1	Capas de una arquitectura del Internet de las cosas	130

Universitat d'Alacant
Universidad de Alicante

Índice de cuadros

1.1	Áreas con la mayor densidad de población en un campus inteligente . . .	20
1.2	Consumo de bebidas en temporada de exámenes	20
2.1	Descripción de las revistas	26
2.2	Descripción de los congresos	27
2.3	Perfiles académicos del autor de la tesis	28



Universitat d'Alacant
Universidad de Alicante

Parte I

SÍNTESIS

Universitat d'Alacant
Universidad de Alicante

1 Introducción

1.1. Motivación

En la actualidad, el análisis de datos se ha convertido en una de las cualidades de las instituciones que desean mejorar y personalizar sus servicios para sus usuarios. La capacidad de análisis consiste en examinar un conjunto de datos con el propósito de sacar conclusiones sobre la información para poder tomar decisiones. El análisis de datos empieza con la necesidad de las empresas de conocer las tendencias de sus clientes y cómo enfocar sus productos con base en esas necesidades (Pompei et ál., 2018).

Para cumplir con estos objetivos, las empresas solían pasar por un proceso complicado para la recolección de datos que es la fuente prima para el análisis. Este proceso, desde la inclusión de las tecnologías de la información (TI) ha sufrido un cambio de enormes magnitudes. Ahora las TI son capaces de generar grandes volúmenes de datos sobre las personas y los ecosistemas donde habitan sin la necesidad de que estas estén al tanto de ello. Es de esta manera que incluso las tecnologías han tenido que evolucionar al punto que ahora estas pueden comunicarse entre sí y controlar los entornos de manera automática (Atzori, Iera, y Morabito, 2010).

La inclusión de las tecnologías en el desarrollo de las sociedades ha sido tan vertiginosa que ahora ya es común hablar de entornos inteligentes basados en la satisfacción de sus miembros. Estos entornos inteligentes comúnmente integran a las nuevas tecnologías como el Internet de las cosas (IoT, acrónimo en inglés de *Internet of things*), computación en la nube (*cloud computing*) y el análisis de datos (Yaqoob et ál., 2017). Estas tecnologías se encargan de la recolección de información, su análisis y la toma de decisiones. Estos son requisitos fundamentales que los entornos deben cumplir para satisfacer las necesidades de sus miembros.

A esta evolución se han unido las instituciones de educación superior (IES). Las IES, como productoras del conocimiento, están en constante crecimiento y actualización de sus métodos y técnicas para mejorar las condiciones de los estudiantes. Sin embargo, no existe una guía o un estándar de cómo hacerlo y cómo utilizar las nuevas tendencias tecnológicas a su favor. Estas tecnologías, además, han creado nuevos paradigmas que se enfocan en satisfacer las necesidades de las personas desde un punto de vista sostenible (Bascopé, Perasso, y Reiss, 2019).

La sostenibilidad es importante, ya que, a medida que las sociedades avanzan, es necesario generar un crecimiento que garantice la igualdad y la prosperidad en todos los lugares del mundo. La Organización de Naciones Unidas (ONU) estableció en el septiembre de 2015 una serie de objetivos globales para poder conseguir un mundo más sostenible (Naciones Unidas, 2015). Uno de estos objetivos es la educación de

1 Introducción

calidad; para ello, las universidades deben establecer arquitecturas y modelos que les permita mejorar todos sus procesos adoptando las TI en todos sus procesos.

Las universidades, desde un punto de vista integral, poseen los requisitos necesarios para crear ambientes basados en las necesidades de los estudiantes y que estos ambientes sean sostenibles (Salmerón-Manzano y Manzano-Agugliaro, 2018). De forma general, cada persona presenta necesidades únicas, lo que conlleva a que los modelos y arquitecturas universitarias deban ser escalables y adaptativas, centradas en las características de los estudiantes.

Los modelos educativos tradicionales se basan en el conocimiento que los tutores y docentes tienen sobre un tema. Pero la globalización y la expansión del conocimiento permiten la transformación de las estructuras socioculturales en una sociedad cada vez más interconectada y competitiva. Es aquí donde las tecnologías cumplen un rol activo que aporta al desarrollo y evolución de las universidades.

Las tecnologías de la información y comunicación (TIC) rompen el paradigma de una educación estática y dependiente de una persona o una infraestructura. Estas características obligan a las universidades a mejorar todos sus procesos para brindar servicios que garanticen la calidad de la educación (Uskov, Bakken, y Pandey, 2016). La principal forma de mejorar estos procesos son automatizarlos y controlarlos a través de las nuevas tecnologías. Una muestra de ello son los campus inteligentes (*smart campus*), donde la integración del IoT, el análisis de datos y la computación en la nube permiten una gestión integral de los componentes que son parte de la educación (Kortuem, Bandara, Smith, Richards, y Petre, 2013).

La importancia de la convergencia e integración de tecnologías radica en su capacidad de solucionar problemas genéricos en las universidades como son la deserción y los bajos porcentajes en la tasa de graduados (Donoso y Schiefelbein, 2009). Estos parámetros son medidos constantemente por los entes reguladores de muchos países con la finalidad de mejorar la calidad de la educación. Para ello, internamente las universidades tienen áreas que se encargan de la calidad y la mejora del aprendizaje (Gairín et ál., 2014). Sin embargo, analizar todos los casos que se presentan en un ambiente tan grande como el universitario consume mucho esfuerzo y recursos de las personas que lo realizan. Otro factor que limita este análisis es que los resultados siempre están sometidos al criterio del analista de datos.

Detectar las causas por la cuales un estudiante fracasa en sus actividades académicas o determinar el porqué de la deserción requiere de la inclusión de un sinnúmero de variables (Palacios-Pacheco, Villegas-Ch, y Luján-Mora, 2019). Estas variables incluyen información académica, financiera, cultural, etc., solo con la inclusión de esta información se puede hablar de una mejora en la calidad de la educación que dé como resultado el éxito estudiantil. Con este panorama, es imposible llevar a cabo un análisis exitoso sin la ayuda de las TIC. Pero ¿cómo las TIC pueden mejorar el aprendizaje? Y ¿cómo estas tecnologías pueden ayudar a todo el proceso educativo de una universidad para convertirla en un campus inteligente? Estas son las preguntas que se hacen las universidades al incluir en sus procesos a las TIC. Para dar respuesta a estas preguntas, es necesario analizar los entornos universitarios, sus fortalezas y debilidades.

En la actualidad, las universidades usan las TIC en sus procesos administrativos aplicados a sistemas informáticos encargados de la parte financiera, administrativa y

académica. Estos sistemas tradicionalmente son aplicaciones informáticas conectadas a una base de datos donde se gestiona la información a través operaciones como consultas, actualizaciones, inserciones, etc. Estas aplicaciones tienen un propósito específico como son el manejo de inventarios, las finanzas, las calificaciones o las asistencias a las clases.

Además de estas aplicaciones, existen otras que ayudan al desarrollo académico de manera más activa como son los sistemas de gestión de aprendizaje (LMS, acrónimo en inglés de *learning management system*). Los LMS permiten interactuar a los docentes y estudiantes por medio del uso de recursos y actividades planteadas en un sitio web (Villegas-Ch y Luján-Mora, 2017a). Además, sirven de repositorios que almacenan información valiosa sobre lo que hacen los estudiantes en cada materia (Villegas-Ch y Luján-Mora, 2017a).

Disponer de toda la información que se genera en los sistemas con los que cuentan las universidades es importante, pero no es suficiente para solucionar los problemas comunes de la educación. Por lo tanto, es necesario implementar sistemas que recojan la mayor cantidad de datos de los estudiantes y su entorno. De manera general, esto se traduce en la implementación de una tecnología que permite la recolección de datos como el IoT (Yashiro, Kobayashi, Koshizuka, y Sakamura, 2013).

El IoT utiliza una variedad de dispositivos que miden y controlan un gran número de variables que pueden ser académicas o administrativas. Los dispositivos de IoT generan un gran volumen de datos que regularmente son almacenados y gestionados por un modelo de computación en la nube. Dependiendo de la arquitectura de IoT, los datos generados pueden ser almacenados de forma local para permitir su análisis (Ray, 2016). El análisis de datos requiere de una tecnología con la capacidad suficiente para tratar el gran volumen de datos, así como dar respuesta a los altos requerimientos de recursos a nivel de procesamiento y flexibilidad en el tipo de datos.

La respuesta a estas necesidades son los *frameworks*¹ de inteligencia de negocios (BI, acrónimo en inglés de *business intelligence*) o el *big data* (Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2018). Estos *frameworks* abarcan la extracción, procesamiento y análisis de datos. Los resultados son presentados a los responsables de la calidad de la educación para que tomen las decisiones necesarias para mejorar el aprendizaje. En este punto, con la ayuda de esta integración de tecnologías, es posible hablar del descubrimiento del conocimiento en relación a las variables que presenta cada estudiante (Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2017). Con este conocimiento es posible recomendar una variedad de actividades y tareas que se alineen a sus necesidades y su forma de aprender.

La capacidad de recomendar actividades o crear ambientes adecuados para cada estudiante se logra a través de una arquitectura robusta que brinda una educación personalizada (Villegas-Ch, Palacios-Pacheco, Buenaño-Fernandez, y Luján-Mora, 2019). La escalabilidad y flexibilidad que la integración de estas tecnologías ofrece permite ir más allá y dotar a toda esta arquitectura de ciertos niveles de decisión con el uso de modelos de inteligencia artificial. Todas estas características embebidas y aplicadas a una universidad permiten transformarla en un campus inteligente (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a). Lo que se busca principalmente con esta trans-

¹Un *framework* o “marco de trabajo”; es un esquema o un patrón para el desarrollo o la implementación de una aplicación.

formación es mejorar el proceso de aprendizaje, sin dejar de lado la gestión del propio campus donde existen temas como la movilidad, la seguridad, la sostenibilidad, etc.

1.1.1. Definición del problema

Esta tesis inicia al definir el problema al cual se enfrentan las universidades y la identificación de variables que causan problemas en la calidad de la educación. Para ello es necesario mencionar que las universidades están sujetas a procesos de evaluación por parte de entidades gubernamentales, así como agencias internacionales que son responsables de garantizar la calidad académica. Los principales factores a los que se hace referencia en los procesos de evaluación son la deserción, la repetición y la efectividad académica.

Para las universidades, la detección de las variables que causan la deserción se ha convertido en una tarea muy compleja por la cantidad de recursos humanos que esto requiere. Hace algunos años estas variables se concentraban en factores simples como la falta de información previa sobre la carrera o la dificultad del estudiante para adaptarse al entorno universitario (Donoso y Schiefelbein, 2009; Palacios-Pacheco et ál., 2019). En la actualidad, identificar los factores que influyen en el aprendizaje requiere la inclusión de nuevas variables que se encargan de la recopilación y evaluación sistemática de los datos que permiten identificar similitudes, definir oportunidades y problemas que contribuyen al aprendizaje.

El análisis de los datos educativos es un medio para lograr identificar de manera real y rápida las variables que afectan al desarrollo satisfactorio de los estudiantes. Con la mejora del proceso de identificación de variables, la toma de decisiones tiene un efecto positivo en los estudiantes al personalizar la educación y mejorar sus modelos educativos. En un ecosistema donde todos los servicios se centran en mejorar la vida, las tecnologías que determinen las tendencias y patrones de los miembros de la comunidad deben integrarse en una arquitectura que conviva con el usuario y el entorno.

Además, es necesario considerar el crecimiento continuo de la población universitaria y como esto obliga a la expansión de los campus universitarios tanto físicamente como en el uso de recursos que necesitan para cubrir las necesidades de sus miembros. El crecimiento, exponencial de las universidades y sus campus, generalmente no es un crecimiento programado o que haya seguido un proceso que permita la convivencia de la universidad con el medio ambiente. Estos son problemas que actualmente son de interés mundial y que se incluyen como temas de la sostenibilidad.

Los campus universitarios que no tienen un plan definido que incluyan pautas y políticas que permitan la gestión adecuada de los recursos económicos, sociales y naturales, afectan de manera directa o indirecta al desarrollo del aprendizaje. Para resolver estos problemas, los campus universitarios buscan alternativas respaldadas por las TIC. Los campus universitarios generalmente tienen sus propias infraestructuras informáticas que les permiten generar, almacenar y procesar grandes volúmenes de datos, que en la mayoría de los casos no son utilizados. Disponer de una plataforma que aproveche todos los sistemas a cargo de adquirir datos y analizarlos en un entorno de nube privada ayuda a la gestión del aprendizaje y la interacción de la población universitaria con las TIC. Este proceso, además de garantizar la mejora en los procesos educativos,

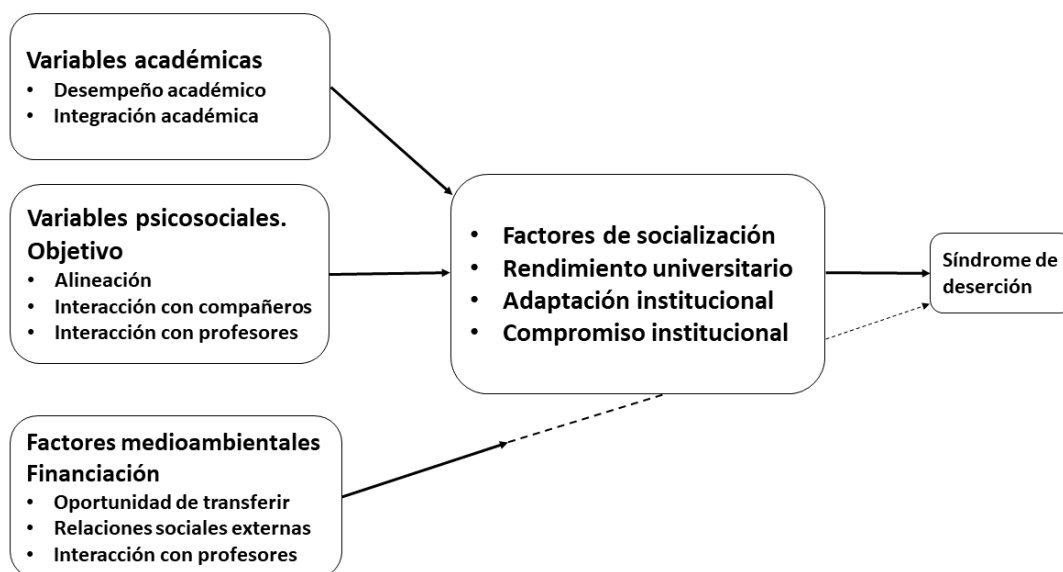


Figura 1.1: Síndrome de la deserción universitaria
(Bean, 1983)

permite a la población universitaria interactuar con el entorno y mejorar temas como la seguridad, el consumo adecuado de recursos, la toma de decisiones, etc.

El desafío para los científicos de datos es unificar todos los datos generados en los sistemas académicos, los sistemas administrativos, los sistemas de gestión de aprendizaje, los sistemas de seguridad, los sistemas autónomos, etc. Los estudios realizados en el campo de las ciudades inteligentes permite la integración de sistemas, aplicando conceptos de IoT y *big data* para transformar un campus tradicional en un campus inteligente (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a). Esta opción tiene un gran valor en la gestión educativa, porque mejora la interacción de estudiantes, docentes y administradores en un entorno sostenible. Un campus inteligente genera ambientes cómodos donde las necesidades de cada individuo se cubren de manera personalizada, incluso antes de que el mismo individuo las tenga. Este ecosistema integra conceptos, tecnologías, datos e individuos para garantizar una sociedad del conocimiento donde el desarrollo del aprendizaje está garantizado por las nuevas tendencias tecnológicas.

1.1.2. Identificación de variables del aprendizaje

En este trabajo, para determinar cuáles son las variables que aportan al aprendizaje se utilizó el modelo de Bean (1983). Bean identifica las variables académicas, psicosociales y ambientales que conducen al síndrome de deserción. Como se detalla en la Figura 1.1, cada una de las variables está constituida por un campo que sirve como guía para la extracción de datos de los diferentes repositorios de la universidad.

En las variables académicas, se consideran el rendimiento y la integración del estudiante; los repositorios comunes que almacenan esta información son los sistemas de

1 Introducción

registro académico y seguimiento académico. Las variables psicosociales incluyen objetivos, alineación, interacción entre iguales e interacción con los docentes. Las variables ambientales se parametrizan de acuerdo con la financiación, las relaciones sociales externas y su relación con el entorno. La interacción con los profesores incluye la interacción con los LMS y todos los sistemas que con las TIC aporten al desarrollo académico. Los factores medioambientales y de financiación influyen de manera directa al síndrome de la deserción por lo que no ingresan a los factores de socialización, rendimiento universitario, etc., en la Figura 1.1 esto se representa a través de líneas entrecortadas.

En el proceso de selección de datos, se pueden incluir todos los registros de una base de datos; sin embargo, esta no es la mejor opción porque corre el riesgo de sobrecargar el proceso y afectan al alcance y el nivel de profundidad deseado en el análisis. Para facilitar el proceso de selección de datos, es común trabajar con preguntas que ayudan a identificar los diferentes campos que comprenderán las variables, por ejemplo:

- ¿Hay información disponible sobre los sistemas que se pueden usar?
- ¿Ayuda esta información al análisis?
- De todos los tipos de información disponibles, ¿cuál nos interesa más?
- ¿Son interesantes los detalles de toda la información disponible o solo los detalles de la información que necesitamos?

Con la identificación clara de los problemas a los que se enfrenta la universidad en el campo de la educación de calidad, se analizaron dos soluciones que la garanticen con base en el análisis de datos (Christozov, 2017). La primera posibilidad se centró en la aplicación de un *framework* de BI para el análisis de datos y la toma de decisiones. Un *framework* de BI tiene varias ventajas, entre estas destaca que son muy utilizados en las empresas y que han demostrado eficiencia en todos los sectores donde han sido aplicados, por lo tanto, existe mucha información y para su implementación.

Un BI trabaja de manera directa con un proceso de extracción, transformación y carga de datos (ETL, del acrónimo en inglés de *extract, transform and load*). Un ETL brinda seguridad y confiabilidad en los datos que son analizados, pues estos pasan por un procesamiento y transformación antes de ser cargados a un almacén de datos (DW, acrónimo en inglés de *data warehouse*). El DW ofrece una base de datos multidimensional para la generación de cubos de procesamiento analítico en línea (OLAP, acrónimo en inglés de *On-Line Analytical Processing*) (Kimball, Ross, y Kimball, 2009).

Esta característica permite generar conocimiento con base en proyecciones y regresiones, al aplicar algoritmos de minería de datos que permiten la extracción del conocimiento de bases de datos (KDD, acrónimo en inglés de *knowledge discovery in databases*). Un BI generalmente forma parte del sistema de información de las universidades y es utilizado para obtener datos estadísticos o financieros que ayuden a una mejor toma de decisiones.

En el desarrollo de esta investigación se implementó un BI como primer experimento para realizar el análisis de datos y poder observar su funcionalidad al aplicar técnicas de minería de datos educativa (EDM, acrónimo en inglés de *educational data mining*)

(Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2018). En la implementación, se consideró el uso de herramientas comerciales como Microsoft SQL Analysis Services que es un motor de datos analíticos y Pentaho en su versión de uso libre. Además de los objetivos ligados al aprendizaje, la arquitectura propuesta debe ser capaz de integrarse a todos los sistemas con los que cuenta la universidad (Villegas-Ch y Luján-Mora, 2016). Un segundo y definitivo experimento se realizó al implementar una arquitectura de gestión de datos basada en un *big data*.

1.1.3. La inteligencia de negocios como arquitectura para la gestión de datos

Un BI combina la tecnología con las herramientas y procesos que permiten transformar los datos almacenados en información, la información en conocimiento y el conocimiento dirigido a un plan o una estrategia comercial (Valdiviezo-Díaz, Cordero, Reátegui, y Aguilar, 2015). Un BI permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de una institución y la capacidad de tomar buenas decisiones para así obtener mejores resultados.

El BI, generalmente utiliza una metodología con base en el modelo KDD. El objetivo de este método es identificar patrones en diferentes fuentes de datos y generar conocimiento. Una arquitectura de BI se compone de seis etapas que se pueden visualizar en la Figura 1.2. La capa de fuentes de datos toma como referencia diversas fuentes y este es el componente más importante de BI a través del uso de un ETL. Un ETL, al incluir una variedad de herramientas puede extraer datos de diferentes fuentes como bases de datos relacionales, archivos de texto, hojas de cálculo, etc. Esto siempre será una ventaja al agregar nuevas fuentes y como un valor agregado se tienen que en su mayoría los ETL, presentan un modelo de programación gráfica.

La capa de selección de datos se realiza de acuerdo a las relaciones y las características de los datos. Para identificar fácilmente esta información es necesario trabajar con un diccionario de datos (Corrales y Corrales, 2016). Este diccionario contiene información que incluye el significado de los datos, su relación con otros datos, su origen, su formato, etc. En el ámbito educativo el diccionario de datos puede contener las características lógicas de los estudiantes, incluidos el nombre, la descripción, el identificador, el contenido y la organización. Con estos detalles, el análisis puede hacer uso de los datos de diferentes repositorios y evaluar sus relaciones con menores costos de procesamiento y mayor efectividad.

En la capa de preprocesamiento es necesario establecer la conexión entre la herramienta de BI y las fuentes de datos incluidas en el análisis, por ejemplo con una base de datos MySQL. Los datos seleccionados en los repositorios generalmente no están limpios o contienen errores. Esto reduce la calidad del análisis al generar reglas inútiles en la etapa de extracción de datos. Los problemas típicos incluyen, datos incompletos, falta de valores en atributos, datos inconsistentes e incluso discrepancias entre los datos. El preprocesamiento de datos aplica varios filtros de forma supervisada y no supervisada. En ambos casos, existe la opción de limpiar el atributo o instancia. La ventaja de la preparación previa de datos es que se genera un conjunto de datos más pequeños para mejorar la eficiencia del análisis de datos.

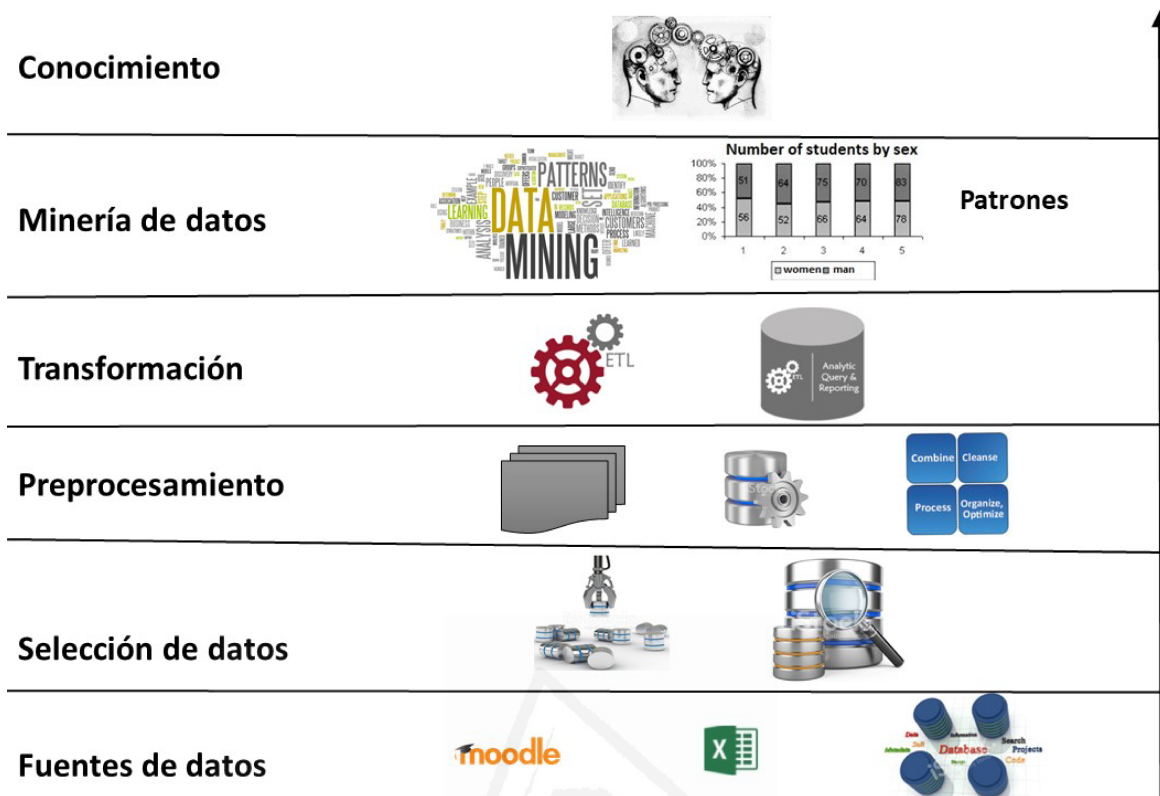


Figura 1.2: Arquitectura de un campus inteligente basado en modelo de inteligencia de negocios

La capa de transformación en la arquitectura de KDD implica tres etapas en el diseño del modelo de BI. La extracción, se encarga de extraer los datos de las diferentes fuentes existentes como son las BDD relacionales. En la transformación se suele utilizar otras fuentes de datos como archivos planos que sirven como datos temporales para la etapa de análisis. Estos contienen las fechas que se utilizan para obtener reportes históricos. En esta etapa no se filtran los datos porque existe información que se relaciona con otras tablas de la fuente. En cuanto a la fuente de datos fecha, se selecciona el mismo objeto para asignar un tipo de dato y longitud máxima. Finalmente, en el proceso de ETL se establece la carga de los datos que consiste en recolectar los datos transformados y con la ayuda de un objeto de inserción se carga los datos a una tabla de dimensional creada en un DW. Cuando se ejecuta el proceso ETL se inicia la carga de los datos a cada tabla de dimensiones y posteriormente se realiza una actualización que se encarga de revisar que cada uno de los registros sea consistente y actualice los que tengan errores.

En la capa de minería de datos los algoritmos de análisis buscan patrones que sean relevantes en un determinado segmento. La elección del algoritmo puede considerarse un desafío ya que cada algoritmo de minería de datos genera un resultado diferente. Sin embargo, esto no significa que en un análisis no sea posible usar más de un algoritmo. Lo importante es determinar las entradas más efectivas para predecir un resultado particular basado en datos existentes.

Finalmente, en la capa de conocimiento se interpretan los resultados y se valida que estos sean reales y satisfactorios. En la etapa de interpretación, los datos se analizan

de acuerdo con los patrones y modelos que se encontraron en ellos. Los patrones son estructuras locales que hacen declaraciones sobre un espacio restringido por variables. Por ejemplo, una anomalía común es la detección de las ausencias de un estudiante de un curso.

Los modelos son estructuras globales que hacen afirmaciones sobre cualquier punto en el espacio de medición. Por ejemplo, basándose en el comportamiento de un grupo de estudiantes en ciertas materias, se puede proyectar su efectividad en actividades futuras. En la presentación de resultados, las técnicas de visualización son importantes ya que los modelos finales o las descripciones en formato de texto pueden ser difíciles de interpretar para los usuarios finales.

1.1.4. Una arquitectura para la gestión de datos en un campus inteligente

Las universidades son ambientes propicios para la creación de entornos inteligentes donde puedan ser utilizadas todas las ventajas que ofrece una ciudad inteligente. Estas ventajas ayudan a mejorar la movilización, el buen manejo de recursos, la seguridad, la gestión de procesos, etc. Estas características permiten hablar de una evolución de las universidades convirtiéndolas en campus inteligentes donde, la tecnología se concentra en mejorar las condiciones de los miembros del campus (Villegas-Ch, Molina-Enriquez, Chicaiza-Tamayo, Ortiz-Garcés, y Luján-Mora, 2019). Un campus inteligente permite gestionar de forma adecuada todos los recursos que intervienen en el aprendizaje, de esta manera se puede dotar de mayor granularidad al análisis de datos.

Para cumplir con esta propuesta, el primer paso fue analizar las características que una universidad utiliza para su gestión. Para esto, se segmentó la administración en dos grupos que se encargan de todo el componente universitario. En el primer grupo se encuentra toda la parte administrativa y un segundo grupo que abarca la parte académica. Estos grupos tienen tareas muy bien definidas dentro de la universidad, la parte administrativa se encarga de gestionar todos los recursos sean financieros, físicos, tecnológicos, estructurales, manejo de recursos humanos, etc.

El grupo académico, se encarga de la gestión de recursos académicos, manejo de LMS, seguimiento estudiantil, gestión de docentes, etc. Para llevar a cabo esta gestión, las universidades manejan varias aplicaciones de desarrollo propio o comerciales como pueden ser sistemas de planificación de recursos de empresa (ERP, acrónimo en inglés de *enterprise resource planning*) o sistemas de gestión de las relaciones con clientes (CRM, acrónimo en inglés de *customer relationship management*).

Las características de un campus universitario permiten identificar los puntos clave que admite un campus inteligente. La estructura de un campus inteligente debe ser flexible, escalable y evolutiva, donde los procesos se evalúan constantemente y el nivel de reacción a eventos externos es eficiente. Los controles y auditorías, deben garantizar que los componentes sean adecuados e incluir todos los factores que intervienen en una arquitectura de campus inteligente (Villegas-Ch, Palacios-Pacheco, Ortiz-Garcés, y Luján-Mora, 2019).

Los componentes del campus inteligente se alinean con la seguridad, el manejo de los recursos, una mejor gestión energética, entornos climatizados, sistemas domóticos,

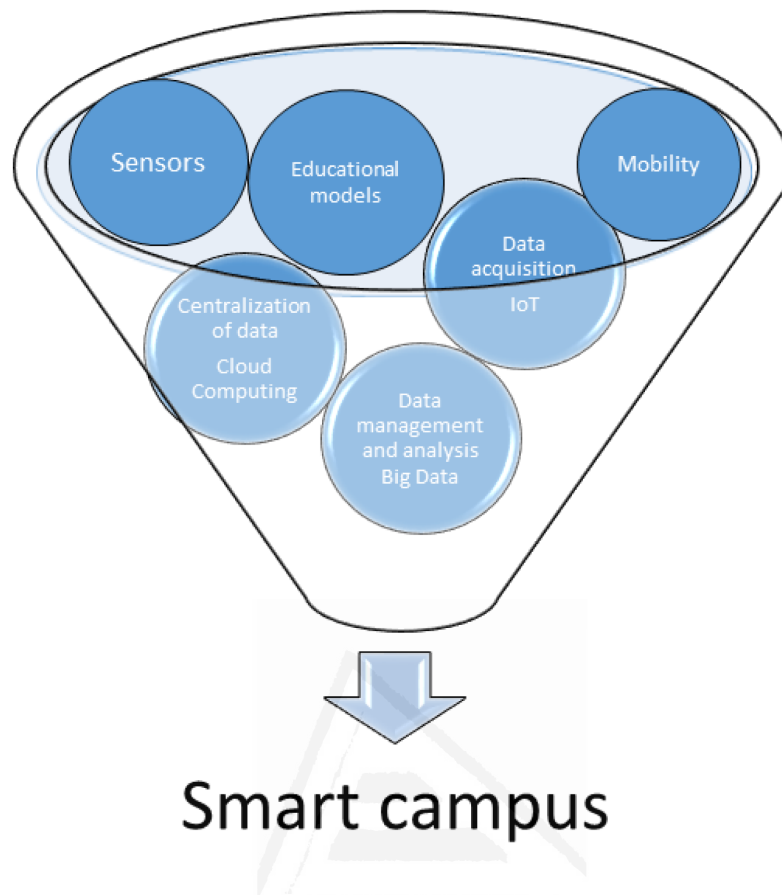


Figura 1.3: Componentes de un campus inteligente

LMS, etc. Para esto, es importante que cada sistema se integre en una plataforma de servicios. La plataforma de servicios es la encargada de gestionar cada sistema y centralizar toda la información y evento que se suscita dentro del campus. Por ejemplo, si ocurre una violación de seguridad en el ingreso a una zona restringida, los sistemas incluidos en este evento no solo se encargan de la detección y la notificación de lo sucedido. Todos los sistemas reportan el evento a la plataforma de servicios, donde toda la información se registra y se almacena para su análisis. Este ciclo permite mejorar cada proceso garantizando que el evento no se vuelva a repetir.

Los componentes de un campus inteligente se alinean al avance continuo de las TI que promueven la mejora en la calidad de la educación y en un desarrollo sostenible. En la Figura 1.3 se consideran los componentes principales de un campus inteligente y que han sido causa de investigación en el compendio.

- Adquisición de datos: en un campus inteligente los datos se obtienen a través de sensores y actuadores que son parte de un ambiente de IoT. El IoT permite obtener información de diferentes áreas y almacenarlas en una nube.
- Computación en la nube: el almacenamiento de los datos es un factor clave en un campus inteligente, pues de ellos depende el descubrimiento de las tendencias de la comunidad, así como sus necesidades.

- Movilidad: los campus universitarios pueden ser tan grandes como ciudades pequeñas, lo que conlleva a buscar modelos de movilidad que satisfagan las necesidades de la población y que permitan disminuir la contaminación del medio ambiente.
- Análisis de datos: el obtener conocimiento sobre los datos generados por los miembros del campus es uno de los principales aportes que hacen las ciudades inteligentes, pues por medio de este conocimiento se pueden satisfacer las diferentes necesidades existentes en el campus.
- Energía: el tema energético es un objetivo en un entorno inteligente que busca transformar el consumo de energía en un consumo eficiente. El consumo eficiente tiene como resultado una mejor convivencia con la naturaleza. En la actualidad, los gobiernos tratan de implementar políticas que disminuyan el consumo energético de la sociedad a través de energías renovables y programas de concientización en su uso.
- Automatización: la automatización dentro de un campus inteligente se genera a través de edificios inteligentes donde la domótica tiene mucha importancia en la manera de gestionar ambientes de confort. Estos ambientes permiten a la población generar conocimiento en ambientes ideales donde el aprendizaje, el confort y la naturaleza conviven en armonía.

Dentro de un campus inteligente se integran las nuevas tecnologías, de tal manera que la adquisición de datos se realiza por medio del IoT que actúa como un sistema de conectividad de una multitud de dispositivos como sensores y sistemas embebidos (Uskov et ál., 2016). Estos dispositivos permiten la recolección de datos incidiendo en la optimización de los procesos. Esta información procesada permite aflorar el conocimiento y controlar el entorno, extrayendo patrones de comportamiento o información relevante para la toma de decisiones.

Los datos generados por los dispositivos de la red del campus son almacenados en el centro de datos para ser procesados y analizados en busca de conocimiento. Dentro del campus inteligente se han considerado los siguientes grupos de sistemas que generan información y alimentan el proceso de *big data*:

- Sistemas de control de acceso: dentro de la división de seguridad e identificación se trabaja con sistemas de identificación por radio frecuencia (RFID, acrónimo en inglés de *Radio Frequency Identification*) que es una tecnología inalámbrica y de captura de datos de los sensores. También se trabaja con los sistemas de localización en tiempo real (RTLS, acrónimo en inglés de *Real Time Location System*), otra tecnología de radio frecuencia utilizada para identificar en tiempo real el movimiento de los objetos etiquetados en el campus.
- Sistemas de automatización: existen varias herramientas para crear ambientes autónomos como la domótica, los sistemas de inteligencia ambiental, el software de control energético inteligente, etc.

1 Introducción

- Sistemas de seguridad: los sistemas de seguridad incluyen cámaras administradas con software especializado para la detección de incidentes. Para asegurar esta información el sistema realiza un escaneo y reconocimiento de las personas que se encuentran en cada área.
- Sistemas informáticos: los sistemas que gestionan tanto la administración como el aprendizaje en el campus universitario generan una gran cantidad de información de la población del campus que luego será tratada para generar conocimiento sobre las personas.

La forma en que los sensores interactúan con el entorno y las personas contribuye a crear una sociedad del conocimiento. Los sensores son responsables de recopilar información del entorno y enviarla a la nube, donde las partes interesadas pueden consumirla. La información convertida en conocimiento ayuda a que las decisiones se tomen de una manera informada, precisa y rápida. La tecnología propuesta por Kamilaris, Pitsillides, Prenafeta-Boldo, y Ali (2017) otorga mayor validez a esta propuesta, ya que la implementación de un ecosistema basado en el IoT garantizará que los estudiantes y los tutores tomen decisiones acertadas en relación con su entorno.

El siguiente componente que debe contemplar un campus inteligente es la computación en la nube. Su concepto es amplio y hace referencia especialmente a nubes públicas donde los datos de una institución son almacenados en infraestructuras contratadas y muchas veces independiente de la infraestructura de la institución. En esta tesis se da mayor aporte a las nubes privadas, puesto que los campus universitarios generalmente cuentan con infraestructura propia para el almacenamiento y gestión de datos garantizando su disponibilidad.

Las infraestructuras están compuestas por uno o varios centros de datos y equipos de comunicación que se encargan de comunicar las diferentes redes con el centro de datos (Nie, 2013). El concepto de un centro de datos concentra y centraliza la información y brinda una ventaja para el despliegue de la arquitectura y garantiza la disponibilidad y la calidad de los datos. Los servicios que brinda la universidad a sus miembros están gestionados a través de una variedad de servidores virtuales alojados en los servidores físicos con los que cuenta el centro de datos. La mayoría de los datos que se generan en el campus se almacenan en una nube privada creada en su propia infraestructura. La nube privada busca garantizar la seguridad en los datos, la calidad, la disponibilidad y flexibilidad ante cualquier evento.

El análisis de datos es el componente de un campus inteligente que mayor importancia tiene, pues sus resultados dotan de conocimiento al campus y que este pueda tomar decisiones con base en los resultados. En una primera etapa y con una prueba experimental se ha definido que en un campus inteligente lo ideal es el uso de *big data* (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a). Sin embargo, para determinar cuál arquitectura de *big data* se ajusta a las necesidades del campus inteligente, se realizó un análisis entre las posibles herramientas existentes. En la actualidad se utilizan dos herramientas destacadas para la implementación de *big data* el ecosistema Hadoop y Apache Spark (Ghazi y Gangodkar, 2015; Shanahan y Dai, 2017).

Para la aplicación de una arquitectura específica es importante tener una visión de lo que cada uno es, Hadoop es un *framework* para almacenar grandes conjuntos de

datos a través de grupos (*clusters*) de computadoras, puede escalar desde un solo sistema informático hasta miles de sistemas y esto da un almacenamiento local y poder de cómputo. En cambio, Apache Spark fue diseñado para un cómputo rápido, porque su característica principal es que procesa todo en la memoria, lo que aumenta la velocidad de procesamiento. Hadoop integra en su procesamiento MapReduce que utiliza almacenamiento persistente mientras que Apache Spark tiene conjunto de datos distribuidos resilientes conocido como (RDD, acrónimo en inglés *Resilient Distributed Datasets*). El rendimiento de Apache Spark es rápido, procesa en memoria y ofrece un análisis en tiempo real. El desempeño de Hadoop fue diseñado para recopilar continuamente los datos de varias fuentes sin tener problemas por el tipo de datos usado en el procesamiento por lotes, por esta razón se entiende que nunca fue construido para procesamiento en tiempo real.

Hadoop no posee modo interactivo, pero tiene complementos como Hive y Pig para que trabajar con MapReduce se convierta en algo más sencillo. Otra característica válida en la comparación de las arquitecturas es el tipo de licenciamiento, en este caso tanto Hadoop como Spark son arquitecturas de código abierto. El procesamiento se puede dividir en dos tipos por procesamiento por lotes y procesamiento por flujo, en el caso de Hadoop es un *framework* de procesamiento por lotes, envía un trabajo, lee los datos, realiza la operación escribe los resultados y los envía al *cluster* de forma sucesiva. Spark cubre algoritmos iterativos de procesamiento por lotes, consultas interactivas y la retransmisión en directo (*streaming*) (Ghazi y Gangodkar, 2015).

Un método eficaz para la tolerancia a fallos es el que utiliza MapReduce mediante el uso de *TaskTrackers*² que emite informes al JobTracker y si se pierde un informe, el jobtracker reprograma todas las operaciones. Mientras que Spark usa RDD que son un conjunto de elementos que toleran fallas y son operados en paralelo (Cohen y Acharya, 2013). Finalmente, en la seguridad, Hadoop posee múltiples formas de proporcionar seguridad, Hadoop admite Kerberos³, también es compatible con otros proveedores como el protocolo ligero/simplificado de acceso a directorios (LDAP, acrónimo en inglés de *Lightweight Directory Access Protocol*, ofrece encriptación con el sistema de ficheros distribuido de Hadoop (HDFS, acrónimo en inglés de *Hadoop Distributed File System*). La seguridad de Spark es un poco escasa en la autenticación, por lo que este necesita de HDFS para ejecutarse y así usar listas de control de acceso (ACL, acrónimo en inglés de *access control list*) y permisos de nivel de archivo.

Además de las razones técnicas que han sido establecidas en el análisis, también se consideraron temas como la información existente y el soporte por técnicos en cada una de los *frameworks*. Esto nos alineó más aún al uso de Hadoop como *frameworks* de *big data* que sería aplicado al campus inteligente. Hadoop permite un análisis efectivo de grandes volúmenes de datos, y los resultados fortalecen la toma de decisiones y mejoran los procesos educativos. Esta arquitectura también permite monitorear las opiniones de los estudiantes, así como la capacidad de sacar conclusiones sobre los problemas de aprendizaje presentados por ciertos grupos de estudiantes. Con Hadoop, las universidades pueden explotar datos complejos, analizarlos y personalizar los resultados

²Es un nodo en el clúster que acepta tareas (operaciones Map, Reduce y Shuffle) de un JobTracker.

³Es un mecanismo de autenticación de terceros, en el que los usuarios y servicios confían en un servidor Kerberos, para autenticarse entre sí.

1 Introducción

adaptando el proceso a las necesidades de la universidad y los estudiantes. Los principales problemas que se resuelven con Hadoop son la captura de datos, el almacenamiento, el filtrado, la transferencia, el análisis y la presentación del conocimiento.

En el análisis de los datos, cada uno de los entornos evaluados se consideran como un proyecto que a su vez se divide en varios subproyectos, lo que facilita la obtención de información para cada sistema incluido en Hadoop. Por ejemplo, un proyecto creado para analizar las bebidas que tienen el mayor consumo; este proyecto se subdivide para que cada nodo de Hadoop se encargue de un determinado proceso como el análisis de fechas, estaciones del año, marcas de las bebidas, ingredientes, etc. La habilidad del científico de datos consiste en plantear las preguntas correctas para ayudar a controlar los parámetros de un evento específico.

- ¿Cuáles son las bebidas que presentan los índices más altos de consumo en las temporadas de exámenes en el campus?
- ¿Cuáles son los lugares en el campus con la mayor densidad de población en invierno y verano?
- ¿Cuáles son las actividades que generan mayor conocimiento en los estudiantes en el campus?

Estas preguntas buscan resolver problemas comunes en una universidad y ayudar a mejorar el uso de los recursos y la comprensión de las tendencias universitarias. Para la primera pregunta, la información generada por el sistema de dispensadores automáticos ingresa al proceso de análisis. La información se envía desde la máquina dispensadora a un servidor virtual. La Figura 1.4 detalla la arquitectura de la adquisición de datos de los diferentes sistemas de sensores y actuadores.

