

Recopilación de Investigaciones en Tecnología 2016

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Omar Otoniel Flores Cortez, German Antonio Rosa

Implementación de un entorno de aprendizaje virtual integrando herramientas de E-learning y CMS

Marvin Elenilson Hernández, Carlos Aguirre

Big data, análisis de datos en la nube

José Guillermo Rivera, Verónica Idalia Rosa Urrutia



Joomla!



COLECCIÓN INVESTIGACIONES
Universidad Tecnológica de El Salvador

59

Recopilación de Investigaciones en Tecnología 2016

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Investigadores

Omar Otoniel Flores Cortez,
Docente investigador Utec. omar.flores@utec.edu.sv
German Antonio Rosa Castellanos,
Docente investigador Utec. german.rosa@mail.utec.edu.sv

Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

Investigadores

Carlos Antonio Aguirre Ayala,
Docente investigador Utec. carlos.aguirre@utec.edu.sv,
Marvin Elenilson Hernández Montoya,
Docente investigador Utec. marvin.hernandez@utec.edu.sv.

***Big data*, análisis de datos en la nube**

Investigadores

Verónica Idalia Rosa,
Docente investigadora Utec. veronica.rosa@utec.edu.sv,
José Guillermo Rivera,
Docente investigador Utec. joseguillermo.rivera@gmail.com

Estas investigaciones fueron subvencionadas por la Universidad Tecnológica de El Salvador. Las solicitudes de información, separatas y otros documentos relativos a este estudio pueden hacerse a la dirección postal: Universidad Tecnológica de El Salvador, Edificio Dr. José Adolfo Araujo Romagoza, Vicerrectoría de Investigación, Dirección de Investigaciones calle Arce y 19 avenida Sur, N° 1045, o al correo electrónico camila.calles@utec.edu.sv

San Salvador, 2016
Derechos Reservados

© **Copyright**

Universidad Tecnológica de El Salvador

004.66

R311 Recopilación de investigaciones en tecnología 2016 ; Internet de las cosas : diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico ;
sv Implementación de un entorno de aprendizaje virtual integrado herramientas de e-learning y CMS ; Big data, análisis de datos en la nube / Omar Otoniel Flores Cortez, German Antonio Rosa, Marvin Elenilson Hernández, Carlos Aguirre, José Guillermo Rivera, Verónica Idalia Rosa Urrutia. -- 1ª ed. -- San Salvador, El Salv. : Universidad Tecnológica de El Salvador, 2016.

100 p. : il. ; 23 cm. -- (Colección investigaciones ; v. 59)

ISBN 978-99961-48-62-0

1. Tecnología-Investigaciones. 2. Sistemas de energía fotovoltaica. 3. Desarrollo científico. 4. Bases de datos en línea. 5. Internet en la educación. I. Título. II. Título: Internet de las cosas : diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico / Omar Otoniel Flores Cortez, German Antonio Rosa. III. Título: Implementación de un entorno de aprendizaje virtual integrado herramientas de e-learning y CMS / Marvin Elenilson Hernández, Carlos Aguirre. IV. Título: Big data, análisis de datos en la nube / José Guillermo Rivera, Verónica Idalia Rosa Urrutia.

BINA/jmh

AUTORIDADES UTEC

Dr. José Mauricio Loucel

Presidente Junta General Universitaria

Lic. Carlos Reynaldo López Nuila

Vicepresidente Junta General Universitaria

Don José Mauricio Loucel Funes

Presidente UTEC

Ing. Nelson Zárate

Rector UTEC

Recopilación de Investigaciones en Tecnología 2016

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Omar Otoniel Flores Cortez, • German Antonio Rosa Castellanos.

Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

Carlos Antonio Aguirre Ayala • Marvin Elenilson Hernández Montoya.

***Big data*, análisis de datos en la nube**

Verónica Idalia Rosa • José Guillermo Rivera.

Vicerrectoría de Investigación y Proyección Social

Licda. Noris Isabel López Guevara

Vicerrectora

Licda. Camila Calles Minero

Directora de Investigaciones

Licda. Evelyn Reyes de Osorio

Diseño y Diagramación

Noel Castro

Corrector

PRIMERA EDICIÓN

100 ejemplares

Mayo, 2016

Impreso en El Salvador

Por Tecnoimpresos, S.A. de C.V.

19 Av. Norte, No. 125, San Salvador, El Salvador

Tel.:(503) 2275-8861 • gcomercial@utec.edu.sv

Índice

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Resumen.....	11
1. Introducción	13
1.1 Problema investigado	13
1.2 Justificación	14
1.3 Objetivos del estudio	15
Objetivo general o principal.....	15
Objetivos específicos.....	15
2. Marco teórico	15
2.1 Internet de las cosas.....	15
¿Cómo funciona el IoT?	15
2.2 Sistemas de generación fotovoltaica	19
2.3 Plataformas para sistemas electrónicos embebidos.....	21
Plataforma de desarrollo Arduino.....	22
Tarjeta Raspberry Pi	23
2.4 Plataforma para Iot: Temboo	24
3. Metodología	26
3.1 Método	26
3.2 Tipo de estudio.....	26
3.3 Sujeto de estudio	26

4. Resultados	27
4.1 Implementación de un sistema IoT	27
4.2 Herramientas utilizadas	28
4.3 Experimentación.....	29
5. Conclusiones y recomendaciones	32
6. Referencias bibliográficas.....	33

Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

Agradecimientos y Resumen	34
1. Introducción	35
2. Objeto de estudio	36
2.1 Planteamiento del problema	37
2.2 Justificación.....	36
2.3 Objetivos	37
2.3.1 General.....	37
2.3.2 Específicos	37
2.4 Alcances.....	37
3. Marco teórico referencial	37
3.1 La educación del futuro.....	37
3.2 Entorno de aprendizaje virtual	39
3.2.1 La educación a distancia	39
3.2.2 <i>E-learning</i>	40

3.2.3 Plataformas de <i>E-learning</i>	42
3.2.3.1 Plataformas de <i>E-learning open source</i>	46
3.2.3.2 Plataformas de <i>E-learning</i> comerciales	51
3.3 Sistemas de administración de contenido.....	55
3.3.1 ¿Qué son?	55
3.3.2 Ventajas de un CMS	56
3.3.3 Principales CMS de código abierto.....	57
3.3.3.1 Drupal	58
3.3.3.2 Wordpress.....	59
3.3.3.3 Joomla	60
4. Metodología de la investigación	61
4.1 Metodología	61
4.2 Sujetos de estudio	61
4.3 Instrumentos que se han de utilizar para recolección de datos.....	61
4.4 Procedimiento	62
5. Resultado.....	62
5.1 Implementación de un entorno de aprendizaje virtual integrando herramientas de <i>E-learning</i> y CMS	62
5.1.1 Selección de las plataformas que se han de utilizar.....	62
5.1.2 Requisitos de instalación	64
5.1.2.1 Para instalar Joomla	63
5.1.2.2 Para instalar Moodle	64
5.1.2.3 Acerca de Joomla	65

5.1.3 Instalación de Joomla	66
5.1.3.2 Creación de base de datos	71
5.1.3.3 Instalación del portal Joomla	74
5.1.4 Instalación de Moodle	78
5.1.4.1 Proceso de instalación.....	79
5.1.5 Instalación de Joomla	86
5.1.5.1 Proceso de instalación.....	86
6. Conclusiones.....	100
7. Bibliografía	102

Big data, análisis de datos en la nube

Resumen.....	105
Abstract.....	105
1. Introducción	106
2. Objeto de Estudio.....	109
2.1 Planteamiento del problema	109
2.2 Justificación	110
2.3 Objetivos.....	112
2.3.1 General	112
2.3.2 Específicos.....	112
2.4 Alcances.....	112
2.5 Delimitación	113
3. Marco Teórico.....	113

3.1 ¿Qué es <i>Big Data</i> ?	113
3.2 Tipos de datos	115
3.2.1 ¿Qué tipo de datos se deben explorar en <i>Big Data</i> ?	117
3.3 Componentes de una plataforma <i>Big Data</i>	119
3.3.1 Principales distribuciones de Hadoop	122
3.4 Investigación y ejemplos de aplicación de <i>Big Data</i>	126
3.4.1 Ventajas de utilizar herramientas de <i>Big Data</i>	129
3.4.2 Desventajas de utilizar herramientas de <i>Big Data</i>	129
3.5 Teorías	131
3.5.1 Potencial <i>Big Data</i>	131
3.6 Contexto	133
3.7 Descripción de las herramientas que han de utilizarse en la investigación	135
3.7.1 Qué es Hadoop	135
3.7.2 Qué es R	141
3.7.3 Herramientas de visualización	142
4 Metodología de la Investigación	143
4.1 Metodología	143
4.2 Participantes	144
4.3 Instrumento	144
4.4 Procedimiento	144
5. Resultados de la Investigación	145
6. Conclusiones	151

7. Propuesta: “Metodología analítica y visualización para el control de productos alimenticios y medicamentos”	152
7.1 Desarrollo y Metodologías utilizadas.....	152
7.1.2 Almacenando y procesando datos con Hadoop	153
7.1.3 Analizando datos con R	154
7.1.4 Visualizando datos	154
7.2. Identificación de requisitos.....	157
7.3 Desplegando Hadoop	159
7.3.1 Trabajando con Hadoop.....	174
7.3.2 Uso de Hive.....	177
7.4 Utilización de R.....	181
7.4.1 Instalación de R.....	181
7.4.2 Utilización de R para el análisis estadístico de los datos	186
7.5 Visualización de la Información.....	193
8. Bibliografía	203
9. ANEXOS.....	206

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Omar Otoniel Flores Cortez,
Docente investigador Utec. omar.flores@utec.edu.sv
German Antonio Rosa Castellanos,
Docente investigador Utec. german.rosa@utec.edu.sv

Resumen

El Internet ha estado dominado por contenidos (correos electrónicos, páginas, blogs, tweet, Facebook, fotografías, etc.) que son generados y posteados por personas, quienes “suben” esta información a la red para ser consumida por otros usuarios. Pero existe una nueva tendencia en internet de que estos contenidos no sean posteados por un humano, sino que por el aparato, maquina o “cosa” que los genera. Esta es la premisa detrás del llamado *internet de las cosas* (IoT, por sus siglas en inglés).

Es decir, un electrodoméstico nos pueda enviar un correo electrónico indicándonos aspectos de su funcionamiento. Imaginemos que el refrigerador nos manda un tuit cuando detecta que se están acabando los huevos y así podremos decidir pasar por el supermercado a comprar más.

Siguiendo la premisa descrita, se formuló la investigación que se basa en el diseño de un sistema electrónico que permita a un panel solar publicar en una página web, accesible desde cualquier punto de internet, aspectos de funcionamiento y generación eléctrica de voltaje.

Se diseñó e implementó un sistema embebido electrónico basado en el microcontrolador ATmega32u4, usando una plataforma de desarrollo Arduino modelo YUN; junto con esta se utilizó un sensor que mide el voltaje generado por el panel solar. El resultado de esta investigación fue el prototipo electrónico junto con la programación en la nube, que permite que el panel solar publique automáticamente su voltaje en un sitio web, o nube, para ser accesible desde cualquier dispositivo con conexión a internet.

Las aplicaciones pueden ser muchas y en otras áreas (aparatos). Este es el inicio de una serie de investigaciones destinadas al estudio de los sistemas embebidos en aplicación del IoT.

Palabras clave: Internet de las cosas, sistemas fotovoltaico, Andruino

Abstract

The Internet has been dominated by content (emails, pages, blogs, tweet, Facebook, photographs, etc.) that are generated and posted by people who “upload” this information to the network for consumption by other users. But there is a new trend in Internet Contender these are not posted by a human but by the device, machine or “thing” that generates them. This is the premise behind the so-called internet of things (IoT, by its acronym in English).

That is, for example, an appliance can send us an email indicating aspects of its operation. Imagine the refrigerator sends us a tweet when it detects that are running eggs and so we decide to go through the supermarket to buy more.

Following the premise described research is based on the design of an electronic system that allows a solar panel post on a web page, accessible from any internet point, aspects of operating voltage and power generation was developed. During the investigation, it was designed and implemented an electronic em-

bedded system based on microcontroller ATmega32u4, using an Arduino development platform YUN model; along with this a sensor that measures the voltage generated by the solar panel was used. It is this information that is automatically published on a website, or cloud to be accessible from any device with internet access.

The result of this research was the prototype electronic programming with the cloud, which allows the solar panel automatically publish your voltage generated on a website, which can be accessed from any internet point.

Forward-looking are many applications in other areas (equipment) and this is the beginning of a series of investigations for the study of embedded systems in application of IoT.

Keywords: application of IoT, photovoltaic systems , Andruino

1. Introducción

1.1 Problema investigado

Durante los últimos años, en El Salvador ha crecido el uso de sistemas de energías renovables para la generación de electricidad, específicamente ha proliferado la instalación de sistemas de generación fotovoltaica, aquellos que están constituidos por paneles que captan la energía luminosa del Sol y la transforman en energía eléctrica.

La instalación y puesta en marcha de un sistema de generación fotovoltaica son complicadas. La mayoría de estos paneles son instalados e inmediatamente comienzan a trabajar y producir energía (voltaje DC), con la salvedad que esta producción está sujeta a la disponibilidad de luz solar (a más irradiación solar mayor voltaje). El Salvador es un país con un porcentaje de irradiación solar alto, por lo que este tipo de sistemas son más eficientes frente a otro tipo de sistemas de producción de energía renovable, como el eólico.

Al instalar un panel solar, una de las características de operación que es muy

útil conocer es el nivel de electricidad producida a lo largo de un periodo, esto con el objetivo de verificar la eficacia y eficiencia del sistema de generación fotovoltaico, y así tomar medidas de mantenimiento correctivo y preventivo. La mayoría de los sistemas instalados en la actualidad, así como de los disponibles en el mercado comercial local, no cuentan con un sistema electrónico que registre e informe sobre los datos de producción de electricidad del panel solar a lo largo del día.

1.2 Justificación

A partir de la problemática anteriormente descrita, se hace necesaria la creación de sistemas electrónicos embebidos que se adapten a la salida de un panel solar y sean capaces de intercambiar información y contenido con otros sistemas y/o usuarios, a través de un medio de acceso remoto, como un sitio o página web, y ya que hoy en día es más factible la disponibilidad de conexión a internet a través de un sinfín de dispositivos, lo es más aún.

Disponer de un reporte de datos de electricidad generada por un panel solar es de suma importancia, ya que permite conocer su desempeño; y más aún si esta información está disponible para ser consultada desde cualquier punto con acceso a internet, usando una computadora, tableta o teléfono móvil, por lo que se necesita tener datos del rendimiento que el sistema instalado tiene y así verificar su eficacia, para la toma de decisiones a corto y mediano plazo.