En el caso específico de los dispensadores, estos contienen diferentes sensores que permiten a los actuadores generar eventos en la capa de datos (*data*) estos datos se envían a la capa de protocolos de comunicación (*communication protocols*). En esta capa, los datos se almacenan en diferentes bases de datos (relacionales o no relacionales) incluso pueden ser almacenadas en las nubes privadas del campus, esto dependerá de la aplicación y el servicio. El protocolo de comunicación de los sensores y actuadores a nivel comercial es variado. Sin embargo, con el propósito de diseñar una arquitectura escalable se estandarizó el uso de las tecnologías y la manera en que estos se comunican dentro del campus inteligente, para lo cual se utilizó el Protocolo de Control de Transmisión (TCP, acrónimo en inglés de *Transmission Control Protocol*).

En la capa de conocimiento (*knowledge*), los datos son analizados a través de diferentes procesos de análisis y minería de datos. El conocimiento generado se aplica en el buen uso de los recursos. La capa de servicios (*services*) es la encargada de presentar los resultados a las diferentes áreas o miembros del campus quienes son los que consumen el conocimiento generado.

La arquitectura por medio de Hadoop almacena los datos y cada vez que se necesita analizar un caso específico, el *cluster* activa al nodo maestro para que divida el proceso y asigne subprocesos a los nodos esclavos para reducir el tiempo de procesamiento y el consumo de recursos. Dentro del proceso de almacenamiento de los datos, se agregaron

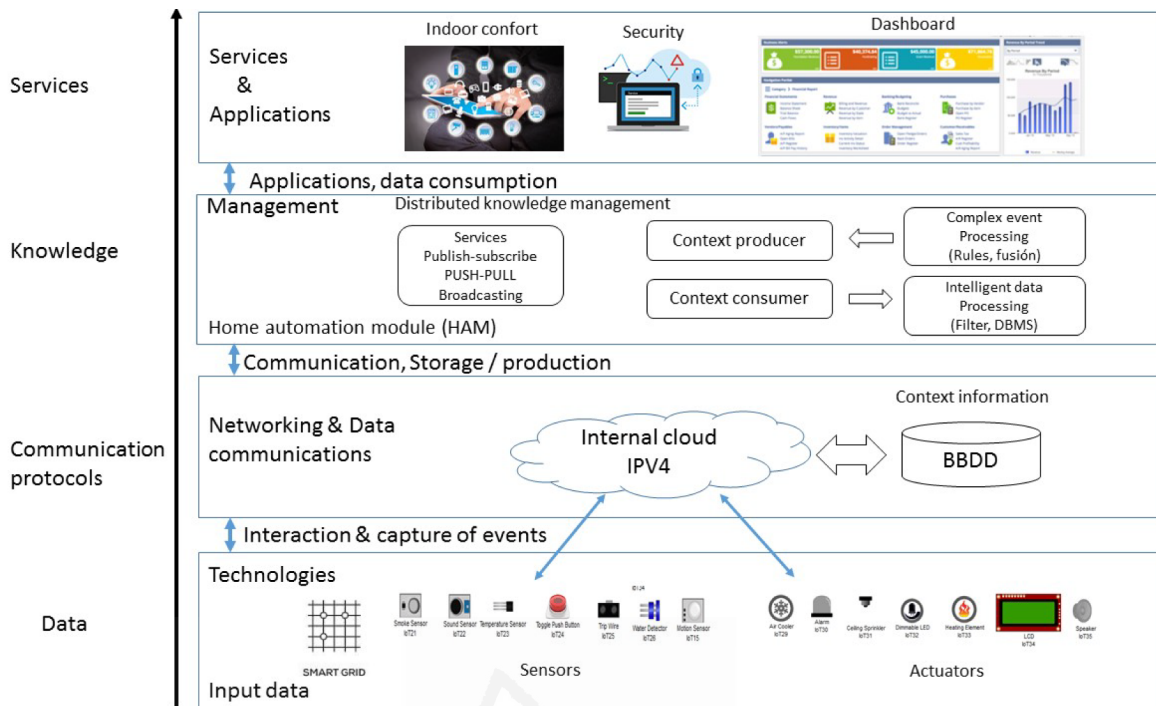


Figura 1.4: Arquitectura de un campus inteligente basado en el análisis de datos

en varios campos adicionales como, la fecha, la hora, el tipo de bebida, la ubicación. Estos datos llegan en texto plano y en tiempo real; por lo tanto, el proceso de análisis tiene información precisa para la toma de decisiones (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a).

1.2. Objetivos

La presente tesis tiene como objetivo general desarrollar una arquitectura para la gestión de datos en un campus inteligente. Los objetivos específicos son:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O2. Identificar los componentes y las tecnologías que son parte de un campus inteligente.
- O3. Diseñar una arquitectura que convierta un campus universitario tradicional en un campus inteligente.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

1.3. Método de trabajo

Para cumplir con el objetivo 1 se considero metodologías como las propuestas por Turner et ál. (2008) y Petersen, Vakkalanka, y Kuzniarz (2015) que brindaron el fundamento necesario para realizar la revisión informal y sistemática de la literatura. El análisis de las herramientas que permiten la gestión de datos se centró en el análisis de las diferentes arquitecturas que podrían integrarse a un campus inteligente. Es así que, en una primera etapa nos concentramos en analizar las herramientas que ofrece un BI para el análisis de datos. La segunda etapa consistió en realizar un análisis de las herramientas que ofrecieron las arquitecturas de *big data* y como se aplicaron a un campus inteligente.

Para el desarrollo del objetivo 2 se utilizan técnicas de análisis y validación que permitieron identificar cada uno de los componentes que debe incluir un campus inteligente. Además, cada uno de los componentes responde a procesos y tecnologías que permiten que estos convivan entre sí.

Para dar respuesta al objetivo 3 se aplican modelos para integrar las nuevas tecnologías a la infraestructura de un campus tradicional. Esta integración busca dotar al campus de “sentidos” para que este pueda interactuar con las personas o entre dispositivos con el único objetivo de mejorar la calidad de vida de todos los miembros del campus. Dentro de los trabajos realizados se incluyen conceptos y técnicas como la inteligencia artificial, el IoT, campus inteligentes, *big data*, etc. Estos conceptos permiten sin duda transformar un campus tradicional en un campus inteligente pensado en satisfacer las necesidades de los miembros de manera general o personalizada, ahí es donde se centra el concepto de entornos inteligentes.

El objetivo 4 se desarrolla en función de mejorar tanto el entorno del estudiante, así como en mejorar los métodos de aprendizaje. Para ello se considera una arquitectura para la gestión de datos en un campus inteligente, donde, se adopten técnicas de análisis de datos para generar conocimiento sobre los miembros. Este conocimiento ayuda a identificar las necesidades existentes en cada una de las áreas del campus y de sus usuarios y pone a su disposición toda la tecnología para mejorar la experiencia en sus actividades, facilitando los procesos administrativo y académicos.

1.4. Resultados

El primer sistema que se consideró para ser analizado fue el LMS de la universidad que participó en esta investigación, el análisis se lo realizó utilizando una arquitectura BI. Durante el proceso de implementación se realizaron varias pruebas de funcionamiento de esta arquitectura, en estas pruebas se tomó en cuenta el número de fuentes, el tipo de datos, la velocidad de procesamiento y la escalabilidad.

Los resultados obtenidos permitieron evidenciar el desempeño de los estudiantes con base en el tiempo de interacción en la plataforma, las calificaciones obtenidas y las actividades propuestas. Se pudo establecer la actividad que mejor se adaptaba a las necesidades de cada estudiante de un grupo de 55 personas. De las 55 personas se procesaron 20 GB, de información en un período de tiempo de 12 meses.

Sin embargo, al ingresar al BI un mayor número de variables desde diferentes fuentes, el BI empezó a presentar problemas en rendimiento debido a los altos requerimientos que este tenía en procesamiento (Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2018). Otra deficiencia encontrada fue el manejo en las diferentes fuentes y formatos de los datos, el ETL, aunque mostró robustez para la extracción de datos difícilmente soporta diferentes formatos de los datos. La solución práctica para convertir los formatos de los datos fue la generación de *script* lo que obliga a realizar un preprocesamiento de los datos antes de que ingresen al ETL. Este proceso llevaría mucho tiempo y convertiría al proceso de BI en un proceso manual, algo que penaliza a la funcionalidad de la propuesta.

La aplicación de un BI para garantizar la calidad en la educación con el uso de las TI fue muy limitado, pero nos dio una guía sobre el potencial del análisis de datos y la integración de las nuevas tecnologías. Para ello fue necesario buscar una tecnología que soporte una gran cantidad de datos, indistintamente de la fuente y el formato. Además, la escalabilidad es otro de los factores buscados para la integración de otras variables que de manera directa o indirecta afectan al aprendizaje.

La tecnología que presenta estas cualidades es el *big data*, además de su capacidad de procesar un gran volumen de datos en períodos de tiempo relativamente cortos. El disponer de una tecnología que soporta un gran volumen de datos permite integrar varios sistemas y que sus datos sean analizados y relacionados entre sí. Esta idea nos llevó a concebir la creación de un entorno inteligente que acoja los conceptos de las ciudades inteligentes donde la tecnología se enfoca en mejorar la calidad de vida de los miembros de la comunidad.

La siguiente prueba realizada fue la implementación de la arquitectura para la gestión de datos por medio de un *big data*. Esta arquitectura se puso a prueba al analizar los lugares con la mayor densidad de población dentro del campus. Para dar solución a este problema, se consideraron los datos generados por la red de área local inalámbrica (WLAN, acrónimo en inglés de *wireless local area network*). Los puntos de acceso (AP, acrónimo en inglés de *access point*) proporcionan información sobre la cantidad de *hosts* que están conectados, y el controlador que administra los AP puede emitir rastros de estos *hosts* que incluyen información del tiempo que se conectaron a la red y el identificador de AP al que están conectados. Las WLAN se han convertido en uno de los medios más utilizados en cualquier entorno, su alta disponibilidad y la movilidad son aprovechados para obtener información valiosa de las personas.

El Cuadro 1.1 muestra los resultados obtenidos en el análisis de densidad, en el que se consideran los datos de cuatro semanas aleatorias tanto de la temporada de verano como del invierno. Con los resultados de este análisis se busca mejorar la distribución de los AP dentro del campus. Esto optimiza los recursos tecnológicos y fundamentalmente recoge información importante sobre el entorno, como la cantidad de dispositivos que se conectan a la red en un período específico. Esta información es útil para determinar en qué áreas se necesitan más recursos de infraestructura o anchos de banda. El análisis también proporcionó información relevante sobre los puntos donde a menudo los estudiantes se encontraban. Esto es aprovechado para realizar actividades y charlas informativas en los lugares preferidos por los estudiantes obteniendo mejor acogida a la recibida en otros lugares. Los resultados que se muestran en el cuadro son los espe-

1 Introducción

rados; sin embargo, la herramienta permite un análisis cuantitativo, que mejora el uso de los recursos y genera objetividad en su uso.

Número de usuarios en verano por semana					Número de usuarios en invierno por semana			
Áreas	S1	S2	S3	S4	S5	S6	S7	S8
Recreativa	6,450	6,098	5,493	6,4536	4,839	3,927	2,983	1,826
Verdes	2,983	4,997	5,382	6,113	5,487	4,387	3,762	2,873
Cafeterias	502	493	650	528	638	936	1,182	1,382
Laboratorios	392	182	293	387	508	1,029	1,603	2,083
Bibliotecas	932	670	398	732	736	1,283	1,893	3,072
Edificios	1,982	805	1,038	963	973	1,562	1,793	1,923
Total	13,241	13,245	13,254	13,259	13,181	13,124	13,216	13,159

Cuadro 1.1: Áreas con la mayor densidad de población en un campus inteligente

El segundo análisis realizado con la arquitectura de gestión de datos se presenta en el Cuadro 1.2, donde el período, es el tiempo durante el cual se tomaron las muestras para el estudio. El período consistió en dos semanas, que es el tiempo regular que duran los exámenes. Se tomaron muestras cada cuatro horas; se verificó la cantidad de bebidas disponibles en las máquinas dispensadoras y se las clasificaron en cuatro clases. El porcentaje está relacionado con la cantidad consumida y la cantidad total de bebidas que posee una máquina expendedora, que es 288 unidades. El Cuadro 1.2 muestra un alto consumo de café, seguido de bebidas de cola, que de una forma u otra contienen cafeína.

Período	Tipo de bebida	Porcentaje
07:00 - 11:00	Café	60 %
07:00 - 11:00	Coca-Cola	25 %
07:00 - 11:00	Jugos	10 %
07:00 - 11:00	Otros	5 %
11:00 - 15:00	Café	53 %
11:00 - 15:00	Coca-Cola	30 %
11:00 - 15:00	Jugos	11 %
11:00 - 15:00	Otros	6 %
15:00 - 19:00	Café	42 %
15:00 - 19:00	Coca-Cola	39 %
15:00 - 19:00	Jugos	15 %
15:00 - 19:00	Otros	4 %

Cuadro 1.2: Consumo de bebidas en temporada de exámenes

Con los resultados obtenidos, se pueden hacer varios ajustes dentro de la optimización de los recursos. Por un lado, tiene la capacidad de proyectar la cantidad de bebidas requeridas según su clasificación en diferentes estaciones. Por otro lado, con los resultados del análisis, es posible complementar un estudio sobre el estrés en los estudiantes

al realizar un examen. El estudio permite crear campañas de concientización sobre la dependencia de la cafeína y el tratamiento del estrés en la población estudiantil.

La movilidad es otro punto importante en un campus inteligente, donde, uno de los principios fundamentales es la convivencia en un entorno sostenible. Para ello es necesario reducir la emisión de CO_2 que producen los vehículos dentro del campus. Como punto de partida, se consideró que un campus universitario geográficamente puede ser tan grande como una ciudad pequeña.

Para resolver este problema, hay opciones como implementar un sistema de transporte interno que realiza los recorridos cada cierto período a las diferentes áreas del campus. Esta opción resuelve hasta cierto punto que los habitantes no utilicen sus vehículos y, por ende, se reduce la emisión de gases. Sin embargo, no es una solución óptima porque su implementación se basa en la experiencia del personal a cargo de la movilidad o transporte interno.

Nuestra arquitectura cubre estas necesidades definiendo los tiempos, unidades y rutas de cada transporte en función del análisis de los datos existentes en cada área del campus. La Figura 1.5 muestra el diagrama de flujo bajo el cual se concibe el proceso para resolver este problema. La primera etapa es la responsable de recopilar los datos que provienen de los sistemas de ubicación a través de la identificación de áreas con mayor densidad poblacional expuesto en el primer análisis. Si los datos existen y son apropiados, la arquitectura de *big data* asigna los procesos a los nodos para realizar el procesamiento de la información.

El siguiente flujo conduce al almacenamiento de los datos y su análisis a través de Hadoop. Hadoop realiza el proceso en función de varios parámetros que se establecieron previamente. Por ejemplo, al identificar patrones en los estudiantes, se determinan los lugares que estos frecuentan. Además, la arquitectura toma información de análisis ya existentes y aprende de los resultados. Esta información permite establecer la distancia que recorren los estudiantes en sus actividades habituales que incluyen el transporte entre facultades o áreas del campus.

Los sistemas de videovigilancia envían información sobre los puntos de parada de autobuses; si existen personas a la espera, el sistema de análisis identifica y alerta al administrador para que envíe unidades para el traslado de usuarios. Los sistemas de gestión académica brindan información sobre los horarios de los estudiantes para que el *big data* proyecte posibles horarios en los que las paradas de autobuses se saturan.

Una vez que se notifica al administrador, este envía la unidad y se ejecuta el traslado. Si por alguno motivo el envío de transporte no se realiza, el ciclo vuelve a la fase de recopilación de datos y se repite el proceso. Si los datos no existen o no son suficientes para el análisis, el proceso se detiene para buscar información que le ayude a resolver el problema.

1.5. Estructura de la tesis

Esta tesis se estructura en tres partes: la Síntesis ubicada en la Parte I; los Trabajos publicados que contribuyeron al compendio de publicaciones en esta tesis se presentan en la Parte II, finalmente, en la Parte III se extraen las conclusiones de esta disertación y se discuten las propuestas de posibles investigaciones futuras.

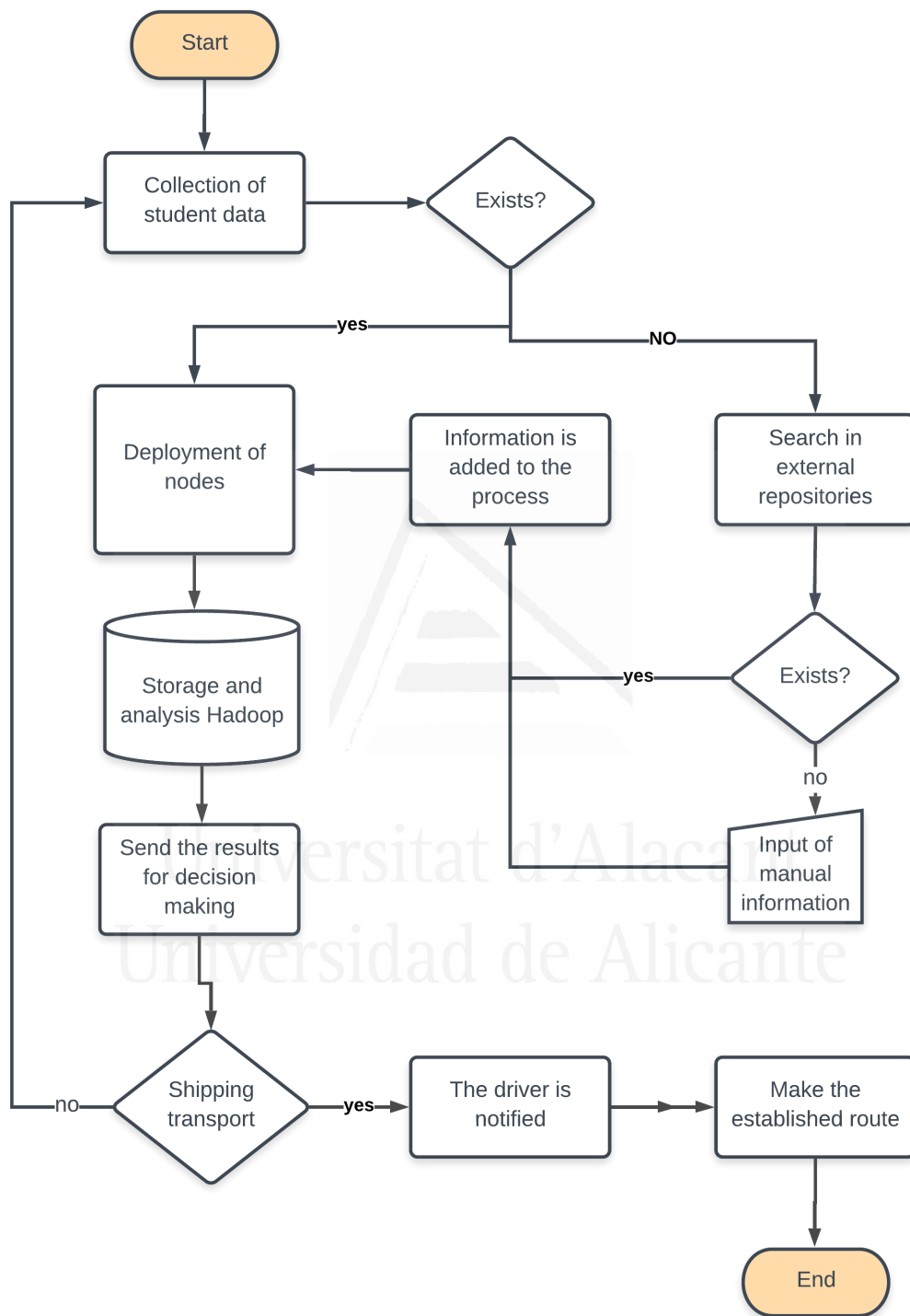


Figura 1.5: Diagrama de flujo de un proceso de transporte interno mediante *big data*

La tesis consta de los siguientes capítulos:

- El capítulo 1 Introducción**, incluye motivación, objetivos, método, resultados, estructura de la tesis y convenciones de escritura.
- El capítulo 2 Publicaciones y visibilidad**, incluye publicaciones en revistas, congresos y otras publicaciones.
- El capítulo 3 Compendio de publicaciones**, incluye todos los artículos publicados durante los estudios de doctorado..
- El capítulo 4 Big Data, the Next Step in the Evolution of Educational Data Analysis**, es un artículo publicado en los *Proceedings of the International Conference on Information Technology & Systems*, ICITS que incluye referencia, contribución y texto completo.
- El capítulo 5 Comprehensive Learning System Based on the Analysis of Data and the Recommendation of Activities in a Distance Education Environment**, es un artículo publicado en la revista *International Journal of Engineering Education*, IJEE que incluye referencia, contribución y texto completo.
- El capítulo 6 Management of Educative Data in University Students with the Use of Big Data Techniques**, es un artículo publicado en la *Revista Ibérica de Sistemas y Tecnologías de la Información*, RISTI que incluye referencia, contribución y texto completo.
- El capítulo 7 Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus**, es un artículo publicado en la revista *Sustainability* que incluye referencia, contribución y texto completo.
- El capítulo 8 Application of a Big Data Framework for Data Monitoring on a Smart Campus**, es un artículo publicado en la revista *Sustainability* que incluye referencia, contribución y texto completo.
- El capítulo 9 Conclusiones**, incluye las conclusiones del trabajo y las contribuciones.
- El capítulo 10 Trabajos futuros**, incluye los trabajos en los que se está trabajando actualmente y las ideas que el autor tiene para continuar con la línea de investigación.

1.6. Convenciones de escritura

Durante el desarrollo de este trabajo se utilizan varios acrónimos para hacer referencia a diferentes tecnologías, esto a causa que el público en general reconoce los términos. Por ejemplo TIC para referirse a las tecnologías de la información y comunicación. El formato utilizado la primera vez que se utilizan es la definición y entre paréntesis el acrónimo. Por ejemplo, Tecnologías de la información y comunicación (TIC).

1 Introducción

Se ha optado por usar algunos acrónimos en inglés porque así se utilizan en español. Por ejemplo, sistemas de gestión de aprendizajes (LMS, acrónimo en inglés de *learning management system*).

Las citas se reproducen en el idioma original de la referencia de donde provienen.

Las Figuras y Cuadros son de elaboración propia del autor, a menos que se indique lo contrario en el título.

Algunas de las figuras que se incluyen provienen de las publicaciones que conforman el compendio. Por esta razón, varias de las figuras se presentan en inglés.

Las palabras en un idioma distinto al castellano se presentan en letra cursiva. Por ejemplo, *English, français*.

Debido al origen del autor, cuando existan sinónimos se prefieren los vocablos más utilizados en Sudamérica. Por ejemplo, computador por ordenador.

Las cifras numéricas de miles están separadas por coma y las cifras decimales están separadas por punto, siguiendo las normas internacionales.



Universitat d'Alacant
Universidad de Alicante

2 Publicaciones y visibilidad

2.1. Publicaciones

A lo largo del desarrollo de esta tesis doctoral, y como resultado de la investigación realizada, se han publicado diferentes artículos en revistas, un capítulo de libro y varios artículos de congresos. Todas las publicaciones presentan los resultados y las contribuciones hechas a la comunidad científica durante este trabajo de doctorado. Algunos de los artículos se desarrollaron para lograr los objetivos de esta disertación, mientras que otros se desarrollaron en colaboración con otros investigadores para alcanzar objetivos de investigación paralelos. El total de artículos publicados durante el período de doctorado fue de 11, que incluyeron cuatro artículos de revistas y uno de congreso (que se presentan en el compendio), además, seis artículos de congresos que se presentan en el apéndice. Las siguientes secciones enumeran estos trabajos.

2.1.1. Revistas

Esta subsección describe las publicaciones en revistas científicas. Estas publicaciones representan el contenido principal de este trabajo de investigación. Por lo tanto, fueron incluidos en el compendio de publicaciones de esta tesis. Los detalles de las publicaciones se muestran en el Cuadro 2.1, donde la primera columna muestra la identificación de la revista; la segunda columna, muestra el nombre de la revista con su ISSN; la tercera columna muestra el factor de impacto del Journal Citations Report (JCR); la cuarta columna describe el factor de impacto del Scimago Journal Ranking (SJR); la última columna presenta la indexación de la revista, ya sea que el artículo haya sido indexado en Scopus (SCO), Web of Science (WOS) o en el Directorio de revistas de acceso abierto (DOAJ).

1. “Comprehensive Learning System Based on the Analysis of Data and the Recommendation of Activities in a Distance Education Environment” (Villegas-Ch, Palacios-Pacheco, Buenaño-Fernandez, y Luján-Mora, 2019). Este artículo fue publicado en la revista J1. En el capítulo 5 se detalla lo presentado en este artículo.
2. “Management of Educative Data in University Students with the Use of Big Data Techniques” (Villegas-Ch, Palacios-Pacheco, Ortiz-Garcés, y Luján-Mora, 2019). Este artículo fue publicado en la revista J2. En el capítulo 6 se detalla lo presentado en este artículo.

2 Publicaciones y visibilidad

3. “Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus” (Villegas-Ch, Molina-Enriquez, et ál., 2019). Este artículo fue publicado en la revista J3. En el capítulo 7 se detalla lo presentado en este artículo.
4. “Application of a Big Data Framework for Data Monitoring on a Smart Campus” (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a). Este artículo fue publicado en la revista J4. En el capítulo 8 se detalla lo presentado en este artículo.

Id.	Revista	JCR IF	SJR	Indexado
J1	International Journal of Engineering Education. ISSN: 0949-149X. Irlanda	0.611	0.43	DOAJ, WOS, SCO
J2	Revista Iberica de Sistemas e Tecnologias de Informacao. ISSN: 1646-9895. Brasil	S/N	0.22	DOAJ, SCO
J3	Sustainability. ISSN: 2071-1050. Suiza	2.592	0.55	DOAJ, WOS, SCOPUS
J4	Sustainability. ISSN: 2071-1050. Suiza	2.592	0.55	DOAJ, WOS, SCOPUS

Cuadro 2.1: Descripción de las revistas

2.1.2. Congresos

Los artículos publicados en memorias de congresos se detallan en el Cuadro 2.2, incluyendo la identificación del congreso, nombre, indexación en Scopus, país, ciudad y fechas de realización del congreso. Todos los congresos en que se ha publicado tienen procesos de revisión por pares y han sido un punto clave en el proceso de investigación. Estos trabajos han permitido establecer los lineamientos y relaciones existentes entre las nuevas tecnologías y la calidad de la educación. Por este motivo varias de estas publicaciones se han considerado como apoyo fundamental al compendio.

1. “Analysis of Data Mining Techniques Applied to LMS for Personalized Education” (Villegas-Ch y Luján-Mora, 2017a). Este artículo fue publicado en el congreso C1. En el apéndice B se detalla lo presentado en este artículo.
2. “Systematic Review of Evidence on Data Mining Applied to LMS Platforms for Improving E-Learning” (Villegas-Ch y Luján-Mora, 2017b). Este artículo fue publicado en el congreso C2. En el apéndice C se detalla lo presentado en este artículo.
3. “Data Mining Toolkit for Extraction of Knowledge from LMS” (Villegas-Ch et ál., 2017). Este artículo fue publicado en el congreso C3. En el apéndice D se detalla lo presentado en este artículo.

4. “Big Data, The Next Step in The Evolution of Educational Data Analysis” (Villegas-Ch, Luján-Mora, Buenaño-Fernandez, y Palacios-Pacheco, 2018). Este artículo fue publicado en el congreso C4 y es parte del compendio, su aporte se lo detalla en el capítulo Cuatro.
5. “Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining” (Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2018). Este artículo fue publicado en el congreso C5. En el apéndice E se detalla lo presentado en este artículo.
6. “Application of Data Mining for the Detection of Variables that Cause University Desertion” (Palacios-Pacheco et ál., 2019). Este artículo fue publicado en el congreso C6. En el apéndice F se detalla lo presentado en este artículo.
7. “Artificial Intelligence as a Support Technique for University Learning” (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019b). Este artículo fue publicado en el congreso C7. En el apéndice G se detalla lo presentado en este artículo.

Id.	Congreso	Indexado	País/Ciudad	Fecha
C1	IEEE World Engineering Education Conference (EDUNINE 2017).	SCO, WOS	Brasil / Santos	Marzo 19-22, 2017
C2	In International Technology, Education and Development Conference (INTED 2017).	WOS	España / Valencia	Marzo 6-8, 2017
C3	International Conference on Education Technology and Computers (ICETC 2017).	SCO, WOS	España / Barcelona	Diciembre 20-22, 2017
C4	International Conference on Information Technology & Systems (ICITS 2018).	SCO, WOS	Ecuador / Santa Elena	Enero 10-12, 2018
C5	IEEE World Engineering Education Conference (EDUNINE 2018).	SCO, WOS	Argentina / Buenos Aires	Marzo 11-14, 2018
C6	International Conference on Technology Trends (CITT 2018).	SCO, WOS	Ecuador / Babahoyo	Agosto 29-31, 2018
C7	IEEE World Engineering Education Conference (EDUNINE 2019).	SCO, WOS	Perú / Lima	Marzo 17-20, 2019

Cuadro 2.2: Descripción de los congresos

2.2. Visibilidad

La visibilidad científica es imprescindible para que los investigadores muestren sus resultados, reciban comentarios y críticas e intercambien ideas con sus colegas, además aumenten las citas en su trabajo. Por lo tanto, durante el período de estudios de doctorado, se tomaron algunos cursos en esta materia y se crearon diferentes perfiles académicos para aumentar el impacto de la investigación realizada. El Cuadro 2.3 muestra los perfiles académicos del autor de esta tesis. Además, una de las decisiones tomadas para ganar mayor visibilidad en los productos de la investigación científica realizada fue el envío a revistas de acceso abierto (*open access*). Este tipo de revistas presentan una opción excelente para que los interesados puedan obtener todo el trabajo sin costo alguno que generalmente lo asume el lector. Por este motivo, tres de los cinco artículos del compendio se publicaron como documentos de acceso totalmente abierto (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019; Villegas-Ch, Palacios-Pacheco, Buenano-Fernandez, y Luján-Mora, 2019; Villegas-Ch, Molina-Enriquez, Chicaiza-Tamayo, Ortiz-Garcés, y Luján-Mora, 2019).

La estrategia de visibilidad seguida para difundir el trabajo científico ha aumentado el impacto de la investigación realizada. A partir del 2017, hay 36 citas en Google Scholar (GS), 28 citas en SCO y 6 citas en WOS.

Id.	Perfil Académico	URL
P1	ORCID	https://orcid.org/0000-0002-5421-7710
P2	Google Scholar	https://scholar.google.es/citations?user=89Uix-Y4AAAAJhl=es
P3	Researchgate	https://www.researchgate.net/profile/William-Villegas4
P4	Scopus	https://www.scopus.com/authid/detail.uri?authorId=57194408086

Cuadro 2.3: Perfiles académicos del autor de la tesis

En la Figura 2.1 se muestran los trabajos del autor indexados en Scopus, de los 16 trabajos indexados se han recibido 28 citas y en estos trabajos se ha colaborado con ocho coautores.

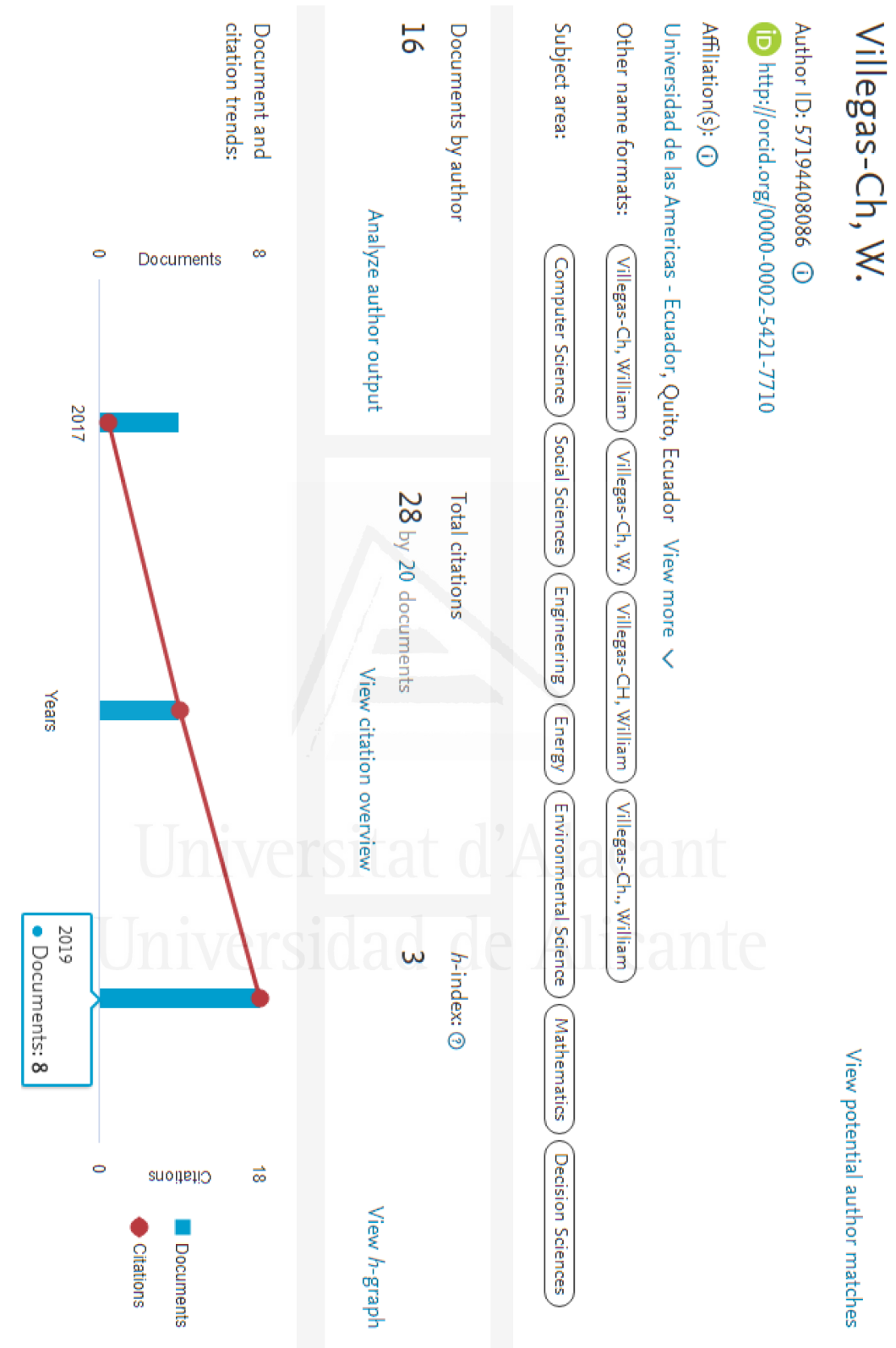


Figura 2.1: Detalles del autor en Scopus

Parte II

TRABAJOS PUBLICADOS



Universitat d'Alacant
Universidad de Alicante

3 Publicaciones

Este capítulo presenta las principales publicaciones de la investigación realizada durante los estudios de doctorado en orden cronológico. Cuatro de las publicaciones incluidas en el compendio son artículos de revistas y una de congreso (Villegas-Ch, Luján-Mora, Buenaño-Fernandez, y Palacios-Pacheco, 2018). Cuatro de las publicaciones tuvieron un factor de impacto. Dos de ellas fueron indexadas en una revista clasificada en el segundo cuartil (Q2) (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a; Villegas-Ch, Molina-Enriquez, et ál., 2019), una en el cuarto cuartil (Q4) del Journal Citation Report de WOS (Villegas-Ch, Palacios-Pacheco, Buenaño-Fernandez, y Luján-Mora, 2019) y una en el tercer cuartil (Q3) del Scimago Journal Rank de SCO (Villegas-Ch, Palacios-Pacheco, Ortiz-Garcés, y Luján-Mora, 2019).

Los detalles de las publicaciones se muestran en la Figura 3.1, se han considerado los artículos que son parte del compendio, incluido el artículo de congreso “C4” del cuadro 2.2. Los artículos de revistas son colocados de acuerdo a su identificador que consta en el Cuadro 2.1. El periodo establecido va desde el 2017 hasta 2019, considerando que en estos años se obtuvieron todos los trabajos para presentar el compendio de publicaciones. Además, en la columna de artículos se incluyó el objetivo al que aportan; se hizo a través de la numeración que le corresponde “O1-O4”. La líneas entrecortadas marcan el mes en el que cada artículo fue publicado.

Universitat d'Alacant
Universidad de Alicante

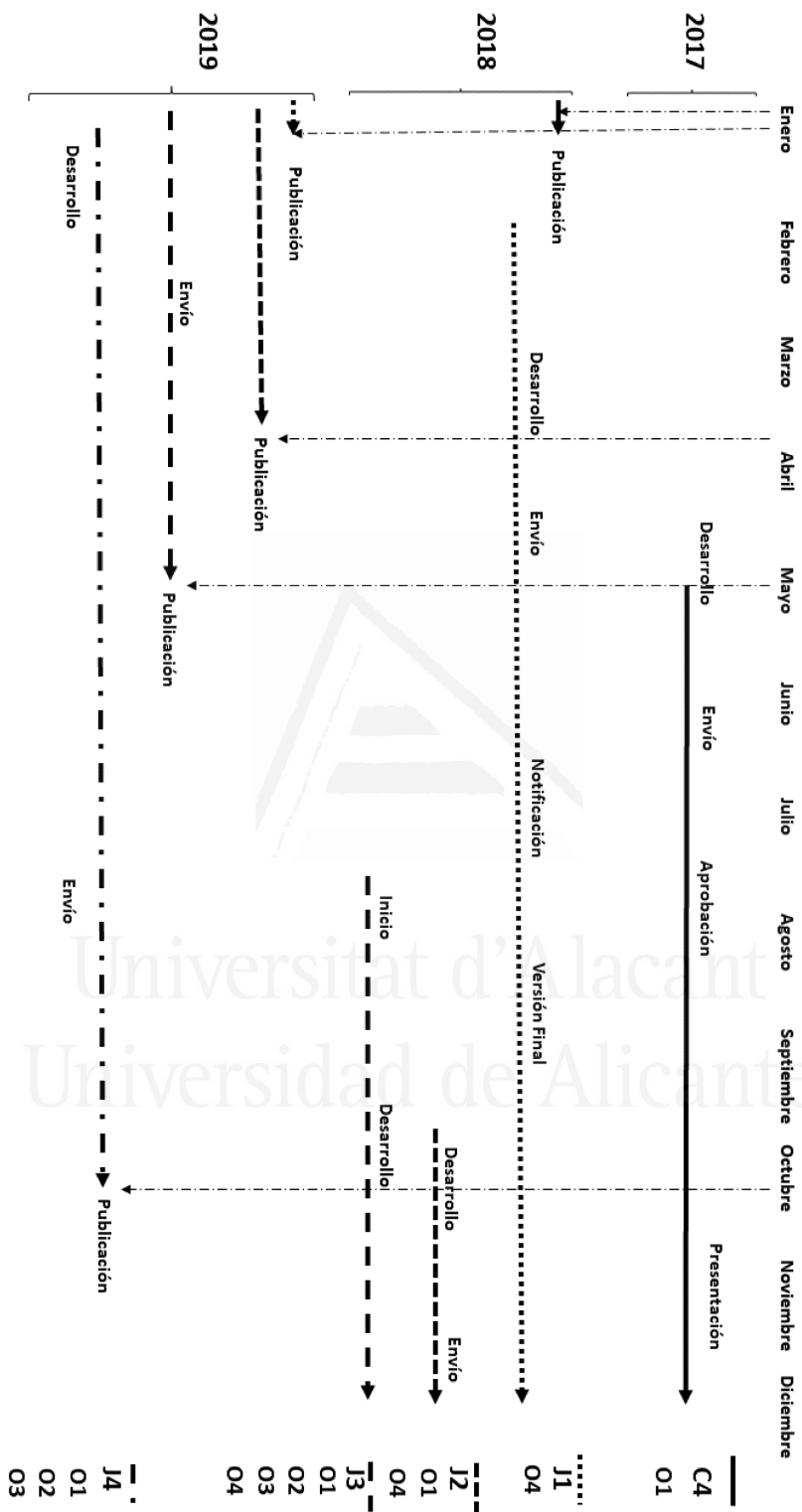


Figura 3.1: Línea de tiempo del compendio de publicaciones

4 Big Data, the Next Step in the Evolution of Educational Data Analysis

Villegas-Ch, W., Luján-Mora, S., Buenaño-Fernandez, D., y Palacios-Pacheco, X. (2018). Big Data, the Next Step in the Evolution of Educational Data Analysis. Proceedings of the International Conference on Information Technology & Systems (ICITS 2018). Advances in Intelligent Systems and Computing, 721, 138-147. (Villegas-Ch, Luján-Mora, Buenaño-Fernandez, y Palacios-Pacheco, 2018)

Disponible en:

URL: https://link.springer.com/chapter/10.1007/978-3-319-73450-7_14

DOI: https://doi.org/10.1007/978-3-319-73450-7_14

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.

Universitat d'Alicant
Universidad de Alicante

Big Data, the Next Step in the Evolution of Educational Data Analysis

W. Villegas-Ch¹(✉), Sergio Luján-Mora²,
Diego Buenaño-Fernandez¹, and X. Palacios-Pacheco³

¹ Facultad de Ingeniería y Ciencias Agropecuarias,
Universidad de Las Américas, Quito, Ecuador
{william.villegas, diego.buenano}@udla.edu.ec

² Departamento de Lenguajes y Sistemas Informáticos,
Universidad de Alicante, Alicante, Spain
sergio.lujan@ua.es

³ Departamento de Sistemas, Universidad Internacional del Ecuador,
Quito, Ecuador
xpalacio@uide.edu.ec

Abstract. This paper presents an analysis of new concepts such as big data, smart data and a data lake. It is to sought integrate learning management systems with these platforms and contribute to education by making it personalised and of quality. For this study, the data and needs of a university in Ecuador have been considered. This university has set its goals to the discovery of patterns, using data mining techniques applied to cubes generated in a data warehouse. However, the institution wants to integrate all the systems and sensors that contribute to the educational development of the student. Integrating more systems into the data warehouse has compromised the veracity of the data and the processing capabilities have been surpassed by the volume of data. The paper proposes the use of one of the platforms analysed and its tools to generate knowledge and to help the students to learn.

Keywords: Analysis of data · Big data · Data lake · Data mining
Data warehouse · Smart data

1 Introduction

Education currently uses learning platforms, information and communications technology (ICT) to manage learning. The aim of this integration of pedagogy and ICT is to create learning methods that are accessible and used by students. The integration uses, as its main tool, the learning management system (LMS) [1]. LMSs have become the main repository of student performance information. In order to take advantage of this information, data mining techniques are used to obtain patterns in student performance [17]. For this work the Moodle platform of a university in Ecuador is used. In this university, the use of Moodle has been institutionalised and policies of use have been created for a standard management of the courses of each one of the teachers. The Moodle platform has been customised according to the policies and needs presented by

the academic department in charge of evaluating student learning [7]. The customisation of the platform is based on the integration of modules and links that allow the proper management of academic resources which students can use for the development of activities. The large amount of data generated has surpassed the analysis capabilities to be processed in a conventional way. To supplement this processing, it is necessary to work with new concepts such as big data, smart data, data lake and data mining. The convergence of these concepts with the educational systems will allow evaluation of the data and transform it into useful information. This paper defines which of the platforms analysed, based on the needs of an educational institution, gives greater benefits for decision making and contributes to the improvement of education.

The work is composed as follows: Sect. 2 presents the concepts used in the development of this work; Sect. 3 determines the method used for the analysis of big data, smart data and the data lake; Sect. 4 presents the analysis of results to choose the platform that gives the solution to the problems raised, and Sect. 5 presents the conclusions that have been reached from the work done.

2 Preliminary Concepts

This article takes into account several key concepts that help the management and application of different analyses that help clarify new processes within the educational field.

2.1 Big Data

Big data is born from data sets or combinations of data sets whose size, complexity (variability) and speed of growth (velocity) make it difficult to capture, manage, process or analyse using conventional technologies and tools [10]. Big data requires a combination of different tools such as relational and non-relational databases [5]. It makes use of analytical tools to transform data into value information. Big data analysis is done on hundreds or thousands of blade servers. The tasks are distributed in intelligent networks of parallel processing that allow the use of analyses in a real time to customise, segment, optimise prices and relate in with customers [6].

Big data allows the management and analysis of huge volumes of data that cannot be processed in a conventional way. The size used to determine if a dataset is considered big data is not defined and keeps changing over time [9]. However, as a benchmark, analysts and professionals refer to datasets ranging from 30–50 TB to several petabytes.

Big data is a complex data set; this is mainly due to the unstructured nature of much of the data generated by modern technologies. These technologies are: web logs, radio frequency identification and built-in sensors in devices, vehicles, Internet searches, social networks, smart phones, GPS devices and call centre records.

Organizations have handled large volumes of data for a long time and have developed data warehouses and powerful analytical tools. These tools allow the adequate handling of large volumes of data [2]. The goal of big data is to turn the data into information that facilitates decision making in real time. However, more than a matter

of size, it is a business opportunity. Organizations use big data to understand the profile, needs and feelings of their customers regarding the products or services offered. In order to use big data effectively, it must integrate structured data from a conventional business application, such as enterprise resource planning (ERP) or customer relationship management (CRM).

2.2 Smart Data

The concept smart data appears following the big data and focuses on processing data, in order to convert them into statistics. These statistics serve to find the content of higher value (separates the useful from the useless), smart data is a complex data filter [5]; it is a concept that revolves around mass information management but only of the one that has a real value.

To understand the smart data easily, the big data would be information gathering, processing and filtering. The smart data would act once all that processed information is available and use mathematical formulae to convert data into “axiomatic” responses on a market [4].

2.3 Data Lake

A data lake is a storage repository that contains a large amount of raw data and is kept there until needed [8]. A data lake is a concept close to data warehouse that allows for the storage and processing of large volumes of data. They are used to collect raw data before the data sets pass into a production analytical environment, such as a data warehouse [11].

The main benefit of a data lake is the centralisation of disparate content sources. Once assembled, these sources can be combined and processed using big data (searches and analyses). The disparate content sources often contain confidential information that will require the implementation of appropriate security measures in the data lake.

The content of the data lake can be normalized and enriched. This may include extracting metadata, format conversion, augmentation, entity extraction, crosslinking, aggregation, de-normalization or indexing. Scattered users around the world can have flexible access to a data lake and its content from anywhere. Accessibility increases re-use of the content and helps the organisation gather the data needed more easily to drive business decisions.

2.4 Data Mining

Data mining is the analysis stage of knowledge discovery in databases (KDD) [3]. It is a field of statistics and computer science that attempts to uncover patterns in large volumes of data. It uses the methods of artificial intelligence, machine learning, statistics and database systems. The general objective of the data mining process is to extract information from a set of data and transform it into useful information. It finds repetitive patterns, trends or rules that explain the behaviour of data in a given context. The data are the raw materials, the user attributes some special meaning to them and they become information; the specialists elaborate or manage a model, so that the

interpretation that arises between the information and that model represents an added value, which is called knowledge.

3 Method

For the development of this work, the analyses of big data, smart data and the data lake have been undertaken. The objective is to specify how these three concepts can be applied to educational institutions and which will help to improve education and the generation of knowledge. For this analysis, technical data such as data volume, historical management needs, pre-processing of data, etc. are considered.

Once the results are obtained, a method is proposed that helps to discover which of the platforms provides the best solution for the needs of the knowledge analysis within the LMS. The tool should detect patterns in student behaviour that help define strategies for resource improvement, as well as educational activities provided on LMS platforms.

3.1 Description of the Problem

For this work we analyse the data of a university in Ecuador. Since 2010, this educational institution has worked with the Moodle platform as an e-learning tool. The institution has eight thousand students in different programs. The students, to finish their programs, must culminate ten educational periods. Each period consists of four months and on average six subjects are given. The use of the Moodle platform is obligatory in support of the teaching-learning activity. For the control of the use of the platform, the administrators generate reports of the activities performed at the end of each period.

At the beginning of each educational period, a new virtual classroom is generated for each subject within the LMS. The generation of virtual classrooms has the purpose of removing the virtual classrooms from past periods and, in such a way, maintains backups and records of the data. With the passage of time, standards have improved the use of the platform, in the same way that modules have been created to help teacher management. These improvements have allowed the creation of new techniques and resources that help the development of the learning in the students. Currently the platform has links to different websites, multimedia material and games. It has even been integrated with tools that allow virtual tutorials, as well as systems that allow the detection of similarity between documents or any file that is on the internet.

The registry of activities in the platform has made it the most important repository that handles educational data within the institution. The constant growth of data volume has forced to implement a data warehouse for data analysis. The analysis is based on the conventional data mining application adapted from business analysis to data cubes. Data mining has revealed patterns in the behaviour of educational data and improved learning techniques. However, the data warehouse tool has fulfilled its life cycle and exceeded the processing capabilities compromising the accuracy of the results. To solve these problems, there needs to be an improvement in the design of the data

warehouse or, in turn, to apply one of the new trends in data analysis and to take advantage of the information that has been generated during this time.

For our work we consider the following data from an Ecuadorian university. For the calculation, the number of hours that an average student with 7/10 grades devotes to the Moodle platform has been considered, as Table 1 shows.

Table 1. Calculation of hours of use of the Moodle platform

# hours of an average student on platform per week	Days per week	# weeks per period	# of periods per year	Total hours per year
5.5	5	16	2	880
Total number of students			8000	7040000

Table 2 calculates the average storage utilization used by each student over the course of a year. These data are considered important for analysis because they give us an exact figure on the consumption that has been generated in the database of the Moodle platform. The calculation has been made with the known data that are: the 5 TB of storage used from 2010 to 2017 and the 8000 students with which it counts, the institution until the indicated year. The result is 0.62 GB for each student since 2010 then we have detailed the storage consumption of each student per year.

The amount of storage shown in Table 2 indicates that it is a very low growth. However, due to data analysis needs, academic authorities have proposed increasing the management of the LMS platform and generating data by 50% with respect to megabyte used per year and per student.

Table 2. Calculation of average Moodle storage by year and student

Megabyte used per year and per student	Gigabyte consumed by students since 2010	Terabyte consumed since 2010
89.3 MB	0.62 GB	5 TB

3.2 Characteristics of Big Data, Smart Data and the Data Lake

Table 3 shows a comparison of the characteristics of big data, smart data and the data lake. It has taken as a reference the capacity of these platforms for data management where big data handles a large volume of data [14]. The Smart data, with respect to capacity, acts as a data filter taking only the most important ones and, on those, applies the techniques of data analysis. The data lake, with its capacity, shares the characteristics of a data warehouse with the advantage that it does not need a process of extraction, transformation and load (ETL) that is in charge of a pre-processing of the data [18].

Table 3. Analyses of datasets

Platform	Capacity	Sources	Storage cost
Big data	Large dataset	Complex, structured or unstructured data set	High
Smart data	It is a data filter	Raw data	Medium
Data lake	Improved version of the data warehouse	Raw data	Low

With regard to the sources each of these supports, all three have similar characteristics, they feed on structured or unstructured datasets. Another feature that is important is the cost of storage where the big data has the highest cost due to the large volume of data and the infrastructure it supports for data acquisition. On the other hand, smart data, acting as a data filter, manages the storage cost better. Finally, a data lake is based on technologies that allow the storage of raw data and then apply incrementally the structure, as defined by the analytical requirements.

So far, we have detailed the characteristics of each of the data platforms which are important for the current situation of our platform, Moodle. The storage of the LMS of the university that has been used so far, since 2010, is 5 TB. The concept of big data is not considered as a fixed parameter from which it can be adopted as a big data platform. However, we will adopt as a reference a starting point of 30 TB [14]. If we consider the level of processing, speed and analysis, it is an advantage to use big data. However smart data and the data lake that have been considered in this work can also offer these characteristics without oversizing the resources. Therefore, categorisation with respect to storage is sufficient to rule out the adoption of big data for the needs of the institution.

The need is to convert the data into useful information. Considering this objective, we can better analyse both the adoption of smart data and a data lake. We begin this analysis by indicating that, in the case study of the Moodle platform, data mining techniques have been applied with the help of a data warehouse and the generation of cubes in the past. However, the life cycle of this technique has come to an end. With this consideration, at present the University has an original Moodle repository (MySQL) and a repository with clean data in an SQL database engine. In addition to these data sources should be considered external sources that are spreadsheets, plain text and even independent databases. These databases may contain relevant information from several of the courses and have been handled internally by teachers either as learning activities or records. As an additional point, the big data in our case is oversized by the volume of data available on the LMS platform, but we can benefit from the tools that big data offers for the analysis of the information.

3.3 Processing in Smart Data and a Data Lake

The smart data is not focused on storing or processing information, but on extracting value from it. This is where the human factor, the business knowledge and the expertise are most relevant. This task is achievable by data scientists who know the data they have, what they need to know and how to obtain it [13]. The questions to be answered in the analysis are: what do we do with all this volume?, what is the relevant

information of all that we have collected?, what level of aggregation is needed?. With regard to speed, we must know precisely what actions make sense in real time, which in near-real time and which can be performed every hour, every day, every month or every year. It does not make sense to analyse the information every second if we can only act every twenty-four hours, we should simply store it to analyse after.

In data lake the access to the original information is direct and reduces the intermediate steps for its processing [12]. Sometimes, when a record is deleted, it may not be needed immediately, but after a while. Data that may not be useful today may be needed after a few months or even years. A data lake marks the differences as a non-pre-processed data storage system. However, it is necessary to consider a higher cost, both of technical means and of professional profiles that are able to manage it.

Table 4 describes the parameters evaluated in the smart data and data lake platforms. The storage parameter refers to the capacity of the platform to provide the user with the storage of the data. The smart data does not act as a storage platform because it is focused on extracting useful information. Their accuracy depends on the analysis of the data scientists in determining the exact times for the execution of processes. The data lake, on its own, acts as a data repository since its processing uses raw data, its operation is based on the conservation of the data. For cost parameters at the technical level, it is considered that the smart data does not need a great infrastructure since it uses the operational data sources to extract the information; by contrast, the data lake, requires greater infrastructure for data storage as well as systems for processing. In human cost, the two platforms require highly trained analysts and data scientists with extensive knowledge of the business and the data generated.

Table 4. Technical analysis of smart data and the data lake

Platform	Storage	Prosecution	Technical costs	Human cost
Smart data	Low	High availability	Medium	High
Data lake	High	High availability	High	High

4 Analysis of Results

The analysis made in Sect. 3 gives us a broader picture of the tool that we can use considering the needs that are presented by the educational institution used as a case study. It is worth mentioning that any change considered for improvement in education should be attached to the economic reality of the institution. With this clarification, the tool that is sought must meet the technical and quality requirements, as well as the technical and human costs.

In the first analysis carried out in Sect. 3.2, big data has been discarded: although its functionality is broad and applicable to each organization often it is not an optimal solution, since it is possible to oversize the utility of the organizations technical resources. Excessive resources allocated to the system will affect implementation costs; however, we can make use of the analysis tools for our process.

In reviewing the needs as explained in Sect. 3.1, we note that the LMS used manages a data warehouse. This data store can be reused as an external source and be

managed by both a smart data as a data lake. The results indicate that the data lake provides advantages to the educational environment if it meets several characteristics, such as high availability in storage resources. In return, it offers us the conservation of data which, if not needed at this moment, but it may be important in the future. With this option, the analysis of several sensors or systems that help the discovery of trends or patterns of the students can be integrated. For example, whether it is necessary to increase the consumption of coffee in the period of examinations, or how many times a student enters the university, the data collected from his access card. The qualities offered by a data lake are interesting and give long-term advantages. However, at the moment our study only covers the LMS platform, so our application focuses on the use of smart data by reducing costs.

The use of smart data focuses on how we can integrate our data warehouse into its processing. The ideal for this tool is that it can make use of the cubes that are available in the current system without the need to process the information. It will simply extract the value of it into the process. Keeping data that has gone through a previous processing ensures the accuracy of data in the same way as it reduces processing. Another feature of smart data is the configuration of processes at specific times. For example, every 24 h, once a month or every 4 months, depending on the need of the organization.

The benefits offered by data mining for the analysis of data will be used, because this converges, without any problem, with the data and information generated by smart data. In this work, we do not perform an analysis of the algorithms to be used. However, from experience of the authors in previous works [15, 16], it can be mentioned that both the search algorithm and the cluster have sufficient characteristics to solve our needs.

On average one student generates 90 Mb per year, this volume of information is only from the use of the Moodle platform. The amount of information is very low in consideration of common enterprises, but this figure will increase if more sensors and systems are integrated into our analysis platform. The description of the problem mentions the need for integration of the systems that the university manages in order to carry out an in-depth analysis of the students. These systems manage the student's attendance, financial situation, qualifications, even there are printing systems that will indicate the trend in each student's reading.

5 Conclusions

This work includes new concepts that can be considered as a component that helps to improve education through the use of information and communication technologies. What has been sought during this development is to qualify the various platforms based on the needs of a particular educational institution considering, as the main base, the multiple sources of data.

Most organizations currently seek to take advantage of the information that is generated daily in the interaction with customers, defining what their interests are and being able to generate more profits based on these statistics. The same concept can be replicated in the educational field and, in this way, a personalised education can be offered based on the characteristics or patterns presented by each individual student.

The use of data mining on educational platforms every day has greater depth. However, it is important that the evaluation environment goes beyond an LMS. It is important that all the sensors or systems surrounding the student converge into one, so that the trends, problems or help that each student requires may be detected in a timely manner. This processing capacity exceeds the typical data warehouse so we have considered it important to scale to other types of tools.

Using a smart data will have the ability to analyse these systems in depth and establish patterns that tell us how the learning outcomes of a specific student can be improved. At the moment, a test on the operation of a smart data has been carried out using Microsoft power BI tool. The results will be presented in a future work since the data obtained are in validation stage. The power BI tool allows an ad hoc analysis and until now, four systems that control the student's activity when he or she is at university have been integrated into the test.

References

1. Dalsgaard, Ch.: Social software: e-learning beyond learning management systems. *Eur. J. Open Distance E-Learn.* **9**(2), 1–7 (2006)
2. Davenport, T.H., Barth, P., Bean, R.: How big data is different. *MIT Sloan Manage. Rev.* **54**(1), 4346 (2012). <https://search.proquest.com/docview/1124397830?accountid=33194>
3. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* **17**(3), 37 (1996)
4. Higdon, S.J., Devost, D., Higdon, J., Brandl, B., Houck, J., Hall, P., Green, J.: The SMART data analysis package for the infrared spectrograph* on the spitzer space telescope. *Publ. Astron. Soc. Pac.* **116**(824), 975 (2004)
5. Lavallo, S., Lesser, E., Shocley, R.: Big data, analytics and the path from insights to value. *MIT Sloan Manage. Rev.* **52**(2), 21 (2011)
6. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.: Big data: the next frontier for innovation, competition, and productivity, pp. 27–36 (2011). <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
7. Dougiamas, M., Taylor, P.: Moodle: using learning communities to create an open source course management system. In: *Proceedings of ED-MEDIA World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp. 171–178. Association for the Advancement of Computing in Education, Honolulu (2003)
8. O'leary, D.: Embedding AI and crowdsourcing in the big data lake. *IEEE Intell. Syst.* **29**(5), 70–73 (2014)
9. Sagiroglu, S., Sinanc, D.: Big data: a review. In: *International Conference on Collaboration Technologies and Systems (CTS)*, pp. 42–47 (2013)
10. Snijders, C., Matzat, U., Reips, U.: Big Data: big gaps of knowledge in the field of internet science. *Int. J. Internet Sci.* **7**(1), 1–5 (2012)
11. Terrizzano, I.G., Schwarz, P.M., Roth, M., Colino, J.E.: Data wrangling: the challenging Journey from the wild to the lake. In: *Conference on Innovative Data Systems Research (CIDR)*, pp. 1–9 (2015)

12. Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sen Sarma, J., Liu, H.: Data warehousing and analytics infrastructure at Facebook. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pp. 1013–1020. ACM (2010)
13. Trautsch, F., Herbold, S., Makedonski, P., Grabowski, J.: Addressing problems with external validity of repository mining studies through a smart data platform. In: Proceedings of the 13th International Conference on Mining Software Repositories MSR, pp. 97–108. ACM (2016)
14. Villars, R.L., Carl, W., Matthew, E.: Big data: what it is and why you should care. White Paper IDC **14**, 1–14 (2011)
15. Villegas-Ch, W., Luján-Mora, S.: Systematic review of evidence on data mining applied to LMS platforms for improving e-learning. In: International Technology, Education and Development Conference (INTED), pp. 6537–6545 (2017)
16. Villegas-Ch, W., Luján-Mora, S.: Analysis of data mining techniques applied to LMS for personalized education. In: World Engineering Education Conference (EDUNINE), pp. 85–89. IEEE (2017)
17. Walker, J.S.: Big data: a revolution that will transform how we live, work, and think. *Int. J. Advertising* **33**(1), 181–183 (2014)
18. Widom, J.: Research problems in data warehousing. In: Proceedings of the Fourth International Conference on Information and Knowledge Management (CIKM), pp. 25–30 (1995)

5 Comprehensive Learning System Based on the Analysis of Data and the Recommendation of Activities in a Distance Education Environment

Villegas-Ch, W., Palacios-Pacheco, X., Buenaño-Fernández, D., y Luján-Mora, S. (2019). Comprehensive Learning System Based on the Analysis of Data and the Recommendation of Activities in a Distance Education Environment. *International Journal of Engineering Education*, 35(5), 1316-1325. (Villegas-Ch, Palacios-Pacheco, Buenaño-Fernandez, y Luján-Mora, 2019)

Disponible en:

URL: https://www.ijee.ie/latestissues/Vol35-5/06_ijee3795.pdf

Temas a los que contribuye:

- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment

Villegas-Ch, William¹; Palacios-Pacheco, Xavier²; Buenaño-Fernández, Diego¹; Luján-Mora, Sergio³

¹ *Universidad de Las Américas, william.villegas@udla.edu.ec; diego.buenano@udla.edu.ec*

² *Universidad Internacional del Ecuador, xpalacio@uide.edu.ec*

³ *Universidad de Alicante, sergio.lujan@ua.es*

Abstract

Traditional teaching, based on techniques in which students develop a passive function, has proven to be an inefficient method in the engineering learning process. Universities have been forced to improve their teaching methods and have found a partial solution in open source platforms; these platforms have allowed a greater collaboration between institutions that improve the contribution of technology to education. There are cases of collaboration between universities where their sole objective is to promote student learning and the automation of educational processes. The massification of this type of technological tools allows the use of systems and platforms commonly used in the business world. This adoption of open source tools has proven to be very effective in educational environments and has offered several benefits such as the reduction of costs and the constant updating of information systems. One of the frequent cases in which there are collaborative projects based on learning is the analysis of educational data that seek to detect students' deficiencies and to take actions before they abandon their studies. In this work, we propose the design of an integral learning system in which business intelligence, expert systems, learning management systems and different learning techniques converge. This integration seeks to create a system capable of recommending different activities that focus on the needs of students.

Keywords: Open source; expert system; engineering education; project based learning; data mining

1 Introduction

A model of distance education is part of current educational trends since it has advantages over a traditional modality where the student fulfills a classroom schedule that is already defined [1]. In the face-to-face model, traditional lectures encourage students to become passive individuals, with low levels of commitment, concentration, participation and motivation towards the subject [2]. Distance education allows the educational act through different methods, techniques, strategies and means in a situation in which students and teachers are physically separated or interact occasionally [3]. These conditions, if not accompanied by a plan to follow up and provide the student with resources that meet their needs can become a problem rather than a solution. In a model of distance education, integrating concepts such as personalized education, active learning and the recommendation of activities will contribute effectively to learning. This process gives the student time flexibility that facilitates the organization of personal time respecting family life and work obligations, in addition to allowing a personal pace of study.

In distance education, the student becomes the protagonist of their learning and it is responsibility of universities to turn this experience into a personalized education. Implementing these advantages presents technical and administrative difficulties, since talking about personalized education in an environment where there is no feedback from the student affects the measurement of learning. The solution is the creation of integrated learning systems that rely on information and communication technologies (ICT), and where the follow-up is continuous based on the information that the student generates. With the knowledge extracted about the students, the system will be able to recommend activities that suit each one of their needs. To guarantee and prioritize learning, the recommendation of activities must be within an active learning model. Active learning consists of the use of a set of more effective and interesting experimental methods, whereby students assume greater responsibility for their own education [4]. Active learning has many benefits for students, including a deeper understanding of the concepts of a certain subject and promoting the student's positive attitude towards learning and, consequently, a greater motivation towards the subject.

In this sense, there have been many proposals as to how to plan educational activities, the strategies that can be applied and the models that can be adopted. However, this process continues to be a difficult path, given the number of variables to consider when appropriating some of these methodological proposals. At present we can see the contributions of the open source tools to the enrichment of educational work [5]. Most research and proposals, especially in the field of artificial intelligence (AI) and education, are usually dedicated

5 Comprehensive learning system based on the analysis of data

to the learning process and to the academic administration related to the management of information about students [6]. The AI provides tools and techniques that allow a knowledge-based system to face problems associated with decision making. In this work, the development of an expert system (ES) is sought; the ES, after evaluating different criteria or variables, will propose applicable activities to the student in a learning situation [7]. A comprehensive system, with systems based on knowledge or ES, comes to constitute permanent support for the student. They represent a response that is oriented to the efficient decision making on the teaching models and the didactic activities to be implemented in a learning management system (LMS). This condition makes it necessary to strengthen and potentiate the autonomy of the LMS with the aim that it works in its entirety to the ES and can recommend the exact activities to groups of students already defined.

Regarding the ideas revealed, ES in the educational field can be considered to have certain advantages, particularly those created for pedagogical and instructional purposes [8]. An ES can diagnose, debug and correct the development of student learning in a particular area of knowledge. In addition, the system determines the cognitive level of the student and helps him/her improve their weaknesses to reach a higher level of learning. The work is distributed in the following sections; Section 2 contains the theoretical foundation that contributes to the design and implementation of the system. Section 3 contains the method where the whole development is explained step by step. Section 4 presents the discussion based on the results obtained.

2 Method

The process through which the integral system proposed for the recommendation of activities guarantees learning can be observed in Fig. 1. In the first phase, students are identified and classified by means of patterns that they share. The identification is carried out by a business intelligence (BI) subsystem that has the necessary capacity to perform the analysis of student data. Prior to this process, there is a whole set of tools that extract information from different data sources and turn it into knowledge. In the second phase, the resulting data are presented to the ES which, like a human expert, will identify each case and cross the information obtained from the BI with the information obtained from the interaction with the student. The analysis carried out by the ES based on its knowledge allows us to establish which activities are aligned to the needs of the students. In order to provide the system with greater possibilities and to obtain greater results with respect to learning, the activities it recommends are part of an active learning method. The ES is integrated into an additional module of the LMS, which also provides a system for continuous monitoring of learning. In the following sections, each of the stages of the integrated learning system is described in detail.

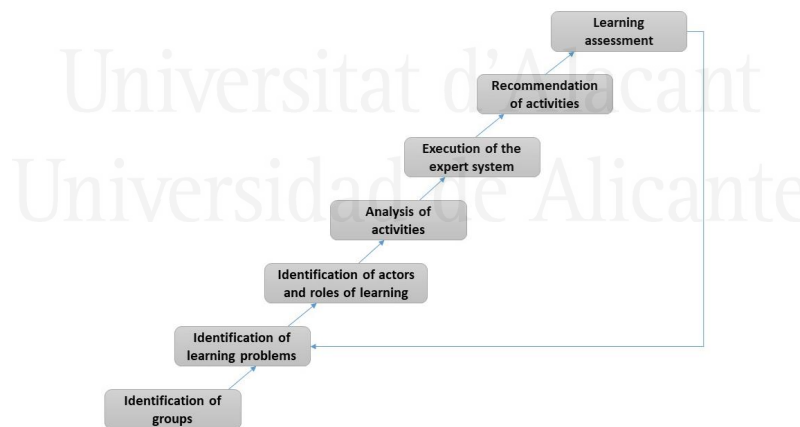


Fig. 1. Stages of the method of an activity recommendation system through an ES.

2.1 Identification of groups

In the identification of student groups, the BI module is used. This process applies data mining algorithms. The process is defined in context; however, the execution is very complex since a whole BI platform is used that considers all the stages for the extraction of knowledge of databases (KDD) [9]. The BI module keeps track of the students based on the data stored in the different repositories. The groups considered for the development of the system are the students that have a high percentage of repetition and those that have been

the cause of a study of university desertion. The selection of the data is based on the study variables; these variables are chosen according to the Bean model [10]. The model identifies the academic, psychosocial and environmental variables that cause the dropout syndrome. In the academic variables, the student's performance and integration are considered; the common repositories that store this information are the systems of academic registration and academic monitoring. Psychosocial variables include objectives, alignment, peer interaction and interaction with teachers. This work, being focused on distance learning models, extracts information from the LMS; for this reason, the interaction with peers and the interactions of teachers in the LMS are modified in the Bean model. Environmental variables are parameterized according to funding, external social relations, transfer opportunities and interaction with the LMS and the use of ICT [11].

To facilitate the process of data selection, several questions are asked to help identify the fields that will be part of the variables, for example:

- Is there information available on the systems that can be used?
- Does this information help the analysis?
- Of all the available information, which one interests us?
- Are the details of all the available information interesting or just the detail of the information we need?

The data used belong to the university that participated in this work, specifically of a distance study modality. The careers included are engineering in information technology, engineering in communications networks and mechatronics engineering. The choice was mainly due to the availability of information provided by the university. Access was obtained from the data on the socioeconomic characteristics of the students that are collected in the student registration systems [12]. These data are important to carry out the academic follow-up of the students and detect their condition of defector, graduate or student active. Students who have similar learning patterns and who belong to the 2013 cohort are included in the analysis; a cohort being defined as the group of students who enroll in a university career in a given year. The advantage, unlike working with the entire population, is to identify a specific group of students who are in the same initial condition and face the same academic and social aspects. According to the data collected from the different systems, the total number of people registered in the distance education modality was 3,207 students across all the cohorts up until the year 2017. The sample considered for this study was the 2013 cohort composed of 208 students. The analysis process identified two groups of students; the sample responds to the following characteristics:

- Group 1. Students in a range of 35 - 44 years of age that have a grade lower than the average of the course considered in 6/10.
- Group 2. The low percentage of hours of significant interactivity.

2.2 Identification of learning problems

Table 1 identifies eight common problems within a distance education model that each group of students faces [13]. The data were obtained through the use of surveys carried out with the population included in the research.

Table 1. Identification of relevant problems in a distance education modality

Problems identified in a distance education modality	Group 1. Age between 35-44 years	Group 2. Low percentage of hours of interactivity
Absence of significant interactivity	X	XXX
Use of materials not designed for online education	XX	X
The lack of evaluations and exercises	X	XXX
The use of irrelevant content	X	XXX
Complex navigation	XXX	X
Instructions for insufficient or unclear use	XXX	X

Criterion of maximum impact XXX

5 Comprehensive learning system based on the analysis of data

Criterion of medium impact XX
Criterion of minimum impact X

Overcoming the problems found in each of the groups is a topic of analysis on which several areas of the university have worked. When activities are suggested, these are adjusted to the needs of each student, considering the time they spend on academic activities, what they want and should learn, and the process of evaluating learning. The evaluation of the learning process is given by the results, processes and conditions that can be represented by questions about any learning situation [14]. For example, what do we want to learn? How do we learn that? What conditions favor this learning?

2.3 Identification of the actors and their roles in learning

It is important to clearly establish the actors and their roles in education: the student, who is the owner of their learning, and the teacher, who guides the student in how to get to the learning. Considering that the student owns their learning, they fall back on what they want; for example, there are students who simply set their goal to pass a certain course and others have a real desire to generate knowledge [15]. Here the teacher enters into action because the task falls on them to propose strategies that help to define what they want the student to learn. Once have defined what to learn, they have to look for techniques that enable doing it and that are within the learning processes. Finally, we must meet the conditions for the student to learn. In any learning scenario, teachers can vary conditions to meet the learning objective. Modifying the conditions is to modify the tactics and the scenarios of learning and teaching. Next, the roles in learning are presented according to their participation guidelines.

Participation of students:

- They move from a passive listening role to active involvement in learning (readings, discussions, reflections, etc.).
- They are involved in higher-order thought processes such as analysis, synthesis and evaluation.
- They learn through dialogue and through interaction with the content and development of competencies.
- Students receive immediate feedback from the teacher and their classmates.

Participation of teachers:

- They design activities according to their discipline and at the current curricular moment needs of their students.
- They adapt the learning activity to the possibilities and needs of the group.
- They facilitate the process of the activity taking care of the extension and the depth of knowledge that is approached.
- Feedback in a timely manner on the performance of the group and individual students.

2.4 Analysis of activities

Established the groups and the participation of each actor in the learning is analyzed the activities that contribute to the learning considering that they must be designed to be presented in an LMS. Under these considerations, the following activities are proposed that can be classified as projects based on learning [16]:

- The "One Minute Paper": This is a highly effective technique to monitor the progress of the student, both in their understanding of the subject and in reacting to the course material. It can be generated by reading a resource within the Moodle activities.
- Reading Contests are a way to motivate students to read the assigned material. Active learning depends on students reviewing the resources available in each module of the LMS. A reading test can also be used as an effective measure of student comprehension of the reading material so that their level of reading sophistication can be measured.
- Puzzles/Paradoxes are one of the most useful means of eliciting insights from students about a topic. It presents a paradox or a puzzle related to the concept in question and guides them towards a solution.

By forcing students to work without authority, it increases the likelihood that they will be able to critically evaluate the theories that are presented.

- Discussion students are asked to pair up and answer a question. This can be easily combined with other techniques such as questions and answers.
- Conceptual maps are a way to illustrate the connections that exist between the terms or concepts covered in the course material. Students construct conceptual maps by connecting individual terms by lines that indicate the relationship between each set of terms connected.

When moving from a normal learning scenario to active learning, one must consider that this focuses on the student by promoting their participation and continuous reflection through activities that promote dialogue, collaboration, development and knowledge construction, as well as skills and attitudes. The activities must be motivating and challenging, and capable of promoting active adaptation to solving problems.

2.5 Execution of the expert system

The ES is in charge of interacting directly with the student; the communication protocol is through questions that will allow solving problems much faster than a human expert. The ES is based on a deterministic methodology that aligns with the exposed needs as well as having a large number of success stories [17]. For the development of ES, the stages proposed by Weiss and Kulikowski [18] are considered; each stage must be properly analyzed and developed since this directly influences the quality resulting from the ES.

a) System tools

For the design of the expert system and the necessary components for the recommendation of learning activities, open source tools are used. These tools are used based on costs, availability of information and the learning curve that is very low. For the design of the expert system, the Scrum methodology is used, which allows continuous adaptation to the different circumstances in the development of the system. The development language used is prolog that allows a logical programming and becomes the processing core. Prolog is based on two elements that must necessarily be identified to meet the expectations of the research. These elements are the atoms that define in a generic way an object of the environment that we want to represent and the predicates that are responsible for specifying the characteristics of the objects in the environment and the relationship between them [19]. For example, the logical representation for the case of students with learning problems. (Student_learning_problems (X) has (X, bad grades): Student_learning_problems (X)). In this rule generalizes the fact that any object that is a student with learning problems, will have bad grades. In development, it must be borne in mind that the fact that an object has poor grades is not a sufficient condition for it to be a student with learning problems.

The development of the interface that interacts with the user is developed in PHP, which is an open source and is widely used in web development environments. It is important to consider that the whole system is in a testing process for which it is accessed via intranet by a sample of the student population [20]. PHP provides the best solution for the environment mentioned by its ease of use, as well as its orientation to the development of dynamic web applications. These tools are easily integrated into the LMS and the BI platform, since the LMS is Moodle with a database in MySQL and for its part the BI platform is developed in Pentaho in its community version [21]. These characteristics make the whole system is based on open source tools acquiring the advantages already mentioned.

b) Problem statement

The objective of identifying the problem for the execution of the ES is to extract general knowledge about the university, its components, characteristics, structures and processes. These parameters contribute to the design of the ES, as well as a global vision of the current state of the university with respect to its environment, its shortcomings, its resources and relationships. Within this context, a university has education as its main activity; this concept integrates a number of variables to consider, as you cannot establish a general method and expect similar results in all students. The variation of learning in students depends on economic as well as psychosocial factors and determining a strategic plan that allows for taking action measures requires much technical and administrative effort. The creation of an ES that is in charge of the analysis of these factors and recommends activities that adjust to each student proposes to solve this problem.

5 Comprehensive learning system based on the analysis of data

Specifically, the problems that are intensive in knowledge are characterized because they are based on facts and rules of a domain, as well as on the experience of the people and organizations that make it up. The sets of problems that accompany universities according to knowledge engineering are:

- Learning problems (Very High).
- Lack of feedback from the student (High).
- Acceptance of the educational model is low (Medium).
- There is no desire to learn in the student (Low), etc.

c) Find human experts

In this phase, several expert teachers work in different areas of engineering. The objective is to establish the questions that the system must ask the students and that they serve as a guide to obtain an accurate conclusion. The design of activities is another task that this group of teachers must perform focusing on active learning. Another working group is made up of experts in educational psychology who seek methods through which the student is interested in generating knowledge and not in obtaining a grade. A third group is made up of experts in the use of open source LMS, specifically Moodle. The objective of this group of experts is to create the activities proposed by the experts in the subject areas and translate them into the platform considering the methodologies proposed by the second group of experts.

d) Expert system design

The design must be flexible and gather characteristics similar to human behavior. This includes the ability to acquire knowledge, the reliability that allows trusting the results, the knowledge domain that provides solidity in a process and the ability to solve problems. The design is done based on a Bayesian theorem, which is an algorithm that is based on probabilistic theory and combines the Bayes theorem with the expressivity of the directed graphs [22]. The directed graphs allow the representation of a causal model by means of the graphic representation of the dependencies between the random variables and the causal influences [23]. The design is constituted by the Bayes theorem (1), whereby given two variables “x” and “y”, such that $P(x) > 0$ for all x and $P(y) > 0$ for all Y is fulfilled:

Equation 1. Bayes theorem

$$P(x|y) = \frac{P(x) \cdot P(y|x)}{\sum_{x^1} P(x^1) \cdot P(y|x^1)}$$

For example, several tests similar to the following process have been carried out for development. For a course with 50 students (23, 18, 9), divided into groups defined by their performance (high, medium, low), it is suggested that they perform some action to improve their learning.

- 30% Games, 40% Quick Tests, 25% Forums and 5% Puzzle.
- 25% Games, 35% Quick Tests, 30% Forums and 10% Puzzle.
- 20% Games, 50% Quick Tests, 10% Forums and 20% Puzzle.

The environment analyzed is as follows: a) is the recommended activity the third?; b) if we find out what the forums are, what would it be?

Solution:

- A priori solution: $P(x) = 9/50 = 0.18 = 18\%$
- To answer this request, we describe the data in Table 2 where $x = \text{activities} \mid y = \text{groups of students}$.

Table 2. Identification of activities and groups of students through Bayes

$P(x y)$	x1	x2	x3
Yc	0.30	0.25	0.20
Yi	0.40	0.35	0.50
Yh	0.25	0.30	0.10
Ye	0.05	0.10	0.20

In the equation (2) the data of the table is processed and the percentage with which the results are processed in the ES is obtained.

Equation 2. Identification of activities and groups through Bayes theorem

$$P * (x^3) = P(x^3|y^e) = \frac{P(x^3) \cdot P(y^e|x^3)}{\sum_x P(x) \cdot P(y^e|x)} = \frac{0.18 \cdot 0.20}{0.46 \cdot 0.05 + 0.36 \cdot 0.10 + 0.18 \cdot 0.20} = 0.379 = 37.9\%$$

The design of the ES, which is based on Bayesian networks, constructs a decision tree that allows combining the judgment of the experts with the available data and making the inference between any set of variables [24]. The architecture of an ES is organized according to three main elements, which are the basis of knowledge, the inference engine and the basis of facts; however, to ensure the dialogue between the machine and the human, additional components are required [25]. The components that act additionally are the human component and the knowledge acquisition systems, etc.

e) Acquisition of knowledge

In this phase, the basic knowledge in the subject is provided and the knowledge engineers transfer this knowledge to a language that the ES understands. This stage requires great dedication and maximum effort due to the different levels of knowledge of those involved and the different experiences they have [26]. For the development of the ES, the technical knowledge of experts in the area of computer tools, networks, computer audit and robotics, among others, is used. These areas of expertise are considered when contributing to engineering careers related to ICT. The experts in teaching crystallize the activities given by the experts with a psychological tinge that contributes to the active learning of the students. Once the experts are selected, and they agree to give their knowledge, they start playing the role of the "Knowledge Engineer". As such, they are in charge of extracting the knowledge from the expert and giving it an appropriate representation, either in the form of rules or another type of representation, thus forming the knowledge base of the expert system.

f) Knowledge base

The knowledge base is the phase where specialists are responsible for providing knowledge engineers with an orderly and structured base and a set of well-defined and explained relationships. This structured way of thinking requires that human experts rethink, reorganize and restructure the knowledge base and, as a result, the specialist becomes a better connoisseur of their own field of expertise [27]. However, we must differentiate between data and knowledge; knowledge refers to statements of general validity such as rules, probability distributions, etc. This is stored in the knowledge base and the data is stored in the working memory. There are three ways of representing knowledge: production rules, semantic rules and frames.

The most commonly used form of representation is production rules, also called rules of inferences [28]. Most ESs are based on this type of representation. For this reason, the design for this work considers this option for success cases in other areas, as well as the amount of existing information that contributes to the development.

The rules of production are of the type:

- IF Premise THEN Conclusion (YES A THEN B)

Where both the premise and the conclusion are no more than a chain of events connected by "Y" or "O", in general it would be:

- IF Done1 AND/OR Done2 AND/OR ... DoneN THEN Done1 AND/OR ... DoneN

The facts are statements that serve to represent concepts, data, objects, etc. The set of facts that describe the problem is the basis of facts.

g) The inference engine

The inference engine is the supervisor; it is responsible for extracting the conclusions based on the symbolic data by applying the rules that govern the system in which it works; thus, a change in the rules will result in different conclusions, which is why it uses data that are facts or evidence, and knowledge that is the set of rules stored in the knowledge base to obtain new conclusions or facts [29].

5 Comprehensive learning system based on the analysis of data

The function of the inference engine is to execute actions to solve the problem from an initial set of facts and, eventually, through an interaction with the user. The execution can lead to the deduction of new facts. This process is done through rules that model the general knowledge of the domain and constitute the knowledge base, long-term memory and implications. For example, if x is 45 years old and studies, “then” x is a university student. The ES will process all the information through the aforementioned stages and reach a conclusion. The percentage of the uncertainty with which it reaches the conclusion will depend on the allocation of activities. Table 3 details the percentages of uncertainties with which the system will work.

Table 3. Percentages of ES uncertainty

% Acceptance	
%	Conclusion
0 – 50	Not feasible
50 – 70	Feasible with uncertainty
70 – 90	Feasible
90 – 100	Sure

The questions that the ES gives to the students are designed in natural language and look for the interaction in a simple and direct way. Below are several questions used to interact with the student; these seek to know if the student is giving a true answer [30]:

- Hello “X”, how old are you?
- Is the time you dedicate to the course sufficient?
- Are the contents of the course easy to understand?
- Do practical activities strengthen your learning?
- Do you want to develop a case applied to business reality?

For example, asking if the student is on a computer tools course allows the ES to validate the information it receives from the student. The validation of information allows it to reach conclusions more accurately. This process is possible by integrating the ES with systems that manage student information, thus allowing advance knowledge of several student data. The ES takes this information and contrasts it with the information received in order to verify that it is the person who claims to be.

The answers in most cases are given by the following four options: Yes; It seems to me that yes; I don't know; I don't think so; No

2.6 Recommendation of activities

The process of recommending activities is based on the three following stages: a) the profile of the user is loaded directly from the data that the ES finds in the LMS, however, the student's name and age are asked as a question of courtesy; b) it extracts information from the available activities and adjusts this to the needs of each student; c) it initiates the recommender with the ES to compare the user's information with the information stored in the knowledge base and recommends an activity. Fig. 2 presents the entire process of the recommender system [31].

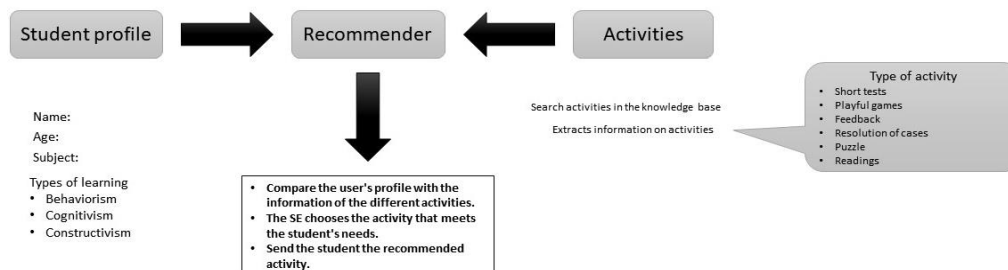


Fig. 2. Process of the recommender system.

For a better understanding of the process carried out by the ES to recommend one or more activities, Table 5 is presented. Here, the weights are established through which the system can perform the calculations to reach a certain conclusion. This process is a basic example of the internal functioning of the ES, whereby five unique values of probability are handled; these are: 1 = yes, 0.75 = I think so, 0.50 = I do not know, 0.25 = I do not think so, and 0 = no.

If universities want to identify recommended student tasks, but they have little available time to develop activities and for students to learn to solve problems of a specific topic, the system eliminates activities that do not meet these criteria and recommends activities with greater weight. For students who meet the indicated conditions after carrying out the process and calculation, the system will recommend play activities. With a lower weight, the system recommends roundtable activities.