Por lo que se plantea el diseño de un sistema electrónico que permita la conectividad de la salida de un panel solar a internet, y así que estos datos de producción de voltaje puedan ser monitoreados desde cualquier punto con acceso a internet.

Con la implementación de esta propuesta se busca lograr los siguientes beneficios:

- Monitorear desde cualquier dispositivo con conexión a internet, el nivel de voltaje DC producido por un panel solar.
- A partir de este monitoreo, tomar de decisiones para el mantenimiento preventivo o correctivo en el desempeño del SGV.

- La ventaja competitiva de disponer de un sistema de bajo costo y eficiente.
- Crear un dispositivo electrónico que se podrá adaptar a otro tipo de maquinarias, para su monitoreo remoto.

1.3 Objetivos del estudio

Objetivo general:

Implementar un sistema electrónico para la interconexión de un sistema fotovoltaico, que permita realizar un monitoreo en tiempo real, desde internet, del rendimiento de este sistema, y así poder establecer factores de desempeño a lo largo de su uso.

Objetivos específicos

- Diseñar un circuito electrónico basado en la aplicación de sistemas embebidos en la solución propuesta.
- Diseñar un *firmware* o programa informático exclusivo para el funcionamiento del prototipo electrónico, escrito en lenguaje C, que sea capaz de controlar de dicho prototipo.
- Construir y acoplar el prototipo diseñado a un sistema fotovoltaico.

2. Marco teórico

2.1 Internet de las cosas

Internet de las cosas es la expresión utilizada para designar el que posiblemente será el siguiente desarrollo tecnológico transcendental de la era de la información. Los usuarios se han acostumbrado a que internet les permita estar conectados y tener acceso desde casi cualquier rincón del planeta, y casi en cualquier momento. Internet de las cosas es la propuesta para conectar a la red un número casi infinito de objetos cotidianos. Edificios que controlan su temperatura y gastos, ropas que informan sobre el estado de su usuario, ciu-

dades con sistemas de aviso y control, enfermedades monitorizadas o incluso ecosistemas y parajes naturales ofrecerán y recibirán ahora información a través de internet. Con la utilización de distintos sensores y dispositivos, y con el crecimiento de las infraestructuras de redes y direcciones de conexión, serán millones los nuevos objetos que podrán conectarse a la red, con las personas y otros objetos.

El IoT se refiere a los objetos físicos y virtuales que tienen identidades únicas y están conectados a internet para facilitar las aplicaciones inteligentes que hacen de la energía, logística, del control industrial, del comercio minorista, de la agricultura y muchos otros ámbitos “más inteligentes” (Madisetti).

Según Osswald (2014), a la IoT la llaman *computación oblicua*, *inteligencia ambiental* o *computación omnipresente*. En realidad todos estos términos son válidos. Sin embargo, una forma simple de describir y comprender el IoT es concebir esta como “sensores conectados”, que juntos proveen la comunicación digital necesaria para un nuevo horizonte de oportunidades tecnológicas, por ejemplo, como los siguientes:

- Sensores conectados al cuerpo. Estos son las computadoras para la salud y el bienestar.
- Sensores conectados a otros objetos. Esta es la forma como el producto provee una retroalimentación de su operación, eficiencia o progreso.
- Sensores que monitorean el medio ambiente. De esta manera se podrá conocer sobre el ambiente en que las personas o cosas se mueven o existen.

Estos sensores están conectados para reportar magnitudes como las siguientes: locación, movimiento, aceleración, temperatura, luz, color, imagen, sonido, calidad del aire, apertura y cierre de interruptores, signos vitales corporales, etc.

Para entender de qué va el IoT, debemos también comprender que sus fundamentos no son en lo absoluto nuevos. Desde hace unos 30 años es que se viene trabajando con la idea de hacer un poco más interactivos todos los objetos

de uso cotidiano. Ideas como el hogar inteligente, también conocido como *la casa del mañana*, han evolucionado antes de que nos demos cuenta en el hogar conectado para entrar al IoT.

El IoT potencia objetos que antiguamente se conectaban mediante circuito cerrado, como comunicadores, cámaras, sensores y demás, y les permite comunicarse globalmente mediante el uso de la red de redes: internet.

Estamos sufriendo una transformación donde “cosas” hablan con otras “cosas”, y estas conversaciones están creando nuevos modelos de negocio, productos y compañías. Los expertos prevén que para el año 2020 habrá un total de 50 mil millones de dispositivos/cosas conectadas a internet. (Madiseti).

Hace 20 años, internet se usaba principalmente como herramienta para buscar información. En los últimos 10 años hemos vivido una nueva forma de uso de internet, donde todo se ha convertido en social, transaccional y móvil.



Figura 2.1- Algunas aplicaciones de los sistemas IoT
(Fuente: <http://www.farodevigo.es>)

¿Cómo funciona el IoT?

Se trata de *chips* y circuitos que comparados con, por ejemplo, un Smartphone, podrían parecernos muy rudimentarios, pero que cuentan con todas las herramientas necesarias para cumplir labores especializadas muy específicas.



Figura 2.2. Infografía que muestra los diversos estados de un sistema IoT

No hay un tipo específico de objetos conectados al IoT, en lugar de eso, a los sí conectados se les puede clasificar como objetos que funcionan como sensores y objetos que realizan acciones activas. Claro, los hay que cumplen ambas funciones de manera simultánea.

En cualquier caso el principio es el mismo, y la clave es la operación remota. Cada uno de los objetos conectados a internet tiene una IP específica; y mediante esta acceder para recibir instrucciones. Así mismo, puede contactar con un servidor externo y enviar los datos que recoja.

Por ejemplo, es el sector privado donde el IoT se está haciendo cada vez más popular.

- *La industria de producción en masa.* La maquinaria que se encarga de controlar los procesos de fabricación, robots ensambladores, sensores de temperatura, control de producción, todo está conectado a internet en cada vez más empresas, lo que permite centralizar el control de la infraestructura.
- *Control de infraestructura urbana.* Control de semáforos, puentes, vías de tren, cámaras urbanas. Cada vez más ciudades implementan este tipo de infraestructuras basadas en el IoT que permiten monitorear el correcto funcionamiento de sus estructuras, además de adaptar más flexiblemente su funcionamiento ante nuevos eventos.
- *Control ambiental.* Esta es una de las áreas en las que está teniendo más éxito el IoT, pues permite acceder desde prácticamente cualquier parte a información de sensores atmosféricos, meteorológicos y sísmicos.
- *Sector salud.* Cada vez más clínicas y hospitales alrededor del mundo confían en sistemas que le permiten al personal de salud monitorear activamente a los pacientes de manera ambulatoria y no invasiva.

También hay aplicaciones del IoT para el transporte, la industria energética y prácticamente para todos los sectores comerciales. Como hemos dicho, el gran pendiente es el mercado de consumo o, lo que es lo mismo, los hogares, un lugar en el que probablemente solo es cuestión de tiempo para que veamos allí la gran explosión del IoT.

2.2 Sistemas de generación fotovoltaica

La energía solar eléctrica, o fotovoltaica, es una energía limpia y renovable, de fácil instalación y mantenimiento. Aunque esta solo representa el 0,001 % del suministro de energía eléctrica que satisface las necesidades de consumo en todo el mundo, se prevé un rápido y significativo crecimiento de su implantación, basado en el actual desarrollo de la tecnología y el compromiso

medioambiental de los países más desarrollados. El sector fotovoltaico se sustenta en una tecnología de vanguardia y en una industria puntera que en los últimos años está teniendo un crecimiento anual medio superior al 30 %. En el medio plazo, se estima que habrá una reducción importante de costes debido a una mejora de la eficiencia de las tecnologías actuales, a la optimización de los procesos de fabricación, a la aplicación de economías de escala y al desarrollo de nuevas tecnologías.

El funcionamiento de un sistema fotovoltaico se logra mediante el siguiente proceso.

La luz solar entra sobre la superficie del arreglo fotovoltaico, donde es convertida en energía eléctrica de corriente directa por las celdas solares, después esta energía es recogida y conducida hasta un controlador de carga con la función de enviar a toda o parte de esta energía hasta el banco de baterías en donde es almacenada, cuidando de que no se excedan los límites de sobrecarga y sobre descarga.

La energía almacenada o enviada a la red se utiliza para abastecer las cargas durante la noche o en días de baja insolación, o cuando el arreglo fotovoltaico es incapaz de satisfacer la demanda por sí solo. Si las cargas que se han de alimentar son de corriente directa, estas pueden hacerse a través del arreglo fotovoltaico o desde la batería. Cuando las cargas son de corriente alterna, la energía proveniente del arreglo y de las baterías, limitadas por el controlador, es enviada a un inversor de corriente, en donde es convertida a corriente alterna.

El elemento principal de un sistema de energía fotovoltaica es el panel solar, que es un dispositivo construido de silicio (extraído de la arena común). La celda solar es un dispositivo de conversión directa, esto es, tiene la capacidad de transformar la radiación solar directamente en electricidad. Las celdas solares funcionan gracias al efecto fotoeléctrico.

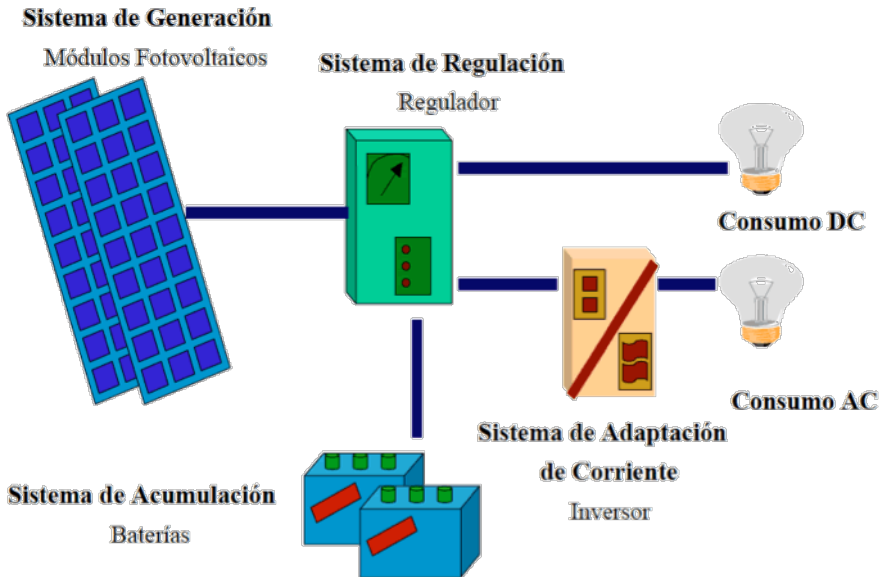


Figura 2.3. Infografía que muestra los diversos componentes de un sistema fotovoltaico

2.3 Plataformas para sistemas electrónicos embebidos

Aunque el ser humano todavía no está arraigado a las propiedades que ofrece la domótica, es un hecho de que en un futuro estará instalada en cualquier vivienda. Pero ¿qué es la domótica? Se podría definir como el conjunto de tecnologías aplicadas al control y la automatización inteligente de la vivienda, que permite una gestión eficiente del uso de la energía además de aportar seguridad, comodidad y comunicación entre el usuario y el sistema. Para conseguir las propiedades comentadas anteriormente, es necesario que los sistemas recojan la información de su entorno con sensores y dispongan de la lógica para actuar en consecuencia utilizando actuadores. Actualmente los sistemas domóticos tienen un precio muy alto de instalación, por lo cual solo es posible verlos en casas de lujo. Estos suelen utilizar buses de transmisión de información que posibilitan una domótica robusta, como son el EIB, X10, CEBus, LonWorks/LongTalk y ZigBee. Una alternativa más barata y casera consiste en la utilización de placas Arduino.

Plataforma de desarrollo Arduino:

En este apartado se describirían los principales elementos que componen una placa Arduino y el entorno de desarrollo en el que se programa el código, es decir, la parte *hardware* y la de *software* que actúan sobre Arduino.

Al ser Arduino una plataforma de *hardware* libre, tanto su diseño como su distribución, puede utilizarse libremente para el desarrollo de cualquier tipo de proyecto sin necesidad de adquirir licencia. Por eso existen varios tipos de placa oficiales, las creadas por la comunidad Arduino; o las no oficiales, creadas por terceros pero con características similares. En la placa Arduino es donde conectaremos los sensores, actuadores y otros elementos necesarios para comunicarnos con el sistema.

Arduino Yun:

Esta es una plataforma electrónica perfecta para utilizar en el diseño de los dispositivos conectados y, más en general, proyectos en IoT. Se combina el poder de Linux con la facilidad de uso de Arduino.

El Arduino Yun es una placa Arduino, a diferencia de cualquier otro. Durante la programación es muy similar a la Arduino Leonardo y utiliza el mismo procesador, el Atmel ATmega32U4. También tiene un procesador adicional, un AtherosAR9331, corriendo Linux y el OpenWrt pila inalámbrica. La programación del 32U4 a través de USB es idéntica a la Arduino Leonardo. Una vez que se haya configurado el Yun para conectarse a su Wi-Fi a la red, se puede programar el 32U4 vía Wi-Fi también.

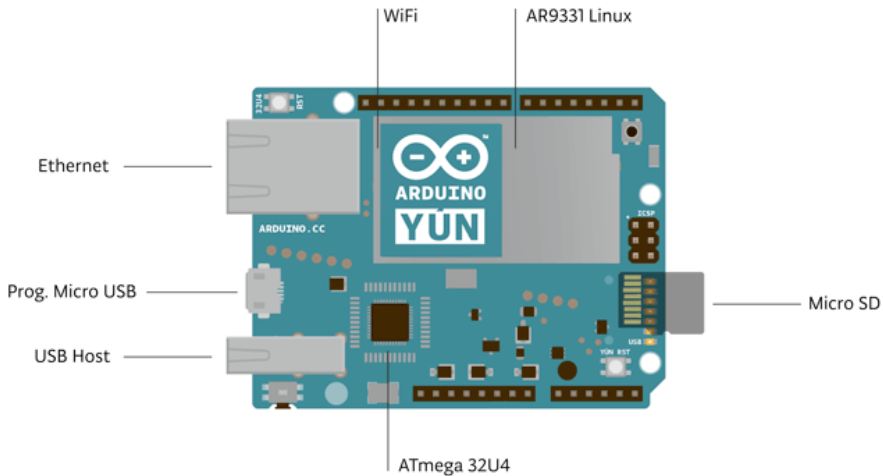


Figura 2.4. Imagen que muestra las diversas partes y componentes de una tarjeta *Arduino Yun*

Tarjeta Raspberry Pi:

Raspberry Pi es un es una computadora del tamaño de una tarjeta de crédito que se conecta a un televisor y a un teclado, Es una pequeña tarjeta computacional que puede ser utilizada por muchas de las cosas que una PC de escritorio hace, como hojas de cálculo, procesadores de texto y juegos. También reproduce vídeo de alta definición (Upton, 2014).