Table 5. Activity recommendation process

Has	Reading control	Rapid test	Playful games	Development of cases	Puzzle	Discussion through forums	Conceptual maps	Roundtables
Little time available to develop activities	0	1	0.75	0	0	0.5	0	0.75
Capture the idea with ease	1	0	1	0.75	0.5	1	1	0
Generates knowledge when solving problems	0	0	0.75	1	1	0	0.75	0.25
Discussion on specific topics	0	0	0	0.5	0	1	0.75	1
Constant review	1	0	0.75	0	0	0.25	0	0
I understand what I do	0	0.25	0.75	1	0.75	0	1	0
I master what I teach to another	0	0	0.25	0.75	0	1	0.5	1
I understand what I talk about with another	0	0	0	1	0	1	0.75	1
Recognizes different appearances and circumstances	0.75	0.25	0.75	1	1	0	0.75	0.5

The functional model is more sophisticated as it includes rules that help to deduce answers from the previous ones, allowing solutions to be found more quickly and efficiently. The system asks very generic questions that allow a significant reduction in the recommendation of activities.

3 Results

Through the use of the BI platform, a group of 10 students of the engineering degree in information technologies has been selected. The filters of group is the low percentage of hours of interactivity in the LMS platform and the low qualifications of the partial I. To this group the ES recommended the activities of case and puzzle developments with the purpose that they generate knowledge by solving problems. The activities have been planned within the LMS with the objective that encourage the use the platform and that the interaction time is effective generating the desired knowledge in the student. Table 6 presents the interaction data of the

5 Comprehensive learning system based on the analysis of data

students and the grades obtained. Each partial consists of eight weeks therefore they contain information on the performance of students before applying the recommendation system. This information is compared to the second eight weeks that are part of the same academic period. The second part contains the academic data of the system with the functioning of the recommendation system.

Table 6. Partial I without ES of activities vs Partial II with ES of activities

Student	Partial I without ES of activities		Partial II with ES of activities	
	Rating / 10	Interaction time hrs/month	Rating / 10	Interaction time hrs/month
ID2431	3.4	2.36	7.6	5.01
ID5246	1.1	0.35	4.3	4.37
ID4876	2.7	4.21	5	6.35
ID7548	3.2	1.52	4.0	2
ID6584	2.9	2.65	7.1	6.35
ID8954	3.3	3.12	6.5	7.32
ID5842	2.2	0.59	8.6	12.25
ID6547	3.1	3.24	1	0.25
ID9854	4	2.54	6.1	4.54
ID8774	5	1.51	2	2.35

The BI platform identified students with low interaction with the platform and low scores in the first part. The activities developed in this time are shared by the teacher without there being any difference between students. With this knowledge, it is contrasted with the data of the second part where, the recommender system has followed the students assigning tasks that motivate the use of the platform. These activities are, the resolution of practical cases on the corresponding topics in each week, as well as the resolution of puzzles on these subjects. The differences are visible since 80% of the sample has improved the times of interaction with the platform and the ratings. The data presented is general, since the system is barely functioning and it is expected that in the period 2020-20 the first cut will be made, which will allow a deeper analysis regarding the efficiency of the activities in each case.

4 Discussion

The problem that is presented to the designers of educational environments is the number of variables that interact with each other and which must be taken into account for an effective design that includes an integral learning system. The strategies of instruction embedded in a highly effective design consist of the adequate interaction between the elements that intervene in the instruction scenario in order to optimal achievement of educational objectives. The proposed system assists designers and tutors in identifying the best activities that are determined by the scenario and defined by the characteristics of the learner and the learning context. The ES to identify the best instructional strategies receives the following scenario data as input: the characteristics of the domain of knowledge to be transmitted, the profile of the student who will act as a user, a description of the technological environment that will act as a facilitator of learning and the description of skills that the individual is expected to achieve as a result of the instruction. Once all these data are known, the ES identifies the pattern or combination of patterns that best suits the situation presented by the user, thus recommending the activities to be used by the designer.

5 Conclusions

The integration of an ES with an LMS that interacts directly with the student guarantees the constant accompaniment of the expert with the student during the learning process. The result of this integration brings more benefits and not only to the student but also to the teacher. With this application, the teacher better uses his time devoted to the construction of knowledge through the design of new activities. For the teacher, the ES becomes a teaching assistant that facilitates learning and assesses the learning of each activity proposed by the teacher. This contribution optimizes the time of dedication of the teacher in the design of interactive resources that accompanied by the recommendation of activities by the ES greatly improves the learning of a student. An important point that allows measuring the effectiveness of the ES are the percentages of desertion that in the short time of execution of the system has been able to reduce.

Technically the solution must go through a time of maturity during which the results in the students must be analyzed. However, the process so far is satisfactory considering that its design is scalable and that its construction is based on the use of an open source language. These features will allow its evolution at a low cost, and it can also include the community that uses this type of language for its continuous improvement.

The next step to follow in this work is to increase the number of variables that the ES manages to improve their learning about how students learn in the proposed ecosystem. The obligation of those in charge of administering the ES is to evaluate the effectiveness of the proposed activities by means of data analysis systems. The validity of the results obtained highlights the role of experts who contributed with their knowledge to deduce the relevant variables of each subject and its importance in the learning. The result of the use of the ES allows recommending a variety of activities that contribute to learning using a multi-criteria concept.

References

- [1] R. M. Bernard, P. C. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, et al. How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of educational research*, **74**(3), pp. 379-439, 2004.
- [2] J. H. Mcmillan, and S. Schumacher. Research in Education: Evidence-Based Inquiry, MyEducationLab Series, Pearson, 7, pp. 528, 2010.
- [3] M. G. Moore and G. Kearsley. *Distance education: A systems view of online learning*, Cengage Learning, pp. 45-293, 2011.
- [4] M. Simonson, S. Smaldino and S. M. Zvacek. *Teaching and learning at a distance: Foundations of distance education*. IAP, pp. 162-323, 2014.
- [5] D. E. Drummond. Open sourcing education for Data Engineering and Data Science, in *Education Conference (FIE)*, pp. 1-1, 2016.
- [6] B. Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **6**(1), pp. 1-114, 2012.
- [7] R. O. Tandogan, and A. Orhan. The Effects of Problem-Based Active Learning in Science Education on Students' Academic Achievement, Attitude and Concept Learning, *Online Submission, Eurasia Journal of Mathematics, Science & Technology Education*, **3**(1), pp. 71-81, 2007.
- [8] R. J. Spiro, B. C. Bruce and W. F. Brewer. (eds), *Theoretical Issues in Reading Comprehension*, vol. 11, 1ra edn, London: Routledge, 2017.
- [9] W. Villegas-Ch, S. Luján-Mora and D. Buenaño-Fernández, Data mining toolkit for extraction of knowledge from LMS. In *Proceedings of the international Conference on Education Technology and Computers*. ACM, pp. 31-35, 2017.
- [10] J. P. Bean. The application of a model of turnover in work organizations to the student attrition process. *The review of higher education*, **6**(2), pp. 129-148, 1983.
- [11] S. M. Weiss, and C. A. Kulikowski. *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. Morgan Kaufmann, 1991.
- [12] W. Villegas-Ch and S. Luján.Mora. Analysis of data mining techniques applied to LMS for personalized education. In *World Engineering Education Conference, IEEE*, pp. 85-89, 2017.
- [13] J. R. Savery. Overview of problem-based learning: Definitions and distinctions, *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, **9**, pp. 5-15, 2015.
- [14] R. E. Mayer, *Learning strategies: An overview*. In Learning and study strategies, pp. 11-22, 1988 Academic Press.
- [15] E. Wenger. *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*, Morgan Kaufmann, 2014.
- [16] J. L. Jensen, T. A. Kummer, and P. D.D. M Godoy. Improvements from a flipped classroom may simply be the fruits of active learning. *CBE-Life Sciences Education*, **14**(1), pp. ar5, 2015.
- [17] D. J. Power, R Sharda, and F. Burstein. *Decision support systems*. John Wiley & Sons, Ltd, 2015.
- [18] D. Heckerman and B. N. Nathwani. Toward Normative Expert Systems Part II, *Methods of Information in medicine*, **31**, 2016.
- [19] C. Dunchev, F. Guidi, C. C. Sacerdoti, E. Tassi, *ELPI: Fast, Embeddable, λProlog Interpreter*, in M. Davis, A. Fehner, A. McIver, A. Voronkov (eds), *Logic for Programming, Artificial Intelligence, and Reasoning*, vol 9450, Springer, Berlin, Heidelberg, p. 460-468, 2015.
- [20] B. Wang, P. S. L. Liu and X. F. Wang, Research on the Application of Flipped Classroom Model Based on MOOC in the Course PHP Dynamic Website Development. In *International Symposium on*

- Educational Technology*, pp. 200-203, 2018.
- [21] V. Vargas, A. Syed, A. Mohammad and M. Halgamuge, Pentaho and Jaspersoft: a comparative study of business intelligence open source tools processing big data to evaluate performances. In *International Journal of Advanced Computer Science and Applications*, **10**(14569), pp. 1-10, 2016
- [22] G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, **9**(4), pp. 309-347, 1992.
- [23] D. E. Heckerman, E. J. Horvitz, and B. N. Nathwani. Toward normative expert systems part i. *Methods of information in medicine*, **31**, 2016.
- [24] A. Okutan and O. T. Yıldız. Software defect prediction using Bayesian networks. *Empirical Software Engineering*, **19**(1), pp. 154-181, 2014.
- [25] M. Monsion, B. Bergeon, A. Khaddad and M. Bansard. An expert system for industrial process identification, In *Artificial Intelligence in Real-Time Control*, pp. 85-89, 1988.
- [26] J. H. Boose. A knowledge acquisition program for expert systems based on personal construct psychology, *International Journal of Man-Machine Studies*, **23**(5), pp. 495-525, 1985.
- [27] J. F. Sowa. (ed). Principles of semantic networks: Explorations in the representation of knowledge. *Morgan Kaufmann*, 2014.
- [28] H. Liu, A. Gegov and M. Cocea. Rule-based systems: a granular computing perspective, *Granular Computing*, **1**(4), pp. 259-274, 2016.
- [29] J. Pearl. Probabilistic reasoning in intelligent systems: networks of plausible inference, *Elsevier*, 2014.
- [30] U. Lindqvist and P. A. Porras. Detecting computer and network misuse through the production-based expert system toolset (P-BEST), In *Security and Privacy, Proceedings of the Symposium on. IEEE*, pp. 146-161, 1999.
- [31] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook. Springer, MA*, pp. 191-226, 2015.

Biographies of the authors:

William Villegas-Ch is professor in the area of Electronics and Information Networks of the University of the Americas. He obtained his masters in communications networks and is an engineer in systems with mention in robotics in artificial intelligence. His main research topics include web applications, accessibility and web usability, data mining, e-learning. He has taught courses in "Ofimatics", "Operating Systems", "Network Security", "Network Quality Service" and "E-bussines" at the International University of Ecuador and at the Universidad de Las Américas.

Xavier Ivan Palacios Pacheco is Engineer in Computer Systems and Computer Science, Master in Connectivity and Telecommunications Networks, both degrees obtained at the National Polytechnic School. He has been teaching since 2000 at the International University of Ecuador, in face-to-face and distance learning, in Commercial Engineering, Computer Engineering and Multimedia, Engineering in Risk Management and Emergencies. He has held the position of Director of the Systems Department of the International University of Ecuador since 2000. He is currently developing research on learning analytics and data mining in the academic field.

Diego Buenaño-Fernández is an Assistant Professor of Computer Systems and Computer Science at the University of the Americas (Quito, Ecuador). Engineer in computer systems and computer science (1999) by the National Polytechnic School (Quito, Ecuador). Coordinator of Quality Assurance of the Faculty of Engineering and Agricultural Sciences of the Universidad de Las Américas. PhD student in the PhD program in computer science at the University of Alicante (Spain). His research line is related to Data Mining in educational environments and Learning Analytics. At the level of undergraduate professor of the subjects of "Operating Systems" and "Electronic Business" and at the graduate level, professor of "Workshops titling".

Sergio Luján-Mora received the Ph.D. degree in Computer Engineering from the Department of Software and Computing Systems at the University of Alicante, Alicante, Spain, in 2005. He received the Computer Science and Engineering degree at the University of Alicante in 1998. He is currently a senior lecturer of the Department of Software and Computing Systems at the University of Alicante. His main research interests include web applications, web development, and web accessibility and usability. In recent years, he has focused on e-learning, Massive Open Online Courses (MOOCs), Open Educational Resources (OERs), and the accessibility of video games. He is the author of several books and many published articles in various conferences (ER, UML, DOLAP) and high-impact journals (DKE, JCIS, JDBM, JECR, JIS, JWE, IJEE, UAIS).

6 Management of Educative Data in University Students with the Use of Big Data Techniques

Villegas-Ch, W., Palacios-Pacheco, X., Ortiz-Garcés, I., y Luján-Mora, S. (2019). Management of Educative Data in University Students with the Use of Big Data Techniques. Revista Ibérica de Sistemas y Tecnologías de la Información, nº E19, p. 227-238. (Villegas-Ch, Palacios-Pacheco, Ortiz-Garcés, y Luján-Mora, 2019)

Disponible en:

URL: <http://www.risti.xyz/issues/ristie19.pdf>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Management of educative data in university students with the use of big data techniques

William Villegas-Ch.¹, Xavier Palacios-Pacheco², Iván Ortiz-Garcés³, Sergio Luján-Mora⁴

william.villegas@udla.edu.ec, xpalacio@uide.edu.ec, ivan.ortiz@udla.edu.ec, sergio.lujan@ua.es

^{1,3} Universidad de Las Américas, 170523, Quito, Ecuador.

² Universidad Internacional del Ecuador, 170411, Quito, Ecuador.

⁴ Universidad de Alicante, 03080, Alicante, España.

Pages:227–238

Abstract: The large volumes of data that exist in the universities keep important information of each student. Analyzing this data represents a challenge for data scientists due to the number of resources they consume. Many of the universities do not have the capacity of infrastructure as well as human resources to do it for this reason they desist from the analysis of data depriving themselves of generating knowledge about their students. The range of sensors that generate data in a university is so wide that doing an analysis of data through a traditional method such as business intelligence does not provide accurate results and their response times are not as expected. This work proposes the use of big data techniques in a university to obtain accurate results in real time that will help in making decisions improving education and learning.

Keywords: big data; Hadoop; analysis of data.

1. Introduction

Universities, like companies with their clients, make decisions for their students based on the data they have about them. However, this process increasingly requires methods that allow an analysis of the data superior and that the results appear at appropriate times. This does not mean that the universities have not already been doing work on the data and getting knowledge. The problem lies in the large volume of data generated from student activities that greatly exceed the capabilities of classical analysis platforms such as business intelligence (BI). The variety of sources and that these do not respond to a strictly structured data model leads to the search for other alternatives, as well as concepts in data analysis (Cheng & Cheng, 2011).

One of these alternatives and that is marked, as a trend in the analysis is the use of big data. These platforms offer alternatives for the treatment of data and that obtaining knowledge is more flexible, with lower costs and in shorter times. For example, in a common and very important analysis for universities is to detect and classify students who have learning problems and which leads to high dropout rates. With the use of

business intelligence and data mining it can be done, however, many data that are not structured are left aside (Villegas-Ch, Lujan-Mora & Buenano-Fernandez, 2018). The experience in the management of these tools suggests that the greater the sources considered in the analysis closest to reality are the results. These unstructured sources usually come from the students' navigation log, their Internet searches, and their interaction in social networks. All this data is impossible to leave out since the use of information and communication technologies (ICT) are an active part of the students (Li, Zhang & Wang, 2013). Being able to draw trends or determine the best way to learn from a student benefits all the components within an educational environment.

This paper considers the use of big data as an alternative to data analysis of a university where there are great diversity sources. The purpose is to consider the steps that an institution should consider to include in its processes the integration of these tools, as well as the use of the Hadoop framework as a manager in the analysis of data. The work is divided as follows: section 2 presents the concepts used for the development of the method; section 3 contains the method where the different phases to be considered for the implementation of a big data platform is established; section 4 presents the conclusions found in the development of the work.

2. Preliminary concepts

2.1. Big data

Big data is the massive data analysis, an amount of data, so large, that traditional data processing software applications are not able to capture, process and present results in a reasonable time (Shah, Soriano, & Coutroubis, 2017). Big Data was born with the aim of covering needs not met by existing technologies. In education, it can have an important impact on teachers, school systems, students and curricula. The analysis of big data can identify students at risk, ensure that students have adequate progress and can implement a better system for the evaluation and support of teachers and principals (Villegas-Ch et al., 2018). To comply with this process, big data techniques work in the storage and processing of large volumes of data that have specific characteristics such as:

- Volume refers to the size of the data that can come from multiple sources.
- Speed defines the speed with which the data arrive using units such as tera, peta or exabytes.
- Variety, we speak of data, structured, Semi-structured, Unstructured.

2.2. Data sources

The information available in universities has grown exponentially due to the inclusion of ICT in education (Cong & Xiaoyi, 2009). The data of this interaction is stored in multiple data sources that support academic management. Next, the following stand out:

- Produced by people. Send an email, write a comment on Facebook, answer a survey, enter information in a spreadsheet, use the learning management systems (LMS) and click on an Internet link. These actions, which are basic and carried out on a daily basis, represent an immense source of data.

- Between machines. The machines share data directly; this action is known as machine-to-machine (M2M). Thus, the parking meters mobile phones, vending machines for drinks and food in the university, to put a few examples, communicate through devices with other machines (Datta & Bonnet, 2014). The transmitted data is stored in different repositories. The communication networks to carry out these actions are very varied. Among the best known are Wifi, ADSL, fiber optics and Bluetooth.
- Biometrics. The data may originate from fingerprint sensors, retinal scanners, DNA readers, face recognition sensors or speech recognition. Its use is common in terms of safety in all its variants.
- Web marketing. All movement in the network is subject to all types of measurements that have marketing studies and behavioral analysis. With the analysis of these data, one can conclude the trends of each student (Bengel, Shawki, & Aggarwal, 2015). For example, the websites most accessed by students or places where they spend more time.

2.3. Type of data

In addition to their origin, the data can have classified into three classes according to their structure: Unstructured data types: documents, videos, audios, etc.; semi-structured data types: software, spreadsheets, reports; types of structured data. Only a small percentage of the information is structured and that can cause many errors if a data quality process is not applied (Gubanov, Stonebraker & Bruckner, 2014).

3. Method

The university that participates in this study stores a large amount of data. The majority of this data comes from the activities carried out by students in the LMS, academic management systems and sensors located in the university facilities. The sensors acquire and store all kinds of information about the activities that students perform within the university. This information in conjunction with all systems that are responsible for academic management can promote an important redesign in the learning methods used today.

When a big data process is included for the first time in a university it is important to generate working groups that prepare all the actors for the future change. The characteristic of this change is that the main value of big data does not come from the data in its raw form, but from its processing and analysis and from the insights, products, and services emanating from the analysis (Laigner et al., 2018). Radical changes in technologies and management methods are similarly dramatic changes in the way data supports decisions and innovation in products/services.

3.1. Phases of big data

The analysis prior to the implementation of a big data model is to determine the budget allocated during the process and the resources that will intervene considering the following parameters (Mohammed, Humbe & Chowhan, 2016):

- Managers are the sponsors, project managers, coordinators and quality managers immersed in an educational environment.
- Designers and data architects have technical profiles with clear objectives regarding the implementation of the project.
- Implementers, is the qualified personnel, analysts and developers, with knowledge of the sector and technology.
- Data operators are the analysts in charge of data at the entry, intermediate and result level.

Once the budget is determined, it passed to the design phase according to the needs of the university and it optimized considering the cost, the scalability and the different options of the market. It consists of two stages:

- Infrastructure, are networks, computers or servers, that is, the physical support of the solution.
- Architecture is the logical support of the solution, formed by protocols, communications or procedures.

For the implementation of the big data model in addition to the phases indicated, it is important to define aspects such as administration, maintenance, and security. The steps, for this, are:

- Installation of servers, components and start-up of the infrastructure.
- Configuration of the infrastructure for its correct functioning.

Figure 1 shows the phases that allow the execution of big data where two processes are considered. The first process is the engineering of big data that is composed of the acquisition and preparation of data (Chen, Mao, & Liu, 2014). The second process is the big data analytics, which consists of analyzing the data, reporting and acting.

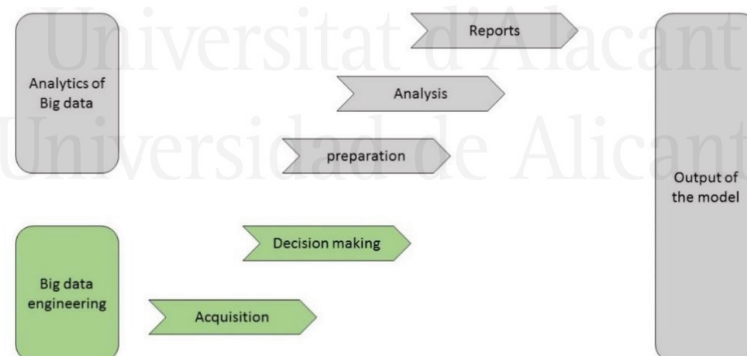


Figure 1 – Phases of execution of big data

3.2. Acquisition (big data engineering)

The first step in the application of big data in a university environment is to understand where the data comes from (Arruda & Madhavji, 2017). For which, the selection of sources considers three categories that fit this type of environment:

- Streaming data. It includes data that reaches the ICT systems of a network of connected devices. The data is analyzed as they arrive, this process determines which data to keep and which requires a deeper analysis.
- Social media data. Social interaction data is an attractive set of information, particularly for marketing, academic monitoring, and support functions. These data often have unstructured or semi-structured formats, so they present a unique challenge when it comes to consumption and analysis.
- Fonts machine to machine. The data of the different sensors contribute to the process identifying tendencies, places and times of permanence of the students in the university.

3.3. Preparation (big data engineering)

For the preparation of the data, certain guidelines are considered that make it possible to take full advantage of the information. The clarity in the guidelines supports the preparation of the data, the amount of data considered and how to use the knowledge discovered.

Raw data extracted directly from the sources are never in the format needed to analyze them. To solve the problem, it is important to prepare the data by applying two main objectives (Taleb & Serhani, 2017). The first is to purge the data to address data quality problems, and the second to transform the raw data to adapt it to the analysis.

The refinement identifies the erroneous data, corrects them or eliminates them; this process allows to improve the data and to consider only those that present a certain level of quality. There are quality problems with data from applications that are in production that include inconsistent data, duplicate records, missing data, etc. For example, the address of a student registered in two different places. This is an example of records that do not match; the lack of demographic data, unavailable values such as lack of a student's age, invalid data. For example, a telephone number, and outliers that cause values to be much higher or lower than expected. The quality problems in the data are solved by detection and correction of errors. In the correction, there are several methods; all depend on a previous analysis that allows applying to correct them. One of these methods is the elimination of records with unavailable values. Another method is to combine the duplicate records presenting unique and validated information.

The correction of the invalid values is made by replacing these values with the best estimate of a fair value. For example, for a missing value in the field of a student's age, the semester the student attends is estimated as a fair value. Outliers can have deleted as long as they are not important for the analysis. An effective process in the correction of errors implies having the clear knowledge of the application (Sehgal & Agarwal, 2016). For example, how the data has been collected, the population and the intended use of the application. This knowledge of the domain is essential to make decisions about how to handle incomplete or incorrect data.

The second part of the data preparation is to manipulate the purified data to convert it to the format needed for the analysis (Londhe & Rao, 2017). Some operations in this phase include scaling, transformation, feature selection, dimensionality reduction, and data manipulation.

- Scaling involves changing the range of values so that it is within a specific range, such as from zero to one. This is done to prevent certain features with large values from dominating the results.
- The transformation in the data eliminates noise and variability. The result of the transformations are considered as aggregate data. Adding data to the process results in a decrease in variability, which helps the analysis.
- The selection of data involves the elimination of redundant or irrelevant features, the combination or creation of new features. During the data exploration phase, it is possible to discover that some characteristics are correlated. In this case, it is possible to eliminate one of the characteristics without adversely affecting the results of the analysis.
- The reduction of dimensionality is useful when the data set has a large number of dimensions. The reduction allows finding a smaller subset of dimensions that captures most of the variation in the data.
- The manipulation is a way to verify that the raw data is in the correct format for the analysis. For example, from samples that record semi-annual changes in student grades, it is possible to capture changes in student performance for a given cycle. With this information, they can be grouped and calculate the mean, range and standard deviation for each group.

The preparation stage includes the socialization of the solution to the different departments responsible for the educational management and obtains the relevant authorizations to be able to carry it out. This stage includes:

- Detect the needs, have to do with the volume of data to be stored, its variety, the speed of collection, processing, and horizontal scalability. This process also reveals shortcomings when confronting the new technology with the existing one in the university.
- Justify the investment, big data improves the identification of both academic and technical problems and this by creating a high-performance environment that enables cost savings and improvement in academic quality.
- Evaluate the limitations; consider the infrastructure, technological maturity, resources, even the legal aspects in relation to data privacy.

3.4. Analysis (analytics of big data)

Data analysis involves constructing a model from the input data using an analysis technique to generate the output data (Eluri et al., 2016). There are different types of problems and different types of analysis techniques such as classification, regression, clustering, association analysis, and graphical analysis (Agnihotri & Sharma, 2015).

- The classification predicts the category of the input data. For example, predict the possible problems that a student encounters in the development of mathematical exercises.
- The regression predicts a numerical value instead of a category. For example, the prediction of the qualification of a questionnaire. The rating is a numerical value, not a category, so it is a regression task instead of a classification task.

- Clustering is organizing similar elements in groups. For example, group the base of students in different segments to recommend activities in such a way that learning is more effective.
- The association consists of developing a set of rules to capture associations within elements or events. Rules are used to determine when elements or events occur at the same time. For example, the association analysis may reveal that students who have good grades also tend to be interested in extracurricular activities.
- Graphical analysis to analyze data occurs when there is a large number of entities and connections between these, as in social networks. For example, in the graphic analysis, it may be useful to study the performance of a student over a period.

3.5. Reports (analytics of big data)

The potential of Big Data lies in the analysis and in converting the data into relevant information. Here comes into play the use of different techniques such as data mining and methodologies based on machine learning. These techniques and methodologies are those that can extract the true value to the information (Cao & Gao, 2018). The reports and the way in which the information is presented manage to complete the transformation of the commercial model created in previous phases into one or more representations of specific data of the university. Once the reports are obtained and with the complete modeling, it is necessary to evaluate and validate the results. The evaluation of the results depends on the type of analysis techniques used. For example, to get an idea of the model's performance with the data, it is evaluated based on questions such as:

- Is it necessary to perform the analysis with more data in order to obtain a better performance of the model?
- Does it help to use different types of data?
- Is it difficult to distinguish students with different needs in the results of clustering?
- Does adding the zip code to the input data help generate more granular student segments?
- Do the results of the analysis suggest a more detailed view of some aspects of the problem?

For example, the prediction of the percentage of students who pass mathematics gives good results, but the predictions of the results in the subject of calculation are not good. In the second case, the samples of the academic activities in the matter of calculation need deeper analysis. The factors that influence the results can be as diverse as for example; the existence of anomalies in the sample or the need to include additional data to fully capture the students' performance. What is sought is that the model is effective with respect to the success criteria defined at the beginning of the project. In this case, communication and action must be prepared on the results obtained in the analysis.

3.6. Decision making (analytics of big data)

The decision making works in tandem with the previous phase since after the analysis the conclusions come to carry out actions and make decisions. The final goal of data

analysis is to execute new strategies that improve academic management and student learning. The premise is that the analysis is in real time and as quickly as possible. The results obtained in the analysis convert the raw data into “actionable knowledge”. For an adequate and understandable management of the different areas, it is necessary to integrate visualization tools that conceive the reality of the study environment or be able to predict the future.

3.7. Data management with Hadoop

For data management, considering the different existing sources it is important to use tools capable of carrying out the process at the appropriate times. For this work, Hadoop is used as an open source framework to store data and run applications in clusters. Provides massive storage for any type of data, enormous processing power and the ability to process virtually unlimited concurrent tasks or works (Mazumdar & Dhar, 2015).

The architecture of Hadoop allows carrying out an effective analysis of large volumes of data, adding a value helps to make strategic decisions, to improve educational processes. This architecture allows to monitor what the students think or to draw scientific conclusions about the learning problems presented by different groups of students. With Hadoop, universities can explore complex data through customized analysis tailored to their students and needs.

The architecture of Hadoop is composed of three fundamental pillars that make it a versatile tool, flexible and fault tolerant. Its pillars are a distributed file system, called HDFS for its operation. The Hadoop engine consists of a MapReduce job scheduler, as well as a series of nodes responsible for executing them (Bhandarkar, 2010). A set of utilities that make possible the integration of subprojects.

On the file system is the MapReduce engine, which is a job scheduler, called JobTracker that is responsible for sending jobs to the nodes. MapReduce sends the incoming workflow to the TaskTracker nodes available in the cluster that are responsible for executing the map functions and reduces in each node. The planner keeps those jobs as close to the machine that has issued that information as possible. If the work cannot be located in the current node in which the information resides, the nodes in the same rack are given priority. This allocation reduces network traffic on the cluster’s core network. If a TaskTracker fails or suffers a waiting time, that part of the work is reprogrammed. Hadoop responds to a master-slave structure where the JobTracker is located in the master while there is a TaskTracker for each slave machine as shown in Figure 2. The JobTracker records the pending jobs that reside in the file system. When a JobTracker starts, it looks for that information, so that it can start the work again from the point where it was left.

The Hadoop file system (HDFS) handles two fundamental elements in the architecture: The NameNode and the DataNode. The NameNode is only found in the master node and is responsible for keeping all the stored data indexed. That is, the application needs specific information about the location of the data. The NameNode is found in the slaves and is responsible for storing the data.

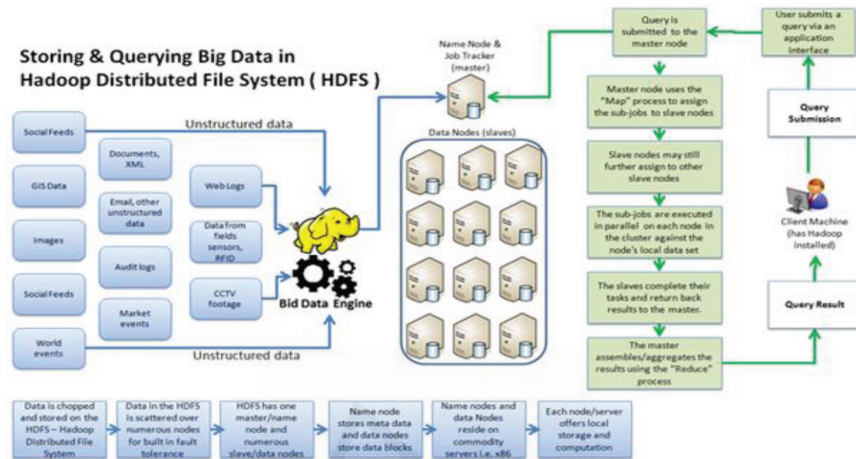


Figure 2 – Hadoop architecture (Ticout Outsourcing Center, 2013)

3.8. Hadoop MapReduce

Hadoop provides an execution environment oriented to applications developed under the MapReduce programming paradigm (Merla & Liang, 2017). Under this model, the execution of an application presents two stages:

- Map: where ingestion and transformation of the input data takes place, in which the input registers are processed in parallel.
- Reduce: aggregation or summary phase, where all the associated records are processed by the same entity.

The main idea, on which the Hadoop MapReduce execution environment revolves is that the entry is divided into fragments and, each fragment, is treated independently by a map task. The results of processing each fragment are physically divided into different groups. Each group is sorted and goes to a task reduced.

The execution cycle of an application in Hadoop is shown schematically. The developer only provides four functions to the Hadoop framework: the function that reads the input records and transforms them into tuples (RecordReader), the map (Mapper) function, the reduce (Reducer) function, and the function that transforms the pairs generated by the function reduces in output registers (RecordWriter).

3.9. Application cycle

The execution of an application in Hadoop consists of the presentation of the application, the generation of the ApplicationMaster for the application and the execution of the application managed by the ApplicationMaster. For example, a client program sends the request, including the necessary specifications to launch the application ApplicationMaster itself. The ResourceManager assumes the responsibility

of negotiating a container in which the ApplicationMaster must start, and then executes the ApplicationMaster. The ApplicationMaster, once started, is registered with the ResourceManager.

The registry allows the program to consult details of the resources to the ResourceManager. During normal operation of the execution, the ApplicationMaster negotiates containers of appropriate resources through the resource protocol. When the container assignment is satisfactory, the ApplicationMaster launches the container, providing the specifications of the container execution to the NodeManager. The execution information, in general, includes the necessary information to allow the container to communicate with the same ApplicationMaster.

The code of the application that runs inside the container provides the necessary information (progress, status, etc.) to your ApplicationMaster through a specific protocol of the application. During the execution of the application, the client that presented the program communicates directly with the ApplicationMaster to check the status, progress updates, etc. Through a specific protocol of the application. Once the application is completed and all work is completed, ApplicationMaster cancels the registration with the ResourceManager and closes, allowing the container to be reused for another application (Verma & Pandey, 2016).

4. Conclusions

This work is a detail of the steps that must be followed for the application of a big data platform. This method is in the testing phase where a large amount of data from different types of sources has been added. The exercise before going to production requires the comparison of results and the evaluation of all phases. So far it has not been possible to perform the evaluation because Hadoop needs deep knowledge in Java to be able to correctly integrate structured data. However, it has been possible to carry out tests in the different phases and the speed of processing, as well as the high savings in infrastructure, are details that allow us to continue with the investigation.

Several methods and models help the implementation of big data in a university institution. The first and most important step in the process is the location, analysis and data cleansing, this has taken more than 60% of the project's execution time.

Having a tool that allows knowing the students' tendency and the way they learn in a university environment is necessary for decision making. It must be borne in mind that in order to carry out this analysis what data scientists are looking for is that there is no ambiguity in the data. For this reason, consider the data that students generate in their normal activities without these being the results of surveys or questionnaires where the answers can be biased, it is an advantage over traditional data analysis platforms.

References

- Agnihotri, N., & Sharma, A. K. (2015). Proposed algorithms for effective real time stream analysis in big data. In *2015 Third International Conference on Image Information Processing (ICIIP)* (pp. 348–352). IEEE. DOI: 10.1109/ICIIP.2015.7414793

- Arruda, D., & Madhavji, N. H. (2017). Towards a requirements engineering artefact model in the context of big data software development projects: Research in progress. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2314–2319). DOI: 10.1109/BigData.2017.8258185
- Bengel, A., Shawki, A., & Aggarwal, D. (2015). Simplifying web analytics for digital marketing. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 1917–1918). IEEE. DOI: 10.1109/BigData.2015.7363968
- Bhandarkar, M. (2010). MapReduce programming with apache Hadoop. In *2010 IEEE International Symposium on Parallel & Distributed Processing (IPDPS)* (p. 1). IEEE. DOI: 10.1109/IPDPS.2010.5470377
- Cao, R., & Gao, J. (2018). Research on reliability evaluation of big data system. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 261–265). IEEE. DOI: 10.1109/ICCCBDA.2018.8386523
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171–209. DOI: 10.1007/s11036-013-0489-0
- Cheng, L., & Cheng, P. (2011). Integration: Knowledge Management and Business Intelligence. In *2011 Fourth International Conference on Business Intelligence and Financial Engineering (BIFE)* (pp. 307–310). IEEE. DOI: 10.1109/BIFE.2011.172
- Cong, P., & Xiaoyi, Z. (2009). Research and Design of Interactive Data Transformation and Migration System for Heterogeneous Data Sources. In *2009 WASE International Conference on Information Engineering (ICIE)* (pp. 534–536). IEEE. DOI: 10.1109/ICIE.2009.222
- Datta, S. K., & Bonnet, C. (2014). Smart M2M Gateway Based Architecture for M2M Device and Endpoint Management. In *2014 IEEE International Conference on Internet of Things(iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing(CPSCom)* (pp. 61–68). IEEE. DOI: 10.1109/iThings.2014.18
- Eluri, V. R., Ramesh, M., Al-Jabri, A. S. M., & Jane, M. (2016). A comparative study of various clustering techniques on big data sets using Apache Mahout. In *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1–4). IEEE. DOI: 10.1109/ICBDSC.2016.7460397
- Gubanov, M., Stonebraker, M., & Bruckner, D. (2014). Text and structured data fusion in data tamer at scale. In *2014 IEEE 30th International Conference on Data Engineering (ICDE)* (pp. 1258–1261). IEEE. DOI: 10.1109/ICDE.2014.6816755
- Laigner, R., Kalinowski, M., Lifschitz, S., Salvador Monteiro, R., & de Oliveira, D. (2018). A Systematic Mapping of Software Engineering Approaches to Develop Big Data Systems. In *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 446–453). IEEE. DOI: 10.1109/SEAA.2018.00079
- Li, X., Zhang, F., & Wang, Y. (2013). Research on Big Data Architecture, Key Technologies and Its Measures. In *2013 IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC)* (pp. 1–4). IEEE. DOI: 10.1109/DASC.2013.28

- Londhe, A., & Rao, P. P. (2017). Platforms for big data analytics: Trend towards hybrid era. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 3235–3238). IEEE. DOI: 10.1109/ICECDS.2017.8390056
- Mazumdar, S., & Dhar, S. (2015). Hadoop as Big Data Operating System -- The Emerging Approach for Managing Challenges of Enterprise Big Data Platform. In *2015 IEEE First International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 499–505). IEEE. DOI: 10.1109/BigDataService.2015.72
- Merla, P., & Liang, Y. (2017). Data analysis using hadoop MapReduce environment. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4783–4785). IEEE. DOI: 10.1109/BigData.2017.8258541
- Mohammed, A. F., Humbe, V. T., & Chowhan, S. S. (2016). A review of big data environment and its related technologies. In *2016 International Conference on Information Communication and Embedded Systems (ICICES)* (pp. 1–5). IEEE. DOI: 10.1109/ICICES.2016.7518904
- Sehgal, D., & Agarwal, A. K. (2016). Sentiment analysis of big data applications using Twitter Data with the help of HADOOP framework. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)* (pp. 251–255). IEEE. DOI: 10.1109/SYSMART.2016.7894530
- Shah, S., Soriano, C. B., & Coutroubis, A. D. (2017). Is big data for everyone? the challenges of big data adoption in SMEs. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 803–807). IEEE. DOI: 10.1109/IEEM.2017.8290002
- Taleb, I., & Serhani, M. A. (2017). Big Data Pre-Processing: Closing the Data Quality Enforcement Loop. In *2017 IEEE International Congress on Big Data (BigData Congress)* (pp. 498–501). IEEE. DOI: 10.1109/BigDataCongress.2017.73
- Ticout Outsourcing Center. (2013). Introducción a Hadoop y su ecosistema. *Ticout*. Retrieved from <http://www.ticout.com/blog/2013/04/02/introduccion-a-hadoop-y-su-ecosistema/>
- Verma, C., & Pandey, R. (2016). Big Data representation for grade analysis through Hadoop framework. In *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)* (pp. 312–315). IEEE. DOI: 10.1109/CONFLUENCE.2016.7508134
- Villegas-Ch., W., Lujan-Mora, S., & Buenano-Fernandez, D. (2018). Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining. In *2018 IEEE World Engineering Education Conference (EDUNINE)* (pp. 1–5). IEEE. DOI: 10.1109/EDUNINE.2018.8450954
- Villegas-Ch., W., Luján-Mora, S., Buenaño-Fernandez, D., & Palacios-Pacheco, X. (2018). *Big data, the next step in the evolution of educational data analysis. Advances in Intelligent Systems and Computing* (Vol. 721). DOI: 10.1007/978-3-319-73450-7_14

7 Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus

Villegas-Ch, W., Palacios-Pacheco, X., y Luján-Mora, S. (2019). Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus. *Sustainability*, 11(10), 2857. (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019a)

Disponible en:

URL: <https://www.mdpi.com/2071-1050/11/10/2857>



DOI: <https://doi.org/10.3390/su11102857>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O2. Identificar los componentes y las tecnologías que son parte de un campus inteligente.
- O3. Diseñar una arquitectura que convierta un campus universitario tradicional en un campus inteligente.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Article

Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus

William Villegas-Ch ^{1,*} , Xavier Palacios-Pacheco ² and Sergio Luján-Mora ³ 

¹ Escuela de Ingeniería en Tecnologías de la Información, FICA, Universidad de Las Américas, 170125 Quito, Ecuador

² Departamento de Sistemas, Universidad Internacional del Ecuador, 170411 Quito, Ecuador; xpalacio@uide.edu.ec

³ Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, 03690 Alicante, Spain; sergio.lujan@ua.es

* Correspondence: william.villegas@udla.edu.ec; Tel.: +593-098-136-4068

Received: 23 April 2019; Accepted: 13 May 2019; Published: 20 May 2019



Abstract: Currently, the integration of technologies such as the Internet of Things and big data seeks to cover the needs of an increasingly demanding society that consumes more resources. The massification of these technologies fosters the transformation of cities into smart cities. Smart cities improve the comfort of people in areas such as security, mobility, energy consumption and so forth. However, this transformation requires a high investment in both socioeconomic and technical resources. To make the most of the resources, it is important to make prototypes capable of simulating urban environments and for the results to set the standard for implementation in real environments. The search for an environment that represents the socioeconomic organization of a city led us to consider universities as a perfect environment for small-scale testing. The proposal integrates these technologies in a traditional university campus, mainly through the acquisition of data through the Internet of Things, the centralization of data in proprietary infrastructure and the use of big data for the management and analysis of data. The mechanisms of distributed and multilevel analysis proposed here could be a powerful starting point to find a reliable and efficient solution for the implementation of an intelligent environment based on sustainability.

Keywords: sustainability; IoT; big data; smart cities; smart campus; Hadoop

1. Introduction

The population of cities is growing through birth as well as migration from rural areas. It is expected that the proportion of the population living in cities and towns will “rise from 54 percent in 2015 to 60 percent by 2030, and to 66 percent by 2050” [1]. This unstoppable growth poses many environmental concerns, mainly an increase in the consumption of natural resources. The consumption of resources has been studied for several years, where its orientation has been based on the sustainability of the environments where people perform their different activities. Conducting a study and a proposal for sustainability supported by the use of information and communication technologies (ICT) is a heavy task for the extensive areas covered by a city, as well as the variations in society. This study addresses sustainability from a technological point of view, developed in a scaled environment such as a university campus. University campuses can be as large as cities and represent environments that are normally difficult to reproduce in another ecosystem. The idea is conceptualized by converting a traditional campus into a smart campus based on the concepts and experiences of smart cities where the integration of systems satisfies the needs of citizens with control over the consumption of resources.

7 Application of a Smart City Model to a Traditional University Campus

Smart cities are partly the result of advances in the field of ICT where, thanks to the Internet of Things (IoT), multiple devices can connect to the internet and generate information that allows them to interact effectively with other members [2]. However, moving from a traditional city to an intelligent city entails a great technical and socio-cultural effort, in addition to a high investment in physical and economic resources. The questions to answer, before considering a smart city environment, are broad and require specialists in different areas to work together under the objective of optimizing resources [3].

A university campus conforms to the previous definition, so having this type of ecosystem is an ideal starting point for this study. The geographical distribution, administration and the number of people who frequent them constitute ideal environments for the demonstration of techniques or processes of a smart campus [4]. However, the question of what a “smart and sustainable campus” is and of what it comprises remains. From this initial question, other more specific questions arise: is ICT able to solve sustainability issues in a smart campus? What improvements does a smart campus offer over a traditional campus in terms of sustainability? Moreover, do traditional campuses meet, at least partially, the requirements of smart and sustainable campuses? Not all these questions can be answered based on a single experience; therefore, they should be addressed in cooperation with the different areas that make up the administrative and academic part of a university campus and based on experiences of related works. Several of these works do not cover a study like the one proposed, where the objective of creating a smart campus goes hand in hand with the sustainability of the campus with the environment.

Related research proposes policies or reviews on how to educate students to encourage good use of resources. This is different from the proposed work, because the basis is the use of ICT to provide the necessary means to meet the needs of students with the adequate use of resources. This study begins by defining university campuses as places where hundreds or thousands of people study or work, and their work depends on the infrastructure to house them. This infrastructure is a network of communications, transport and services that are required for the management of university life [5]. As a result, the processing of each activity generates large volumes of information. This information is stored with the purpose of consulting it later and generating knowledge that is useful for the user. Among the systems and devices that generate information within a university campus are security systems, biometric access devices or card readers, wireless systems, automatic dispensers, domotic buildings, academic management systems, financial systems and so forth [6]. In traditional university campuses, these systems are independent and do not have processes that centralize the information in such a way that they can be used to optimize resources and take advantage of knowledge to improve student learning.

The integration of these different systems is the basis for a better interpretation of why things happen within a university environment. For example, the latest generation of wireless systems provide relevant information about the places that students prefer in certain seasons [7]. In the same way, automatic dispensers can reveal information about consumables that teachers prefer in their free time [8]. The access sensors provide important information with which to determine the time a person stays on campus [9]. The application of domotic systems allows the adequate management of the consumption of energy. These are several of the systems that can be found on a university campus, and that allow the supposition that the transformation to a smart campus is feasible [10]. The IoT concept provides the necessary support for all devices to connect to the Internet, and the data they generate is stored in centralized places such as the cloud. At the university campus level, there are technological advantages, since most have private data centers that would replace a public cloud, reducing the costs of external services and ensuring the high availability of the data [11].

The centralization and analysis of data are important in a smart campus for their contribution to the processes of identification of events and needs, considering that universities make decisions based on the data they have about students and their administrative structures. However, current decision making requires improving methods of data analysis and results in shorter periods. The problem is

that the volume of data greatly exceeds the processing capabilities of the classic analysis platforms [12]. Having techniques that detect the needs of the university population and generate results based on trends is the basis of a smart campus. Therefore, the alternative, which currently represents a trend in data science for its superior results, is the use of big data. These platforms offer alternatives for the management of data and obtaining knowledge about students that are flexible, with lower costs and in shorter periods. Traditional techniques such as business intelligence (BI) and data mining can perform data analysis. However, due to their nature and the limitations of the data, many of these are omitted or eliminated in the preparation process; in a smart campus, it is important to keep most of the data while looking for alternative techniques for cleaning, so that it does not affect the efficiency of the decision making [13].

This work proposes, taking as a guide, the smart cities and moving from a traditional campus to a smart campus that works with three fundamental axes: the acquisition of data through the IoT, the centralization of data in a proprietary infrastructure and the management and analysis of data with the use of big data. This integration allows proper management of the information that is generated within the campus. The sensors are responsible for monitoring all the events that arise in their environment and sending the information to a storage system. The system stores the information in a private cloud where all the data are processed and transformed to present quality data for the next phase. The big data architecture is in charge of selecting the data and, through data analysis processes, it gives the campus the knowledge necessary to make decisions. For example, on a hot day, the sensor system monitors the temperature and the environment of each of the classrooms. According to this information, the management systems determine what action they should take, such as opening windows or turning on the air conditioning system. Being able to manage the different devices, by means of a previous analysis of the acquired data of the environment, allows the adequate use of the resources in a sustainable environment [14]. The results of this work allow the generation of comfortable environments where the students, teachers and administrative staff have their needs met in total harmony with the environment. The depth of the analysis should make it possible to determine exactly the places most frequented by students through the data obtained from the wireless system. Knowing the exact location of students makes it possible to take advantage of these spaces to generate awareness campaigns about the use of resources.

This work is divided as follows: Section 2 presents the concepts used for the development of the method; Section 3 contains the method, where the different phases to be considered for the implementation of a big data platform are established; and Section 4 presents the conclusions found in the development of this work.

2. Preliminary Concepts

2.1. Big Data

Big data is massive data analysis, referring to an amount of data so large that traditional data processing applications are not able to capture process and present the results in a reasonable period. Big data was born with the aim of covering needs not met by existing technologies [15]. In education, it can have an important impact on teachers, school systems, students and curricula. The analysis of big data can identify students at risk, ensure that students are making adequate progress, and can support the implementation of a better system for the evaluation and support of teachers and principals [16]. To comply with this process, big data techniques work around the storage and processing of data that have specific characteristics, such as:

- The content format;
- The type of data;
- The frequency with which the data is made available;
- The intention: how the data should be processed (ad hoc query in the data, for example);
- Volume: the size of the data that can come from multiple sources;

- Velocity: the speed with which data arrives using units such as terabytes, petabytes or exabytes;
- Variety: structured, semi-structured, and unstructured.

2.2. Smart Campus

Technological advances modify the immediate future and create new paradigms on human interactivity with things. The integration between technologies and their applications in social environments promotes the generation of intelligent environments, which support the automation of processes, remote control, and decision making in their environment. University campuses are places where thousands of people study or work daily. The university campuses are in communication with the cities in which they are located on tangible issues related to infrastructure, and on intangible issues such as social relations or innovation [17]. The use of ICT, as in cities, improves educational campuses and the quality of life of the inhabitants [18]. It allows areas of educational control to monitor their students and those involved in university education. A smart campus allows a better coexistence between the university population and its surroundings, adequately manages the resources within the campus, and provides favorable places for learning.

2.3. Smart Cities

An intelligent city is able to take advantage of the data it produces in its daily operations to generate new information that allows its management to be improved and to be more sustainable, more competitive, and offer a better quality of life, thanks to the participation and collaboration of all the citizen actors. A smart city detects the needs of its members and reacts to these demands by transforming the interactions of citizens with public knowledge systems and elements of knowledge [19]. The intelligent city bases its actions and its management on this knowledge, in real time, or even anticipating what may happen.

2.4. Internet of Things

There is data everywhere in the home, at work, and in practice in all facets of life. Communication devices generate large volumes of data that allow managers or data scientists to establish research on user trends, as well as the status of machines and their products. The connection of physical things to the internet makes it possible to access data from remote sensors and control the physical world remotely. The combination of captured data and data retrieved from other sources, for example data on the Web, results in new synergistic services that go beyond the services that an isolated integrated system can provide. This technological innovation aims to connect the items we use daily to the internet, with the aim of bringing the physical world closer to the digital world [20].

2.5. Hadoop

Hadoop is an open-source framework for storing data and running applications in basic hardware clusters. It provides massive storage for any type of data, as well as enormous processing power and the ability to handle virtually unlimited tasks or work. Hadoop applied in technological environments is not only able to acquire and manage large volumes of data, but it does so regardless of its format. Hadoop can store all kinds of data: structured, unstructured, and semi-structured; log files, images, video, audio, communication, and more. Hadoop stands out for presenting an architecture capable of ensuring high availability and recovery of the data it ingests [21].

The large volumes of data generated in the systems included in this study require a platform that allows their processing, guaranteeing speed, effectiveness, and data quality. In this proposal, the Hadoop architecture is used as a platform due to the advantages it offers in the distribution of nodes for all sub processes. Hadoop allows improvement of the processing without punishing the use of the infrastructure that the campus has. Another advantage is the versatility that it presents in

the processing of data that is in different formats. An example of this is the events generated by the personnel access systems; these systems send the data in plain text files or .xml files.

In a campus, as well as in a smart city, finding a diversity of sources forces us to look for solutions that go beyond the traditional BI platform. Hadoop assigns massive collections of data across multiple nodes within a cluster of servers, while indexing and keeping track of the data and preparing it for the analytical process. It is necessary for the implementation of a smart campus to integrate several data sources and apply a process of analysis on these, because the behavior and the patterns that a person presents depend on many variables, unlike any other process of data analysis.

For example, if a student consistently fails a multiple-choice exam, it is important to recognize why he has this difficulty. The data that is available in the different computer systems of the university campus contain substantial information that, by means of preprocessing and analysis, can explain the factors relevant to why the student fails. These factors often do not depend on a linear analysis, since the variables are many and depend on one another. In the case that is presented, a socio-academic analysis of the student, whereby applying a regression of the data, it is possible to detect whether the student is carrying this problem from the high school, may be necessary. If this process is satisfactory, it will not be necessary to include more variables. Otherwise the analysis is extended where systems are added and thus there are more variables. This example gives a simple output if it is considered that there are cases where the analysis process needs to include the data of all the systems, with the volume of data, as well as the processing, increasing exponentially. For this reason, it is important to select a big data architecture that provides enough resources for the process, both in memory and in storage. Hadoop adapts to the aforementioned needs for the analysis of data, which collects information from IoT systems, systems with transactional as well as non-transactional databases, plain text files, and so forth.

2.6. Definition of the Problem

The continual growth of the university population forces the expansion of university campuses both physically and in the expansion of resources as they need to cover the needs of their members [22]. To cover these needs, it is important to establish a structured plan that allows students to live with the environment where issues of global interest, such as sustainability, are raised. University campuses that do not have a defined plan, which includes guidelines and policies that allow the proper management of economic, social and natural resources, affect the development of learning. To solve these problems, university campuses look for alternatives supported by ICT. The consideration that is made is that university campuses generally have their own computer infrastructures that allow them to generate and store large volumes of data, which in most cases is used. Having a platform that takes advantage of all the systems in charge of acquiring data and analyzing them in a private cloud environment helps the management of learning and the interaction of the university population with ICT. This process, in addition to guaranteeing improvement in educational processes, allows the university population to interact with the environment and improve situations of comfort, safety and decision-making. The challenge for data scientists is to unify all the data generated in academic systems, financial systems, enterprise resource planning, learning management systems, security systems, sensor systems and actuators, among others.

The experience of studies carried out in the field of smart cities allows systems to be integrated, applying IoT and big data concepts to transform a traditional campus into a smart campus. This option has great value in educational management because it improves the interaction of students, teachers and administrators in a sustainable environment. It generates comfortable environments where the needs of each individual are covered in a personalized way, based on a projection of their tendencies. This ecosystem integrates concepts, technologies, data and individuals to guarantee a knowledgeable society where the development of learning is guaranteed by new technological trends [17].

Another important aspect of the definition of the problem is the importance of the security and privacy of the data. The ability to extract and analyze the data that big data achieves will probably

generate controversies and negatives in the campus population. To solve these problems, it is important to have clear security and privacy policies in the handling of data [23].

3. Related Work

The process of the selection of related works was made by taking the model proposed by Barbara Kitchenham as a guide [24]. This method details the steps to follow to obtain a review with research value. In the process, we searched and classified more than six thousand articles from four sources related to smart cities and smart campuses. The scientific bases considered for the search were Scopus, Springer, Elsevier and IEEE. The process of inclusion and selection is detailed in Figure 1. The figure shows the phases used for the filtering of the articles and the number of works reviewed. In the first stage, a search was made on scientific bases, simply relating to works in the smart cities and smart campus area published between 2014 and 2018 and which have open access, with the result that six thousand and seventy-five works were found.

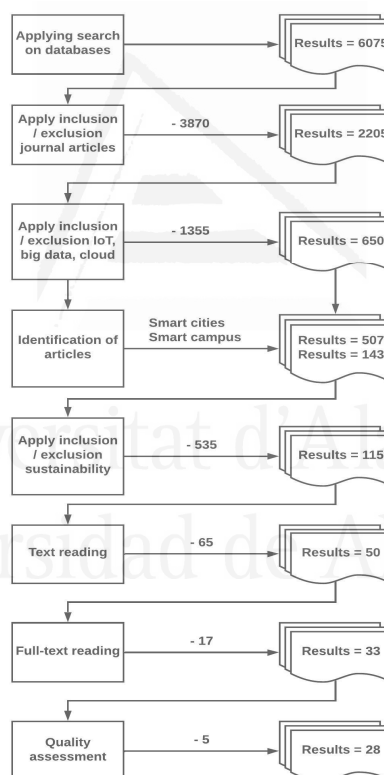


Figure 1. Process of the selection of previous works. Source: authors.

In the second stage, the works were filtered and only journal articles were considered, resulting in two thousand two hundred and five articles. In the third phase, the works that were related to the use of IoT, big data and cloud computing were filtered, resulting in six hundred and fifty articles. These articles were processed to identify those that deal with smart cities, obtaining five hundred and seven results and those that deal with smart campuses, obtaining one hundred and forty-three

articles. The articles were processed to identify those that addressed issues such as sustainability, resulting in one hundred and fifteen related articles. The next stage was to make a reading of the summary, introduction and conclusions focused on verifying the guidelines of the articles with the proposed work, of which sixty-five articles were excluded. Of the fifty remaining articles, seventeen dealt with the issue of system integration in a smart campus or smart city environment in a superficial manner, so they were excluded by selecting thirty-three articles for the next stage. The thirty-three selected articles went through a new review where the quality of the article was evaluated specifically within the proposed method and the relevance that it represented for this work and five articles were excluded. The twenty-eight articles included in this study contribute directly to the objective of this work that seeks to establish an intelligent campus, integrating innovative technologies that guarantee the coexistence of the population with the environment in a sustainable model.

Table 1 describes the twenty-eight works considered, the type of research, and the type of contribution. These works were reviewed in detail, since they were those that directly influenced the research.

Table 1. Distribution of publications by type of research. Source: authors.

REF	Title	Type of Research	Type of Contribution
[25]	A smart campus prototype for demonstrating the semantic integration of heterogeneous data	Proposed solution	Process
[26]	From the university to smart cities—how engineers can construct better cities in BRIC's countries: a real case from smart campus FACENS	Evaluation research	Model
[27]	Social activities recommendation system for students in smart campus	Proposed solution	Tool
[28]	OnCampus: a mobile platform towards a smart campus	Proposed solution	Process
[4]	What smart campuses can teach us about smart cities: user experiences and open data	Evaluation research	Model
[29]	Data acquisition and analysis of smart campus based on wireless sensor	Proposed solution	Process
[17]	Learning analytics for smart campus: data on academic performances of engineering undergraduates in Nigerian private university	Evaluation research	Model
[30]	The construction of smart campus in universities and the practical innovation of student work	Proposed solution	Process
[31]	Evaluation of the smart campus information portal	Evaluation research	Model
[32]	Intelligent campus (ICampus) impact study	Evaluation research	Model
[33]	A study on association algorithm of smart campus mining platform based on big data	Evaluation research	Model
[34]	The development of a smart campus—African universities point of view	Proposed solution	Process
[35]	Composite indicators for smart campus: data analysis method	Proposed solution	Model
[36]	Student perception of smart campus: a case study of Czech republic and Thailand	Evaluation research	Process
[37]	Building a smart campus to support ubiquitous learning	Evaluation research	Model
[38]	Smart campus: fostering the community awareness through an intelligent environment	Evaluation research	Process
[5]	Constructing smart campus based on the cloud computing platform and the internet of things	Evaluation research	Process
[39]	A smart, caring, interactive chair designed for improving emotional support and parent-child interactions to promote sustainable relationships between elderly and other family members	Evaluation research	Process
[40]	E-Learning and Its Effects on Teaching and Learning in a Global Age	Evaluation research	Process
[41]	Adding the "e-" to learning for sustainable development: challenges and innovation	Proposed solution	Model
[42]	Corporate attitudes towards big data and its impact on performance management: a qualitative study	Proposed solution	Model
[14]	Designing an efficient cloud management architecture for sustainable online lifelong education	Evaluation research	Process
[43]	Sustainability and energy efficiency research implications from an academic	Proposed solution	Model
[44]	Exploring factors, and indicators for measuring students' sustainable engagement in e-learning	Evaluation research	Process
[45]	Operating charging infrastructure in China to achieve sustainable transportation: the choice between company-owned and franchised structures	Proposed solution	Model
[46]	Systematic review of education for sustainable development at an early stage: cornerstones and pedagogical approaches for teacher professional development	Evaluation research	Process
[47]	Definitions and frameworks for environmental sustainability in higher education	Proposed solution	Model
[48]	The impact of modern markets on the performance of micro, small and medium enterprises	Proposed solution	Model

The first consideration was that the works related to smart cities addressed the problem from an urbanistic point of view; the variables that were part of the studies represented the care of the environment and the proper use of energy resources. Another factor of analysis was mobility and security; these factors include geographic problems and video surveillance systems [5,29,37]. Here, domotics and the IoT are considered stronger, as the objective they focus on is to implement models that allow the use of energy and the proper use of energy. The IoT, for its part, manages the deployment of sensors and devices that generate information that is stored in the cloud through the internet. For the process to be complete, the information stored must generate knowledge. Big data is responsible for the analysis through the processing of data, and the results respond to the needs of people in real time. Figure 2 shows the basic pillars of an intelligent city that contribute to competitiveness and safeguard a sustainable future in the symbiotic link of networks of people, companies, technologies, infrastructure, consumption, energy, and space [49].

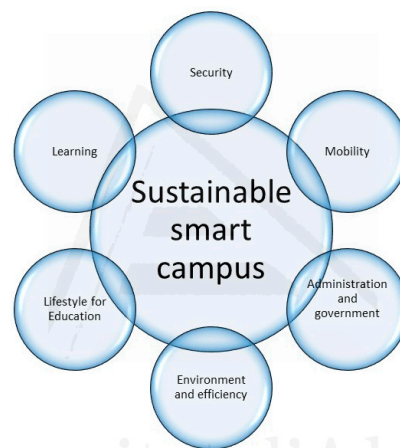


Figure 2. Pillars of a sustainable smart campus. Adapted from: [50].

Secondly, the works related to smart campuses are considered and 85% were studies that recommend the use of different models of smart cities applied to a university campus. They performed an analysis of the characteristic features of these environments, focused mainly on the management of services through mobile or desktop applications [51]. The main objective was to exploit the information generated by the different systems responsible for academic management. For example, through applications users could know exactly where they were, or easily find an office or person. Other applications focused on a specific area of the campus, such as a faculty or a building. The purpose was to centralize the largest number of devices and sensors that provided relevant data on the activities developed in that area. The ten works considered as the main ones to address the sustainability issue were directly related to specific areas that were included in the study, for example, a sustainable environment which supports improvement to the learning or the administration of the cloud in an efficient way [14,40]. These works guaranteed the functioning of the architecture proposed in this investigation. The guideline that was carried out with these works was to take the experiences and proposed designs and compare them, in certain cases, with the architecture of this work and to improve the design of sustainability in a campus.

4. Method

The concept of smart campuses is not new; however, the conception and integration of new technologies make it truly innovative. The results of the analysis of related works determined that in the

field of smart campuses, there are still many topics to be addressed, and problems that require scientific contributions. The method of this research establishes a process that includes all the components that contribute to the transformation of a traditional campus into a smart campus. The existing models and processes in smart cities act as a guide for this work. Based on this, several pillars that are part of smart cities are included in the analysis of university campuses, maintaining the difference in scales as a reference [37]. Therefore, a university campus is considered a controlled representation at the scale of a city. University campuses have always displayed a recurrent ambivalence. On the one hand, they have acted as a powerful exchanger of knowledge and opportunity for collaboration; while on the other hand, they have provoked conflicts and isolation in students with academic problems. The explanation of these phenomena is complex and there are different understandings, depending on the point of view of the analysis. From a social point of view, university campuses are a place of crisis for certain populations [52]. Their complex and concentrated socio-academic activities often give rise to problems of exclusion, segregation, and educational polarization. However, learning generates wealth and well-being for those involved. In fact, university campuses are a space of coexistence and connection for individuals and social groups, which stimulates and facilitates the development of learning and knowledge [53].

In summary, a university campus is a generator of social problems and, at the same time, an institution that provides solutions. However, student agglomerations produce important negative externalities, such as greater responsibilities and a lack of resources, which reduce academic efficiency. Solving these negatives requires considerable effort towards their correction [54]. From the point of view of innovation, university campuses are excellent places to generate inventions, develop new technologies and disseminate knowledge. From an environmental point of view, the artifacts and physical devices incorporated in the university campuses transform and model the natural environment where it is located. The transformation of the physical environment has to do with the construction of infrastructure and buildings that can generate significant environmental impacts [55].

4.1. Characteristic Features of University Campuses

By thinking of a university campus as a small city, it serves as a test bed for the integration of techniques that make up a smart campus. This concept allows identification of the characteristic features of a campus based on those of a city [56]. For example:

- **Complexity:** One of the main challenges facing university campuses is the level of complexity of the processes that take place within their limits, and in their area of closest influence. So-called “complexity science” is understood as a set of ideas about the self-organizational capacity and the adaptive nature of some complex systems. Complex systems include the climate, natural ecosystems, the economy, and in this case, university campuses.
- **Diversity:** The campuses differ among themselves because of their geographical location, their academic vocations, or their socio-economic structure. At the same time, very different spaces coexist within each campus. The more sophisticated and disparate the functions of a campus, the more diverse the agents involved in them will be. If the dimension factor is added to this condition, then the greater the size and functional complexity, the greater the number of agents that will have to be counted when formulating policies.
- **Uncertainty:** For planners, the uncertainty that surrounds the future of university campuses is a constant consideration. Any projection is faced with the task of foreseeing the future of a campus in ten or twenty years. The current limitations of forecasting tools have greater weight in situations that worsen if one operates in a constantly changing environment.
- **Sustainability:** A university campus generates a significant change in an environment, the installed infrastructure; the consumption of energy, the interaction of people with the environment must be controlled through institutional policies. New technologies provide sufficient means and processes to find a balance between the environment and the consumption of natural resources [43]. Sustainable development must be included in the design of smart campuses, where the needs

of all its stakeholders are met without compromising the development of new generations and natural resources.

The characteristic features of a university campus allow identification of the key points that support a smart campus. The structure of a smart campus must be flexible, scalable and evolutionary, where processes are evaluated constantly and the level of reaction to external events is efficient [57]. The controls and audits, considering the characteristics of the campus, must guarantee that its components are adequate and include all factors of influence in the daily life of the campus.

4.2. Components of a Smart Campus

This research is carried out in a university campus that chose to participate in the study. In Figure 3, the components that are considered for the development of the method are presented. These components internally govern diverse systems that comply with a specific process, and the integration of these and the general components results in a smart campus. A smart campus aims to improve the quality of life of its community by applying ICTs in a sustainable manner [35]. For this purpose, the components included in this work are related to three main axes: the IoT, in charge of generating and obtaining environmental data; cloud computing, which centralizes data in internal or external infrastructure; and big data, which comes with the analysis and management of data [58].

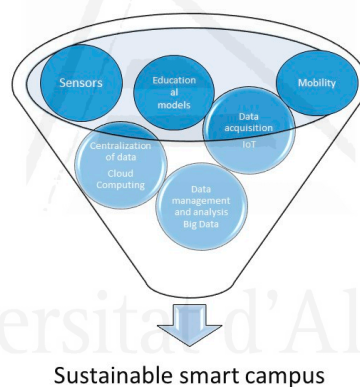


Figure 3. Components considered for a sustainable smart campus. Source: authors.

4.2.1. IoT Data Acquisition

The IoT allows information to be stored on the internet, thanks to the connectivity that different devices and sensors have. This information is processed, and the results allow measurements or control of different environments. Within a smart campus, this concept is maintained but on a small scale. This feature has advantages with respect to the handling and control of information [59]. In this work, the data of all the identified components of the campus is part of an internal network and are stored in a data center infrastructure where the best physical characteristics for the storage and processing of information are provided [60]. The campus has integrated several systems considered as part of the IoT, and these systems acquire information on activities and events that are a daily part of university life. Among these are:

- Access control systems: Biometric sensors or RFID readers control the security of the university campus. When detecting any event, the sensors activate the different actuators located strategically in the smart campus [29].
- Automation systems: The university infrastructure is complemented directly with autonomous systems. These systems seek to reduce energy waste, as well as improve the quality of life of the

university population. Several of these systems are integrated into the data analysis architecture that helps manage resources efficiently [42].

- Security systems: Security is vital in a university campus, and for this, there are systems of sensors or video surveillance systems that generate information about any event 24 h a day. The potential of these systems allows their data to be used for other activities, such as detecting trends or identifying special needs of the population. The monitoring capacity added to the analysis process allows autonomous control of the services offered by the university.
- Automatic dispensing system: Automatic dispensers are widely used in university campuses and are devices that have relevant information on the trends of the population that consumes these products. The analysis of these data can reveal which product is the most consumed in different periods.

The way in which sensors interact with the environment and people contributes to creating a knowledge society. The sensors are responsible for collecting information from the environment and sending it to the cloud, where interested parties can consume it. Information converted into knowledge helps decisions to be made quickly and accurately. This contribution is ideal for the development of smart cities; however, in an environment of scale, the results are measurable and quantifiable in much shorter periods. The aforementioned does not imply a leap in the processes or in the architecture of an IoT system. The technology proposed by Kamilaris and Pitsillides [61] gives this proposal greater validity, since implementing an ecosystem based on the Internet of Things will ensure that students and teachers make sound decisions in relation to their environment. This ecosystem contributes to managing hours of classroom use and managing virtual environments with interactive learning devices. The decision-making system can automatically detect which equipment or system is not in use and put it in standby or turn off as appropriate. In the area of academic management, this technology contributes to the allocation of physical spaces that meet the needs of each subject, as well as the number of participants.

The system, with the use of IoT, reduces the consumption of energy resources by generating friendly environments. These environments are based on the automation of events such as opening blinds, turning lights on or off, etc. Controlling a large number of events has greater advantages and encompasses a set of systems that allow for the control of more complex and susceptible environments such as a data center [47].

With the security that the integration of cameras will bring to the system, the teacher will be able to verify that each student is who they say they are. In addition, tasks assigned to the teacher are automated, which, in traditional education models, are still practiced manually, such as taking attendance. With this technology, acting through cameras, facial recognition that determines which student is present or not, is possible. Having technology as the main assistant in the educational environment breaks the paradigms of learning and opens greater possibilities to generate knowledge.

Another important point is the acquisition of data where the protection of personal data of the individuals that make up the smart campus is considered. On this subject, there are works [62] that study the regulatory framework applicable to the use of personal data in smart cities that serves as a reference for the development of a smart campus. The controlled ecosystem where this work is carried out allows the total availability of the data. However, access to them is done with the appropriate permission of the population. Another measure of control is the ability to persuade the university community to participate in the use of the data that, through its analysis, seeks to satisfy its own needs.

4.2.2. Cloud Computing and Centralization of Data

Previously, the systems that are part of the data acquisition were described; the objective is to improve the management of the information. For this, it is necessary to store the data in a centralized place with access to all the systems that are part of the university campus. For example, smart cities upload and store data to the cloud in order to manage the data in an adequate and fast way [63]. Using the cloud in the architecture of smart cities does not represent major inconvenience due to the

management capacity, as well as the availability of greater economic resources. In a smart campus, storage becomes an infrastructure problem and a cost-benefit analysis. To solve these problems, the use of internal data centers that are part of the university infrastructure is considered, essentially transforming its intranet into a private cloud [64].

The ability to manage data internally improves control by implementing quality processes in storage and high availability when consuming data. This research uses a data center architecture that allows the generation of a private cloud to provide secure data storage. Figure 4 details the topology within the data center. This topology provides the capacity to store 60 terabytes of information. It also has redundancy systems and backup equipment that guarantee the availability and quality of the data. In addition, energy backup systems are available in case of any electrical failure. The devices that comprise this private cloud architecture guarantee availability on the proposed platform. The VNX3200 is responsible for storage in solid-state disks. It is connected to the management and processing unit via fiber optics that allow high-speed data transmission. It is composed of two internal controllers that, in addition to managing the storage system and meeting the requests of the processing unit, allow high availability and redundancy of the data. The second element is a Cisco UCS MINI chassis that integrates eight blade servers that allow the creation, administration and distribution of virtual machines. In addition to these tasks, the servers also handle computer processes, attends to customer requests, and are responsible for processes such as load balancing and virtualization tasks. There are elements of redundancy that allow the system to guarantee availability in case of any incident. Finally, the N3K-DC devices are three-layer switches; each one integrates 48 fiber optic ports that are connected in a redundant architecture which guarantees communication. Through this equipment it is possible to communicate with the campus intranet, and the services can even be published to the internet.

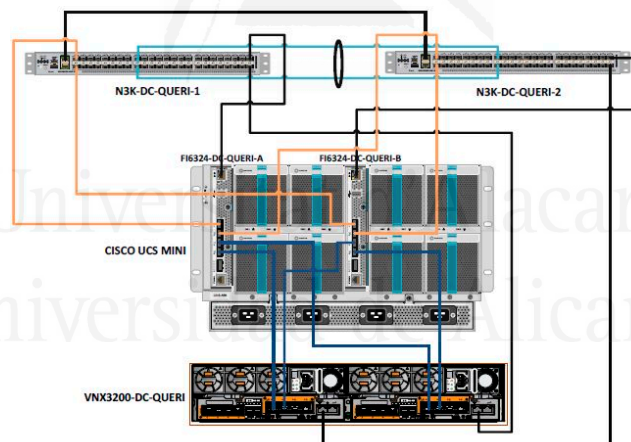


Figure 4. Topology of a private cloud. Source: authors.

All systems are connected directly to the computing center, which is the intelligent part and is responsible for processing tasks. The data center is connected through two-network equipment, that works in layer three, to the data network of the entire campus. This part of the topology guarantees high availability and redundancy in case of an accident in one of the pieces of equipment. This architecture allows the creation and management of different virtual servers, facilitating management in the processing of information according to the concept of big data.

4.2.3. Data Management and Analysis of Big Data

Before starting the process of data analysis and management, work must be aligned to the regulatory framework that applies to the use of data in the smart campus. A specific framework, which applies to environments as proposed, has not been possible to locate in the analysis of previous works; however, at this level the laws generated by each state also govern. These problems are better known in the use of the internet as new factors undermining the right to privacy generate serious problems of personal and commercial security. For the development of this work, the right to privacy for each person has been considered due to the law established by the country where the study originates [65]. In addition, the data is duly protected during the process and is used only for educational analysis, maintaining the right to privacy of the individual protected by the legal tradition that protects and preserves the inviolability of the home, papers and documents. This ensures that others can use none of these elements without the consent of the individual to whom they belong. During the execution of the method, sensitive data that affect the most intimate details of a person's privacy have not been used, eliminating the possibility of creating an ideological, racial, sexual, health, economic or any other profile that becomes a threat to the individual. Therefore, the regulations on data protection will not apply to information such as files kept by individuals in the exercise of exclusively personal or domestic activities, nor those data made anonymous in such a way that it is no longer possible to identify the interested party [66]. In a smart campus like the one proposed, there are processes where the identification of individuals is not necessary. The opposite happens in ecosystems such as medicine where the application of big data requires identifying each individual to associate an illness or disease. In this phase, it is important that the data of each person can be associated as belonging to him or her.

The existing data sources in a smart campus directly influence data management. For this reason, it is important to use tools capable of performing a quality process in the extraction, transformation and loading of data (ETL), considering adequate processing times. Studies on educational data analysis use BI techniques applied to educational data in order to discover patterns in students that allow the detection of how they learn [67]. The results are used to make corrective decisions in teaching methods, as well as to make projections and anticipate a possible event such as student desertion [68]. These studies use ETL processes, the processed data is stored in a data warehouse, where data is queried through data mining algorithms and a conclusion is reached. These techniques generally take as a guide the application of the processes of data analysis of a company and, when passing them to a wider environment, they present certain technical and knowledge difficulties. On the one hand, a BI system is developed in an environment where the analysis objective is unique and, in many cases, can focus on a part of the business, for example, the detection of variables affecting sales in certain quarters. If the study needs more scalability, it is necessary to go back to the design process, add the variables to generate cubes of online analytical processing (OLAP) and design the dashboards to present the information. For the development of processes, both commercial and open source tools can be used. The decision to use commercial or open source tools depends on the economic limitations of the organizations and knowledge level of the tools.

Another important factor that intervenes in the use of BI platforms is the data sources they can handle and which are usually structured databases [69]. When processed by the ETL they are stored in a multi-relational database. The management of these platforms is increasingly versatile and staff with the required knowledge are not difficult to find. If we compare the needs in the aforementioned environment vs. what a smart campus implies, good results are not expected immediately in the volume of data. The first comparison is because, in the extraction of data from the different sources, an ETL does so on an individual basis based on the variables that it needs for the analysis, which means that if these variables change, it will be necessary to create a new connection to the correct source and start the process again. In a smart campus model, the idea is to consider all sources and that the platform in charge of data analysis manages the variables regardless of the sources. Having a large number of variables exponentially increases the volume of data considered for the analysis.

Another point to consider is the ability to use data that is not precisely in a specific format: an ETL has the ability to work with several formats, extract them and convert them to a specific one. This task, although it seems an advantage when we talk about a large volume of data and variety in the formats of the data, can become a problem when consuming many storage and processing resources. In a smart campus, the management of several systems is considered both in structured databases as well as in unstructured sources.

The analysis of the possible usefulness of a data analysis tool based on BI platforms confirms that the conditions of the data in an intelligent campus exceed its functionalities. In an environment such as the one proposed, it is necessary to have processing and storage that guarantees high availability, safety and quality in the data. For this reason, in a smart campus, it is important to think about platforms that have been used in large companies that work with large volumes of data like Hadoop [70]. Companies of the scope of Google, Yahoo, Amazon, and so forth, have used this tool, and this guarantees the treatment of the data at the required level within the guidelines established in the study. Another advantage of Hadoop is the ease it has to skillfully manage any type of file or format; something that is very difficult to obtain with a traditional BI approach [71]. Other architectures such as Apache Spark apply to big data platforms. This architecture in relation to Hadoop presents better response time in processing and is used in real-time analysis. Apache Spark is more expensive than Hadoop for the deployment of infrastructure that it needs. Although this architecture is novel and has been put on par in the market with Hadoop, it does not meet the conditions presented in this study [72].

The method we used based on cost and the availability of the infrastructure is Hadoop, which is an open-source framework for storing data and running applications in clusters [73]. Hadoop provides massive storage for any type of data, has an enormous processing capacity, and is able to process virtually unlimited concurrent tasks or work.

The architecture of Hadoop allows an effective analysis of large volumes of data, and the results can strengthen decision-making and improve educational processes. This architecture also allows monitoring of the opinions of students, as well as the ability to draw conclusions about learning problems presented by certain groups of students. With Hadoop, universities can exploit complex data, analyze it, and customize results by adapting the process to the needs of the university and the students.

Hadoop is composed of three fundamental pillars: versatility, flexibility, and fault tolerance. Among the components that allow the execution of the architecture is the distributed file system (HDFS). The Hadoop engine consists of a MapReduce work scheduler, as well as a series of nodes responsible for executing it [74]. These characteristics are presented as a set of utilities that enable the integration of subprojects. It is important to consider that MapReduce also provides retrospective and complex analysis capabilities that can touch most or all of the data [75]. MapReduce provides a method of data analysis that is complementary to the functions provided by SQL.

The main problems that are resolved in Hadoop are

- Data capture;
- Storage;
- Filtering out;
- Transfer;
- Analysis;
- Presentation.

These problems are classic in a traditional campus, and the way to solve them is through client-server storage methods where the user interacts with the application, which in turn controls the storage and analysis through relational databases. This method works well with applications that do not generate a large volume of data and that are processed by traditional servers, or that do not exceed the limit of their processor. In summary, this method depends on the applications and available computing resources. An example of this is the process used by BI. However, when it comes to dealing

with huge amounts of scalable data, processing the data through a single database is a frantic task, which makes the process a bottleneck.

MapReduce divides the task into small parts and assigns them to many teams, collects the results, and forms the resulting data set. Figure 5 shows the traditional method of data analysis in a university campus, where the user interacts with an application and the data is stored in relational BDD. On the other hand, MapReduce manages the data generated in several processes; the centralization system processes this data and interacts with the user when presenting the results.

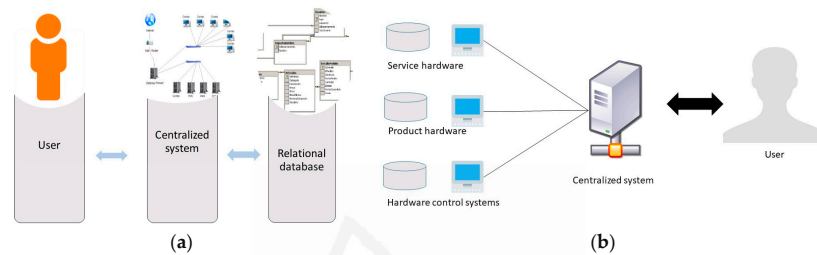


Figure 5. (a) Traditional analysis system vs. (b) MapReduce. Source: authors.

The architecture shown in the previous figure represents an improvement in the management and processing of the data; Hadoop integrates the MapReduce algorithm, which is responsible for the processing of the data in parallel [76]. Figure 6 presents the architecture of Hadoop, which in its core has two main layers; the first is responsible for the computation, and the second is responsible for distributed storage. Thus, the base apache Hadoop framework is composed of the module of Hadoop common that contains libraries and utilities needed by other Hadoop modules. There are Java libraries and utilities required by different Hadoop modules. These libraries provide file systems and OS level abstractions and contain the necessary Java files and scripts required to start Hadoop. The next module is the Yarn framework that is a resource-management platform responsible for managing computing resources in clusters, providing very high aggregate bandwidth across the cluster.

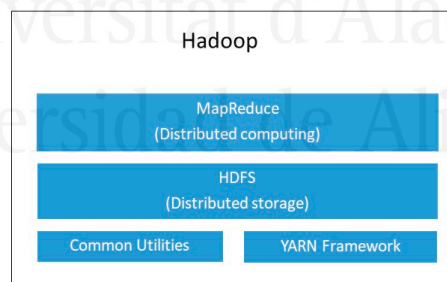


Figure 6. Hadoop architecture. Source: authors.

According to the characteristics and needs presented in a smart campus, Hadoop has the best architecture when applied in a fully distributed mode [77]. This mode of operation requires that a defined number of clusters be deployed that are responsible for processing all assigned work. The management cluster carries out the assignment of tasks, and both the management and the processing are virtual machines assigned in the data center of the campus. The advantage of having the architecture in the intranet is network management, which leads to the best use of resources. Being integrated with the internal network, the communication is transparent and there are no critical problems, as can be the case when creating the clusters in an external cloud [78].

The installation of the architecture is done on a Linux platform then the functionality in the workflows is checked. MapReduce plans the tasks through a JobTracker that is responsible for sending the works to the nodes. MapReduce sends the incoming workflow to the available TaskTracker nodes in the cluster, handling the map functions and reducing them in each node [79]. The planner keeps tasks as close to the machine that has issued that information as possible. If the work cannot be located in the current node in which the information resides, the nodes in the same rack are given priority. This allocation reduces network traffic on the cluster's core network. If a TaskTracker fails or suffers a timeout, that part of the work is rescheduled. Hadoop responds to a master–slave structure, where the JobTracker is located in the master and there is a TaskTracker for each slave machine, as shown in Figure 7. The JobTracker records the pending works that reside in the file system. When a JobTracker starts, it looks for that information, in such a way that it can start the work again from the point where it was left [80].

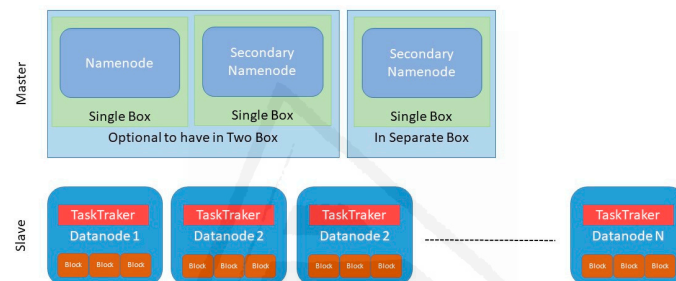


Figure 7. Hadoop master–slave structure. Adapted from: [81].

The HDFS handles two fundamental elements in the architecture: the NameNode and the DataNode [82]. The NameNode is only found in the master node and is responsible for keeping all the stored data indexed. That is, it informs the application where the searched data is found. The NameNodes are found on the computers of the slaves and are responsible for storing the information.

With the architecture mounted and with the data acquisition process executed, it starts its analysis. This is done through Hadoop, which allows visualization of the different nodes in a graphical interface. In the analysis of the data, the project is divided into several subprojects, which facilitates obtaining information for each system included in Hadoop. For example, it is possible to analyze the drinks that have the highest consumption at certain dates or seasons. The skill of the data scientist is in posing the right questions to help to control the parameters of a specific event.

The Hadoop interface contains all the works that are in process and stores the corresponding information in files. To verify the functioning of the architecture in this research, the following conditions for the analysis are presented:

- Which are the beverages that present the higher indices of consumption in examination seasons in the campus?
- Which are the places in the campus with the highest population density in winter and summer?
- What are the activities that generate greater knowledge in students in the campus?

These questions seek to solve common problems in a university and help improve the use of resources and the understanding of university trends. For the first question, the information generated by the automatic dispensing system enters the analysis process. The information is sent from the dispensing machine to a virtual server. Figure 8 details the architecture of the data acquisition of the different sensor and actuator systems. In the specific case of the dispensers, they contain different sensors that allow the actuators to generate a specific event that is translated into data, which is sent to the information layer. In this layer, the data is stored in relational or non-relational BDD, depending

on the application. The communication protocol of the sensors and actuators at the commercial level is varied. However, the university that participated in the study, with the purpose of designing a scalable architecture and guaranteeing communication, standardized the use of the technologies and with them the protocol of communication within the smart campus using the TCP protocol. In the knowledge layer, the data sent by the dispensing machines is acquired through the different data mining processes, or what is defined within the big data platform. The knowledge generated is applied in the good use of resources for the first case, and the improvement of services. Hadoop stores the data and the administration cluster assigns small analysis processes to the 120 nodes implemented within the architecture. The storage format of the data is simple and contains fields such as date, time, type of beverage, location within the campus, and its identifier. This data arrives in plain text and in real time; therefore, the analysis process has accurate information. The Hadoop analysis is based on applications developed in Java, where the Hadoop libraries are imported and a .jar file is generated that, when executed, starts the analysis process.