La tarjeta Raspberry PI tiene varios puertos y entradas, dos USB, uno de Ethernet, y salida HDMI. Estos puertos permiten conectar el miniordenador a otros dispositivos, teclados, ratones y pantallas. También posee un *System on Chip*¹ que contiene un procesador ARM que corre a 700 MHz, un procesador gráfico Video Core IV y hasta 512 MG de memoria RAM. Es posible instalar sistemas operativos libres a través de una tarjeta SD.

1 Tecnologías de fabricación que integran todos o gran parte de los componentes de un computador electrónico en un único circuito integrado o chip.

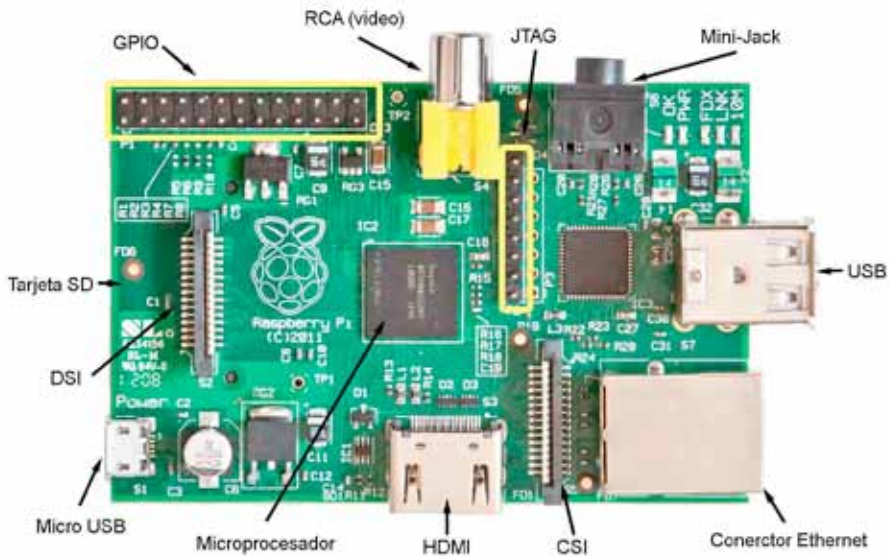


Figura 2.5. Imagen que muestra las diversas partes y componentes de una tarjeta Raspberry Pi (Fuente: zonamaker.com)

2.4 Plataforma para Iot: Temboo

Temboo es una interface de desarrollo que provee una serie de API,² que permite conectar la placa Arduino a internet y brindar un acceso desde este.

La tecnología de Temboo permite ejecutar cualquiera de estos procesos con solo unas pocas líneas de código, un número mucho menor de las que normalmente se necesitan para la misma tarea al usar implementación con lenguajes de programación de alto nivel. Aún mejor, el código que se utiliza con Temboo se estructura consistente y se puede generar en vivo en nuestra página web para copiar y pegar en su IDE, por lo que es fácil e intuitivo para probar y utilizar todos nuestros procesos inteligentes.

2 Es el conjunto de subrutinas, funciones y procedimientos que ofrece cierta biblioteca para ser utilizado por otro *software* como una capa de abstracción.

La plataforma de Temboo hace que sea increíblemente fácil desarrollar y mantener aplicaciones de *software* avanzadas; y es especialmente adecuado para la programación para IoT.

El almacenamiento en nube (del inglés *cloud storage*), es un modelo de almacenamiento de datos basado en redes de computadoras, en las que los datos están alojados en espacios de almacenamiento virtualizados, por lo general aportados por terceros.

Las compañías de alojamiento operan enormes centros de procesamiento de datos. Los usuarios que requieren estos servicios compran, alquilan o contratan la capacidad de almacenamiento necesaria. Los operadores de los centros de procesamiento de datos, a nivel servicio, virtualizan los recursos según los requerimientos del cliente; solo exhiben los entornos con los recursos requeridos. Los clientes administran el almacenamiento y el funcionamiento de los archivos, datos o aplicaciones. Los recursos pueden estar repartidos en múltiples servidores físicos.

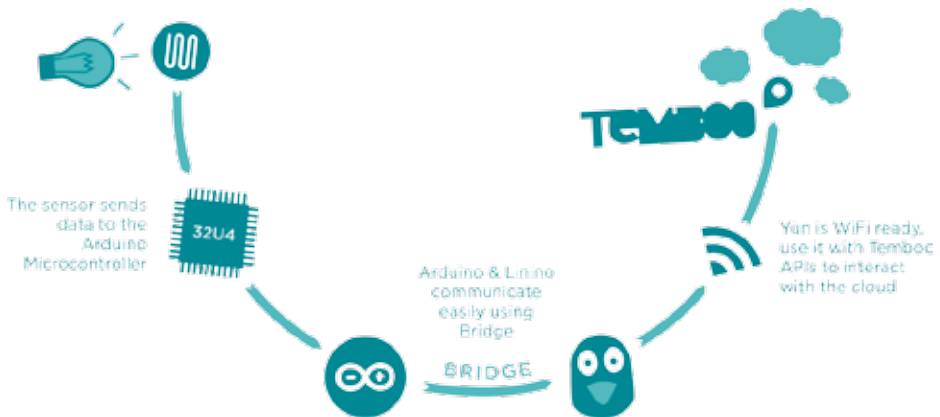


Figura 2.6. Esquema simplificado de una aplicación IoT utilizando la plataforma Temboo (Fuente: temboo.com)

3. Metodología

3.1 Método

La investigación fue desarrollada usando el método experimental, el cual es “un método empleado en casos donde la investigación tiene por objeto el provocar determinados fenómenos que no se presentan usualmente en la naturaleza y cuyo conocimiento puede ser interesante o importante en el avance de la ciencia y la tecnología” (Cegarra, 2004).

Específicamente en esta investigación, y a partir de la revisión teórica, se seleccionaron diversas herramientas y componentes electrónicos para el diseño de un prototipo electrónico que se apegara a los objetivos del estudio.

3.2 Tipo de estudio

Este estudio fue del tipo investigación aplicada tecnológica o investigación técnica. Según Cegarra (2004), este tipo de investigación “tiende a la resolución de problemas o al desarrollo de ideas, a corto o mediano plazo, dirigidas a conseguir innovaciones, mejoras en procesos o productos, incrementos de calidad y/o productividad, etc.”.

3.3 Sujeto de estudio

Se utilizó el panel solar instalado en el sistema de generación fotovoltaico del laboratorio de electrónica de la Utec, con la posibilidad de que el prototipo desarrollado puede ser adaptado fácilmente en cualquier panel solar.

Dicho panel ya está trabajando (instalado y produciendo energía eléctrica) y no dispone, de fábrica, de una conexión a internet, por lo que, a partir de la investigación bibliográfica realizada, se decidió diseñar un sistema electrónico embebido que se adaptara a la salida de producción de voltaje del panel y que fuera capaz de enviar esta información vía internet a un sitio web para su monitoreo desde cualquier punto con acceso.

4. Resultados

4.1 Implementación de un sistema IoT

A partir de lo descrito anteriormente, técnicamente existen las siguientes dos opciones para permitir que un aparato, dispositivo, máquina, etc., pueda estar conectado a internet e intercambiar datos con otros usuarios: 1) incluir en el diseño primario y en la fabricación del propio dispositivo de estos circuitos e interfaces que permitan dicha conexión a internet y 2) dotarlo de un circuito electrónico externo pos fabricación del dispositivo o máquina que le brinde la capacidad de comunicación vía internet.

Todo el sistema de monitoreo del panel se diseñó en las siguientes cuatro etapas:

1. Adquisición de datos de producción del voltaje del panel (*hardware*: sensor electrónico).
2. Interface de procesamiento y conexión a internet (*hardware*: micro-controlador).
3. Plataforma informática en internet para la colección de estos datos (*software*: servicio de almacenamiento en internet o en la “nube”).
4. Sitio web para la presentación de los datos (*software*: *web hosting* de Google Sites).

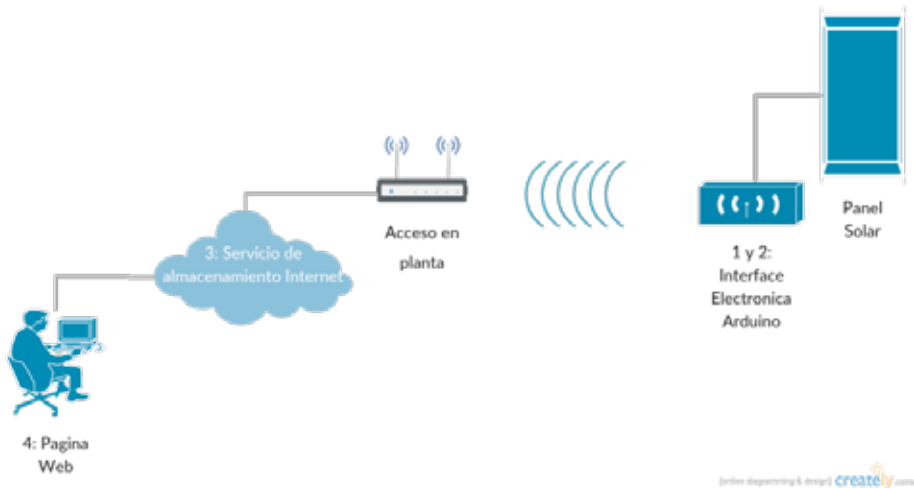


Fig. 4.1. Esquema de bloques que conforman el sistema diseñado para el monitoreo desde un sitio web del voltaje producido por el panel solar

4.2 Herramientas utilizadas

Para la implementación del sistema diseñado se utilizaron las herramientas de última generación y que, a su vez, son de bajo costo, sin dejar de lado aspectos como la familiaridad y experiencia con estas.

Para el bloque de *Adquisición de datos* se utilizó un sensor de voltaje de tipo activo- analógico apto para la medición de la DC en el rango de 1 a 15 voltios. Específicamente se usó el sensor con referencia Pow05161p, cuya salida es un nivel de voltaje apto para ser llevado a la siguiente etapa.

En el bloque de *Procesamiento y conectividad* se utilizó una tarjeta Arduino modelo YUN, la cual incluye un microcontrolador ATmega32u4 junto a un transeptor o modem para conexiones Wi-Fi, además, se diseñó un *firmware* en lenguaje Arduino C para que la tarjeta cumpliera la función de procesar los datos del sensor y al mismo tiempo hiciera enlaces vía conexión Wi-Fi con el servicio de almacenamiento en internet (*cloud computing*³).

³ Concepto que se basa en que las aplicaciones *software* y los equipos *hardware* con capacidad de proceso y almacenaje; no están en el equipo del usuario, sino que están ubicados en un centro de datos que permite a los usuarios acceder a las aplicaciones y

Para implementar el tercer bloque del sistema se utilizó la plataforma Temboo como “servicio en la nube”, para el almacenamiento en internet de los datos del voltaje producido por el sistema de procesamiento y captura de voltaje. Esta plataforma permite la creación de una base de datos y de una interface entre estos datos almacenados y la página web destinada a la visualización, además, Temboo permite el uso de las API de las aplicaciones de los servicios de Google, lo cual es muy beneficioso para el sistemas, ya que la Utec cuenta con la plataforma Google para sus servicios estudiantiles y docentes. El servicio de Temboo se programó en conjunto con la plataforma Arduino para capturar, almacenar y transferir la información, a través de la API de Google, hacia la última etapa del sistema.

El sitio web para la presentación de los datos de voltaje del panel solar fue diseñado y programado usando la herramienta Google Sites. Como se mencionó anteriormente, se eligió esta opción por ser los servicios que usa la Utec.

4.3 Experimentación

El sistema completo fue implementado con éxito en el panel solar del Laboratorio de Electrónica de la Utec, se programó para generar una medición cada media hora, y así poder visualizar cómo varía su producción de voltaje a lo largo de un día, una semana, una estación, etc.

El sistema implementado permite que un usuario que acceda a <https://sites.google.com/a/mail.Utec.edu.sv/labeleUtec/panelsolar> y pueda observar en tiempo real una gráfica con los datos del voltaje producido por el panel solar del Laboratorio de Electrónica de la Utec.

A continuación se muestran algunos gráficos y fotografías del sistema electrónico implementado.

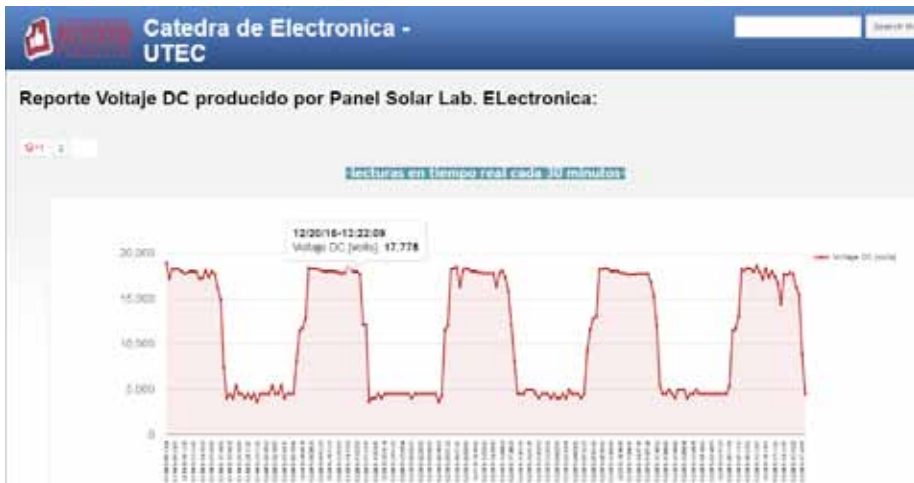


Fig. 4.2. Captura del gráfico de reporte de voltaje producido en un día típico, (disponible en <https://sites.google.com/a/mail.Utec.edu.sv/label/Utec/panelsolar>)



Fig. 4.3. Prototipo de monitoreo (medio) instalado junto a la acometida del panel solar(izq.)



Fig. 4.4. Detalle del prototipo de monitoreo instalado.



Fig. 4.5. Vista frontal prototipo de monitoreo instalado junto a la acometida del panel solar

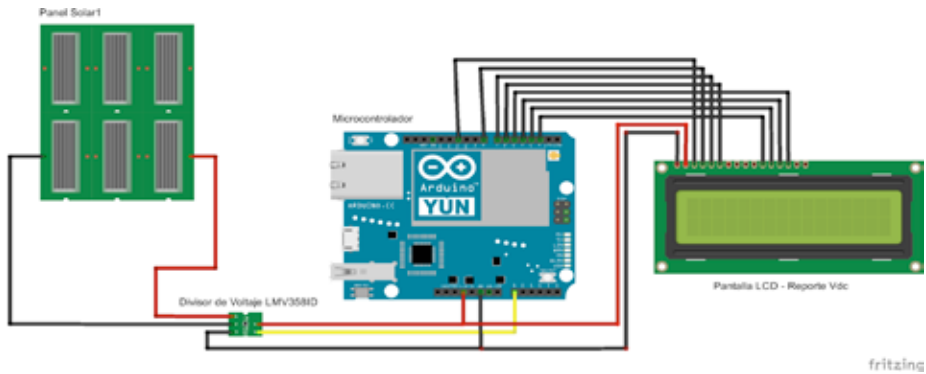


Fig. 4.6. Detalle de la conexión de los componentes del prototipo de monitoreo.

5. Conclusiones y recomendaciones

El sistema cumple con el objetivo de permitir al panel solar ser monitoreado desde cualquier punto o dispositivo con acceso a internet, con lo cual se ha demostrado que se pueden implementar soluciones electrónicas de este tipo con el objetivo de monitoreo remoto del desempeño de instalaciones similares a la estudiada en esta investigación.

Dentro de los logros de esta investigación se pueden citar los siguientes:

- Se realizó una investigación teórica sobre el estado del arte de las tecnologías utilizadas en el desarrollo de sistemas aplicado al IoT.
- Se hizo una selección de los componentes que mejor se adaptan a los objetivos de funcionamiento del sistema, que cumplen con ser de bajo costo, funcionales y actualizados.
- Se diseñó, programó e implementó un sistema electrónico embebido que se adaptó a la salida de voltaje del panel solar.
- Se diseñaron y programaron las interfaces en *software* necesarias para poder acceder desde internet al monitoreo de voltaje producido por el panel solar.

Se recomienda continuar con el apoyo a investigaciones de este tipo, ya que es un campo poco explorado dentro de la Utec, del que se vislumbra que tiene mucho potencial dentro de nuestro ámbito.

Se propone continuar con la realización de más investigaciones dentro el campo de los sistemas electrónicos embebidos en aplicaciones en el área del IoT, por ejemplo: control de sistemas de asistencia estudiantil, monitoreo de pacientes en hospitales, alertas de posibles fugas en cámaras de suministros, etc.

6. Referencias bibliográficas

- Adamo, F.; Attivissimo, F.; Cavone, G.; Guarnieri Calo Carducci, C., & Lanzolla, A.M.L. (2015, May). "New technologies and perspectives for laboratory practices in measurement science". In *Instrumentation and Measurement Technology Conference (I2MTC), 2015 IEEE International* (pp. 1-6). IEEE.
- Cegarra Sánchez, J. (2004). *Metodología de la investigación científica y tecnológica*. Madrid, Díaz de Santos.
- Da Costa, F. (2014). *Rethinking the Internet of Things*. Santa Clara: Apress.
- Gustafson, S., & Sheth, A. (2014). Web of Things. *Computing Now*, 7(3).
- Griffiths, J.R. and Brophy, P. (2005), "Student searching behaviour and the web: use of academic resources and Google", *Library Trends*, Vol. 53 No. 4, pp. 539-54.
- Bahga, A., & Madiseti, V. (2014). Internet of Things: A Hands-on Approach. VPT.
- Mercado, G.; Diedrichs, A., & Aguirre, M. (2011). "The Wireless Embedded Internet". *Annals of CASE*.
- Radoš, I. (2007). *Google API Maps*. Ekscentar, (9), 52-53.
- Osswald, M. (2014, febrero 6). "Hanson Inc.". Retrieved: agosto 10, 2015, from <http://www.hansoninc.com>
- Upton, E., & Halfacree, G. (2014). *Raspberry Pi user guide*. John Wiley & Sons.
- Vega, A.M.; Santamaría, F., & Rivas, E. (2014). "Internet de los objetos empleando arduino para la gestión eléctrica domiciliaria". *Revista EAN*, (77), 24-40.

Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

Carlos Antonio Aguirre Ayala,
Docente investigador Utec. carlos.aguirre@utec.edu.sv,
Marvin Elenilson Hernández Montoya,
Docente investigador Utec. marvin.hernandez@utec.edu.sv.

Agradecimientos

Estamos muy agradecidos con la Licda. Noris Isabel López Guevara, vicerrectora de Investigación y Proyección Social; y con la Licda. Camila Calles, directora de Investigaciones, de la Universidad Tecnológica de El Salvador, por el patrocinio por parte de la universidad de esta investigación y por la confianza y apoyo para que este trabajo pudiera realizarse. No fue un trabajo fácil, pero con el esfuerzo del equipo se logró finalizar.

Resumen

El documento que se presenta a continuación tiene como objetivo describir las diferentes plataformas de *E-learning*, así como las soluciones de administradores de contenido (CMS) existentes en el mercado. Se implementará un portal virtual integrando Moodle y Joomla a través de una extensión llamada *Joomla*, ya utilizada en el mundo. Con este tipo de soluciones se podrá proporcionar al usuario más facilidad y una mejor experiencia de aprendizaje al momento de realizar cursos virtuales.

Palabras claves: plataformas virtuales, *E-learning*, Moodle, Joomla, CMS, Joomla.

Abstract:

The document presented below aims to describe the different platforms *E-learning*, as well as solutions content managers (CMS) on the market. Also, a virtual portal integrating Moodle and Joomla through an extension called *Joomla* be implemented. With this type of solutions, you can provide the user more easily and better learning experience when making online courses.

Keywords: virtual platforms, *E-learning*, Moodle, Joomla, CMS, Joomla.

1. Introducción

La educación es un pilar fundamental para las sociedades en vías de desarrollo, ya que por medio del conocimiento se obtienen los escenarios reales, que se convierten en la fuerza impulsora para la generación de innovación que, a su vez, tiene su origen en la investigación. Gracias a los aportes tecnológicos, las barreras que reducen las posibilidades de acceder a cursos de capacitación y carreras profesionales ahora son menores, ya que los medios en la administración de contenido brindan una serie de servicios enfocados en la oferta de programas curriculares que permiten incrementar el conocimiento y desarrollar competencias en ambientes flexibles de aprendizaje, puesto que no dependen de horarios ni de la ubicación física.

Por ello se estudió la plataforma de E-learning Moodle y el CMS Joomla, integrando la potencialidad de ambos a través de una extensión llamada Joomla, ya utilizada para varias aplicaciones, con el objetivo de mostrar una nueva plataforma educativa con mayor potencial, para el manejo de cursos virtuales, en un entorno más dinámico y amigable al usuario.

2. Objeto de estudio

2.1 Planteamiento del problema

En esta investigación se tendrán como objeto de estudio la plataforma de *E-learning* Moodle y el CMS Joomla. Muchas instituciones que han implementado Moodle como gestor de cursos virtuales han experimentado diversidad de problemas, entre estos se destacan los siguientes:

- ✓ Poca flexibilidad en el diseño de la plataforma Moodle, en el portal de la institución.
- ✓ Los usuarios manejan diferente claves.
- ✓ La información de la institución pasa desapercibida por los usuarios, ya que el aula virtual no está enlazada al portal institucional.

2.2 Justificación

Con la implementación de un portal en Joomla se integra Joomla y Moodle; dos potentes herramientas de código abierto que han evolucionado los entornos web. Joomla es una extensión que permite aprovechar al máximo las virtudes de un administrador de contenido robusto y potente, así como las características de un gestor de cursos virtuales; y al fusionarlas se convierte en un portal institucional que ofrezca servicios de capacitación virtual.

Al integrar ambas plataformas, el usuario del aula virtual podrá:

- ✓ Visualizar sitio web en cualquier plataforma (dispositivos móviles, PC, *laptop*, entre otros).
- ✓ Tener una sola clave de ingreso y facilidad para recuperar su clave si la olvida.
- ✓ Acceder al sitio web, flexible, liviano y vistoso.
- ✓ Visualizar toda la información del portal institucional, dado que ambas plataformas estarán integradas.

2.3 Objetivos

2.3.1 General

Implementar un entorno virtual de aprendizaje integrando herramientas de *E-learning* y CMS, para mostrar sus virtudes en ambientes de aprendizaje.

2.3.2 Específicos

- a) Identificar las características técnicas de Moodle y Joomla para su instalación.
- b) Instalar y configurar Moodle y Joomla.
- c) Instalar y configurar Joomla.
- d) Implementación del portal.

2.4 Alcances

- a) Se instala y configura Moodle y Joomla.
- b) Se instala y configura Joomla.
- c) Se implementa el portal.

3. Marco teórico referencial

3.1 La educación del futuro

Con el cambio de siglo, la tecnología ha evolucionado produciendo una gran transformación en la humanidad. Sin embargo, la implementación de nuevas herramientas tecnológicas en algunos campos produce, en ocasiones, miedo en la mayoría de las personas. Por ejemplo, en el área de educación, a pesar del temor, se están introduciendo de manera paulatina técnicas innovadoras para el aprendizaje en colegios y universidades. Se vive en un tiempo en el que es

inevitable seguir avanzando hacia una educación puramente tecnológica (Faver, 2015).

Estos cambios que se están dando impactan directamente en el modelo educativo, ya que la educación que antes se recibía en las aulas era basada en “no cambio”, con relación a utilización solo de la pizarra, el cuaderno, la explicación con fichas, algo que no tiene congruencia con la forma en la que se vive hoy día (Faver, 2015).

La revolución tecnológica supone cambios en todas las esferas de la vida social, entre estas, en la de la educación. Según la Cumbre Mundial para la Innovación en Educación (Wise, por sus siglas en inglés), en los próximos 15 años los colegios se convertirán en entornos interactivos y el prototipo de escuela tal como la se conoce hoy cambiará. La clase magistral desaparecerá y el profesor se convertirá en guía del alumno. Wise considera que el aprendizaje será personalizado, permanente y más caro. Según la investigación publicada en las redes, la internet será la principal fuente y el inglés la lengua más hablada, y los sistemas educativos de todo el mundo sufrirán grandes modificaciones por la revolución tecnológica (Wise, 2015).

En el futuro próximo, el papel de los profesores será guiar al estudiante por su propia vía de conocimiento. Serán facilitadores y orientadores, más que transmisores del saber. Hasta hace muy poco eran la única fuente de información disponible, pero internet lo ha acaparado todo y los alumnos pueden encontrar en la red buena parte de lo que se explica en clase. El papel de los profesores va a ser aún más relevante. Van a tener que mostrar a los alumnos que hay que ser críticos con la información; que no todo lo que encuentran en internet es correcto; que deben seleccionar y acudir a las fuentes más fiables. En el nuevo escenario de educación se utilizarán nuevas metodologías, como el *flipped classroom*, que consiste en que los estudiantes preparan por su cuenta las clases y realizan una exposición en el aula mientras el profesor realiza una labor de acompañamiento. Este sistema permitirá que una parte del trabajo se haga en casa y que cada alumno siga su ritmo. La clase magistral no será relevante, dado que los contenidos podrán proporcionársele al estudiante en vídeos o en audio (Wise, 2015).

El papel del estudiante cambiará. Este tendrá muchas más posibilidades de acceso a fuentes de conocimiento, con una mentalidad más universal y menos localista; será el protagonista indiscutible de su aprendizaje, un ciudadano global que busca a través del aprendizaje un modo de responder a alguna necesidad del entorno. Todo esto tendrá consecuencias en la relación jerárquica con el profesor; se desdibuja la línea que separa a quiénes son los que enseñan y quiénes son los que aprenden (Wise, 2015).

De acuerdo con estos escenarios las instituciones educativas deben estar preparadas para enfrentar estos nuevos desafíos. El interés por mejorar los portales de educación virtual debe ser una prioridad, ya que el índice de la cantidad de población estudiantil en modalidad virtual aumentará, y las plataformas utilizadas deben reunir los requisitos necesarios para que los participantes de los cursos puedan tener las herramientas necesarias para aprender.

3.2 Entorno de aprendizaje virtual

3.2.1 La educación a distancia

Definir la educación a distancia ha sido complejo para distintos autores, por varias razones: primero, el concepto *distancia* tiene muchos significados; segundo, el concepto *educación a distancia* ha sido utilizado en una gran variedad de programas que han servido a audiencias y públicos numerosos, vía una variedad de gamas de medios. Finalmente, los cambios tan rápidos en la tecnología han establecido un reto a los medios tradicionales, en los que la educación a distancia ha sido definida (Bargas, 2011).

A continuación se presentan tres definiciones de autores que incluyen varios de los tópicos principales de la educación a distancia (Bargas, 2011).

- a) Michael G. Moore: “La enseñanza a distancia es el tipo de método de instrucción en que las conductas docentes acontecen aparte de las discentes, de tal manera que la comunicación entre el docente y el alumno pueda realizarse mediante textos impresos, por medios electrónicos, mecánicos o por otras técnicas” (Bargas, 2011).

- b) Otto Peters: “La enseñanza a distancia es un método de impartir conocimientos, habilidades y actitudes, racionalizando mediante la aplicación de la división del trabajo y de principios organizativos, y el uso extensivo de medios técnicos, sobre todo con el objetivo de reproducir material de enseñanza de alta calidad, lo cual hace posible instruir a un gran número de estudiantes al mismo tiempo y donde quiera que ellos vivan. Es una forma industrial de enseñar y aprender” (Bargas, 2011).
- c) Börje Holberg: “El término de *educación a distancia* cubre las distintas formas de estudio en todos los niveles que no se encuentran bajo la continua, inmediata, supervisión de los tutores presentes con sus estudiantes en el aula, pero que, sin embargo, se benefician de la planificación, guía y seguimiento de una organización tutorial” (Bargas, 2011). [Cursivas añadidas]

Los orígenes de la educación a distancia son muy remotos, no es algo nuevo en el mundo. Podríamos decir —aunque hay distintas versiones— que se originó a principios del siglo XIX, pero algunos autores encuentran orígenes de esta modalidad ya en la civilización griega (EducAr, 2015).

Internet proporcionó un nuevo paso con respecto a la educación a distancia, convirtiendo a la educación en un espacio social-virtual, posibilitando nuevos procesos de aprendizaje y la transmisión del conocimiento a través de las redes de comunicación. Utilizar internet como herramienta para capacitar facilitará esta constante actualización, ya que las tecnologías posibilitan nuevas propuestas (EducAr, 2015).

3.2.2 E-learning

E-learning es una de las palabras actualmente de moda en educación. En muy pocos años este término ha pasado del vocabulario utilizado por una minoría de expertos en las aplicaciones de la tecnología en la enseñanza a ser empleado por múltiples instituciones, empresas y agentes educativos (Moreira, 2015).

El concepto de *E-learning* es una modalidad de enseñanza-aprendizaje que consiste en el diseño, puesta en práctica y evaluación de un curso o plan formativo desarrollado a través de redes de ordenadores y puede definirse como una educación o formación ofrecida a individuos que están geográficamente dispersos o separados o que interactúan en tiempos diferidos, empleando los recursos informáticos y de telecomunicaciones. Lo característico del *E-learning* es que el proceso formativo tiene lugar totalmente o en parte a través de una especie de aula, o entorno virtual, en la cual tiene lugar la interacción profesor-alumnos, así como las actividades de los estudiantes con los materiales de aprendizaje (Moreira, 2015).