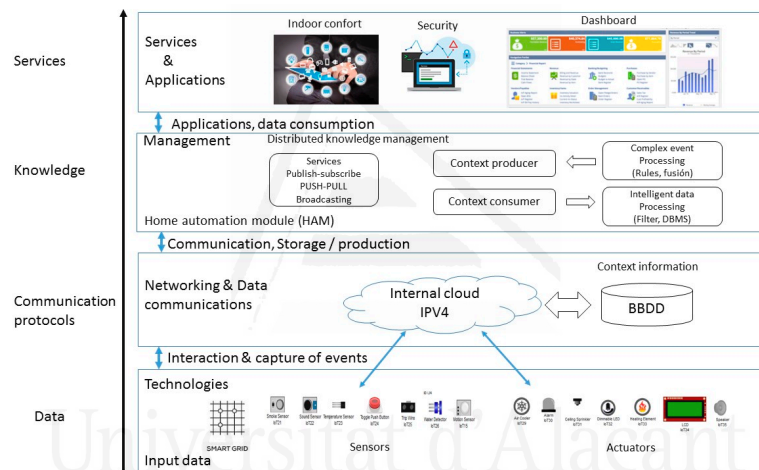


Figure 8. Internet of Things (IoT) architecture within a sustainable smart campus. Source: authors.

Table 2 presents the results of the first analysis, where the study period consisted of 2 weeks, which is the usual period of exams. Samples were taken every four hours, and drinks available in each machine were classified into four classes. The percentage is related to the amount shipped and the total amount of beverages owned by a vending machine of 288 units capacity. The table shows a high consumption of coffee, followed by coca cola drinks, which in one way or another contain caffeine. With the results obtained, several adjustments can be made within the optimization of resources. On the one hand, it has the capacity to project the number of beverages required according to their classification in different seasons. On the other hand, with the results of the analysis, it is possible to complement a study on stress in students when performing an exam. The study makes it possible to create awareness campaigns on the dependence on caffeine, and the treatment of stress in the student population.

The analysis of the places with the highest population density within the campus requires accurate information on the location of the students. In order to comply with the requirements of the analysis, the data generated by the wireless local area network (WLAN) is considered. The university campus has an integrated WLAN system that handles load balancing device identification, enabling this type of study. Access-point (AP) devices provide information about the number of hosts that are connected, and the controller that manages the APs can emit traces of these hosts that include information of

the time they connected to the network, as well as the identification of the AP to which they are connected [83]. The potential of wireless systems promotes the use of information to the inhabitants of the smart campus, based on the conditions of the environment. This consideration is applicable according to work that has been done in urban sectors, and that can be adapted to the needs presented in this study [84]. This monitoring is continuous; therefore, the volume of data is high, and the consequence is that the processing of the file takes more time. To reduce the processing time, we work with algorithms developed in Java that perform a pre-filtering, whereby the sample is segmented from months to weeks or days.

Table 2. Consumption of drinks in examination seasons. Source: authors.

Period	Type of Drink	Percentage
07:00–11:00	Coffee	60%
07:00–11:00	Coca-cola	25%
07:00–11:00	Juices	10%
07:00–11:00	Others	5%
11:00–15:00	Coffee	53%
11:00–15:00	Coca-cola	30%
11:00–15:00	Juices	11%
11:00–15:00	Others	6%
15:00–19:00	Coffee	42%
15:00–19:00	Coca-cola	39%
15:00–19:00	Juices	15%
15:00–19:00	Others	4%

Table 3 shows the results obtained in the density analysis, in which the gross data in of four random weeks of both the summer period and the winter period are considered. The distribution of the AP within the smart campus is by area, and the amount of equipment assigned depends on the analysis of the population density and the optimization methods that the wireless system allows. The APs, in addition to providing access to the network, generate important information such as the amount and detail of the hosts that connect to the network in a specific period. This information is useful to determine in which areas more infrastructure resources or bandwidths are needed, with the objective of optimizing these resources. Another derivative of this analysis is that academic authorities can provide relevant information about the university at the points where students usually meet. Chatting or informative activities can be held by academics taking advantage of the places preferred by the students. The results shown in the table are as expected; however, the tool allows a quantitative analysis, which allows the use of resources to be improved and generates objectivity in its use.

Table 3. Areas with the highest population density in a smart campus. Source: authors.

Areas	Number of Users in Summer per Week				Number of Users in Summer per Week			
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Playground	6450	6098	5493	4536	4839	3927	2983	1826
Green areas	2983	4997	5382	6113	5487	4387	3762	2873
Coffee shops	502	493	650	528	638	936	1182	1382
Laboratories	392	182	293	387	508	1029	1603	2083
Libraries	932	670	398	732	736	1283	1893	3072
Buildings	1982	805	1038	963	973	1562	1793	1923
Total	13,241	13,245	13,254	13,259	13,181	13,124	13,216	13,159

The implementation of a data analysis architecture such as Hadoop allows for better decision making regarding the management of natural resources. Through information on places where there are a greater concentration of students, it is possible to carry information about the proper use of resources. Generating awareness campaigns becomes a way of education and even more so when the

smart campus has a system that learns from the data generated by each user. For the deployment of network equipment or AP, sectors that do not have a minimum number of users are restructured immediately to take advantage of resources and promote use in the places that need it.

To respond to the academic activities that generate learning in students, the structure of a single system that provides data to multiple systems is changed. This coupling of more data sources is required because analyzing performance in people requires greater effort, as well as a greater number of variables. The variables contain the student's socio-academic information, academic record, financial situation, interaction with learning management tools, and so forth. There are several methods through which the analysis can be focused; one is to create a sub process that filters the information of each system, and then unites it in one to present the information. Another method is to perform the sequential analysis of each system, where the results are stored in variables and then presented as a common result.

When creating sub processes in charge of analyzing each system, the processing time is optimized. This functionality is found in the capacity of Hadoop when assigning and coupling a certain number of cores to each task. The inclusion of data mining algorithms allows deepening of the analysis to create clusters and identify the patterns in each student. This analysis allows determination of how students learn, as well as evaluation of the teaching methods of each teacher. This type of study is a true contribution to a smart campus, since it allows the improvement of learning, which is the main objective of a university. By improving the quality of learning, the results of the environment improve, as this conveys a vision of excellence internally and externally. A good image as a university helps improve student income rates, and this goes hand in hand with the economic growth of the campus. Therefore, economic well-being improves the quality of life of those involved, and budgets are increased in all areas.

Mobility is another important point that directly affects sustainability within the smart campus and one of the objectives of this architecture is to reduce the CO₂ emission that this causes. In order to meet this objective, it is important to analyze the information that exists within the campus and why the problem arises. As a starting point, we must consider that a university campus can be geographically as large as a small city, therefore, mobility must be considered in a sustainable environment. To solve the problem there are options, such as implementing an internal transport system that carries out the tours every so often. This option solves the problem, to a certain extent, of the inhabitants of the campus not using their vehicles and therefore reducing the emission of gases. However, it is not an optimal solution because its implementation is based on the experience of the administrative staff of that area.

The architecture proposed in this work covers these needs by defining the times, units and routes of each transport based on the analysis of the students' existing data. Figure 9 shows the flow diagram under which the process to solve this problem is conceived. The first stage is responsible for collecting the data that comes from systems such as the location system through the wireless, video surveillance systems and academic management systems. If the data exist and are appropriate, the big data architecture assigns the necessary nodes to perform the processing of the information. The following flow leads to the storage of the data and its analyses through Hadoop that performs the process based on the research parameters. For example, when identifying patterns in students, the most frequent places are determined, even the big data process takes data from previous processes such as the location of students through the wireless system. This information is selected according to the students who have traveled the greatest distance, in this way it is possible to determine if the students move over short or long distances. As a fixed variable, the system detects students who travel more than two kilometers, with this data, there is enough information to define the places to which, and times in which buses should be sent.

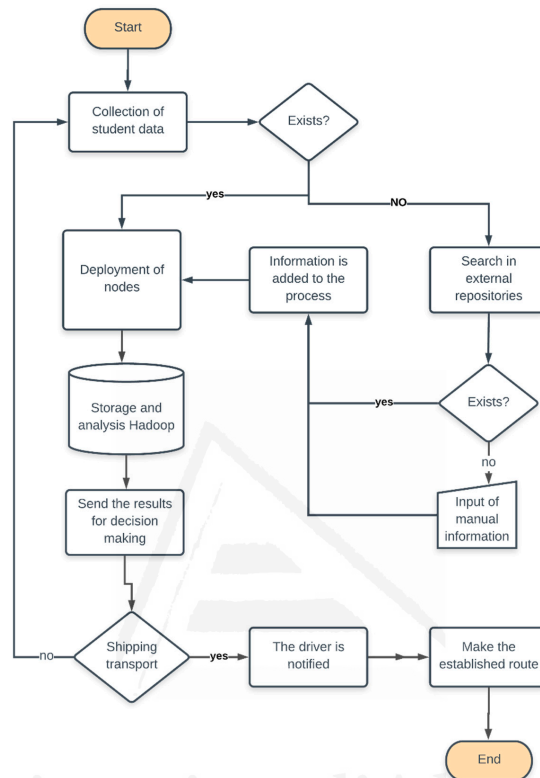


Figure 9. Flow diagram of an internal transport process by means of big data. Source: authors.

The video surveillance systems, as an integral part of the proposal, send information about bus stop points. If there is saturation; the analysis system detects and alerts the administrator to send buses. Academic management systems provide information about student schedules so that big data identifies all the variables that contribute to solve the problem. Once the transport administrator is notified, they arrange the driver and the route. If the transported shipment is not made, the cycle returns to the data collection phase and the process is repeated. If the data does not exist or is not sufficient for the analysis, the process stops and looks for information that helps to solve the problem in external sources such as alternative systems or spreadsheets. If the data found is sufficient, the data is added to the process and the Hadoop nodes are implemented according to the request made. Otherwise, the data is entered manually, this corresponds to scenarios where the system does not find enough data to make a decision and an operator has to join the process to verify or correct it if it is the case.

5. Discussion

Transforming traditional campuses into smart campuses not only improves the management of processes that contribute positively to student learning, but the approach carried out in this work also allows the integration of state-of-the-art technologies to solve problems related to the use of resources, as well as to improve the environment where learning takes place. The use of technologies, such as IoT, allows the automation of different environments within the campus, enabling the interaction of individuals directly with the environment. This technology, taken from smart cities, allows the

campuses to acquire information on all the events that each student, teacher or member of the administrative staff performs [47]. At the same time, it is capable of executing actions that improve the lives of people, no matter how basic they may seem. For example, the use of dispensers that send data on the preferred choice of individuals at each time of the year show the population trends with the possibility of generating healthy eating habits in individuals.

The wireless service, when implemented, allows us to measure the population density in each area of the campus, offering even detailed information about the movement of each device and therefore of each person. This ability can be exploited, in addition, to offset charges in the AP and opens the possibility of detecting these places and bringing important information to the university, such as waste management, services provided by the university, etc. All this is possible with the proper use of resources and the most important thing that these technologies allow is the reuse of the devices that the campus has to take advantage of the technological and economic resources [43]. These possibilities, which are framed in these systems, directly support the sustainability within the environment. In the work developed, the systems, through the information acquired, processed and analyzed, are able to execute actions that align with the proper use of resources. For example, in the energy part, the smart campus acts directly on areas or stations that do not require energy by directly turning off any device that is located in it. Additionally, the interactive equipment that has been integrated into the system shows whether a teacher uses it or not, in such a way that, at a certain time of inactivity, the device generates an event and it enters into energy saving mode.

In terms of security, the smart campus architecture allows integration of all the means used for perimeter security, turning them into the eyes of the campus, where each event is monitored in real time. Internally, the architecture is able to recognize people through intelligent services implemented as an additional component of big data. This service improves the management of each class by eliminating processes that a teacher is used to doing, such as taking a register. With facial recognition, the smart campus is responsible for this task, becoming the assistant in each class. Another factor that is possible to improve is the mobility within the campus. This study has been developed in a campus that has more than thirty hectares and there are several disadvantages when trying to move from one faculty to another. The use of vehicles has been restricted and emphasis is placed on the use of internal transport, reducing the pollution emitted by private vehicles [44]. However, how is it possible to cover the mobility needs of each student? The responsibility rests again with the monitoring systems, but especially with the results obtained from Hadoop. Data analysis allows the system to improve the allocation of schedules for both teachers and students so that students with similar schedules are classified and located in specific areas. These results reduce the students' needs to move long distances within the campus. However, the analysis goes further and has the ability to identify the trends of each student, for example, the hours and places he usually visits during his/her free time. By identifying this trend, Hadoop classifies and effectively issues the number of buses, their capacity and hours that these must travel through the different areas.

The combination of technologies demonstrates that generating ecosystems, where people, technology and the environment can coexist, is possible and that their results can be applied to larger environments. Smart campus is the ideal test bench where new technologies can be evaluated and their architectures can be replicated in cities, improving the way of life for individuals.

The use of data can become a topic of discussion in the application of big data within a smart campus; however, it is necessary to motivate the population about their participation in this process. The objective of this project is to improve the wellbeing of the people who are in this environment, for which the use of data is indispensable. With this method, it has been explained about their use and how they are protected. However, there are processes where identifying people will be essential, specifically in the analysis of the data that determines how students learn. This process, which is specific to identifying patterns and classifying students, recommends activities that help improve student learning. Another point is security. The smart campus has several systems that are able to identify who enters or leaves each of the facilities, guaranteeing the trust of each person in the system.

In this work, it has been considered as an internal policy of the smart campus that not all information is published and the current law of the country where it carries out its activities governs the use of the data.

Another necessary aspect to consider is technological acceptance. This issue directly involves the population who are part of the study, because, in the design of the smart campus, the changes to the different areas of the campus are considered. At this stage, it is possible to detect negatives to the implementation due to ignorance or to a technological change, which brings with it a rejection of new processes. To counteract this, a model capable of mediating between technology and people is required. The technology acceptance model (TAM) will be considered as a model in future work and provides the correct path for the population to accept a technological change that adds new concepts and automation of processes [85]. In many cases, people consider new technologies as a direct competitor, arguing that many of these technologies can replace people in their activities. Awareness and education are the keys to this process and those people who will be responsible for motivating others to accept the technologies in their environment and, even more so, the proposals in this work that interact with the university population. The importance of the TAM makes it possible to prepare students to live in societies that support good communities, living with nature and generating sustainable environments such as smart cities.

6. Conclusions

The main question stated at the beginning of this work was what makes a university campus a smart campus? To address the issue, it has been necessary to perform a specific analysis of the characteristics they present. A university campus is a set of infrastructures such as classrooms, libraries, laboratories, faculties and computer systems where the university community can develop activities for their learning. This concept is broadened by Sections 2 and 3 of this work. A university campus should not focus solely on infrastructure, it needs to interact more with its community. This interaction helps to guarantee learning by offering services that guarantee the comfort and safety of each person. In addition, our proposal aims to facilitate management in all areas of a traditional campus through data analysis processes that have been described in this method.

The next question that was addressed in the study was to define what problems traditional university campuses are facing; the answer is part of the definition of the problem. However, it is important to mention the experience that large companies have gone through, where all those that have not evolved and integrated ICT in their processes are destined to disappear. This same concept applies to university campuses, which generally assume that learning is linear. This perception has changed and it is the duty of its authorities to improve learning methods using ICT. It is not enough to maintain large buildings and the best computer systems if each of these works individually. The integration of data that seeks to discover how a student learns and generate completely suitable environments that adapt to those needs is the key in education. Once it is defined that each student has specific needs and that these can be measurable in variables, a smart campus can be created that learns about their needs and supports the management of those involved.

A smart campus brings several improvements to the management of a traditional campus, this is in response to the third question raised in the development of this work. However, it is important to consider the depth of the idea; this improvement is not simply based on an organizational guideline. On the contrary, the response covers topics such as security, the discovery of trends in the university population, improvement of learning and, above all, that systems and services are used to improve the condition of their community. In several works that have been the subject of analysis for this proposal, it has been possible to show that the improvements on a traditional campus are many, but the continuing work of all the sectors that support this evolution is important. If we consider technical issues, it is important to mention that independently a university campus through a BI platform can analyze the academic data of students and make decisions that contribute to learning. However, an analysis on this scale is often not enough, since an analysis of behavior includes many variables that

range through sociological, academic and financial aspects. The inclusion of variables sacrifices storage and processing, so applying technologies such as big data are presented as a solution that guarantees the inclusion of diverse sources and effective processing which considerably improves decision-making.

Traditional campuses have certain components that have been detailed in Section 4 and serve as a platform to create a smart campus. The fundamental component when implementing this solution is the technological reuse that gives greater value to the capacity of the platform that allows integrating the data generated by all the systems. The university campus where the study was developed has a diverse infrastructure; there are systems and devices of the last generation as well as not so new technology. The benefit in this evolution to a smart campus is given by three main components that are obtaining data through IoT-based technologies, its storage in a private cloud and finally its treatment using a robust platform such as Hadoop. By detailing these components as the main basis of this work, the requirements that a traditional campus has to meet can be divided into two parts. In the first place, there are the technical aspects that integrate the aforementioned, such as, for example, video surveillance systems, personnel access systems, computer systems, and so forth. The second part is the qualified personnel, this implementation needs a great knowledge in different areas, and the study to focus on a university defines that researchers, teachers, technical, administrative, and so forth, will support it. That guarantee the applicability of the proposal.

Converting a traditional campus to a smart campus is the obligatory step that universities must undergo, as the results are clearly positive and the resources in these environments exist. In this work, Hadoop is implemented as the big data architecture that supports the current needs of a smart campus. For the implementation, it is important to know the different tools that allow the inclusion and analysis of the variables that need to be evaluated in a smart campus. Considering smart campuses as test banks for smart cities is completely valid, as demonstrated in this study.

For a transformation of this type, it is necessary to prioritize the work in common with the different areas that contribute to the success of the data analysis. Generating knowledge from light and critical data contributes to the optimization of resources. In the three cases of analysis that have been presented in this study, it is found that each time a result is obtained it affects different areas, which directly compromises the effectiveness and efficiency of the decision making.

A critical point of a smart campus is to create optimal learning environments with tools and academic methods that match the needs of students. With a big data architecture, this is possible through having the ability to analyze a huge amount of data from different sources, and then correlating the data and presenting it to the interested parties so that they can take preventive and corrective measures as appropriate. The optimization of energy resources is another controllable area because; by having an integrated system where sensors and actuators generate information, big data analyzes the data and based on these applications can regulate consumption. An example of this is the results generated by the analysis of the places with the highest university population.

With these results, it is possible to determine the need and establish priorities for these places and others, to put them in waiting for an event to happen so that they can shift into an active mode. On the other hand, access control systems allow improved security and reduced human resources. This is thanks to the security systems that can track individuals through cameras and facial recognition applications. These allow a person in any area of the campus to be located in the same way, as the person's access is analyzed each period, generating reports of the activity of each person.

The contribution of this work is the improvement of learning through the integration and analysis of data in a smart campus. All the systems of a university campus generate data that is available in different repositories. The acquisition mode is based on the IoT systems, which, in this work, connect to a private cloud. The implemented distribution presents the essential processing capabilities for the production workflow to establish metrics that promote the availability, scalability, and reliability of the services. The availability of information converted into knowledge helps detect patterns in the way students learn, and what their needs are. Once these needs are met, all those involved in learning

within the smart campus assume their role with effectiveness, thanks to an environment that is adapted to their needs.

Sustainability in a traditional university campus requires a great deal of effort and even generates more consumption of resources when looking at students' awareness of the environment. Our proposal includes a process of transformation to a smart campus that has many advantages with respect to the responsible management of resources, considering that each action follows a process of analysis and that each decision depends on the integration of many variables. At the beginning, it can be considered that it is a system that requires many trained personnel; however, the implemented method is scalable and it is possible to create scenarios where the results can be measurable and compared with traditional models. The potential advantages compared with a traditional model focus on the development of learning and generate comfortable ecosystems, which are friendly to the environment.

The big data architecture deployed contributes to the treatment and analysis of data to such an extent that it is possible to determine the needs of each person and customize the services to suit their needs. In specific cases, where the academic activities proposed by the teachers do not conform to the student learning model, the analysis of the data allows for identifying and recommending the best activities for each student, improving learning in the community. This process, simple as it seems, is more important since, by improving learning, student dropout rates are reduced and the academic effectiveness indexes known as the graduation rate are improved.

Author Contributions: W.V.-C. contributed to the following: the conception and design of the study, acquisition of data, analysis, and interpretation of data, drafting the article and approval of the submitted version. The authors X.P.-P. Moreover, W.V.-C. Contributed to the study by design, conception, interpretation of data, and critical revision. S.L.-M. made the following contributions to the study: analysis and interpretation of data, approval of the submitted version. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Swilling, M.; Hajer, M.; Baynes, T.; Bergesen, J.; Labbé, F.; Musango, J.K.; Ramaswami, A.; Robinson, B.; Salat, S.; Suh, S.; et al. The Weight of Cities Resource Requirements Of Future Urbanization. Available online: <https://europa.eu/capacity4dev/unep/documents/weight-cities-resource-requirements-future-urbanization> (accessed on 1 April 2019).
- Atzori, L.; Iera, A.; Morabito, G. The Internet of Things: A survey. *Comput. Netw.* **2010**, *54*, 2787–2805. [\[CrossRef\]](#)
- Morales Lucas, C.; de Mingo López, L.; Gómez Blas, N. Natural Computing Applied to the Underground System: A Synergistic Approach for Smart Cities. *Sensors* **2018**, *18*, 4094. [\[CrossRef\]](#)
- Vasileva, R.; Rodrigues, L.; Hughes, N.; Greenhalgh, C.; Goulden, M.; Tennison, J. What Smart Campuses Can Teach Us about Smart Cities: User Experiences and Open Data. *Information* **2018**, *9*, 251. [\[CrossRef\]](#)
- Nie, X. Constructing Smart Campus Based on the Cloud Computing Platform and the Internet of Things. In Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), Hangzhou, China, 22–23 March 2013; pp. 1576–1578.
- Hannan, A.; Arshad, S.; Azam, M.A.; Loo, J.; Ahmed, S.H.; Majeed, M.F.; Shah, S.C. Disaster management system aided by named data network of things: Architecture, design, and analysis. *Sensors* **2018**, *18*, 2341. [\[CrossRef\]](#)
- Wang, K.; Wang, Y.; Hu, X.; Sun, Y.; Deng, D.-J.; Vinel, A.; Zhang, Y. Wireless Big Data Computing in Smart Grid. *IEEE Wirel. Commun.* **2017**, *24*, 58–64. [\[CrossRef\]](#)
- Britt, J.; Matsumura, S.; Forood, H.; Zimmerman, S.; Myles, P.; Zawicki, S.; Kutami, D. Internet of Things Platforms, Apparatuses, and Methods. Available online: <https://patents.google.com/patent/US20160147506> (accessed on 17 May 2019).
- Sotres, P.; Lanza, J.; Sánchez, L.; Santana, J.R.; López, C.; Muñoz, L. Breaking Vendors and City Locks through a Semantic-enabled Global Interoperable Internet-of-Things System: A Smart Parking Case Pablo. *Sensors* **2019**, *19*, 229. [\[CrossRef\]](#)

10. Morales, R.; Badesa, F.J.; García-Aracil, N.; Perez-Vidal, C.; Sabater, J.M. Distributed smart device for monitoring, control and management of electric loads in domotic environments. *Sensors (Switzerland)* **2012**, *12*, 5212–5224. [CrossRef]
11. Ray, P.P. A survey of IoT cloud platforms. *Futur. Comput. Inform. J.* **2016**, *1*, 35–46. [CrossRef]
12. Villegas-Ch, W.; Luján-Mora, S. Analysis of data mining techniques applied to LMS for personalized education. In Proceedings of the 2017 IEEE World Engineering Education Conference (EDUNINE), Santos, Brazil, 19–22 March 2017.
13. Hsinchun, C.; Roger, H.L.C.; Veda, C.S. Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Q.* **2018**, *36*, 1293–1327.
14. Kim, T.; Lim, J. Designing an Efficient Cloud Management Architecture for Sustainable Online Lifelong Education. *Sustainability* **2019**, *11*, 1523. [CrossRef]
15. Provost, F.; Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **2013**, *1*, 51–59. [CrossRef] [PubMed]
16. Ellaway, R.H.; Pusic, M.V.; Galbraith, R.M.; Cameron, T. Developing the role of big data and analytics in health professional education. *Med. Teach.* **2014**, *36*, 216–222. [CrossRef]
17. Popoola, S.L.; Atayero, A.A.; Badejo, J.A.; John, T.M.; Odukoya, J.A.; Omole, D.O. Learning analytics for smart campus: Data on academic performances of engineering undergraduates in Nigerian private university. *Data Br.* **2018**, *17*, 76–94. [CrossRef] [PubMed]
18. Cormack, A.N. See No . . . , Hear No . . . , Track No . . . : Ethics and the Intelligent Campus. *J. Inf. Rights Policy Pract.* **2019**, *41*, 1–23.
19. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518. [CrossRef]
20. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [CrossRef]
21. Batty, M. Big data, smart cities and city planning. *Dialogues Hum. Geogr.* **2013**, *3*, 274–279. [CrossRef] [PubMed]
22. Sundorph, E.; Mosseri-Marlio, W. Smart campuses: How big data will transform higher education. *Accenture* **2016**, 1–8. Available online: https://reform.uk/sites/default/files/2018-10/Smart%20campuses_WEB%20final.pdf (accessed on 20 May 2019).
23. Boiten, E. 9 Squares: Framing Data Privacy Issues Eerke. *J. Inf. Rights Policy Pract.* **2016**, 1–10. [CrossRef]
24. Turner, M.; Bailey, J.; Linkman, S.; Budgen, D.; Pearl Brereton, O.; Kitchenham, B. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2008**, *51*, 7–15.
25. Boran, A.; Bedini, I.; Matheus, C.J.; Patel-Schneider, P.F.; Keeney, J. A smart campus prototype for demonstrating the semantic integration of heterogeneous data. In Proceedings of the International Conference on Web Reasoning and Rule Systems, Berlin/Heidelberg, Germany, 29–30 August 2011; pp. 238–243.
26. Pinto, L.G.P.; Romano, R.R.; Tomoto, M.A. From the University to Smart Cities—How Engineers Can Construct Better Cities in BRIC’s Countries: A Real Case from Smart Campus FACENS. *Adv. Hum. Side Serv. Eng.* **2018**, 347–354.
27. Abdrabbah, S.B.; Ayachi, R.; Amor, N.B. Social Activities Recommendation System for Students in Smart Campus. *Smart Innov. Syst. Technol.* **2015**, *40*. [CrossRef]
28. Dong, X.; Kong, X.; Zhang, F.; Chen, Z.; Kang, J. OnCampus: A mobile platform towards a smart campus. *Springerplus* **2016**, *5*, 974. [CrossRef]
29. Luo, L. Data Acquisition and Analysis of Smart Campus Based on Wireless Sensor. *Wirel. Pers. Commun.* **2018**, *102*, 2897–2911. [CrossRef]
30. Liu, M.; Li, L. The construction of smart campus in universities and the practical innovation of student work. In Proceedings of the 2018 International Conference on Information Management & Management Science, Chengdu, China, 25–27 August 2018; pp. 154–157.
31. Zhicheng, D.; Feng, L. Evaluation of the Smart Campus Information Portal. In Proceedings of the 2018 2nd International Conference on Education and E-Learning, Bali, Indonesia, 5–7 November 2019; pp. 73–79.
32. Aion, N.; Helmandollar, L.; Wang, M.; Ng, J.W.P. Intelligent campus (iCampus) impact study. In Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), Macau, China, 4–7 December 2012; pp. 291–295.

7 Application of a Smart City Model to a Traditional University Campus

33. Yan, H.; Hu, H. A study on association algorithm of smart campus mining platform based on big data. In Proceedings of the 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 17–18 December 2016; pp. 172–175.
34. Malatji, E.M. The development of a smart campus—African universities point of view. In Proceedings of the 2017 8th International Renewable Energy Congress (IREC), Amman, Jordan, 21–23 March 2017.
35. Pompei, L.; Mattoni, B.; Bisegna, F.; Nardecchia, F.; Fichera, A.; Gagliano, A.; Pagano, A. Composite Indicators for Smart Campus: Data Analysis Method. In Proceedings of the 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Palermo, Italy, 12–15 June 2018; pp. 1–6.
36. Pibyl, O.; Opananon, S.; Horak, T. Student perception of smart campus: A case study of Czech Republic and Thailand. In Proceedings of the 2018 Smart City Symposium Prague (SCSP), Prague, Czech Republic, 24–25 May 2018; pp. 1–7.
37. Atif, Y.; Mathew, S.S.; Lakas, A. Building a smart campus to support ubiquitous learning. *J. Ambient Intell. Humaniz. Comput.* **2015**, *6*, 223–238. [[CrossRef](#)]
38. Prandi, C.; Monti, L.; Ceccarini, C.; Salomoni, P. Smart Campus: Fostering the Community Awareness Through an Intelligent Environment. *Mob. Netw. Appl.* **2019**. [[CrossRef](#)]
39. Tseng, W.S.W.; Hsu, C.W. A Smart, Caring, Interactive Chair Designed for Improving Emotional Support and Parent-Child Interactions to Promote Sustainable Relationships Between Elderly and Other Family Members. *Sustainability* **2019**, *11*, 961. [[CrossRef](#)]
40. Jethro, O.O.; Grace, A.M.; Thomas, A.K. E-Learning and Its Effects on Teaching and Learning in a Global Age. *Int. J. Acad. Res. Bus. Soc. Sci.* **2012**, *2*, 203–210.
41. Barth, M.; Burandt, S. Adding the “e-” to Learning for Sustainable Development: Challenges and Innovation. *Sustainability* **2013**, *5*, 2609–2622. [[CrossRef](#)]
42. Pugna, I.B.; Duțescu, A.; Stanila, O.G. Corporate attitudes towards Big Data and its impact on performance management: A qualitative study. *Sustainability* **2019**, *11*, 684. [[CrossRef](#)]
43. Van Hoek, R.; Johnson, M. Sustainability and energy efficiency Research implications from an academic. *Int. J. Phys. Distrib. Logist. Manag.* **2010**, *40*, 148–158. [[CrossRef](#)]
44. Lee, J.; Song, H.D.; Hong, A.J. Exploring factors, and indicators for measuring students’ sustainable engagement in e-learning. *Sustainability* **2019**, *11*, 985. [[CrossRef](#)]
45. Ma, L.; Zhai, Y.; Wu, T. Operating Charging Infrastructure in China to Achieve Sustainable Transportation: The Choice between Company-Owned and Franchised Structures. *Sustainability* **2019**, *11*, 1549. [[CrossRef](#)]
46. Bascopé, M.; Perasso, P.; Reiss, K. Systematic review of education for sustainable development at an early stage: Cornerstones and pedagogical approaches for teacher professional development. *Sustainability* **2019**, *11*, 719. [[CrossRef](#)]
47. Wright, T.S.A. Definitions and frameworks for environmental sustainability in higher education. *High. Educ. Policy* **2002**, *15*, 105–120. [[CrossRef](#)]
48. Sitepu, R.K.-K. The Impact of Modern Markets on The Performance of Micro, Small and Medium Enterprises. *J. Ekon. Bisnis* **2011**, *16*, 10–24.
49. Trindade, E.P.; Hinnig, M.P.F.; da Costa, E.M.; Marques, J.S.; Bastos, R.C.; Yigitcanlar, T. Sustainable development of smart cities: A systematic review of the literature. *J. Open Innov. Technol. Mark. Complex.* **2017**, *3*, 11. [[CrossRef](#)]
50. Nam, T.; Pardo, T.A. Conceptualizing Smart City with Dimensions of Technology, People, and Institutions. In Proceedings of the 12th Annual International Conference on Digital Government Research, DG.O 2011, College Park, MD, USA, 12–15 June 2011; pp. 282–291.
51. Angelidou, M. Smart city policies: A spatial approach. *Cities* **2014**, *41*, S3–S11. [[CrossRef](#)]
52. Cole, L. Assessing Sustainability on Canadian University Campuses: Development of a Campus Sustainability Assessment Framework. Unpublished Thesis, Royal Roads University, Victoria, BC, Canada, 2003; pp. 1–66.
53. Charles, D. Universities as key knowledge infrastructures in regional innovation systems. *Innovation* **2006**, *19*, 117–130. [[CrossRef](#)]
54. Shannon, T.; Giles-Corti, B.; Pikora, T.; Bulsara, M.; Shilton, T.; Bull, F. Active commuting in a university setting: Assessing commuting habits and potential for modal change. *Transp. Policy* **2006**, *13*, 240–253. [[CrossRef](#)]

55. Cook, D.J.; Augusto, J.C.; Jakkula, V.R. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive Mob. Comput.* **2009**, *5*, 277–298. [CrossRef]
56. Pagliaro, F.; Mattoni, B.; Gugliermi, F.; Bisegna, F.; Azzaro, B.; Tomei, F.; Catucci, S. A roadmap toward the development of Sapienza Smart Campus. In Proceedings of the International Conference on Environment and Electrical Engineering, Florence, Italy, 7–10 June 2016; pp. 1–6.
57. Alvarez-Campana, M.; López, G.; Vázquez, E.; Villagrà, V.A.; Berrocal, J. Smart CEI moncloa: An iot-based platform for people flow and environmental monitoring on a Smart University Campus. *Sensors* **2017**, *17*, 2856. [CrossRef] [PubMed]
58. Montori, F.; Bedogni, L.; Bononi, L. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. *IEEE Internet Things J.* **2018**, *5*, 592–605. [CrossRef]
59. Talari, S.; Shafie-Khah, M.; Siano, P.; Loia, V.; Tommasetti, A.; Catalão, J.P.S. A review of smart cities based on the internet of things concept. *Energies* **2017**, *10*, 421. [CrossRef]
60. Tan, L.; Wang, N. Future Internet: The Internet of Things. *ICACTE 2010—2010 3rd Int. Conf. Adv. Comput. Theory Eng. Proc.* **2010**, *5*, 376–380.
61. Kamilaris, A.; Pitsillides, A.; Prenafeta-Bold, F.X.; Ali, M.I. A Web of Things based eco-system for urban computing—Towards smarter cities. In Proceedings of the 2017 24th International Conference on Telecommunications (ICT), Limassol, Cyprus, 3–5 May 2017.
62. Li, Y.; Dai, W.; Member, S.; Ming, Z.; Qiu, M.; Member, S. Privacy Protection for Preventing Data Over-Collection in Smart City. *IEEE Trans. Computers* **2016**, *65*, 1339–1350. [CrossRef]
63. Chen, Y.S.; Tsai, Y.T. A mobility management using follow-me cloud-cloudlet in fog-computing-based RANs for smart cities. *Sensors* **2018**, *18*, 489. [CrossRef]
64. Jadeja, Y.; Modi, K. Cloud computing—Concepts, architecture and challenges. In Proceedings of the 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), Kumaracoil, India, 21–22 March 2012; pp. 877–880.
65. Weber, R.H. Internet of things: Privacy issues revisited. *Comput. Law Secur. Rev. Int. J. Technol. Law Pract.* **2015**, *31*, 618–627. [CrossRef]
66. White, M. Protection by Judicial Oversight, or an Oversight in Protection? *J. Inf. Rights Policy Pract.* **2016**, 1–42. Available online: https://www.google.com.tw/url?sa=t&rcrt=j&q=&esrc=s&source=web&cd=2&ved=2ahUKewjNp9WwoaniAhWRdHAKHXOLCu0QFjABegQIARAC&url=https%3A%2F%2Fjirpp.winchesteruniversitypress.org%2Farticles%2F10%2Fgallery%2F10%2Fdownload%2F&usg=AOvVaw1PuaR_LjxtPwARKuyX5sw (accessed on 20 May 2019).
67. Buenaño-Fernandez, D.; Villegas-CH, W.; Luján-Mora, S. The use of tools of data mining to decision making in engineering education—A systematic mapping study. *Comput. Appl. Eng. Educ.* **2019**, *10*, 4040. [CrossRef]
68. Palacios-Pacheco, X.; Villegas-Ch, W.; Luján-Mora, S. Application of Data Mining for the Detection of Variables that Cause University Desertion. *Commun. Comput. Inf. Sci.* **2019**, *895*, 510–520.
69. Villegas-Ch, W.; Lujan-Mora, S.; Buenano-Fernandez, D. Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining. In Proceedings of the IEEE World Engineering Education Conference, Santa Cruz de Tenerife, Spain, 17–20 April 2018; pp. 1–5.
70. Silva, B.N.; Khan, M.; Jung, C.; Seo, J.; Muhammad, D.; Han, J.; Yoon, Y.; Han, K. Urban planning and smart city decision management empowered by real-time data processing using big data analytics. *Sensors* **2018**, *18*, 2994. [CrossRef]
71. Debortoli, S.; Müller, O.; Vom Brocke, J. Comparing business intelligence and big data skills: A text mining study using job advertisements. *Bus. Inf. Syst. Eng.* **2014**, *6*, 289–300. [CrossRef]
72. Khalifa, S.; Elshater, Y.; Sundaravarathan, K.; Bhat, A. The Six Pillars for Building Big Data Analytics Ecosystems. *ACM Comput. Surv.* **2016**, *49*, 1–36. [CrossRef]
73. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop Distributed File System. In Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010; Volume 26, pp. 1–10.
74. Patel, A.B.; Birla, M.; Nair, U. Addressing big data problem using Hadoop and Map Reduce. In Proceedings of the 2012 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, 6–8 December 2012; pp. 6–8.

75. Blanas, S.; Patel, J.M.; Ercegovac, V.; Rao, J.; Shekita, E.J.; Tian, Y. A Comparison of Join Algorithms for Log Processing in MapReduce. In Proceedings of the SIGMOD'10, Indianapolis, IN, USA, 6–11 June 2010; pp. 1–12.
76. Shan, Y.; Yan, J.; Wang, Y.; Xu, N. FPMR: MapReduce Framework on FPGA A Case Study of RankBoost Acceleration. In Proceedings of the 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays, Monterey, CA, USA, 21–23 February 2010; pp. 93–102.
77. Wang, J.; Crawl, D.; Altintas, I. Kepler + Hadoop: A General Architecture Facilitating Data-Intensive Applications in Scientific Workflow Systems. In Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science, Portland, OR, USA, 16 November 2009; pp. 1–8.
78. Dean, J.; Ghemawat, S. MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM* **2008**, *51*, 107–113. [[CrossRef](#)]
79. Dai, W.; Ji, W. A mapreduce implementation of C4.5 decision tree algorithm. *Int. J. Database Theory Appl.* **2014**, *7*, 49–60. [[CrossRef](#)]
80. Hammoud, S.; Li, M.; Liu, Y.; Alham, N.K.; Liu, Z. MRSim: A discrete event based MapReduce simulator. In Proceedings of the 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, Yantai, China, 10–12 August 2010; Volume 6, pp. 2993–2997.
81. Chauhan, A. Master Slave Architecture in Hadoop. Available online: <https://blogs.msdn.microsoft.com/avkashchauhan/2012/02/24/master-slave-architecture-in-hadoop/> (accessed on 5 February 2019).
82. Cohen, J.; Acharya, S. Towards a more secure Apache Hadoop HDFS infrastructure: Anatomy of a targeted advanced persistent threat against HDFS and analysis of trusted computing based countermeasures. In *Network and System Security. NSS 2013*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7873, pp. 735–741.
83. Toutouh, J.; Arellano, J.; Alba, E. BiPred: A Bilevel Evolutionary Algorithm for Prediction in Smart Mobility. *Sensors* **2018**, *18*, 4123. [[CrossRef](#)]
84. Kamilaris, A.; Pitsillides, A. The impact of remote sensing on the everyday lives of mobile users in urban areas. In Proceedings of the 2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU), Singapore, 6–8 January 2014; pp. 153–158.
85. Fathema, N.; Shannon, D.; Ross, M. Expanding The Technology Acceptance Model (TAM) to Examine Faculty Use of Learning Management Systems (LMSs) In Higher Education Institutions. *J. Online Learn. Teach.* **2015**, *11*, 210–232.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Universidad de Alicante

8 Application of a Big Data Framework for Data Monitoring on a Smart Campus

Villegas-Ch, W., Molina-Enriquez, J., Chicaiza-Tamayo, C., Ortiz-Garcés, I., y Luján-Mora, S. (2019). Application of a Big Data Framework for Data Monitoring on a Smart Campus. *Sustainability*, 11, 20, 5552. (Villegas-Ch, Molina-Enriquez, et ál., 2019)

Disponible en:

URL: <https://www.mdpi.com/2071-1050/11/20/5552>




DOI: <https://doi.org/10.3390/su11205552>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O2. Identificar los componentes y las tecnologías que son parte de un campus inteligente.
- O3. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Article

Application of a Big Data Framework for Data Monitoring on a Smart Campus

William Villegas-Ch ^{1,*}, Jhoann Molina-Enriquez ¹, Carlos Chicaiza-Tamayo ¹,
Iván Ortiz-Garcés ¹ and Sergio Luján-Mora ²

¹ Escuela de Ingeniería en Tecnologías de la Información, FICA, Universidad de Las Américas, 170125 Quito, Ecuador; jhoann.molina@udla.edu.ec (J.M.-E.); carlos.chicaiza@udla.edu.ec (C.C.-T.); ivan.ortiz@udla.edu.ec (I.O.-G.)

² Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, 03690 Alicante, Spain; sergio.lujan@ua.es

* Correspondence: william.villegas@udla.edu.ec; Tel.: +593-098-136-4068

Received: 27 August 2019; Accepted: 25 September 2019; Published: 9 October 2019



Abstract: At present, university campuses integrate technologies such as the internet of things, cloud computing, and big data, among others, which provide support to the campus to improve their resource management processes and learning models. Integrating these technologies into a centralized environment allows for the creation of a controlled environment and, subsequently, an intelligent environment. These environments are ideal for generating new management methods that can solve problems of global interest, such as resource consumption. The integration of new technologies also allows for the focusing of its efforts on improving the quality of life of its inhabitants. However, the comfort and benefits of technology must be developed in a sustainable environment where there is harmony between people and nature. For this, it is necessary to improve the energy consumption of the smart campus, which is possible by constantly monitoring and analyzing the data to detect any anomaly in the system. This work integrates a big data framework capable of analyzing the data, regardless of its format, providing effective and efficient responses to each process. The method developed is generic, which allows for its application to be adequate in addressing the needs of any smart campus.

Keywords: smart campus; big data; Hadoop

1. Introduction

At present, university campuses recognize the integration of technologies as a path that leads them to digital transformation. Many of these campuses even work to consolidate information and communication technologies (ICT) on this basis, leading them to be considered as intelligent campuses [1]. Creating a comfortable, sustainable, and safe environment is characteristic of a smart campus, which efficiently contributes to the development of learning and the administrative management of the campus [2]. To meet these characteristics, ICTs include data acquisition systems, cloud computing, and data analysis on smart campuses [3]. The integration of these technologies allows for the improvement of the processes of smart campuses and their ability to respond to problems found in this type of environment. One of the problems that a smart campus goes through that has more relevance from an environmental and economic approach is energy consumption. Generally, a smart campus does not have established policies and applications that allow good management of energy resources. This is largely due to the high costs involved in the acquisition of technologies with the ability to integrate all equipment into a single monitoring system. The lack of a system that alerts about alterations in energy consumption results in that, first, there is an incident or damage to a device,

then corrective measures are taken, which generates a greater expenditure of resources on the smart campus. An adequate response to this problem is the implementation of a big data framework to monitor the conditions of each team. However, the framework must include versatility and scalability to allow adaptation for any campus, regardless of the infrastructure that it has.

Previous studies have focused on improving learning conditions through the detection of trends in students [4]. These trends depend on systems capable of detecting any type of event on campus. Such events include sensor and actuator data, safety devices, wireless devices, or learning management systems (LMS) [5]. The framework monitors and analyses of data provide information that helps decision-making. This work responds to the need from an administrative point of view in situations where energy management acquires greater value due to the importance of its good use [6]. It is common on university campuses to observe the misuse of energy resources; this causes economic and technical problems in these environments. In a smart campus, this perspective changes dramatically because having a diversity of systems that continuously monitor each area of the campus is feasible in the detection of any anomaly and thus allows for the ability to take action on it [7]. However, technically this process is more complicated than it seems, as the variables behave dynamically, including a substantial increase in the volume of data [8]. In addition, the results must be validated on the basis of strict standards established by the manufacturers for each of the pieces of equipment immersed in the energy issue.

This research answers questions such as, “is there enough information within a smart campus to detect the energy consumption of the equipment that is part of the data center?” Moreover, we asked, “does the big data framework provide valid answers to smart campus energy consumption?” For this, the work provides a description of the technological concepts and the main characteristics that serve as a preamble to the development of the analysis. In addition, a brief analysis of the existing research in similar environments is carried out. A detailed description of the existing big data frameworks is made, focusing on which and why it is the most used in the proposed environment [9]. The development of this work is done with the use of open-source tools on the basis of the method and architecture of the cluster. The use of these types of tools makes the method generic by enabling its implementation on any type of campus [10]. In addition, the method is validated and refined in the treatment and monitoring of data that identify the voltage variation of the smart campus data center.

The rest of this paper is structured as follows: In Section 2 the concepts used for the development of the method are presented; Section 3 describes the analysis of the previous research that has contributed to the development of this proposal; Section 4 includes a method for establishing the different phases to be considered when implementing a big data platform; Section 5 presents the results applied to the monitoring and analysis of data; Section 6 presents the discussion and Section 7 presents the conclusions.

2. Preliminary Concepts

2.1. Big Data

The concept of big data refers to the analysis of large volumes of data stored in various sources. The data are not necessarily structured or in any specific format. The objective of big data are to meet the needs not previously met by existing technologies, such as business intelligence platforms or statistical data analysis [11]. When this technology is used in educational environments, it has an impact on academic management. It helps in the use and management of resources by providing knowledge about the data stored in these environments. This has an impact on real-time decisions, which contributes to successful management [12]. To comply with this process, big data techniques involve the storage and processing of data with specific characteristics, such as

- The content format.
- The type of data.
- The frequency with which the data is made available.

- The intention: how the data should be processed (for example, an ad hoc query on the data).
- Volume: the size of the data that can come from multiple sources.
- Velocity: the speed with which data arrive using units such as tera, peta, or exabytes.
- Variety: structured, semi-structured, and unstructured.

2.2. The Smart Campus

Technological advances have allowed societies to talk about concepts such as smart cities that can modify people's futures. These technologies support the resolution of problems related to the management of environmental resources, the reduction of energy consumption, mobility, and waste management, among others. [4]. When technology is applied in a university environment, the resulting smart campuses are an illustration of what smart cities could be like. University campuses are places where thousands of people study or work every day [2]. The use of ICT in smart campuses improves the quality of life of its inhabitants and enhances the coexistence between the university population and their surroundings, properly managing resources within the campus, and providing favorable environments for learning [13].

2.3. The Internet of Things (IoT)

The Internet has gone from being a network of computers and servers to including a variety of devices that interact with each other and with users [14]. The devices have the ability to generate data and transfer it through a network automatically without the indispensable interaction of people or computers [15].

All the devices that make up the IoT capture data from the real environment and send it to be processed and provide a better user experience. The main objective of the IoT is the digitalization of the physical world so that all traditional devices are connected to the network and synchronized with each other provide a better and efficient service to the user [16].

3. Previous Research

Previous research has allowed us to identify different works aligned with the use of big data in controlled environments, focusing specifically on the use of Hadoop as an analysis architecture [17,18]. The research carried out is divided into two groups, and the first provides a perspective on the application of big data in Hadoop architecture [19,20]. Much of this architecture focuses on processes that do not necessarily look for results in real time; however, it needs high availability and great processing capacity [21]. The guidelines define the correct architecture according to the requirements of this study.

The second group of research was carried out in university environments and focuses on meeting the needs of students [22]. These needs range from improving administrative processes to analyzing the data of the inhabitants [23]. In this group, there was little research into the architecture or integral analysis where all the data generated in a campus are taken advantage of through a process of big data analysis [20] to help in efficiently managing processes and resources [24].

The review of existing research highlights that most of the work solves problems separately, for example, the acquisition of data by IoT, the management of cloud computing, and the use of big data, among others. However, few works have addressed the integration of technologies for the creation of intelligent environments. The integration of technologies allows converting traditional environments into intelligent ones, thus contributing to the implementation of an intelligent campus [25]. In previous work, the authors propose an intelligent campus model aligned to meet the needs of all residents through data acquisition, storage, and analysis through emerging technologies. The work presented in this paper takes as reference the proposed model and generates a big data framework that adds to previous works in the following ways:

- The integration of the equipment in the model so that the framework can access the data they generate.
- The capacity of big data nodes being improved to ensure the processing and analysis of variables.
- In previous works, the framework with data was considered critical in the administration of campus infrastructure. The inclusion of the data center team in the big data framework offers the possibility of improving the use of each of these resources by supporting the management of an intelligent campus and its sustainability.

4. Method

The objective of the smart campus is to manage the technological, human, and environmental resources properly so that the activities of the inhabitants are balanced with the campus components. To achieve this goal, it is important to analyze the factors that make a campus smart and how they interact with each other to improve the processes and management.

In previous works [13,17], the authors defined in detail how to build all the necessary infrastructure to convert a traditional campus into an intelligent campus. This work provides the basis and support to complement the big data framework, focusing on more technical problems, such as the monitoring of voltages in data center equipment. The previous works analyzed the data generated by several systems, such as the mobility system through the wireless system and the vending machine system [5]. This analysis helped to detect existing trends in the campus population by grouping it in specific periods. However, the model proposed in [17] was put to the test with the data generated by the equipment that were part of the data center by monitoring more technical variables such as the variation of voltages. These data, unlike those analyzed above, are of high flow, since the collection is constantly carried out, which compromises factors such as processing and storage capacity. In response, the authors have seen the need to adjust the big data framework to perform a much more technical work, including nodes with greater capacity and segmenting tasks to guarantee results when there are simultaneous processes [9].

The proposed framework is based on the model proposed in [17], which allows defining what is the best option for the integration of technologies and how they can improve administrative and academic management in these environments. Figure 1 defines the architecture of the intelligent campus; this architecture considers all the phases and equipment that contribute to the development of the framework proposed in this study.

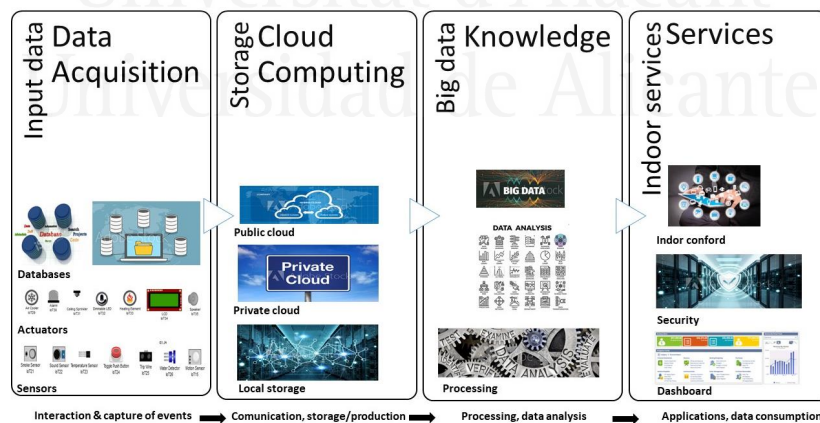


Figure 1. Architecture of a smart campus, composed of four stages, each stage containing a representation of the key elements and artefacts that are used.

4.1. Tools for Developing a Big Data Framework

For the development of the method, it is important to define the big data tool to use, as this constitutes the basis for the development of the framework. Currently, two frameworks are used as big data ecosystems; these are the Hadoop ecosystem and Apache Spark. For the application of a specific framework, it is important to have an understanding of the ecosystem. Hadoop is a framework for storing large datasets using computer clusters. It can scale from a single computer system to thousands of systems, and it provides local storage and computing power [26]. Apache Spark was designed for quick calculations, and its main feature is that it processes everything in its own memory, which increases the processing speed. Hadoop is integrated with MapReduce processing and uses persistent storage, whereas Apache Spark has a resilient distributed dataset known as resilient distributed datasets (RDD). Apache Spark performance is fast, processed in memory, and offers real-time analysis. Hadoop was originally designed to continuously collect data from various sources without hitting problems caused by the type of data used in batch processing; thus, it was never built for real-time processing.

Hadoop does not have an interactive mode, but it has add-ons such as Hive and Pig that make working with MapReduce easier [27]. Hadoop and Spark are projects hosted by Apache, meaning that they are open-source and do not have licensing costs. Processing can be divided into two types, batch-processing and processing by the flow. Hadoop uses a batch-processing framework—it sends a work, reads the data, performs the operation, writes the results, and sends them to the cluster successively. Spark encompasses iterative batch application algorithms, interactive queries, and streaming. MapReduce uses an effective method for fault tolerance by using TaskTracker that issues reports to JobTracker. If this is lost in the report, the JobTracker reprograms all the operations. Spark uses RDD, which is a set of elements that tolerates faults and operates in parallel.

Hadoop has multiple ways to provide security. Kerberos supports and is compatible with other providers such as lightweight directory access protocol (LDAP), and offers encryption with Hadoop distributed file system (HDFS). Spark's security is partly in the authentication process, which is one reason why it needs HDFS to be executed as it uses, accesses, and gives permission to files.

4.2. Data Acquisition

Within a smart campus, new technologies are integrated so that the IoT acts as a connectivity system and includes a multitude of devices, such as sensors and embedded systems, as it carries out the acquisition of data [28]. These devices allow data to be collected, which affects the optimization processes [5]. The processed information allows knowledge to emerge and control the environment, extracting patterns of behavior or information relevant to decision-making. The data generated by the devices on the campus network is stored in the data center and processed and analyzed in search of knowledge [29]. The way in which sensors interact with the environment and people contributes to creating a knowledge society. Sensors gather information from the environment and send it to the cloud, where interested parties can consume it. Information that is converted into knowledge helps decisions to be made quickly and accurately. The technology proposed by Kamilaris [30] and Pitsillides [31] gives greater validity to this work, since the implementation of an ecosystem on the basis of the IoT ensures that students and teachers make sound decisions in relation to their environment. In environments that monitor energy consumption in a data center, it is essential to have a data analysis platform that constantly monitors any variations in voltage and in the equipment, which could prolong its useful life.

4.3. Cloud Computing

University campuses generally have their own infrastructure for data storage and management. They tend to be composed of a data center and communication equipment, which are responsible for unifying the different areas within the data center. A data center, where the information is concentrated and centralized, is advantageous for the deployment of the framework and guarantees the availability

and quality of the data [32]. The data stored comes from the different systems within the university, in addition to the IoT devices deployed throughout the campus. The services offered by the university to its residents are managed through a variety of virtual servers hosted in the different server blades. In addition, some services, such as the Office 365 mail service, are hosted in a public cloud.

In a smart campus, it is not advisable to migrate all services to a public cloud because of issues of control and information security [33]. For this reason, the data on the campus that are part of this study are stored in a private cloud generated from the information available in the campus infrastructure [34]. The advantages of this are data security, quality, availability, and flexibility.

4.4. Knowledge

In order to obtain knowledge on a specific topic, it is necessary to analyze the data generated by each of the systems. In many cases, in order to determine the knowledge about an object, it will be necessary to interrelate all the data available on the campus; this implies a great technical and physical processing capacity. Our previous work [17] specifies the use of Hadoop as a framework used for data processing and analysis [35]. The knowledge generated at this stage allows for making decisions that contribute to improving processes, as well as implementing systems that act directly on the results and interact with the environments automatically.

4.5. Services

Within the smart campus, technological integration with the framework provides several services that contribute to the operation of the campus [36]. These services are classified into two groups: the first group is responsible for everything related to administrative management and the second group is responsible for academic management. The framework contributes to administrative management in environments such as human resources, where it provides more efficient management in organizational processes. In addition, it can be used to develop tests for the evaluation and selection of candidates or the creation of resources aligned to the development of transversal competences [37]. In the energy environment, monitoring and analysis of voltage variation allow the creation of predictive maintenance programs to improve electricity consumption in a safer, more economical and sustainable way [38].

In academic management, the framework offers new information about students and teachers to the areas of marketing and admissions in order to make strategic decisions in the educational model [39]. In the learning environment, it applies its own data analysis techniques to optimize educational management, learning, and attention to the student, giving support to personalized education [40]. The framework is able to offer alert services and generate recommendations, among other benefits [23].

In this work, the framework is applied in an administrative management environment where the energy environment is included. The problem to be overcome is to improve the life and performance of the electrical components of certain equipment in the data center of the smart campus [10]. In addition, it seeks to detect anomalies in the power supply that receive this equipment. Once an anomaly is detected, it is possible to generate early alerts and preventive maintenance that guarantee the proper functioning of the data center [41]. To respond to the problem, the work involves monitoring the electrical consumption of the equipment and analyzing the data in search of any type of harmful event related to this process. In this environment, equipment seen to be critical in the operation of the smart campus data center are considered. The devices are the communication switches responsible for enabling the connection between the data infrastructure of the smart campus and the data center [8]. The other piece of equipment is the uninterruptible power supply (UPS)—this equipment is the main energy backup in case of lack or failure of the public electricity supply. In the following subsections, the monitoring and analysis of energy consumption data in data center equipment are presented.

Analysis of the Energetic Consumption of the Data Center

To evaluate the big data framework, special monitoring was carried out to gauge the energy consumption of the data center equipment of the campus participating in this study. The monitored equipment were the Nexus 3500, the UPS, and the American power conversion (APC) model AP9215RM. The parameters monitored and integrated into the Hadoop environment for analysis are as follows:

- Speeds in each interface (Nexus).
- Administration labels (Nexus—UPS).
- Input voltages (Nexus—UPS).
- Output voltages (Nexus—UPS).
- Battery temperature (UPS).
- Output current (UPS).
- Output and input frequency (UPS).
- Battery capacity (UPS).

The implemented big data framework was prepared to process huge amounts of information. To calculate the volume of information to be processed and loaded, Equation (1) was applied.

The variables established for the data acquisition configuration were

- n = data sampling period.
- y = amount of monitored items (Tags).
- t = days/months of monitoring (Time).
- r = projection.
- p = size.
- s = seconds per minute.
- m = minutes per hour.
- h = hours per day.

The analysis considers the data generated in a period of 3 days, and according to this clarification the following values were obtained:

- $n = 30''$.
- $y = 11$ (UPS) 102 (Nexus) (total of 113 items monitored).
- $t = 3$ days \rightarrow 72 h.
- $x = ?$ (number of records).

$$x = \left\{ \left(\frac{n \times 2}{s} \right) * (m) \right\} \times y \times t. \quad (1)$$

$X = 1.057.680$ (Total records generated in 72 h).

To obtain the projection of records in a period of 6 months, the following were considered:

- N = quantity of months.
- H = hours per day.
- D = days of the month.
- Y = number of records.
- $r = [(H \times D) \times N] \times Y$.
- $r = [(24 \times 30) \times 6] \times 1,057,680$.
- $r = 761,529,600$ (total records generated in 6 months).

Once the number of records was identified, it was then important to know their weight in megabytes and gigabytes to ascertain the volume of information being processed by the big data framework. For this calculation, it was first necessary to know the fields consulted the type of data and the size by data type. Table 1 lists the names of the fields selected for monitoring and includes their data type and size in bytes. The values specified in the table were considered in order to find the size of the query per record, all based on the size of each field corresponding to the item generated in the monitoring.

Table 1. Size of the fields for monitoring.

Name	Datatype	Value/Size
Host	Bigint (4)	8 bytes
Name	Varchar (125,6)	125 bytes
itemName	Varchar (125,6)	125 bytes
Data sampling period	Bigint (4)	8 bytes
Description	Varchar (125,6)	125 bytes
LifeTimeData	Bigint (4)	8 bytes
Date	Bigint (12)	8 bytes
Values	Bigint (4)	8 bytes

The variables and values were

- v = weight in bytes by data type.
- n = number of records.
- m = type of measurement (megabyte MB).

Established values:

- v = byte.
- n = 1,507,680.
- m = 1,048,576.
- x = weight of total registrations.

x = 418.60 MB (this result is the weight of the records currently monitored). To calculate the total size with a projection of six months the value of the records was changed as follows:

- n = 761,529,600.
- r = 301,394.256591796875 MB.
- r = 294.33 (gigabyte GB).

5. Results

In order to evaluate the developed framework, the exercise was applied in the monitoring and analysis of the data generated by the communication teams, as well as the UPS. This process tested the entire distribution of the framework from data acquisition to data analysis. The results obtained were compared with the energy consumption data of each of the equipment provided by the manufacturer. Next, the whole process generated for the analysis was detailed, wherein each of the stages and the specific variables in the electrical consumption were included.

5.1. Analysis of Voltage Variations

On the basis of an analysis of the campus infrastructure and data center equipment for six months, an equivalent of 300 GB of information was processed. An important condition to consider in the process was the monitoring of items by devices or new equipment income, as the volume of data can be rapidly increased until it reaches the surrounding terabytes or petabytes of information. It

is necessary to emphasize that the data was acquired from the network equipment and the energy control system where information is processed during a 24 h period. To perform a history of events, a regression was carried out that involved processing large amounts of information at the terabyte level. For this reason, the Hadoop framework was configured to process batch information to solve the problem of an environment that required a regression or projection analysis. For the analysis, the framework worked with the input voltage data in the Nexus 3000 family and the input and output voltages of the APC UPS and AP9215RM models. Table 2 shows the unification of the voltage values of the network equipment, computing, storage, and UPS, which are provided in the datasheets.

Table 2. Values of data center equipment. UPS: uninterruptible power supply.

Name	Voltage Input	Voltage Output	Efficiency of the Energy Supply	Source
Cisco Nexus 3124(1)	100 to 240 VAC		89 a 91% at 220 VCA	Datasheet
Cisco Nexus 3124(2)	100 to 240 VAC		89 a 91% at 220 VCA	
Cisco UCS Chasis 5108	100 a 120 VAC 200 a 240 VAC		94% to 240 VCA	
Cisco UCS B200M4	100 to 240 VAC 90 to 264 VAC		92% to 95 VCA	
EMC VNXe3200	100 to 240 V		92% to 95 VCA	
UPS—APC AP9215RM	200, 208, 220, 230, 240 Vac; 60 or 50 Hz,	200, 208, 220, 230, 240 Vac; 50 or 60 Hz,	Approximately 89%	

In the analysis, the default values detailed in the technical specifications of the equipment were established; the values had a direct line in their input voltage. Figure 2 shows the best scenario of input voltage consumption, which was 120 volts—this is the common voltage value of Ecuador, where the work was performed. This can be observed in the left section of the figure, and the real scenario that takes values in relation to those acquired in the monitoring is observed in the right hand section. This ranged from a minimum of 124 to a maximum of 128 volts, considering the average of 124.8 volts on Nexus 3124 devices.

The analysis of the Cisco Nexus 3124 equipment shows that the voltages acquired in the monitoring varied between each piece of equipment. The Nexus (a) had a higher peak voltage, at 128 volts, even so, the values were still in the range specified by the manufacturer. Therefore, on the basis of the technical data sheet, there is an energy efficiency corresponding to the range of 89% to 91% by the equipment. The results imply that the equipment used at the time of monitoring had a performance focused on sustaining the efficient use of energy and was less than or equal to one.

During the analysis, a sample was taken during a period of 6 to 10 hours of monitoring, and as a result, variations of 2 to 8 volts per minute were identified. This means that the equipment underwent a type of overload in short periods, the Nexus equipment (b) in particular. These results indicate that it is important to have constant monitoring policies that verify the values for frequency, power, memory, and processing. The objective was to identify what caused the variations in voltage, even though they were within those established by the manufacturer, as over time they can still cause irreparable damage to the equipment.

5.2. Voltage Monitoring in the Uninterruptible Power Supply System

The UPS was integrated into the monitoring to handle an approximation closer to the real load consumption of all the equipment. The equipment was connected to this alternating power supply, and, through monitoring, it is possible to verify its total voltage when it is at maximum consumption capacity. The input and output voltages were monitored and compared with the datasheet of the equipment that specified a consumption of between 200 and 240 volts, which guaranteed the efficiency of 89%. Once the data were entered into the framework, the voltages obtained ranged from 214 to 226 volts; thus, it was concluded that the equipment was in favorable conditions and contributed to a PUE close to one. Figure 3 displays the monitoring during a particular period, and the parameters monitored

correspond to the input voltage. A minimum of 214 and a maximum of 226 volts was verified in certain minutes; however, these values were within the range established by the manufacturer.

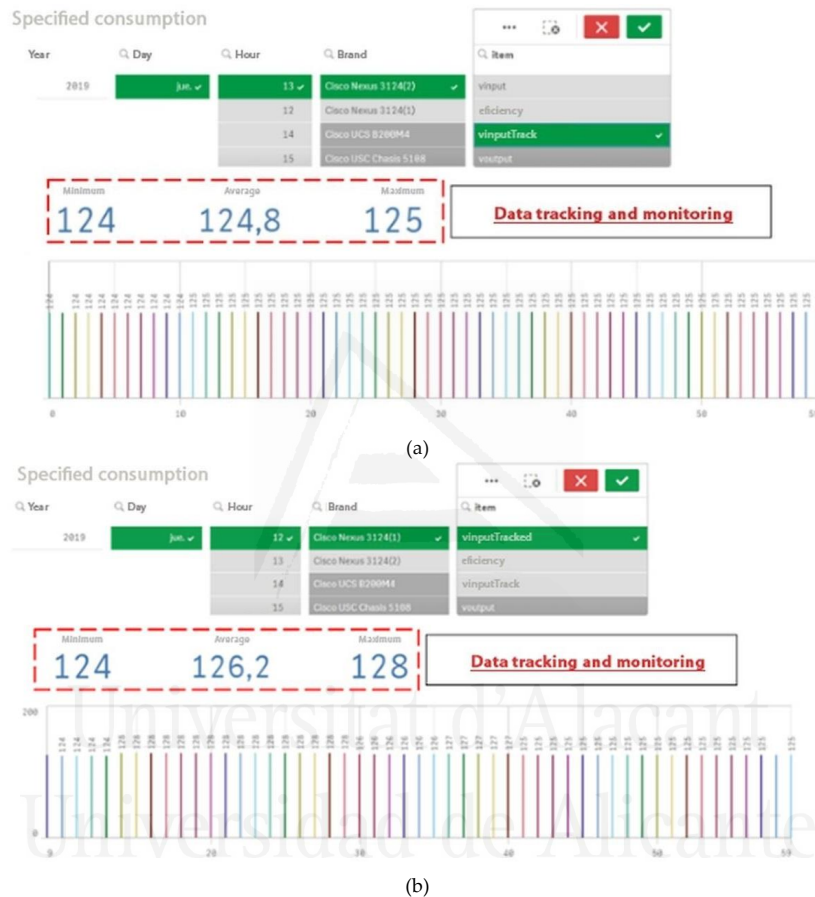


Figure 2. This figure shows a comparative analysis of the Nexus switches; the data center has two of these devices to manage redundancy. (a) variation of voltages in 12 h of monitoring; (b) variation of voltages in 13 h of monitoring.

Figure 3 shows the variations in input voltage, which fluctuated from 11 volts with an average of 220 volts, which was within the range established for the equipment according to its technical data sheet. This measure was taken during a 48 h period, where the load of the equipment was not greater than the standard values. In future, it would be beneficial to analyze the equipment when there is a power cut and when all the devices are connected to the UPS, so as to validate a higher load and determine the minimum and maximum voltages.

Figure 4 illustrates the monitoring and shows two high peaks during two time periods on a normal operating day, the first between 10:00 and 10:30 and the second between 13:45 and 14:30. The first period corresponded to the peak when additional equipment was connected to the power source in

the data center. This type of anomaly explained the higher electrical traffic and increase in voltage. It is important to monitor the equipment during peak hours to verify the minimum and maximum voltages, thus validating that the parameters oscillate in the range determined by the manufacturer.

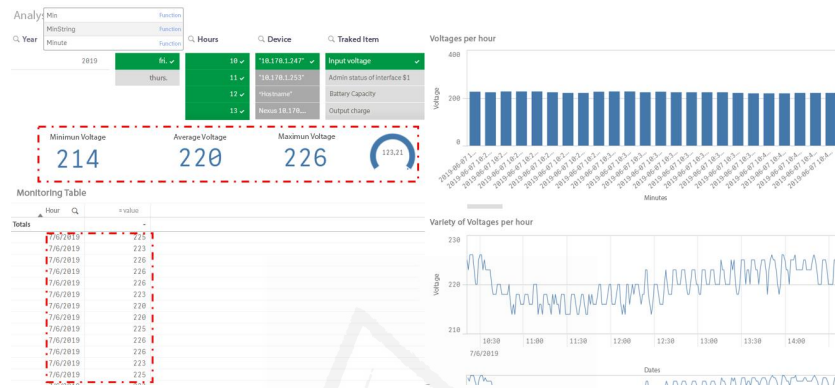


Figure 3. Input voltage values (UPS).

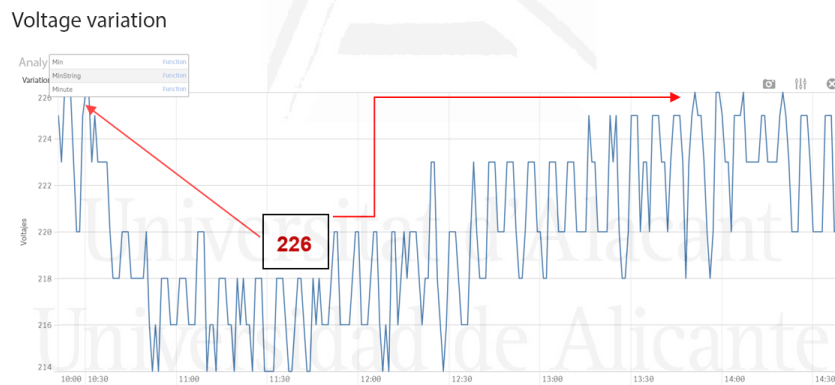


Figure 4. Input voltage spikes (UPS).

In the second period from 13:45 to 14:30, it was validated that the peak voltage was due to the machines inside the data center starting up, because as soon as the classrooms went into operation the voltage stabilized within minutes.

6. Discussion

This work integrated technologies that help improve the processes of an environment to the point of turning them into intelligent environments. Under this concept, the integration carried out in this work contributes to the management of campus resources. Its method is to be conceived in a generic way and through utilizing technologies that are mostly open-source, it is able to be replicated on any campus regardless of the infrastructure.

Given the rise of big data and technologies such as cloud computing, the creation of efficient data centers committed to the environment and energy-saving has become a first-order need. This is not

only the case environmentally but also economically, as a smart campus invests a large amount of its budget in energy jacks and equipment maintenance.

The big data framework uses Hadoop, which provides the most reliable storage layer on the market using HDFS. It has a batch-processing engine, such as MapReduce, and a yet another resource negotiator (YARN) resource management layer that makes it a scalable and reliable tool. In addition, Hadoop is an open-source tool and can be modified according to the requirements of the smart campus. By default, Hadoop stores three replicas of each block in the cluster and can be changed as needed. Then, if any node falls, the data from that node can be easily retrieved from other nodes. Nodes or task failures are automatically recovered by the framework, which makes it extremely fault-tolerant. Hadoop is easy to use, without consuming excessive human resources in the administration of distributed computing; the framework takes care of everything. Hadoop runs on machines with basic hardware without problems, and, for this reason, it is not necessary to invest in specialized machines with advanced hardware. Hadoop also reduces investment by allowing for the ability to add more nodes on the fly and expand the framework according to the needs of the campus. Therefore, if the needs and resource requirements increase, the framework increases the nodes without any downtime and without requiring prior planning.

In energy consumption, the big data framework manages to introduce predictive maintenance in the infrastructure that are able to save costs and improve the quality of the service-avoiding stopping for events related to energy. Through the effective analysis of the data, it is able to calculate the amount of energy that is really required for the optimal operation of the data center. The next step will be to use the right technology in order to obtain the most efficient consumption.

7. Conclusions

Several related studies were reviewed for this research, with some dealing with similar issues but with very few proposing a framework that is applied to a smart campus. Therefore, this research contributes to scientific research that validates data analysis using big data in applying it to larger environments such as smart cities.

The integration of technologies such as IoT, cloud computing, and big data provides the basis of emerging technologies that helps to generate new management methods or improve existing ones. The IoT within the smart campus replaces the human in many activities, such as to monitor or maintain a certain environment. Cloud computing improves the availability of data and provides greater speed when it is exploited. In a smart campus, this technology is integrated into traditional storage models where computer systems store the data generated in relational databases. This integration allows managing a large amount of data, and, if an analysis is dependent on the information of several systems, the availability to use them is accessible.

The data analysis awarded to the Hadoop framework guaranteed the results obtained in the energy consumption analysis of the campus data center. The results were validated with the parameters recommended by the manufacturer, validating the veracity of the analysis. This helps to make decisions about the use of the equipment, determine the life of the UPS batteries, and improve the management of energy consumption.

Achieving the effectiveness of the results depends on the architecture of the Hadoop framework, which stands out from others such as Spark, which is another architecture currently used by large companies. The comparison between these two frameworks focused on the details and characteristics of each architecture and in which environments they obtain the best results.

The applied framework guarantees high availability and scalability, as during its application to energy monitoring, several types of equipment were added and removed respectively to check the measurements. These factors allow the framework to adjust to the needs of any environment and of agile responses to the needs that each of these presents. In addition, the versatility of the framework allows for the addressing of needs such as learning management, administrative processes, mobility, security, and others.

As a future study, the authors are integrating artificial intelligence (AI) with the smart campus. The integration of AI with technologies used in this work (IoT, cloud computing, and big data) is imminent. AI allows us to take the results of the data analysis and learn from them to personalize the learning. In addition, it will allow for the execution of autonomous processes that, up until now, have been done by people, therein combining production and efficiency. AI can resolve issues such as energy efficiency, climate alerts, internal mobilization, and waste management, which will be treated properly and efficiently. In departments such as marketing and admissions, a Chabot will be implemented with the ability to learn from natural language, which leads it to give information on the smart campus in the same way as a person does.

Author Contributions: W.V.-C. contributed to the following: the conception and design of the study, acquisition of data, analysis, and interpretation of data, drafting the article and approval of the submitted version. The authors J.M.-E., C.C.-T., and I.O.-G., contributed to the study in the design, conception, and critical review. S.L.-M. He made the following contributions to the study: analysis and interpretation of data, approval of the version presented. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Boran, A.; Bedini, I.; Matheus, C.J.; Patel-Schneider, P.F.; Keeney, J. A smart campus prototype for demonstrating the semantic integration of heterogeneous data. In *International Conference on Web Reasoning and Rule Systems*; Springer: Berlin, Heidelberg, 2011; Volume 6902, pp. 238–243.
- Aion, N.; Helmandollar, L.; Wang, M.; Ng, J.W.P. Intelligent campus (iCampus) impact study. In *Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Macau, China, 4–7 December 2012; pp. 291–295.
- Nie, X. Constructing Smart Campus Based on the Cloud Computing Platform and the Internet of Things. In *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*; Luo, X., Ed.; Atlantis Press: Paris, Francia, 2013; pp. 1576–1578.
- Sundorph, E.; Mosseri-marlio, W. Smart Campuses: How Big Data will Transform Higher Education. Available online: <http://www.reform.uk/wp-content/uploads/2016/09/Smart-campusesWEB.pdf> (accessed on 29 September 2019).
- Luo, L. Data Acquisition and Analysis of Smart Campus Based on Wireless Sensor. *Wirel. Pers. Commun.* **2018**, *102*, 2897–2911. [[CrossRef](#)]
- Popoola, S.I.; Atayero, A.A.; Okanlawon, T.T.; Omopariola, B.I.; Takpor, O.A. Smart campus: Data on energy consumption in an ICT-driven university. *Data Br.* **2018**, *16*, 780–793. [[CrossRef](#)]
- Lazaroiu, G.C.; Dumbrava, V.; Costoiu, M.; Teliceanu, M.; Roscia, M. Smart campus-an energy integrated approach. In *Proceedings of the 2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, Palermo, Italy, 22–25 November 2015.
- Zhou, K.; Fu, C.; Yang, S. Big data driven smart energy management: From big data to big insights. *Renew. Sustain. Energy Rev.* **2016**, *56*, 215–225. [[CrossRef](#)]
- Yang, C.T.; Chen, S.T.; Liu, J.C.; Liu, R.H.; Chang, C.L. On construction of an energy monitoring service using big data technology for the smart campus. *Cluster Comput.* **2019**, 1–24. [[CrossRef](#)]
- Barbato, A.; Bolchini, C.; Geronazzo, A.; Quintarelli, E.; Palamarciuc, A.; Piti, A.; Rottondi, C.; Verticale, G. Energy optimization and management of demand response interactions in a smart campus. *Energies* **2016**, *9*, 398. [[CrossRef](#)]
- Villegas-Ch, W.; Luján-Mora, S.; Buenaño-Fernandez, D.; Palacios-Pacheco, X. Big data, the next step in the evolution of educational data analysis. *Adv. Intell. Syst. Comput.* **2018**, *721*, 138–147.
- Molinari, A.; Maltese, V.; Vaccari, L.; Almi, A.; Bassi, E. Big Data and Open Data for a Smart City. In *Proceedings of the IEEE-TN Smart Cities White Papers*, Trento, Italy, 10–11 December 2014; pp. 1–8.
- Liu, M.; Li, L. The construction of smart campus in universities and the practical innovation of student work. In *Proceedings of the International Conference on Information Management & Management Science*, Chengdu, China, 24–26 August 2018; pp. 154–157.

14. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [[CrossRef](#)]
15. Britt, J.; Matsumura, S.; Forood, H.; Zimmerman, S.; Myles, P.; Zawicki, S.; Kutami, D. Internet of Things Platforms, Apparatuses, and Methods. U.S. Patent No. 9,497,572, 15 November 2016.
16. Sotres, P.; Lanza, J.; Sánchez, L.; Santana, J.R.; López, C.; Muñoz, L. Breaking Vendors and City Locks through a Semantic-enabled Global Interoperable Internet-of-Things System: A Smart Parking Case. *Sensors* **2019**, *19*, 229. [[CrossRef](#)]
17. Villegas-Ch, W.; Palacios-Pacheco, X.; Luján-Mora, S. Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus. *Sustainability* **2019**, *11*, 2857. [[CrossRef](#)]
18. Shvachko, K.; Kuang, H.; Radia, S.; Chansler, R. The Hadoop Distributed File System. In Proceedings of the 26th Symposium on Mass Storage Systems and Technologies (MSST), Incline Village, NV, USA, 3–7 May 2010.
19. Provost, F.; Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **2013**, *1*, 51–59. [[CrossRef](#)]
20. McHugh, J.; Cuddihy, P.E.; Weisenberg Williams, J.; Aggour, K.S.; Kumar, V.S.; Mulwad, V. Integrated access to big data polystores through a knowledge-driven framework. In Proceedings of the IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; IEEE: Piscataway, NJ, USA; Volume 1, pp. 1494–1503.
21. Osman, A.M.S. A novel big data analytics framework for smart cities. *Futur. Gener. Comput. Syst.* **2019**, *91*, 620–633. [[CrossRef](#)]
22. Gairín, J.; Triado, X.M.; Feixas, M.; Figuera, P. Student dropout rates in Catalan universities: Profile and motives for disengagement. *Qual. High. Educ.* **2014**, *20*, 165–182. [[CrossRef](#)]
23. Abdrabbah, S.B.; Ayachi, R.; Amor, N.B. Social Activities Recommendation System for Students in Smart Campus. *Smart Innov. Syst. Technol.* **2015**, *76*, 461–470.
24. Braganza, A.; Brooks, L.; Nepelski, D.; Ali, M.; Moro, R. Resource management in big data initiatives: Processes and dynamic capabilities. *J. Bus. Res.* **2017**, *70*, 328–337. [[CrossRef](#)]
25. Trindade, E.P.; Hinnig, M.P.F.; da Costa, E.M.; Marques, J.S.; Bastos, R.C.; Yigitcanlar, T. Sustainable development of smart cities: A systematic review of the literature. *J. Open Innov. Technol. Mark. Complex.* **2017**, *3*, 11. [[CrossRef](#)]
26. Borthakur, D. Apache Hadoop 2.6.0 - HDFS Architecture. Available online: <http://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html> (accessed on 27 August 2019).
27. Shanahan, J.; Dai, L. Large Scale Distributed Data Science from scratch using Apache Spark 2.0. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland; pp. 955–957.
28. Li, H.; Shou, G.; Hu, Y.; Guo, Z. WiCloud: Innovative uses of network data on smart campus. In Proceedings of the 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23–25 August 2016; pp. 461–466.
29. Hannan, A.; Arshad, S.; Azam, M.A.; Loo, J.; Ahmed, S.H.; Majeed, M.F.; Shah, S.C. Disaster management system aided by named data network of things: Architecture, design, and analysis. *Sensors* **2018**, *18*, 2413. [[CrossRef](#)]
30. Kamilaris, A.; Pitsillides, A.; Prenafeta-Bold, F.X.; Ali, M.I. A Web of Things based eco-system for urban computing—towards smarter cities. In Proceedings of the 24th International Conference on Telecommunications (ICT), Limassol, Cyprus, 3–5 May 2017.
31. Kamilaris, A.; Pitsillides, A. The impact of remote sensing on the everyday lives of mobile users in urban areas. In Proceedings of the Seventh International Conference on Mobile Computing and Ubiquitous Networking (ICMU), Singapore, 6–8 January 2014; pp. 153–158.
32. Uskov, V.L.; Bakken, J.P.; Pandey, A. Smart University Taxonomy: Features, Components, Systems. *Smart Educ. e-Learn.* **2016**, *59*, 3–14.
33. Ray, P.P. A survey of IoT cloud platforms. *Futur. Comput. Inform. J.* **2016**, *1*, 35–46. [[CrossRef](#)]
34. Jadeja, Y.; Modi, K. Cloud computing - concepts, architecture and challenges. In Proceedings of the 2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET), Kumaracoil, India, 21–22 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 877–880.

35. Verma, C.; Pandey, R. Big Data representation for grade analysis through Hadoop framework. In Proceedings of the 2016 6th International Conference—Cloud System and Big Data Engineering (Confluence), Noida, India, 14–15 January; IEEE: Piscataway, NJ, USA, 2016; pp. 312–315.
36. Hashem, I.A.T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Ullah Khan, S. The rise of “big data” on cloud computing: Review and open research issues. *Inf. Syst.* **2015**, *47*, 98–115. [[CrossRef](#)]
37. Lee, J.; Kao, H.A.; Yang, S. Service innovation and smart analytics for Industry 4.0 and big data environment. *Procedia CIRP* **2014**, *16*, 3–8. [[CrossRef](#)]
38. Londhe, A.; Rao, P.P. Platforms for big data analytics: Trend towards hybrid era. In Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3235–3238.
39. Popoola, S.I.; Atayero, A.A.; Badejo, J.A.; John, T.M.; Odukoya, J.A.; Omole, D.O. Learning analytics for smart campus: Data on academic performances of engineering undergraduates in Nigerian private university. *Data Br.* **2018**, *17*, 76–94. [[CrossRef](#)]
40. Pibyl, O.; Opananon, S.; Horak, T. Student perception of smart campus: A case study of Czech Republic and Thailand. In Proceedings of the Smart City Symposium Prague (SCSP), Prague, Czech, 24–25 May 2018; pp. 1–7.
41. Lu, P.; Zhang, L.; Liu, X.; Yao, J.; Zhu, Z. Highly efficient data migration and backup for big data applications in elastic optical inter-data-center networks. *IEEE Netw.* **2015**, *29*, 36–42. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Parte III
CONCLUSIONES



Universitat d'Alacant
Universidad de Alicante

9 Conclusiones

Las organizaciones buscan aprovechar la información que se genera diariamente de la interacción con sus clientes definiendo cuáles son sus intereses y necesidades. El mismo concepto se puede replicar en el campo educativo y, de esta manera, brindar una educación más eficiente basada en las características o patrones presentados por cada individuo.

Este trabajo incluye nuevos conceptos que pueden considerarse como un componente que ayuda a mejorar la educación mediante el uso de las TIC. Este estudio permite analizar e identificar las arquitecturas de análisis de datos en función de las necesidades de un campus universitario, considerando, como base principal, el volumen, las múltiples fuentes y la variedad de los datos.

Para abordar el problema de la educación y los campus universitarios, ha sido necesario realizar un análisis específico de las características estos campus. Un campus universitario es un conjunto de infraestructuras, como aulas, bibliotecas, laboratorios, facultades, sistemas informáticos, etc., donde la comunidad universitaria puede desarrollar actividades para su aprendizaje. Pero el estudio no se centra únicamente en la infraestructura, sino que hace uso de las tecnologías para garantizar la calidad de la educación al ofrecer servicios orientados en satisfacer las necesidades de cada persona.

La clave de la educación en las universidades está en los nuevos modelos educativos que, a través del análisis de datos, buscan descubrir cómo aprenden los estudiantes y generan ambientes propicios que se adaptan a esas necesidades. Una vez que se define que cada estudiante tiene necesidades específicas y que estas pueden medirse en variables, es posible crear un campus inteligente que aprenda sobre sus miembros generando un ciclo que ayude a superar estos problemas.

Los campus universitarios generalmente comparten una infraestructura genérica, basada en un modelo cliente-servidor y su estructura organizativa responde de manera similar a la de una gran empresa. Estos componentes deben ser analizados, reutilizados y actualizados para crear un campus inteligente. Un campus inteligente aporta varias mejoras a la gestión de un campus tradicional en temas como la seguridad, el manejo de recursos, el aprendizaje, etc. Si consideramos cuestiones técnicas, es importante mencionar que independientemente, un campus universitario a través de una plataforma de BI puede analizar los datos académicos de los estudiantes y tomar decisiones que contribuyan al aprendizaje. Sin embargo, un análisis a mayor escala no es suficiente por el elevado número de variables que incluyen aspectos sociológicos, académicos y financieros. La inclusión de variables sacrifica el almacenamiento y el procesamiento, por lo que la aplicación de tecnologías como el *big data* se presenta como una solución que garantiza la inclusión de diversas fuentes y un procesamiento efectivo que mejora

9 Conclusiones

considerablemente la toma de decisiones.