Las primeras experiencias comenzaron en aquellos países donde la sociedad de la información y de las telecomunicaciones tuvo mayor desarrollo e impacto social (principalmente fueron países anglosajones y del norte de Europa), extendiéndose en muy poco tiempo al resto de países occidentales y de otras partes del planeta (Moreira, 2015).

El *E-learning*, o formación *online*, formación virtual o teleformación, es una nueva modalidad educativa que se apoya en las herramientas que ofrece internet. Un curso con modalidad *E-learning* se desarrolla a través de una plataforma virtual, en la que el alumno puede vincularse a través de una computadora con conexión a internet, obtener los materiales, realizar consultas a sus tutores, ser evaluado y contactarse con otros alumnos (EducAr, 2015).

La plataforma es el conjunto de recursos informáticos (tecnologías, información, personal técnico, etc.) unido a un modelo educativo (contenidos, planificaciones, recursos pedagógicos, filosofía educativa, etc.) para crear un ambiente que facilite el proceso de enseñanza-aprendizaje, impulsando una nueva forma de enseñar y de aprender en la que cada uno de los actores asume nuevos papeles de participación ante la necesidad de actualizarse y capacitarse continuamente (EducAr, 2015).

3.2.3 Plataformas de *E-learning*

La plataforma de *E-learning*, campus virtual o Learning Management System (LMS) es un espacio virtual de aprendizaje orientado a facilitar la experiencia de capacitación a distancia, tanto para empresas como para instituciones educativas. Este sistema permite la creación de aulas virtuales; en ellas se produce la interacción entre tutores y alumnos, y entre los mismos alumnos; en este se realizan las evaluaciones, el intercambio de archivos, la participación en foros, chats y otras herramientas (E-ABC, 2015).

Entre algunos de los beneficios que una plataforma de *E-learning* proporciona se mencionan los siguientes: 1) brinda una capacitación flexible y económica, 2) combina el poder de internet con el de las herramientas tecnológicas, 3) anula las distancias geográficas y temporales, 4) permite utilizar la plataforma con mínimos conocimientos, 5) posibilita un aprendizaje constante y nutrido a través de la interacción entre tutores y alumnos, y 6) ofrece libertad en cuanto al tiempo y ritmo de aprendizaje (E-ABC, 2015).

Característica de una plataforma de *E-learning*:

a) Interactividad

Se relaciona con la conversación bidireccional entre receptor y emisor. El receptor elige la parte del mensaje que le interesa; el emisor establece qué nivel de interactividad le dará a su mensaje y el receptor decidirá cómo utilizarla (Clarenc, 2013).

Las plataformas de *E-learning* a través de sus recursos y características deberían de ofrecer suficiente interactividad, de tal forma que, teniendo el alumno acceso a la diversidad de información, material y recursos, sea él mismo el protagonista de su propio aprendizaje (Clarenc, 2013).

a) Flexibilidad

La flexibilidad es una condición que posee algo material o inmaterial, referida al poder sufrir adaptaciones a los cambios, a ser maleable. Un cuerpo es

flexible cuando es capaz de moverse con facilidad y adoptar posturas diversas sin demasiado esfuerzo. Cuando un LMS ofrece flexibilidad, la plataforma no se mantiene rígida a los planes de estudio, sino que puede adaptarse tanto a la pedagogía como a los contenidos adoptados por una organización (Clarenc, 2013).

b) Escalabilidad

Se refiere a la propiedad de aumentar la capacidad de trabajo de un sistema sin comprometer por ello su funcionamiento y calidad habituales. Es decir, poder crecer sin perder la calidad en sus servicios. En un LMS, la escalabilidad permite que la plataforma pueda funcionar con la misma calidad, independientemente de la cantidad de usuarios registrados y activos (Clarenc, 2013).

c) Estandarización

Un estándar es un método aceptado, establecido y seguido normalmente para efectuar una actividad o función, para lo cual se deben cumplir ciertas reglas (implícitas y explícitas) con el fin de obtener los resultados esperados y aprobados para la actividad o función. Por medio de un estándar se garantiza el funcionamiento y acoplamiento de elementos que fueron generados independientemente. Es importante que un LMS ofrezca estandarización, a los efectos de poder utilizar cursos o materiales que hayan sido realizados por terceros (Clarenc, 2013).

d) Usabilidad

Se refiere a la rapidez y facilidad con que las personas realizan tareas propias mediante el uso de un producto, y se logran objetivos específicos con: (Clarenc, 2013)

- ✓ **Efectividad.** Para que los usuarios logren los objetivos con precisión y plenitud. Aquí cuentan la facilidad de aprendizaje del producto, la facilidad con que puede ser recordado y la cantidad de errores que contiene.

- ✓ **Eficiencia.** Se refiere a los recursos empleados para lograr la precisión y la plenitud.
- ✓ **Satisfacción.** Es el grado de complacencia con el uso del producto. Es subjetivo.

Son los usuarios de las plataformas los que determinan su usabilidad.

f) Ubicuidad

El término *ubicuo* tiene origen latino (*ubique*) y significa “en todas partes”. Se usa en el ámbito religioso indicando la capacidad de Dios de tener presencia simultánea en todos lados al mismo tiempo. La ubicuidad está vinculada con la omnipresencia. Hace un tiempo la novedad dentro de las TIC era lo electrónico (*E-learning*). Ahora se habla de *ubicuidad*. La tecnología nos permite “estar presentes” en diferentes lugares al mismo tiempo, tener la información disponible a cualquier hora y en cualquier lugar, porque los dispositivos tecnológicos modifican la manera de acceder a la información y al conocimiento (Clarenc, 2013).

g) Persuabilidad

Persuabilidad es una palabra compuesta por dos términos (*persuasión* y *usabilidad*) e implica la integración y articulación de cuatro características (funcionalidad, usabilidad, ubicuidad e interactividad). Este concepto se puede sintetizar en la capacidad que tiene una plataforma de convencer, fidelizar o “evangelizar” a un usuario a través de su uso. Es decir, la capacidad de convertirlo en un potencial cliente [persuadir para convertir] (Clarenc, 2013).

h) Accesibilidad

La *accesibilidad* se refiere a los medios que permiten a personas con otras capacidades a acceder a la información *online*. Por ejemplo, las personas con deficiencias visuales usan un mecanismo llamado *screen reader* para leer la pantalla, para lo cual las páginas web necesitan estar diseñadas especialmente para que dichos medios las puedan leer (Clarenc, 2013).

El consorcio W3C tiene en marcha una iniciativa llamada *Web Accessibility Initiative* (WAI), con una serie de normas para que las páginas web sean accesibles. Parece ser la norma más seguida en la actualidad y se puede considerar un estándar de facto. Algunos sistemas que tienen incorporadas estas características para cumplir con alguna norma concreta de accesibilidad son Atutor, Moodle y LRN (Clarenc, 2013).

Funcionalidades esenciales de una plataforma de E-learning (Clarenc, 2013):

a) Gestión administrativa

- ✓ Gestión del estudiante
- ✓ Herramientas de monitoreo
- ✓ Mecanismos de acceso a bases de datos
- ✓ Elaboración de informes
- ✓ Administración cualitativa y funcional de flujos de trabajo
- ✓ Seguimiento de usuarios

b) Gestión de recursos

- ✓ Control de autoría y edición de contenidos
- ✓ *Learning Objects* y otros tipos de gestión de contenidos
- ✓ Plantillas de ayuda en la creación de contenidos
- ✓ Mecanismos de subida y descarga de contenidos
- ✓ Reutilización y compartición de *Learning Objects*

c) Gestión de recursos

- ✓ Foros
- ✓ Chat
- ✓ Pizarra
- ✓ *E-mail*
- ✓ *Wiki*

3.2.3.1 Plataformas de *E-learning Open Source* (de código abierto)

Las plataformas de *E-learning* libres son aquellas que pueden ser descargadas y distribuidas de forma gratuita y libre, las cuales pueden ser instaladas en un servidor local o uno remoto sin restricciones, el usuario puede personalizarla y realizar todos los cambios que se desee sin ninguna restricción (Ntic, 2015). Surgieron como una alternativa para hacer más económico un proyecto de formación en línea. Estos LMS generalmente están desarrollados por instituciones educativas o por personas que están vinculadas con el sector educativo (Clarenc, 2013).

Algunas de estas plataformas son de tipo Open Source (de código abierto), lo que establece que son de libre acceso, permitiendo que el usuario sea autónomo para manipular ese *software*, es decir, que una vez obtenido se pueda usar, estudiar, cambiar y redistribuir libremente. Es variada la gama de funcionalidades que trae cada una de estas plataformas (Clarenc, 2013).

Entre las plataformas de código abierto más utilizadas se mencionan las siguientes:

a) Dokeos

Dokeos es una *suite* de aprendizaje de código abierto profesional. Se pueden crear e-cursos usando plantillas; importar cursos SCORM; evaluar alumnos por medio de test y encuestas; realizar un seguimiento exhaustivo usando las herramientas de interacción; convertir documentos de Word y PowerPoint en e-cursos; organizar videoconferencias, etc. (TES, 2015).

Portada del sitio web



Figura 1. Portal Dokeos

Fuente: www.dokeos.com

El URL de este LMS es <http://www.dokeos.com/es>.

b) Ilias

La plataforma Ilias es un proyecto Open Source y coordinado por la Universidad de Colonia, Alemania. **Ilias** ofrece un conjunto de herramientas para crear ejercicios, evaluaciones y encuestas, foros, chats, RSS y mensajería interna (TES, 2015).

Portada del sitio web



Figura 2. Portal Ilias

Fuente: www.ilias.de

El URL de este LMS es <http://www.ilias.de/docu/>.

a) ATutor

ATutor es un sistema de gestión de contenidos de aprendizaje (Learning Content Management System) de código abierto basado en la web y diseñado con el objetivo de lograr accesibilidad y adaptabilidad. Los administradores pueden instalar o actualizar ATutor en minutos (TES, 2015).

Portada del sitio web

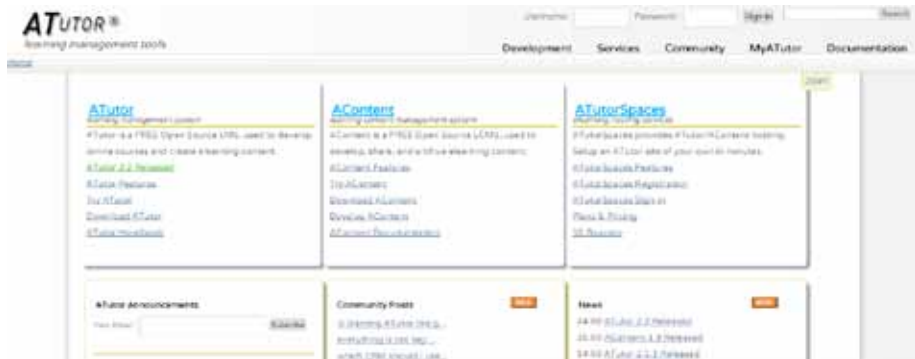


Figura 3. Portal ATutor

Fuente: www.atutor.ca

El URL de este LMS es <http://www.atutor.ca/>.

d) Moodle

Moodle es una plataforma para el manejo, producción y administración de cursos vía web. Creada por docentes para docentes, es de fácil manejo y con Open Source (código fuente abierto). Hoy en día hay muchos docentes que conforman una comunidad Moodle se dedican a asesorar a docentes que quieran iniciarse en el manejo de esta plataforma. Al ser Open Source, esta comunidad mejorará continuamente la plataforma. Fue creada por Martin Dougiamas, que se basó en principios del constructivismo social (TES, 2015).

Portada del sitio web



Figura 4. Portal de Moodle

Fuente: <https://moodle.org>

El URL de este LMS es <https://moodle.org>.

e) Claroline

Claroline se caracteriza por ser una plataforma libre, de fácil manejo y con una potencialidad aceptable en cuanto al manejo de entornos de aprendizajes montados en la web. Ha sido desarrollada siguiendo las necesidades y experiencias de profesores, por lo que ofrece una interfaz intuitiva y clara de administración de espacios, logrando una gestión o manejo diario de la plataforma; no requiere ninguna habilidad técnica especial, y, a través del uso de cualquier navegador web, permite manejar las distintas partes del curso y la admisión de usuarios con fluidez (Carleto, 2015).

Portada del sitio web



Figura 5. Portal Claroline

Fuente: <http://www.claroline.net/>

El URL de este LMS es <http://www.claroline.net/>.

3.2.3.2 Plataformas de *E-learning* comerciales (privativas)

Son de uso licenciado, es decir, que para su uso hay que pagar a alguna empresa, ya sea la que desarrolló el sistema o la que lo distribuye. Son sistemas generalmente robustos y bastante documentados, con diversas funcionalidades que pueden expandirse de acuerdo con las necesidades y presupuesto del proyecto (Clarenc, 2013).

Entre las plataformas comerciales más utilizadas se mencionan las siguientes:

a) e-doceo

E-doceo, compañía francesa dedicada al desarrollo de *software* para la formación a distancia, ha hecho del *E-learning* una herramienta sencilla para la formación por la flexibilidad de sus soluciones, que permiten que cada compañía cree sus módulos de formación, los gestione y los actualice siempre que lo necesite, sin necesidad de otro intermediario ni conocimientos avanzados de informática. Fundada en 1999 por dos pedagogos franceses, E-doceo es hoy en día el primer editor de programas en el sector *E-learning*. E-doceo ofrece una

solución *E-learning* global basada en la calidad de sus servicios, la potencia de su *software*, la facilidad de uso de sus herramientas y la experiencia de su equipo (TES, 2015).

Portada del sitio web



Figura 6. Portal e-doceo

Fuente: <http://es.e-doceo.net/>

El URL de este LMS es <http://es.e-doceo.net/>.

b) Webct

Webct (Web Course Tools, Herramientas para cursos web) es un sistema comercial de aprendizaje virtual *online*, el cual es usado principalmente por instituciones educativas para el aprendizaje a través de internet. La flexibilidad de las herramientas para el diseño de clases hace este entorno muy atractivo, tanto para principiantes como para usuarios experimentados en la creación de cursos en línea. Los instructores pueden añadir a sus cursos Webct varias herramientas interactivas, tales como tableros de discusión o foros, sistemas de correos electrónicos, conversaciones en vivo (chats), contenido en formato de páginas web y archivos PDF, entre otros (TES, 2015).