Durante el análisis de las herramientas que permiten la gestión de datos se abordó dos arquitecturas un BI y un *big data*. Esto nos ayudó a conocer con detalle el tipo de herramientas que existen en el mercado, sus características y donde tienen mayor incidencia cada una de estas. Por ejemplo, si abordamos la arquitectura de BI existen muchas herramientas comerciales su ventaja principal se encuentra en sus ETL, al ser desarrolladas para adaptarse a un entorno específico las empresas de software afinan sus herramientas para satisfacer al cliente.

Las herramientas de acceso abierto requieren un mayor conocimiento de parte de los técnicos para su implementación, específicamente al momento de desarrollar algoritmos para la conexión y extracción de datos. Se hicieron varias pruebas con este tipo de herramientas, se implementó un BI con el uso de Microsoft SQL y sus herramientas de análisis, de igual manera se implementó un BI con el uso de Pentaho en su versión abierta. Los dos BI se probaron en un ambiente académico y su objetivo fue analizar datos exclusivamente relacionados con el desempeño académico de los estudiantes. Los resultados fueron adecuados en relación a lo que esperábamos, se pudieron evaluar distintas variables que se encontraban en sistemas que apoyan a la gestión académica como son los LMS. El factor determinante para que un BI no sea considerado como herramienta para la gestión de datos en un campus inteligente fue la cantidad de datos que se generan en estos entornos, además de la variedad de formatos.

En cambio, las herramientas de las arquitecturas de *big data* tanto Hadoop como Spark comparten varias similitudes entre estas que las dos se utilizan una plataforma Linux. Las dos son de acceso abierto, sin embargo, los costes se generan la buscar soporte por parte de especialistas en la implementación y afinación de cada una de las herramientas. Su diferencia principal se encuentra en el uso y las necesidades que presenta el entorno donde son aplicadas. Si deseamos una arquitectura centrada en la gestión de grandes volúmenes de datos es Hadoop sin duda la mejor opción. En cambio, si las necesidades del entorno requieren un análisis en tiempo real habrá que considerar el uso de Spark.

Otro punto importante en el desarrollo de esta investigación, fue la identificación de componentes y las tecnologías del campus inteligente. Es aquí donde se pudo establecer las variables que existen en el campus y como estas se alinean a los componentes y tecnologías. Esto se relaciona directamente con las nuevas tecnologías como el IoT que en un campus inteligente además de interactuar con los miembros del campus se encarga de extraer la información que en un siguiente paso sirve de materia prima para el análisis.

La computación en la nube permite almacenar de forma eficiente los datos, garantizando su disponibilidad. Otra tecnología es el análisis de datos, para ello el uso del *big data* se transforma en el eje principal de los entornos inteligentes pues, el transformar los datos en información y la información en conocimiento permite personalizar los servicios y que estos respondan a necesidades específicas de los miembros del campus.

El diseñar una arquitectura de gestión de datos acoplada necesariamente a un campus inteligente permite garantizar la calidad de la educación. Al conocer las tendencias y maneras de aprender de los estudiantes es posible crear modelos educativos centrados en mejorar el aprendizaje de los estudiantes. Otra forma de hacerlo es con la inclusión

de la inteligencia artificial, esta tecnología, además de interactuar con los miembros del campus puede convertirse en un sistema recomendador de actividades alineadas a la forma de aprender de cada estudiante.



Universitat d'Alacant
Universidad de Alicante

10 Trabajos futuros

Como continuación existen diversas líneas de investigación que quedan abiertas y en las que es posible continuar trabajando. Durante el desarrollo de este compendio han surgido algunas líneas futuras que se han dejado abiertas y que se esperan desarrollar en un futuro. Algunas de ellas están más directamente relacionadas con este trabajo y son el resultado de cuestiones que han ido surgiendo durante la realización del mismo.

Entre los posibles trabajos futuros se destaca el uso de IoT en la gestión administrativa de las universidades y la inclusión de tecnologías que permiten la interacción entre máquinas y personas. Por este motivo, se trabajarán líneas donde se incluye el machine learning y temas como la seguridad de la información. Al momento nos encontramos trabajando en un artículo donde se considera una arquitectura de IoT aplicada a la mejora de la gestión y el aprendizaje de una universidad. El enfoque de este trabajo se basa en que el IoT es una tecnología emergente que ha tomado fuerza al permitir mejorar la calidad de vida de las personas a través, de la medición de variables. El IoT hace uso de dispositivos como sensores y actuadores que son los que se encargan de medir las variables y controlar el entorno. El internet y los dispositivos de IoT son parte de nuestras vidas, interactuamos con ellos en cada momento de nuestro diario vivir. Pero el IoT va más allá, pues no necesariamente para controlar los dispositivos necesita de una persona, el IoT permite la interacción entre dispositivos para controlar o interactuar con él entorno.

El contar con tecnologías que gestionen los recursos de un ecosistema en base a las necesidades de los usuarios y en total armonía con la naturaleza permite considerar a los entornos como inteligentes. Crear una arquitectura de IoT que ayude a la gestión de una universidad mejora la gestión de los procesos, el manejo de recursos, la mejora en la calidad de vida de los miembros de la universidad. Las universidades generalmente desarrollan sus actividades en grandes áreas geográficas que incluyen la suficiente infraestructura para dotar de servicios a un determinado número de pobladores. A esto se lo conoce como campus universitarios, estos campus están compuestos de diferentes facultades y carreras.

Tradicionalmente, en los campus universitarios para gestionar recursos como, energéticos, infraestructura y humanos se encarga a áreas donde el control o ejecución recae en las personas. Las TIC permiten mejorar estas condiciones y que la tecnología se haga cargo de estos procesos. La arquitectura propuesta ayudará a la gestión de procesos considerados claves en un campus como lo es un manejo adecuado de los activos del campus. Otra de las características que se desea obtener al integrar IoT es la mejora de la seguridad de los pobladores con la inclusión de sistemas de acceso, cámaras o reconocimiento de imágenes. El potencial del IoT es tan amplio que se pueden desple-

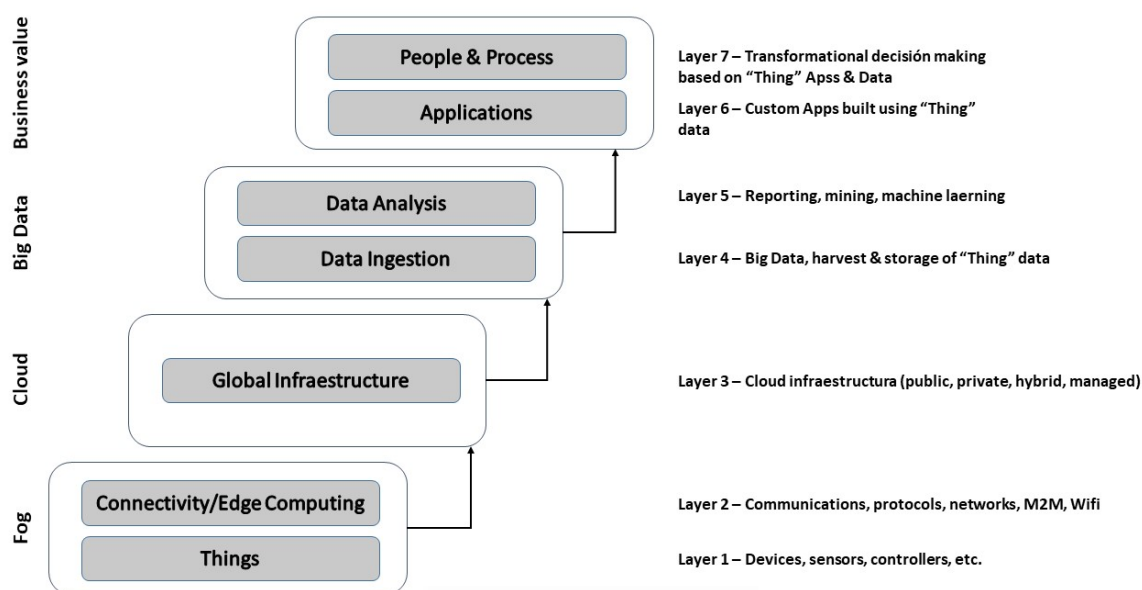


Figura 10.1: Capas de una arquitectura del Internet de las cosas

gar servicios que sean accedidos por los teléfonos móviles de los estudiantes. El uso de los servicios se convierte en una forma de obtener información sobre las tendencias de cada uno de los individuos y al generar conocimiento sobre estas tendencias se puedan ofrecer servicios personalizados.

La arquitectura se ha diseñado para poder gestionar un gran número de dispositivos. Los dispositivos están constantemente enviando datos, esto genera un volumen significativo de información. Este requerimiento se refiere a los sistemas de almacenaje de información con una gran capacidad de escalabilidad, que soporta diversos tipos y volúmenes de datos. Las acciones están consideradas para que sean casi tiempo real. En la Figura 10.1 se presentan las capas generales que maneja una arquitectura de IoT, desplegada en siete capas. La arquitectura describe la estructura de una solución de IoT, lo que incluye los aspectos físicos como los dispositivos y los aspectos virtuales como los servicios y los protocolos de comunicación. Adoptar una arquitectura con múltiples niveles permite concentrarse en comprender acerca de cómo todos los aspectos más importantes de la arquitectura funcionan antes de que los integre dentro de su aplicación de IoT. Este enfoque modular ayuda a gestionar la complejidad de las soluciones IoT.

APÉNDICE



Universitat d'Alacant
Universidad de Alicante

A Otros artículos

Este capítulo presenta otros trabajos realizados durante el transcurso del doctorado, estos trabajos apoyan la investigación y guardan una línea específica en relación al tema. Varios de estos trabajos se encargaron de realizar un análisis de las tecnologías que pueden ser aplicadas al campo educativo. Otros trabajos fueron más específicos y en estos se propusieron arquitecturas, modelos o metodologías para el análisis de datos educativos. Estos trabajos fueron presentados en congresos a nivel nacional e internacional y fueron revisados por pares. Una vez que fueron aceptados, se realizó la presentación del artículo en cada congreso, esta experiencia sirvió para discutir y mejorar el trabajo con base en las observaciones y discusiones recibidas por otros participantes. Los artículos que han sido incluidos en este apéndice son seis y todos se encuentran indexados en Scopus.

Universitat d'Alacant
Universidad de Alicante

B Analysis of Data Mining Techniques Applied to LMS for Personalized Education

Villegas-Ch, W. & Luján-Mora, S. (2017). Villegas-Ch, W., y Luján-Mora, S. (2017). Analysis of Data Mining Techniques Applied to LMS for Personalized Education. En 2017 IEEE World Engineering Education Conference (EDUNINE) (pp. 85-89). (Villegas-Ch y Luján-Mora, 2017a)

Disponible en:

URL: <https://ieeexplore.ieee.org/document/7918188>

DOI: <https://doi.org/10.1109/EDUNINE.2017.7918188>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Analysis of data mining techniques applied to LMS for personalized education

W. Villegas-Ch¹

¹ Universidad de Las Américas
Quito, Ecuador
william.villegas@udla.edu.ec

S. Luján-Mora²
²University of Alicante
Alicante, Spain
sergio.lujan@ua.es

Abstract - This article describes the models and the use of data mining techniques applied to Learning Management Systems (LMS) which allow institutions to offer the student a personalized education. It considers the ways in which the concepts of educational data mining (EDM) are applied to the information extracted from the LMS. The data from these systems can be evaluated to convert the information collected into useful information to provide an education tailored to the needs of each student. This approach seeks to improve the effectiveness and efficiency of education by recognizing patterns in student performance. This article presents an analysis of the data mining techniques that fit LMS, specifically in terms of a case study applied to the e-learning platform Moodle. The objective is to provide stakeholders with guidance on the use of EDM tools.

Keywords - Data mining, e-learning, Moodle.

I. INTRODUCTION

Learning Management Systems (LMS) store large volumes of information that are not usually fully exploited by educational institutions. The use of such information serves for the continuous improvement of the content and structure of the virtual courses contained in LMS.

Under this scheme, the pedagogical model consists of web-based virtual courses supervised by a tutor. This work is developed with the use of online resources, activities, forums and various shared services. The data are gathered in the LMS and are based on the activities and didactic resources that are offered to the students. The data obtained when evaluated with the use of data mining leads to the possibility of choosing the most appropriate resources and adapting it to the characteristics and personal interests of each stakeholder in the educational context.

In this article, we propose the analysis of the different models and tools that allow the exchange of experiences with regard to how students learn: Section II presents several concepts necessary to address the problem based on previous work; Section III details the stage of pre-processing required for the analysis of the data using the main techniques of data mining; Section IV performs an analysis of the results obtained; finally section V presents the conclusions and makes recommendations for future investigations.

II. PRELIMINARY CONCEPTS

In this article, several key concepts that help the management and application of different models and data mining tools are taken into account.

A. E-learning and LMS platforms

E-learning consists of education and training through the web [1]. This online teaching model allows the interaction of the user with the material under consideration through the use of various information and communication technology tools (ICTs).

Such an approach is presented as one of the formative strategies that can be used to solve several educational problems [2]. These problems include the geographical isolation of the student, the access to information, and the need for constant improvement.

B. Data mining

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence and automatic learning to extract and identify useful information from large databases in order to generate knowledge [3].

Based on the concept of data mining, EDM techniques can be used to extract knowledge from e-learning systems. The objective of this tool is to look for behavior patterns in terms of system use by both teachers and students [4]. This helps to customize the virtual environment by identifying useful knowledge, and using techniques such as prediction, classification, clustering, fuzzy logic, etc.

III. METHOD

For the development of this work an analysis of the different techniques and models of data mining applied to the education was realized by taking into account the stages of the knowledge discovery process with regard to databases. The academic data of students on an undergraduate program of a university in Ecuador were selected, in order to follow up their results during 2016. The aim was to determine their strengths and weaknesses in each activity proposed by a tutor within a virtual course generated in the Moodle platform.

Moodle is a modular platform which makes it easy to manage the users and courses, as well as to add content. This tool supports all records of the activities undertaken by both teachers and students in a database engine based on Mysql [5].

The data obtained were included in a SQL data repository. To this data a processing and transformation phase was applied in order to obtain clean data. The purpose of obtaining clean data is to be able to apply a number of different data mining techniques, so that the results can be analyzed, evaluated and interpreted.

In the analysis of the proposed EDM techniques, the method that each uses to generate information that is useful

C Systematic Review of Evidence On Data Mining Applied to LMS Platforms for Improving E-Learning

Villegas-Ch, W., y Luján-Mora, S. (2017). Systematic Review of Evidence On Data Mining Applied to LMS Platforms for Improving E-Learning. En International Technology, Education and Development Conference (INTED) (pp. 6537-6545). (Villegas-Ch y Luján-Mora, 2017b)

Disponible en:

URL: <https://library.iated.org/view/VILLEGASCH2017SYS>

DOI: <https://doi.org/10.21125/inted.2017.1510>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O2. Identificar los componentes y las tecnologías que son parte de un campus inteligente.

SYSTEMATIC REVIEW OF EVIDENCE ON DATA MINING APPLIED TO LMS PLATFORMS FOR IMPROVING E-LEARNING

W. Villegas-Ch¹, S. Luján-Mora²

¹Universidad de Las Américas (ECUADOR)

² Universidad de Alicante (SPAIN)

Abstract

This article examines the work on the application of data mining tools in learning management systems (LMS). The purpose of this paper is to introduce the concept of rigorous reviews of current empirical evidence to the educational data mining community (EDM). This paper proposes a method to extract information from records, analyze them statistically and interpret that information into useful knowledge; this includes how they interact with regard to positive impacts, especially with regard to learning and improving the skills of those involved. This work is aimed at researchers who focus their studies on E-learning and the use of data mining tools.

Keywords: data mining, Moodle, collaborative learning, EDM.

1 INTRODUCTION

This work develops a method to extract information from the works that cover topics, such as educational data mining EDM, which are applied to the Learning Management System (LMS). The criteria are developed based on four questions that allow classifying the different articles throughout the investigation based on the importance that these reflect before the improvement of the E-learning. For this analysis, the evaluation of learning is also considered based on different competencies considered for the selection or exclusion of articles. These learning competencies are divided into two groups: specific and generic. Specific competences are those, which are related directly to the use of concepts, theories or skills specific to a particular area, while generic competences are skills, abilities and knowledge that any student should develop independently of their area of study [24]. LMS stores large volumes of information that are not fully exploited by educational institutions. The use of these volumes of information serves for the continuous improvement of the content, structure and use of the virtual courses.

Under this scheme, the pedagogical model consists of virtual courses supervised by a tutor, who bases their work on the web. This work is developed with the use of online resources, activities, forums and various shared services. The data are generated in the LMS platforms based on the activities and didactic resources that are offered to the students. The data, which were obtained when evaluated with the use of data mining, lead to the possibility of choosing the appropriate resources adapted to the characteristics and personal interests of each of the actors of education. EDM tools create a bridge for the fields of traditional statistics through pattern recognition and learning, which is an aspect that enables data mining to be compatible to modeling techniques with new information technologies (IT) and database technologies. In this article, we propose the analysis of the different works done on the improvement of E-learning with the use of techniques and models of data mining that allow the exchange of experiences on how students learn. Section 2 details the proposed method for the collection of articles, which contribute to the proposed research; Section 3 presents the results obtained from the research, as well as the process applied in the analysis of information, Section 4 presents the conclusions and suggestions for future research.

2 METHODOLOGY

A Systematic Mapping Study (SMS), carries out an in-depth review of the studies carried out in a specific area, the objective is to identify evidence on the subject. This work is based on the guidelines published in the methodology proposed by Kitchenham [25]. This methodology describes how the results of a literature review in software engineering should be planned, executed and presented. For this work, Petersen's proposal was used [26]. To carry out a rigorous review of the literature, it is important to follow a systematic procedure, for which the articles were grouped according to the type of tool, model, paradigm or discussion they posed. The final classification was:

D Data Mining Toolkit for Extraction of Knowledge from LMS

Villegas-Ch, W., Luján-Mora, S., y Buenaño-Fernandez, D. (2017). Data Mining Toolkit for Extraction of Knowledge from LMS. En Proceedings of the 2017 9th International Conference on Education Technology and Computers (pp. 31-35). ACM. (Villegas-Ch et ál., 2017)

Disponible en:

URL: <https://dl.acm.org/citation.cfm?id=3175553>

DOI: <https://doi.org/978-1-4503-5435-6>

Temas a los que contribuye:

- O1. Analizar las herramientas que permiten la gestión de datos en un campus universitario.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Data mining toolkit for extraction of knowledge from LMS

W. Villegas-Ch
Universidad de Las Américas
Av. de los Granados E12-41 y
Colimes esq.
+59323970000, Quito-Ec. 170125
william.villegas@udla.edu.ec

S. Luján-Mora
University of Alicante
Carretera de San Vicente del Raspeig
s/n
+34 965903400, Alicante-Es. 03690
sergio.lujan@ua.es

Diego Buenaño-Fernandez
Universidad de Las Américas
Av. de los Granados E12-41 y
Colimes esq.
+59323970000, Quito-Ec. 170125
diego.buenano@udla.edu.ec

ABSTRACT

Today, information technology (IT) is an active part of education. Its main impact is in the administration of learning management systems (LMS). The support provided by IT in LMS has generated greater dexterity in the evaluation of the quality of education. The evaluation process usually includes the use of tools applied to online analytical processing (OLAP). The application of OLAP allows the consultation of large amounts of data. Data mining algorithms can be applied to the data collected to perform a pattern analysis. The potential use of these tools has resulted in their specialization, both in the presentation and in the algorithmic techniques, allowing the possibility of educational data mining (EDM). EDM seeks to enhance or customize education within LMS by classifying groups of students in terms of similar characteristics that require specific resources. The ease of use and extensive information about some of the EDM tools has caused many educational institutions to consider them for their own use. However, these institutions often make errors in data management. Errors in the use of data mean that the improvements in LMS are inadequate. The work described in this paper provides a guide on the use of applied methodology in the process of knowledge extraction (KDD). It also enumerates some of the tools that can be used for each step of the process.

CCS Concepts

• Information systems → Information systems applications → Data mining.

Keywords

Educational data mining; knowledge extraction process; LMS platforms; analytical learning.

1. INTRODUCTION

At present, education trends are rooted in the quality of teaching, learning-based educational models and the establishment of evaluation processes. To meet these needs, educational institutions make use of information and communication technologies (ICT). The tools that offer ICT and that have become

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICETC 2017, December 20–22, 2017, Barcelona, Spain

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5435-6/17/12...\$15.00

<https://doi.org/10.1145/3175536.3175553>

the most commonly used are platforms of learning [1]. These platforms belong to the group of learning management systems (LMS) also known as virtual learning management (VLE). LMS allow you to create customized learning environments that support both teaching and learning. They are part of a subgroup of content management systems (CMS) [2].

The evaluation processes seek to monitor teachers and their teaching strategies, and whether or not students are learning effectively. With the evaluation of these parameters it is possible to guarantee the academic quality that has become an internal and external evaluation factor for learning institutions.

LMS have integrated new variables such as the time dedicated to the development of activities, historical notes, forums with frequently-asked questions, etc. These variables are measurable and help to process data when assessing the quality of educational provision. The specific tool that provides the necessary features to support this process is data mining. It is so important that developers have created modules that are integrated with the different platforms of e-learning or can be used as an independent system. Data mining allows the analysis of data by processing and applying mathematical and statistical methods that allow tracking trends that relate to an individual or groups that share similar characteristics [3]. The importance of the application of data mining e-learning has contributed to the emergence of educational data mining (EDM).

EDM allows users to evaluate the different activities created in a VLE and to obtain projections of the students' performance in a specific course. These results help the teacher to take corrective action and prevent students from failing an activity, or having to take up re-do the course. In order to fulfill the objective of EDM it is important that support departments in charge of learning evaluation within educational institutions take into account several aspects such as information extraction, information purging, appropriate repositories for information, choice of the best EDM method to be used, results analysis, etc.

2. PRELIMINARY CONCEPTS

In this article, several key concepts that help the management and application of different models and the use of data mining tools are considered.

2.1 E-learning and LMS platforms

E-learning consists of education and training through the web [4]. This online teaching model allows the interaction of the user with the material under consideration through the use of various information and communication technology tools.

Such an approach is presented as one of the formative strategies that can be used to solve several educational problems [5]. These

E Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining

Villegas-Ch, W., Luján-Mora, S., y Buenaño-Fernandez, D. (2018). Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining. En 2018 IEEE World Engineering Education Conference (EDUNINE) (pp. 1-5). (Villegas-Ch, Luján-Mora, y Buenaño-Fernandez, 2018)

Disponible en:

URL: <https://ieeexplore.ieee.org/document/8450954>

DOI: <https://doi.org/10.1109/EDUNINE.2018.8450954>

Temas a los que contribuye:

- O2. Identificar los componentes y las tecnologías que son parte de un campus inteligente.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Towards the integration of business intelligence tools applied to educational data mining

William Villegas-Ch
Facultad de Ingenierías y Ciencias
Aplicadas,
Universidad de Las Américas,
Quito, Ecuador
william.villegas@udla.edu.ec

Sergio Luján-Mora
Department of Software and
Computing Systems,
University of Alicante.
Alicante, Spain
sergio.lujan@ua.es

Diego Buenaño-Fernandez
Facultad de Ingenierías y Ciencias
Aplicadas,
Universidad de Las Américas,
Quito, Ecuador
diego.buenano@udla.edu.ec

Abstract—At present the educational institutions have computer systems that generate information of the educational activity of each student. This large amount of information is often not used to benefit the management of educational processes. Under these conditions the institutions have acquired systems that extract reports general such as student attendance, semester notes or number of students who approve a course. This information is not sufficient to take corrective measures that act in time and avoid actions such as student dropout. This work it is proposed to use the tools of business intelligence in educational data. The aim is to detect groups of students with similar. This behavior will allow to analyze the causes and possible effects on student performance in a given subject. Once the results are obtained, those involved in the educational process will be allowed to make decisions that contribute to the improvement of the educational quality.

Keywords—data mining; business intelligence; analysis of data; e-learning.

I. INTRODUCTION

Educational institutions, like companies, manage an organizational structure whose data can be measurable and quantifiable. Companies for several years have decided to improve their systems and processes in the search to apply a business intelligence based on the information they have from customers. In the analysis of data companies take advantage of educational institutions, since for several years' companies have applied techniques and tools that help management and decision making [1]. The tools used in the majority are responsible for the analysis and representation of data. This analysis allows the managers of the companies or areas like marketing to make decisions about the business. Business intelligence through its process and components presents skills that can be used in education [2]. Educational institutions currently have robust systems that generate data that can be analyzed for the improvement of their responses to a possible change in the educational approach. This new approach establishes its pillars in the use of information technologies for the improvement of the quality of education. To achieve this improvement in educational quality it is necessary to analyze the academic information of the students, in this way can be classified by pattern recognition. The classification allows to offer a personalized or specific education according to the needs of each group.

In order to comply with a data analysis process as detailed above, it is important to know that educational data is found in several storage sources. This will use a process of extraction, transformation and loading that are techniques of a business intelligence often used in companies and applied to educational data. This work proposes a method where business intelligence and data mining techniques are applied to educational environments, using Microsoft SQL. The work has been divided as follows: Section II covers the concepts to be used throughout the research; Section III details the method used to achieve the proposed objectives; Section IV makes an analysis of the results; Finally, section V presents the conclusions and future work.

II. PRELIMINARY CONCEPTS

In this article, several key concepts that support the management and application of different models and of data mining tools are taken into account.

A. Data warehouse

A data warehouse is where you integrate and debug information from one or more data sources. These data are processed and analyzed from an infinite number of perspectives. Creating a data warehouse is the first step in implementing a business intelligence solution. This solution allows company managers to organize, understand and use their data to make strategic decisions that contribute to the evolution of the business [1].

B. Business intelligence

Business intelligence and business analytics are increasingly important in the structure of companies looking for a competitive advantage over their competitors. Business intelligence provides the ability to transform data into information and information into knowledge [2]. This process optimizes the process in decision making.

Technologically a business intelligence is a set of methodologies and applications. These characteristics allow the collection, debugging and transformation of data from transactional systems and unstructured information into structured information for direct exploitation [3].

F Application of Data Mining for the Detection of Variables that Cause University Desertion

Palacios-Pacheco, X., Villegas-Ch, W., y Luján-Mora, S. (2018). Application of Data Mining for the Detection of Variables that Cause University Desertion. En International Conference on Technology Trends (pp. 510-520). Springer, Cham. (Palacios-Pacheco et ál., 2019)

Disponible en:

URL: https://link.springer.com/chapter/10.1007/978-3-030-05532-5_38





DOI: https://doi.org/10.1007/978-3-030-05532-5_38

Temas a los que contribuye:

- O3. Diseñar una arquitectura que convierta un campus universitario tradicional en un campus inteligente.
- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.



Application of Data Mining for the Detection of Variables that Cause University Desertion

X. Palacios-Pacheco¹ , W. Villegas-Ch²  ,
and Sergio Luján-Mora³ 

¹ Universidad Internacional del Ecuador, Quito, Ecuador
xpalacio@uide.edu.ec

² Universidad de Las Américas, Quito, Ecuador
william.villegas@udla.edu.ec

³ Universidad de Alicante, Alicante, Spain
sergio.lujan@ua.es

Abstract. College desertion is one of the problems currently addressed by most higher education institutions throughout Latin America. From different investigations, it is known that a large percentage of students do not complete their studies, with the consequent social cost associated with this phenomenon. Some countries have begun to design deep improvement processes to increase retention in the first years of university studies. The process considered for the improvement of the desertion is through the data mining, the use of its algorithms allows discovering patterns in the students that help to explain this effect. The algorithms also identify the independent variables that influence the desertion and analyze them according to a level of depth previously established by the interested parties. The purpose of this study is to determine a model that explains the desertion of undergraduate students at the university and design actions that tend towards the decrease of the desertion.

Keywords: Data mining · Desertion · Data analysis · Weka

1 Introduction

Higher education institutions are currently subject to evaluation processes by government entities as well as international agencies that are responsible for the assurance of academic quality. The main factors that are referenced in the evaluation processes are the abandonment of studies known as desertion, repetition and academic effectiveness [12]. The World Bank and UNESCO indicate that desertion at the level of university education reaches a percentage of around 40% in Latin America and the Caribbean [1]. For universities, it has become a very complex task to detect the possible causes of desertion. These causes until a few years ago were based on factors such as the lack of prior information about the career and the student's difficulty in adapting to a university environment. Currently, the studies carried out on the subject include new variables that can be analyzed and seek to answer the percentages of influence they have on the student to drop their studies. The discovery of these variables is the main concern of the

G Artificial Intelligence as a Support Technique for University Learning

Villegas-Ch, W., Palacios-Pacheco, X., y Luján-Mora, S. (2019). Artificial Intelligence as a Support Technique for University Learning. En IEEE World Conference on Engineering Education (EDUNINE) (pp. 1-6). (Villegas-Ch, Palacios-Pacheco, y Luján-Mora, 2019b)

Disponible en:

URL: <https://ieeexplore.ieee.org/document/8875833>

DOI: <https://doi.org/10.1109/EDUNINE.2019.8875833>

Temas a los que contribuye:

- O4. Crear una arquitectura para la gestión de datos que se acople a un campus inteligente y garantice la calidad de la educación.

Universitat d'Alacant
Universidad de Alicante

Artificial intelligence as a support technique for university learning

William Villegas-Ch
Facultad de Ingenierías y Ciencias
Aplicadas,
Universidad de Las Américas,
Quito, Ecuador
william.villegas@udla.edu.ec

Xavier Palacios-Pacheco
Departamento de sistemas,
Universidad Internacional del Ecuador,
Quito, Ecuador
xpalacio@uide.edu.ec

Sergio Luján-Mora
Department of Software and
Computing Systems,
Alicante, Spain
sergio.lujan@ua.es

Abstract— Currently, universities seek to improve academic methods that come from a traditional model. Traditional models base learning on the experience of teachers and the development of activities. Studies carried out in the educational field consider that not all activities generate knowledge in students. The activities must adapt to the needs of each student to build an efficient model that contributes to learning. To solve this problem, education experts support academic management in the use of information and communication technologies. The inclusion of techniques such as the analysis of educational data and artificial intelligence identify the deficiencies and needs of students. This work proposes the design of an expert system that interacts with students and evaluates their responses with those of data analysis systems to reach a conclusion and recommend activities that align with the needs of each student.

Keywords— artificial intelligence, expert systems, active learning, intelligent systems

I. INTRODUCTION

The contributions of information and communication technologies (ICT) enrich the educational work. The development of ICT demands a change in traditional learning environments. The use of software and computerized educational materials as a resource to support teaching and learning processes has become a necessity. In addition, they constitute a response to the problem that revolves around the cognitive understanding of concepts and notions in classrooms [1]. Previous works analyze the advances that help to learn by integrating new technological concepts in education. A large percentage of research focuses especially on the integration of artificial intelligence (AI) in educational processes. The inclusion of computer systems in universities is increasingly necessary due to its flexibility and scalability in educational management. An advantage of using ICT is the analysis of this data. This process allows detecting the academic problems that each student presents. Once the problems are identified, methods and models can be generated to improve learning.

Several techniques allow the student to reflect on what they have learned. These techniques can be adapted to satisfy diverse needs, everything depends on the tasks or activities proposed to the student [2]. On the revealed ideas, it can be considered that AI systems such as the expert systems (ES), created for pedagogical and instructional purposes can diagnose, debug and correct the development of student learning. In addition, the system determines the cognitive level of the student and helps him improve his weaknesses to reach a higher level of learning. This work proposes the development of an ES that represents a response-oriented to the efficient decision making on teaching models and didactic activities.

The proposed ES captures the knowledge of experts on different topics, as well as experts in education. This functionality allows us to recommend suitable strategies that facilitate the scope and objectives proposed for a specific university population [3]. The ES is able to interact with each student and integrate into various computer systems to extract information to reach a conclusion. Students develop different activities in learning management systems (LMS) based on the resources provided by each teacher [4].

The information that the ES obtains from the LMS added to the information obtained from the interaction with the student allows him to recommend specific activities to each student. Reaching this level of learning management guarantees effectiveness and can even nullify other problems linked to student performance such as desertion. The paper is divided as follows: Section II contains the theoretical foundations that contribute to the design and implementation of the proposed system. Section III contains the method in which the entire development is systematically explained. Finally section IV contains the conclusions obtained.

II. THEORETICAL FOUNDATION

A. Expert systems

ES use artificial intelligence techniques to respond automatically to questions raised as if they were human experts. These systems are applied in many fields of knowledge due to their own nature and that are increasingly being perfected. In the educational field, they play a leading role due to the trends related to personalized education through technological means [5].

B. Active learning

Active learning involves the use of a set of more effective and interesting experimental methods. With active learning, students assume greater responsibility for their own education. It is characterized by very well structured and challenging activities, with enough flexibility to adapt them to the characteristics of the learning group and even at the individual level. The activities are organized so that the students develop them in face-to-face and virtual spaces or the combination of them [6].

III. METHOD

The process begins with the identification of the problem that lies in the fact that not all students learn in the same way. There are several types of activities that are tailored to the needs of each student, thus personalizing education. This educational method is known as active learning. The objective is to identify the needs of each student and recommend a type of activity that meets these needs [7]. The identification of needs, as well as the recommendation of

Referencias

- Atzori, L., Iera, A., y Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. doi: 10.1016/j.comnet.2010.05.010 (citado en 3)
- Bascopé, M., Perasso, P., y Reiss, K. (2019). Systematic review of education for sustainable development at an early stage: Cornerstones and pedagogical approaches for teacher professional development. *Sustainability*, 11(3), 719. doi: 10.3390/su11030719 (citado en 3)
- Bean, J. P. (1983). The application of a model of turnover in work organizations to the student attrition process. *The review of higher education*, 6(2), 129–148. (citado en 7)
- Christozov, D. (2017). Business analytics as a tool to transforming information into an Informing System: The case of the on-line course registration system. *Informing Science*, 20, 167–178. doi: 10.28945/3764 (citado en 8)
- Cohen, J., y Acharya, S. (2013). Towards a more secure Apache Hadoop HDFS infrastructure: Anatomy of a targeted advanced persistent threat against HDFS and analysis of trusted computing based countermeasures. En *Network and System Security. NSS 2013. Lecture Notes in Computer Science* (Vol. 7873, pp. 735–741). Berlin. doi: 10.1007/978-3-642-38631-2_64 (citado en 15)
- Corrales, D. C., y Corrales, J. C. (2016). Sequential classifiers for network intrusion detection based on data selection process. En *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1827–1832). (citado en 9)
- Donoso, S., y Schiefelbein, E. (2009). Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social. *Estudios pedagógicos (Valdivia)*, 33(1), 7–27. doi: 10.4067/s0718-07052007000100001 (citado en 4, 6)
- Gairín, J., Triado, X. M., Feixas, M., Figuera, P., Aparicio-Chueca, P., y Torrado, M. (2014). Student dropout rates in Catalan universities: Profile and motives for disengagement. *Quality in Higher Education*, 20(2), 165–182. doi: 10.1080/13538322.2014.925230 (citado en 4)
- Ghazi, M. R., y Gangodkar, D. (2015). Hadoop, mapreduce and HDFS: A developers perspective. *Procedia Computer Science*, 48, 45–50. doi: 10.1016/j.procs.2015.04.108 (citado en 14, 15)
- Kamilaris, A., Pitsillides, A., Prenafeta-Bold, F. X., y Ali, M. I. (2017). A Web of Things based eco-system for urban computing - towards smarter cities. En *International conference on telecommunications* (pp. 1–7). doi: 10.1109/ICT.2017.7998277 (citado en 14)
- Kimball, R., Ross, M., y Kimball, R. (2009). *The data warehouse toolkit: the complete*

Referencias

- guide to dimensional modeling* (John Wiley ed.). New York, NY: Wiley. (citado en 8)
- Kortuem, G., Bandara, A. K., Smith, N., Richards, M., y Petre, M. (2013). Educating the Internet of Things Generation. *Computer*, 46(2), 53–61. doi: 10.1109/MC.2012.390 (citado en 4)
- Naciones Unidas. (2015). *Objetivos de desarrollo sostenible*. Descargado 2019-11-04, de <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/> (citado en 3)
- Nie, X. (2013). Constructing Smart Campus Based on the Cloud Computing Platform and the Internet of Things. En *Proceedings of the 2nd international conference on computer science and electronics engineering (ICCSEE 2013)* (pp. 1576–1578). Atlantis Press. doi: 10.2991/iccsee.2013.395 (citado en 14)
- Palacios-Pacheco, X., Villegas-Ch, W., y Luján-Mora, S. (2019). Application of Data Mining for the Detection of Variables that Cause University Desertion. En *Communications in computer and information science* (Vol. 895, pp. 510–520). doi: 10.1007/978-3-030-05532-5 (citado en 4, 6, 27, 151)
- Petersen, K., Vakkalanka, S., y Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. doi: 10.1016/j.infsof.2015.03.007 (citado en 18)
- Pompei, L., Mattoni, B., Bisegna, F., Nardecchia, F., Fichera, A., Gagliano, A., y Pagano, A. (2018). Composite Indicators for Smart Campus: Data Analysis Method. En *IEEE International Conference on Environment and Electrical Engineering* (pp. 1–6). IEEE. doi: 10.1109/EEEIC.2018.8493893 (citado en 3)
- Ray, P. P. (2016). A survey of IoT cloud platforms. *Future Computing and Informatics Journal*, 1(1-2), 35–46. doi: 10.1016/j.fcij.2017.02.001 (citado en 5)
- Salmerón-Manzano, E., y Manzano-Agugliaro, F. (2018). The higher education sustainability through virtual laboratories: The Spanish University as case of study. *Sustainability*, 10(11), 4040. doi: 10.3390/su10114040 (citado en 4)
- Shanahan, J., y Dai, L. (2017). Large Scale Distributed Data Science from scratch using Apache Spark 2.0. En *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 955–957). doi: 10.1145/3041021.3051108 (citado en 14)
- Turner, M., Bailey, J., Linkman, S., Budgen, D., Pearl Brereton, O., y Kitchenham, B. (2008). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. doi: 10.1016/j.infsof.2008.09.009 (citado en 18)
- Uskov, V. L., Bakken, J. P., y Pandey, A. (2016). Smart University Taxonomy: Features, Components, Systems. En *Smart education and e-learning* (Vol. 59, pp. 3–14). doi: 10.1007/978-3-319-39690-3 (citado en 4, 13)
- Valdiviezo-Díaz, P., Cordero, J., Reátegui, R., y Aguilar, J. (2015). A business intelligence model for online tutoring process. En *IEEE Frontiers in Education Conference (FIE)* (Vol. 2014, pp. 1–9). doi: 10.1109/FIE.2015.7344385 (citado en 9)
- Villegas-Ch, W., y Luján-Mora, S. (2016). Análisis de las herramientas de minería de datos para la mejora del E-learning en Plataformas LMS. En *TIC actualizadas*

- para una nueva docencia universitaria (pp. 761–774). McGraw-Hill. (citado en 9)
- Villegas-Ch, W., y Luján-Mora, S. (2017a). Analysis of data mining techniques applied to LMS for personalized education. En *IEEE World Engineering Education Conference (EDUNINE)* (pp. 85–89). doi: 10.1109/EDUNINE.2017.7918188 (citado en 5, 26, 135)
- Villegas-Ch, W., y Luján-Mora, S. (2017b). Systematic Review of Evidence on Data Mining Applied to LMS Platforms for Improving E-Learning [Proceedings Paper]. En I. Chova, LG and Martinez, AL and Torres (Ed.), *International technology, education and development conference* (pp. 6537–6545). (citado en 26, 139)
- Villegas-Ch, W., Luján-Mora, S., y Buenaño-Fernandez, D. (2017). Data mining toolkit for extraction of knowledge from LMS. En *ACM International Conference Proceeding Series* (Vol. Part F1346, pp. 31–35). doi: 10.1145/3175536.3175553 (citado en 5, 26, 143)
- Villegas-Ch, W., Luján-Mora, S., y Buenaño-Fernandez, D. (2018, jan). Towards the Integration of Business Intelligence Tools Applied to Educational Data Mining. En *IEEE World Engineering Education Conference (EDUNINE)* (pp. 1–5). IEEE. doi: 10.1109/EDUNINE.2018.8450954 (citado en 5, 9, 19, 27, 147)
- Villegas-Ch, W., Luján-Mora, S., Buenaño-Fernandez, D., y Palacios-Pacheco, X. (2018). Big data, the next step in the evolution of educational data analysis. En *Advances in intelligent systems and computing* (Vol. 721, pp. 138–147). doi: 10.1007/978-3-319-73450-7_14 (citado en 27, 33, 35)
- Villegas-Ch, W., Molina-Enriquez, J., Chicaiza-Tamayo, C., Ortiz-Garcés, I., y Luján-Mora, S. (2019, oct). Application of a Big Data Framework for Data Monitoring on a Smart Campus. *Sustainability*, 11(20), 5552. doi: 10.3390/su11205552 (citado en 11, 26, 33, 105)
- Villegas-Ch, W., Palacios-Pacheco, X., Buenaño-Fernandez, D., y Luján-Mora, S. (2019). Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment. *International Journal of Engineering Education*, 35(5), 1316–1325. (citado en 5, 25, 33, 47)
- Villegas-Ch, W., Palacios-Pacheco, X., y Luján-Mora, S. (2019a). Application of a Smart City Model to a Traditional University Campus with a Big Data Architecture: A Sustainable Smart Campus. *Sustainability*, 11(10), 2857. doi: 10.3390/su11102857 (citado en 5, 7, 14, 17, 26, 33, 75)
- Villegas-Ch, W., Palacios-Pacheco, X., y Luján-Mora, S. (2019b). Artificial intelligence as a support technique for university learning. En *IEEE World Conference on Engineering Education (EDUNINE)* (pp. 1–6). doi: 10.1109/EDUNINE.2019.8875833 (citado en 27, 155)
- Villegas-Ch, W., Palacios-Pacheco, X., Ortiz-Garcés, I., y Luján-Mora, S. (2019). Management of educative data in university students with the use of big data techniques. *Revista Ibérica de Sistemas e Tecnologias de Informação*(E19), 227–238. (citado en 11, 25, 33, 61)
- Yaqoob, I., Ahmed, E., Hashem, I. A. T., Ahmed, A. I. A., Gani, A., Imran, M., y Guizani, M. (2017). Internet of Things Architecture: Recent Advances, Taxonomy, Requirements, and Open Challenges. *IEEE Wireless Communications*, 24(3),

Referencias

- 10–16. doi: 10.1109/MWC.2017.1600421 (citado en 3)
- Yashiro, T., Kobayashi, S., Koshizuka, N., y Sakamura, K. (2013). An Internet of Things (IoT) architecture for embedded appliances. En *IEEE Region 10 Humanitarian Technology Conference* (pp. 314–319). doi: 10.1109/R10-HTC.2013.6669062 (citado en 5)



Universitat d'Alacant
Universidad de Alicante