Portada del sitio web



Figura 7. Portada Webct

Fuente: Sitio web de Webct

El URL de este LMS es <http://www.webct.com>.

c) Blackboard

Blackboard Inc. (NASDAQ: BBBB) es una compañía de *software* con sede en Washington, DC, EE. UU. Fundada en 1997, Blackboard se formó como firma consultora con un contrato con la organización sin fines de lucro IMS Global Learning Consortium (<http://www.imsglobal.org/>). En 1998, Blackboard LLC se fusionó con CourseInfo LLC, una pequeña compañía proveedora de programas de administración de cursos originaria de la Universidad de Cornell. La nueva compañía se conoció como Blackboard Inc. La primera línea de productos de aprendizaje en línea (*E-learning*) fue llamada *Blackboard Courseinfo*, pero luego el nombre *Courseinfo* fue discontinuado en el 2000. Blackboard se convirtió en una compañía con acciones al público en junio de 2004. En octubre de 2005, Blackboard anunció planes de fusión con Webct, una compañía rival de programas de aprendizaje en línea. La fusión se completó el 28 de febrero de 2006. La empresa resultante retuvo el nombre de *Blackboard*, dirigida por su *president* y CEO, Michael Chasen (TES, 2015).

A fecha del 2005, Blackboard desarrolló y licenció aplicaciones de programas empresariales y servicios relacionados a más de 2.200 instituciones educativas en más de 60 países. Estas instituciones usan el programa de BlackBoard para

administrar aprendizaje en línea (*E-learning*), procesamiento de transacciones, comercio electrónico (*E-commerce*), y manejo de comunidades en línea [*online*] (TES, 2015).

Portada del sitio web



Figura 8. Portal Blackboard

Fuente: Sitio Web de Blackboard

El URL de este LMS es <http://www.blackboard.com>.

d) Skillfactory

Skillfactory es un LMS, también conocido como *plataforma de aprendizaje para internet*. Una de las principales diferencias de este con relación a otros sistemas LMS es su interfaz diferenciada: una para administradores y otra para usuarios. Con este esquema se reduce el nivel de complejidad para el usuario final, ofreciendo un ambiente de trabajo cálido y sencillo de usar (TES, 2015).

Portada del sitio web



Figura 9. Portal de Skillfactory

Fuente: Sitio web de Skillfactory

El URL de este LMS es <http://www.skillfactory.com.ar/>.

3.3 Sistemas de administración de contenido

3.3.1 ¿Qué son?

Un sistema de gestión de contenidos (CMS, por sus siglas en inglés) es una aplicación informática utilizada para la creación, edición, administración y publicación de información de una forma fácil, organizada y que no requiere de conocimientos especializados sobre edición digital. Generalmente la interfaz administrativa es sencilla y amigable, permitiendo que cualquier persona pueda administrar sus propios contenidos web sin necesidad de tener conocimientos técnicos ni saber lenguaje HTML, y sin recurrir a un especialista cada vez que se requiera actualizar los contenidos. Esto permite que los contenidos que produce una organización fluyan de manera más rápida y directa a los usuarios (UNAM, 2015).

Estas herramientas permiten manejar de manera independiente el contenido por una parte y el diseño por otra. De esta manera podemos manejar el contenido y darle cada día un diseño nuevo, además de permitir la publicación en el sitio a varios editores de forma fácil y controlada. Como ejemplo, los editores que cargan el contenido al sistema y otro de nivel superior que permite que aquel sea visible a todo el público. Los CMS se conocen principalmente por ser el *software* que se utiliza para facilitar la gestión de las webs (Parra, 2015).

3.3.2 Ventajas de un CMS

A continuación se enumeran las principales ventajas de los CMS (UNAM, 2015).

- ✓ *Organización del sitio web.* El CMS está preparado para organizar eficientemente los contenidos del sitio en secciones y categorías, lo que facilita la navegación a los usuarios y permite crear una estructura sólida, ordenada y sencilla para los administradores. Desde el panel administrador del CMS se pueden crear, editar y borrar las secciones y categorías del sitio de la manera en que más convenga.
- ✓ *Publicación de contenidos.* Crear páginas ilimitadas y editarlas desde un sencillo editor que permite formatear los textos con los estilos y las imágenes deseados. Los contenidos son totalmente editables y modificables.
- ✓ *Escalabilidad e implementación de nuevas funcionalidades.* Ofrece la posibilidad de instalar, desinstalar y administrar componentes y módulos, que agregarán servicios de valor a los visitantes del sitio web, por ejemplo: galerías de imágenes, foros, boletines, clasificados, otros.
- ✓ *Administración de usuarios.* Permite almacenar datos de usuarios registrados y también la posibilidad de enviar correos electrónicos masivos a todos los usuarios. La administración de usuarios es jerárquica, y los distintos grupos de usuarios poseen diferentes niveles de permisos dentro de la gestión y administración del sitio.

- ✓ *Diseño y aspecto estético del sitio.* Es posible personalizar el aspecto del sitio web gracias a sistemas de plantillas (*templates*).
- ✓ *Navegación y menú.* Totalmente editables desde el panel administrador.
- ✓ *Administrador de imágenes.* Poseen una utilidad para subir imágenes al servidor y usarlas en todo el sitio.
- ✓ *Disposición de módulos modificable.* En un sitio creado con CMS, la posición de módulos puede acomodarse como se prefiera.
- ✓ *Automatización en la publicación.* Las páginas y documentos pueden programarse con fecha de publicación y de caducidad. Es decir, un documento puede programarse para que se publique automáticamente al llegar una determinada fecha, y luego dejar de ser visible también de forma automática en otra fecha.
- ✓ *Archivo e historial.* Las páginas viejas o publicaciones que hayan perdido vigencia pueden enviarse a un “archivo” de almacenamiento, sin necesidad de tener que borrarlas. Esto permite también dar la posibilidad a los navegantes de consultar artículos anteriores.
- ✓ *Formatos de lectura.* Cada CMS ofrece diferentes opciones de lectura, que se adaptan a las preferencias de los usuarios.
- ✓ *Comentarios (opcional).* Los usuarios podrán comentar sus opiniones o expresar sus inquietudes en la misma página de contenidos.

3.3.3 Principales CMS de código abierto

Los principales CMS de código abierto, que son utilizados en el mercado para el diseño de portales corporativos, blog, foros entre otros, son Drupal, Joomla y Wordpress (Bilib, 2015). A continuación se detallan.

3.3.3.1 Drupal



*Figura 10. Logo Drupal
Fuente: Sitio web Drupal*

Drupal es un sistema de administración de contenido web. En sus orígenes, el sistema estaba dirigido a dar soporte a una comunidad de *weblogs*. Su desarrollo fue iniciado en 2009 por Dries Buytaert en 1999, y no fue sino hasta en 2001 cuando se publicó la primera versión del CMS hasta el lanzamiento de la versión 4.0.0 (Bilib, 2015).

Entre 2008 y 2009 Drupal se sitúa como uno de los CMS referentes del sector empresarial, siendo reflejado por Gartner durante varios años dentro del cuadrado mágico de portales web. Drupal no está dirigido a un tipo de escenario específico. El límite de este CMS lo impone el desarrollador. Al igual que ocurre con muchos otros CMS, es necesario disponer de un buen conocimiento y experiencia en dicha solución para sacarle el máximo provecho (Bilib, 2015).

Características de este CMS

- ✓ Dispone de un entorno de personalización robusto, tanto el contenido como la presentación pueden ser tratados de forma individual, de acuerdo con unas preferencias definidas por el usuario (Bilib, 2015).
- ✓ Los mecanismos de actualización de contenidos son realmente sencillos, permiten editar la mayor parte de los contenidos tanto desde el *frontend* como desde el *backend*.
- ✓ Ofrece la posibilidad de gestionar las taxonomías y la estructuración de contenidos de forma personalizable, algo indispensable para sitios de complejidad media-alta.
- ✓ Desde el punto de vista de la seguridad, la gestión de permisos destacaba por encima de cualquier otra característica; ofrece un sistema

muy avanzado y completamente personalizable según el papel y las páginas.

- ✓ El rendimiento y la escalabilidad son otras de sus señas de identidad: sistema de caché avanzado, replicación de base de datos, balanceo de carga, mecanismos de control de congestión configurable para habilitar o deshabilitar módulos, etc.
- ✓ La comunidad de desarrolladores es otro de los puntos fuertes de Drupal, ofreciendo un desarrollo dinámico y un soporte amplio basado en foros web.
- ✓ Dispone de cientos de extensiones, éstas se encuentran agrupadas según su funcionalidad en distintas categorías: administración, control de acceso, eventos, comercio, comunidad, contenidos, gestión de usuarios, búsquedas, otros.

Cabe mencionar que Drupal se encuentra liberado bajo licencia GPL y utiliza PHP como lenguaje de programación, y MySQL como motor de base de datos, aunque también puede funcionar con PostgreSQL o SQLite, y Apache o Microsoft IIS como servidor web (Bilib, 2015).

3.3.3.2 Wordpress



*Figura 11. Logo de WordPress
Fuente: Sitio web Wordpress*

WordPress fue en sus inicios un CMS orientado al desarrollo de blogs, pero ha evolucionado hacia uno de los CMS de uso general más completos del mercado. Su origen se encuentra en la herramienta b2/cafelog, diseñada por Michel Valdrighi y pensada como un sistema para la gestión de blogs personales. La primera versión de WordPress, la 0.7, surge en 2003, encontrándose

actualmente en la versión 3.3.1. El modelo de actualizaciones ha sido hasta la fecha bastante dinámico con 3 o 4 versiones base anuales, y versiones menores con una frecuencia de actualización de aproximadamente un mes. WordPress es uno de los CMS más conocidos, utilizados y descargados del mercado. El motivo se encuentran es la sencillez de uso e implantación; es dirigido a sitios web donde se requiere una presencia básica en internet y de comunicación, como por ejemplo: sitios corporativos, comunidades de usuarios, tiendas *online*, etc. Además, con una buena base de conocimiento y experiencia, este CMS también puede ser utilizado en escenarios más complejos (Bilib, 2015).

La personalización es sin duda uno de los puntos fuertes de WordPress, pues es especialmente sencilla, a través de la gran variedad de temas adaptables y extensiones, desde el enfoque más profesional al más informal (Bilib, 2015).

La protección de la privacidad de los contenidos es otra de las características principales de WordPress, a través de la definición de niveles de usuario, protección de contenidos por contraseña, filtros anti-*spam* o controles de comentarios (Bilib, 2015).

3.3.3.3 Joomla



Figura 12. Logo de Joomla
Fuente: Sitio web Joomla

Joomla es un potente CMS que permite crear sitios web elegantes, dinámicos e interactivos de forma simple. Este gestor de contenidos surge en 2005 como resultado de una división del proyecto Mambo. La primera versión de Joomla integraba el núcleo de Mambo, pero con nuevo *software* libre y muchos cambios importantes en el código. A partir de esta versión, muchos colaboradores, comunidades y diseñadores respaldaron el proyecto, que evolucionó hasta convertirse en lo que es hoy en día: uno de los CMS más usados y conocidos del mercado. Para extraer todo el potencial de Joomla, se requiere cierto cono-

cimiento y experiencia, ya que su máxima versatilidad se obtiene de la integración, adaptación y desarrollo de nuevos módulos (Bilib, 2015).

Uno de los puntos fuertes de Joomla es su magnífica comunidad. Fruto de la gran participación de los usuarios, el sistema se encuentra en continua actualización frente a vulnerabilidad, *bugs*, nuevas funcionalidades y extensiones. La fortaleza de su comunidad dispone de un soporte muy completo a través de webs oficiales, foros y todo tipo de documentación generada. Otra de las características destacadas de Joomla es la versatilidad que ofrece el sistema a través de plantillas, extensiones y adaptaciones (Bilib, 2015).

Existen cientos de módulos, componentes y *plugins* que extienden la funcionalidad original del CMS: gestión de archivos, gestión de contactos, sistema de búsqueda, tiendas *online*, bolsas de trabajo, integración con redes sociales, gestión de noticias, sistemas de encuestas y *newsletters* (Bilib, 2015).

4. Metodología de la investigación

4.1 Metodología

La investigación es descriptiva experimental. Se realizó una descripción de las características de los elementos involucrados en la investigación, tales como Joomla, Moodle y Joomla. Además se preparó un portal integrando ambas plataformas, lo que implicó realizar pruebas experimentales.

4.2 Sujetos de estudio

Plataformas Moodle y Joomla.

4.3 Instrumentos que se han de utilizar para recolección de datos

Análisis documental por medio del cual se buscó información para la implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS.

4.4 Procedimientos

Lo primero que se realizó fue investigar en diferentes medios información relevantes sobre los sujetos de estudio, llámense *Moodle* y *Joomla*, luego se identificaron características, requerimientos de instalación y todos aquellos aspectos que ayudaran a montar el portal colaborativo. Como siguiente paso se compró el *hosting* y se procedió a instalar Joomla y Moodle. Después de realizada la instalación, se verificó que funcionaran correctamente y se procedió a instalar y configurar Joomla.

Finalizando todo el proceso con la personalización del portal, integrando ambas plataformas Joomla y Moodle en una sola.

5. Resultado

5.1 Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

5.1.1 Selección de las plataformas que se han de utilizar

Para la implementación del portal se utilizará Joomla por ser un CMS versátil, multiplataforma y muy potente. Los archivos de instalación se pueden descargar del sitio web oficial de Joomla. <https://www.joomla.org/>

Para la administración de cursos virtuales se utilizará Moodle por ser un LMS potente, multiplataforma y que se integra con facilidad con Joomla. Los archivos de instalación pueden descargarse del sitio web oficial de Moodle: <https://moodle.org/>.

También se descargó un complemento importante para la integración de ambas plataformas: el Joomla. Los archivos de instalación se descargarán <http://www.joomdle.com/>.

5.1.2 Requisitos de instalación

5.1.2.1 Para instalar Joomla

Los requerimientos necesarios que hay que tener en cuenta para instalar Joomla son los siguientes: (Intef, 2015)

- ✓ Servidor http: Apache 2.x o superior (<http://www.apache.org>); o Microsoft IIS 7 o superior (<http://www.iis.net>)
- ✓ Servidor de bases de datos: MySQL 5.0.4 o superior (<http://www.mysql.com>).
- ✓ Intérprete del lenguaje PHP 5.3 o superior (<http://www.php.net>).
- ✓ Soporte XML (siglas eXtensible Markup Language), lenguaje de marcado ampliable o extensible, desarrollado por el sitio web Consortium (W3C).
- ✓ Soporte Zlib, configurado en PHP. Zlib es una librería de compresión gratuita, cuyo propósito es plantear una librería de compresión estable y libre de errores, que pueda ser usada en cualquier máquina o sistema operativo, y que permitirá a PHP leer archivos comprimidos en formato ZIP.

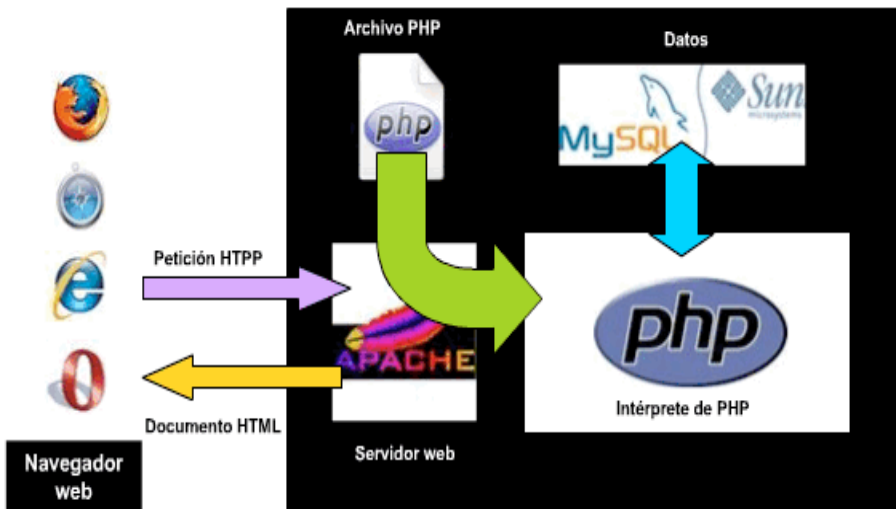


Figura 13. Generación de páginas dinámicas con Joomla

Fuente: Sitio web de Itef.

5.1.2.2 Para instalar Moodle

Los requerimientos necesarios que se tendrán que tener en cuenta para instalar Moodle son los siguientes: (Moodle, 2015)

- Servidor web [Apache](#) o [IIS](#) este debe de estar configurado correctamente para servir archivos PHP.
- Intérprete del lenguaje PHP 5.3 o superior.
- Un gestor de base de datos. MySQL y PostgreSQL son las bases de datos para desarrollo primario, las más completamente probadas y las que tienen abundante documentación y soporte. La base de datos de Oracle no está completamente soportada y tampoco recomendada.
 - [MySQL](#) - versión mínima 5.1.33; Moodle 2.7 requiere MySQL 5.5.31
 - [MariaDB](#) - versión mínima 5.3.5 (reemplazo completo para MySQL)
 - [PostgreSQL](#) - versión mínima 8.3

- [MSSQL](#) - versión mínima 9.0
- [Oracle](#) - versión mínima 10.2 (no recomendada)
- Navegador mínimo para acceder Moodle: versiones recientes de Firefox, Internet Explorer 9, Safari 6, Google Chrome 11 reciente (versiones inferiores podrían funcionar pero no están completamente soportadas).



*Figura 14. Aplicación web Moodle
Fuente: Slide Share (Inserver, 2015).*

5.1.2.3 Acerca de Joomla

La necesidad de integrar servicios en Joomla ha llevado a utilizar puentes (bridges) que realicen esta tarea de forma transparente al usuario que será el destinatario de dichos servicios, pero no siempre con una integración amigable para el desarrollador o implementador del servicio en el *backend* (Alejo, 2015).

Integrar los servicios entre sí permitirá unificar accesos, gestionar de forma común usuarios y, a través de Joomla, administrar desde un solo lugar los cursos, los alumnos y todas las tareas administrativas (Alejo, 2015).

Joomla es una extensión desarrollada por el español Antonio Durán Terrés y liberada para Joomla 1.5 o 1.7, compatible con Joomla 2.5 y 3.x, bajo licencia no comercial, lo que permite que la integración de la herramienta de *E-learning* más conocida y basada en código de fuentes abiertas, Moodle, pueda conectar con Joomla (Alejo, 2015).

5.1.3 Instalación de Joomla

Lo primero que se debe hacer es verificar si en el computador que se utilizará tiene instalado Apache, Php y MySQL.

A continuación se instalará XAMPP, que es un *software* libre que instala el gestor de bases de datos MySQL; el servidor web Apache y los intérpretes para lenguajes de script: PHP y Perl. El instalador puede descargarse del URL <https://www.apachefriends.org/download.html>

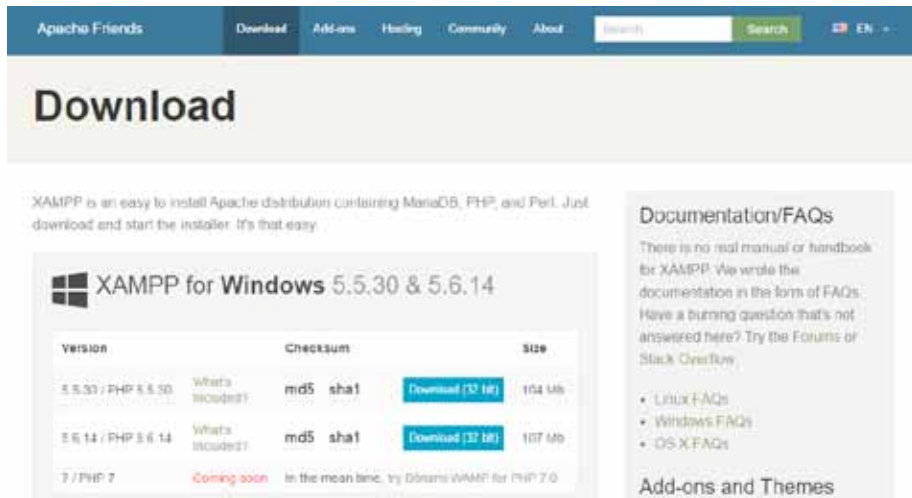


Figura 15. Portal para descargar XAMMP
Fuente: Sitio web de XAMMP

Instalación de Apache

- Después de descargar XAMPP, instale el programa haciendo clic en *Ejecutar*.

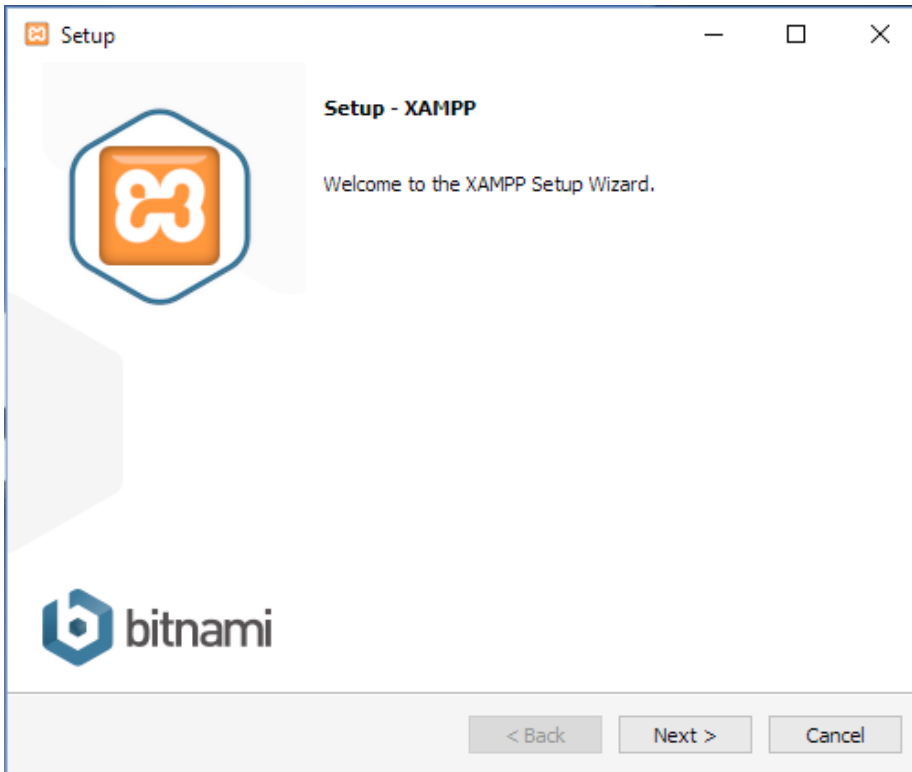
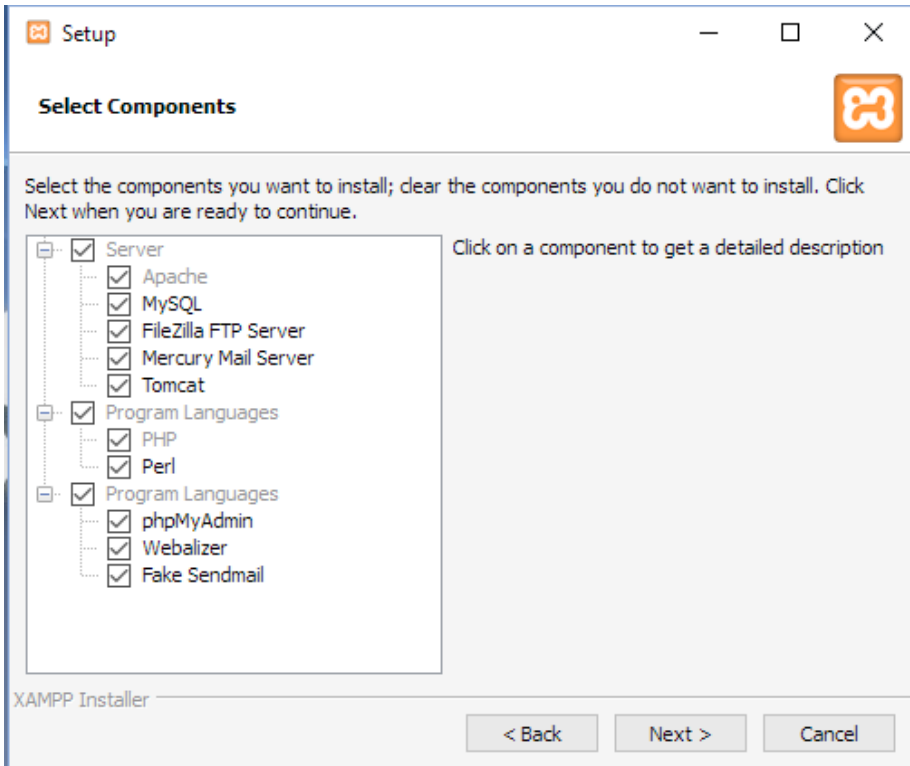


Figura 16. Pantalla inicial instalación de XAMPP
Fuente: Elaboración propia.

- En esta ventana presione clic en *Siguiente*. Comenzará el proceso de instalación y configuración de XAMPP. La configuración puede ser cambiada en cualquier momento en la edición de los archivos de configuración.



*Figura 17. Instalación de XAMMP
Fuente: Elaboración propia.*

- Al finalizar la instalación se mostrará una imagen como la siguiente:

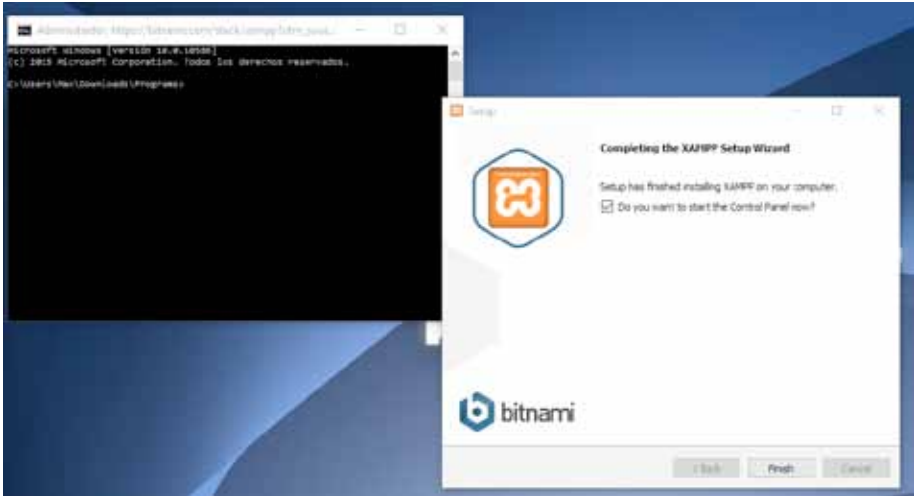


Figura 18. Instalación de XAMPP
Fuente: Elaboración propia.

- Finalizada la instalación se abrirá el panel de control de XAMPP y este inicializará Apache y MySQL.

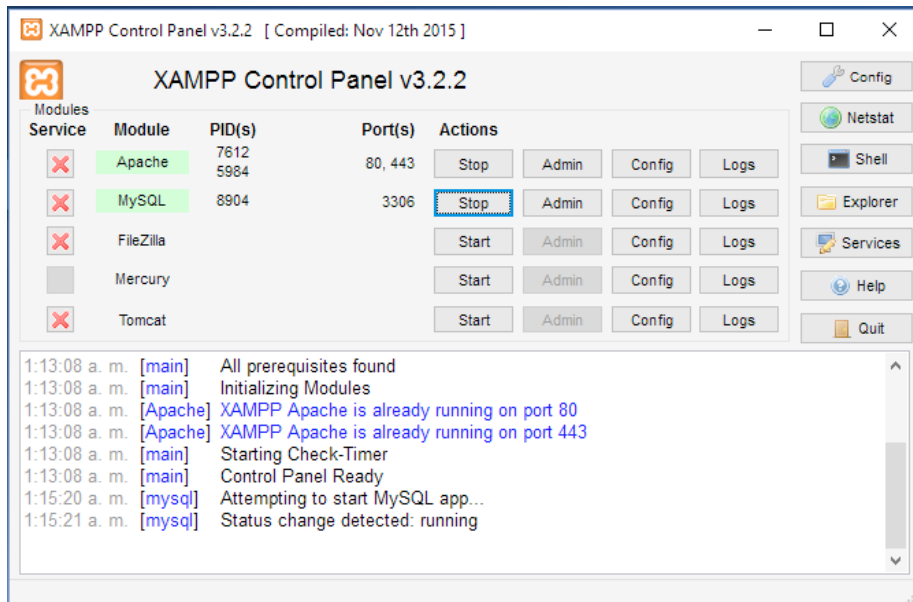


Figura 19. Instalación de XAMMP
 Fuente: Elaboración propia.

- Ahora se deberá abrir la ventana del navegador para verificar si apache y MySQL están funcionando. Escriba en el navegador <http://localhost>; y se deberá mostrar un sitio web similar al de la imagen de abajo.



Figura 20. Instalación de XAMMP
Fuente: Elaboración propia.

Después de instalar XAMMP ya se cumple con los requisitos para la instalación de Joomla, por lo que se procederá a realizarla.

5.1.3.2 Creación de base de datos

- Abrir el navegador y escribir en la barra de direcciones la ruta <http://localhost/phpmyadmin/>. Se abrirá el phpMyAdmin, lugar en el cual se creará la base de datos.

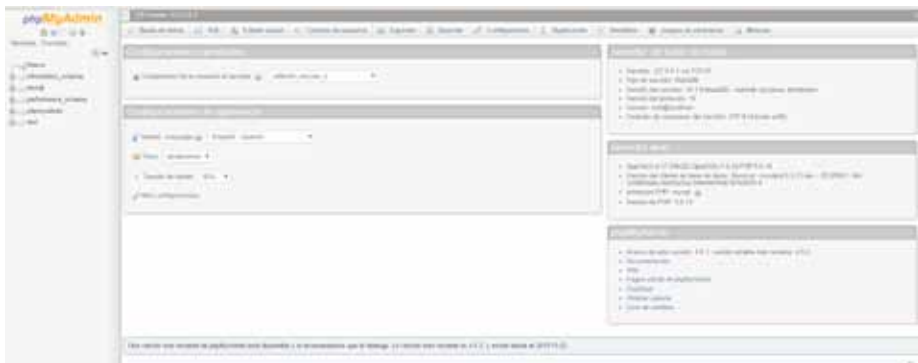


Figura 21. Ventana de phpMyAdmin
Fuente: Elaboración propia.

- Crear la base de datos con el nombre que sea de su preferencia. Para el caso se asignara el nombre *genius*, y en la parte del cotejo se deberá seleccionar `utf8_unicode_ci.`

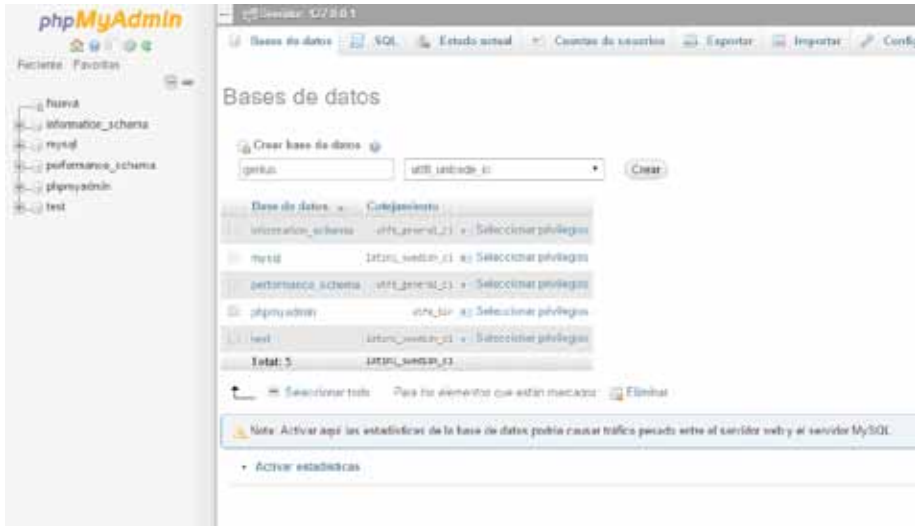


Figura 22. Creación de base de datos
Fuente: Elaboración propia.

- Una vez creada la base de datos, se procede a crear un usuario. Para ello se selecciona en la parte izquierda la base de datos creada, y luego en la parte superior se selecciona la opción *Privilegios* y se crea el nuevo usuario. Para este caso el usuario se llamará *genius*; se agrega el nombre de servidor, el cual será *localhost* y se asigna la contraseña. Posteriormente deberá agregar todos los permisos al usuario, para que tenga privilegios globales y pueda administrar la base de datos, luego dar clic en *Continuar*.

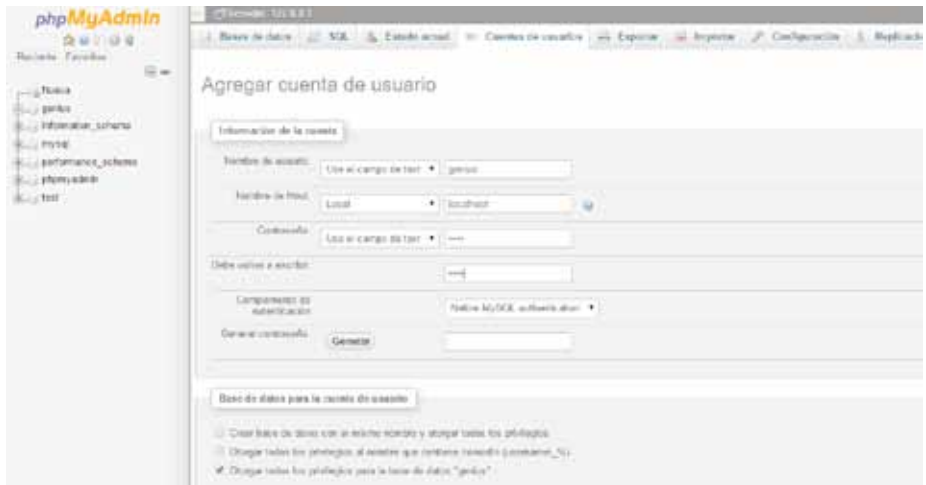


Figura 23. Creación de usuario
Fuente: Elaboración propia.

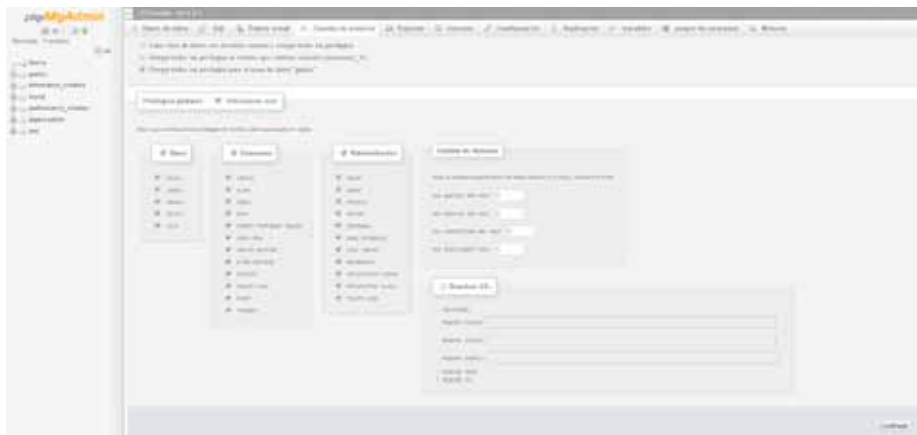


Figura 24. Configurar privilegios de usuario
Fuente: Elaboración propia.

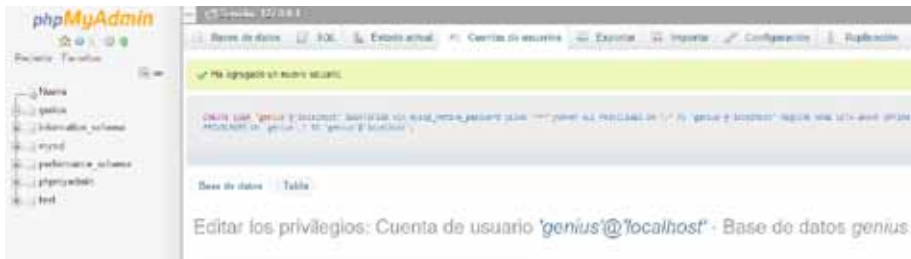


Figura 25. Mensaje presentado confirmando de la creación del usuario
Fuente: Elaboración propia.

5.1.3.3 Instalación del portal Joomla

- Descargar los archivos de instalación del URL <https://www.joomla.org/download.htm>.
- Al finalizar la descarga, se deberán trasladar al directorio `c:\xampp\htdocs`. El archivo ZIP descargado.

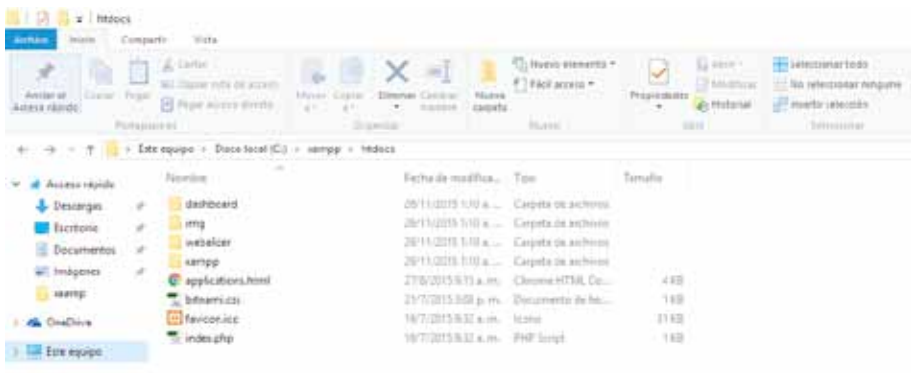


Figura 26. Ruta directorio servidor web
Fuente: Elaboración propia.

- Descomprimir el archivo descargado en un directorio con el nombre de su preferencia, para el caso se llamará *genius*.
- Abrir el navegador y en la barra de direcciones escribir `localhost/genius`, deberá cargar la ventana que inicia el proceso de instalación de Joomla.



Figura 27. Inicia proceso de instalación de Joomla
Fuente: Elaboración propia.

- En cada uno de los ítems del formulario deberá llenar los datos requeridos y luego dar clic en *Siguiente*. En la nueva pantalla deberá configurar la base de datos.

Joomla!® es software libre liberado bajo la GNU General Public License.

1 Configuración 2 Base de datos 3 Vista general

Configuración de la base de datos

← Atrás Siguiente

Tipo de base de datos * MySQL
Probablemente sea "mysql"

Hostedaje * localhost
Normalmente es "localhost"

Usuario * genius
Año como "root" si el nombre de usuario facilitado por quien le sirve el hospedaje.

Contraseña *
Por cuestiones de seguridad, es primordial usar una contraseña para la puerta de su base de datos.

Base de datos * genius
En algunos hospedajes solo se permite el nombre específico de una base de datos por sitio. En otros casos, si le habiéndose instalado más de un sitio, puede usar el prefijo de las tablas para distinguir entre los sitios de Joomla! que usan la misma base de datos.

Prefijo de las tablas * joomla_
Ej: un prefijo para la base de datos o que se genera automáticamente. Lo óptimo es que sea de tres o cuatro caracteres de largo y que comience sólo caracteres alfanuméricos, y DEBE acabar con un guión bajo. **Asegúrese de que el prefijo elegido no esté siendo usado por otras tablas.**

Proceso para una base de datos antigua * Respaldo Borrar
Se respaldará cualquier respaldo existente de sitios pertenecientes a Joomla!

Figura 28, Configuración de base de datos
Fuente: Elaboración propia.

- Si todo este proceso sale bien, se mostrará una ventana la cual indicará que se seleccione el tipo de visualización del portal Joomla. Para el caso de esta instalación, se selecciona *Brochure*. Al finalizar el proceso de instalación, deberá eliminar la carpeta de instalación. Y listo, ya está instalado Joomla.



Figura 29. Finalización de instalación de Joomla
Fuente: Elaboración propia.

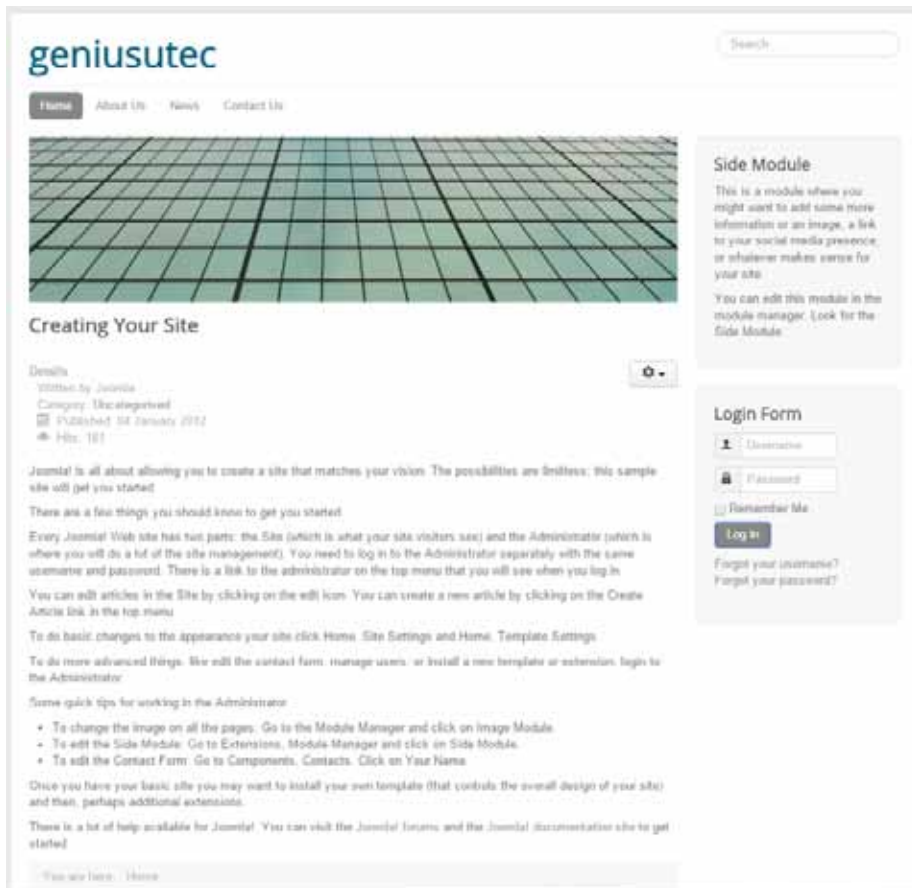


Figura 30. Portal genius sin modificación
Fuente: Elaboración propia.

5.1.4 Instalación de Moodle

El paquete de instalación completo es un archivo ZIP que contiene un servidor web llamado *Apache*, además de la base de datos MySQL y el programa PHP que necesita Moodle.

5.1.4.1 Proceso de instalación

Las siguientes son las tres partes básicas del proceso de instalación:

- Descargue el paquete de instalación completo, desde el URL <https://download.moodle.org/>.
- Al finalizar la descarga, se deberá trasladar el archivo ZIP al directorio `c:\xampp\htdocs`.
- Descomprimir el archivo descargado en un directorio con el nombre de su preferencia, para el caso se llamará *moodle*.
- Inicie el servidor web.
- Abrir el navegador y escribir en la barra de direcciones el URL <http://localhost/moodle>. Se cargará una ventana solicitándole el idioma en el cual se instalar moodle. En esta parte deberá configurar el idioma de su preferencia. Luego dar clic en *Siguiente*.



Figura 31. Idioma de instalación de moodle
Fuente: Elaboración propia.

- En el siguiente paso, el instalador de moodle realizará un diagnóstico del sistema en el cual se instalará moodle.



*Figura 32. Idioma de instalación de moodle
Fuente: Elaboración propia.*

- Ahora deberá configurar el directorio por medio del cual se cargará moodle. Para el caso de esta instalación se definirá al directorio Moodle (<http://localhost/moodle>).



*Figura 33. Instalación moodle
Fuente: Elaboración propia.*

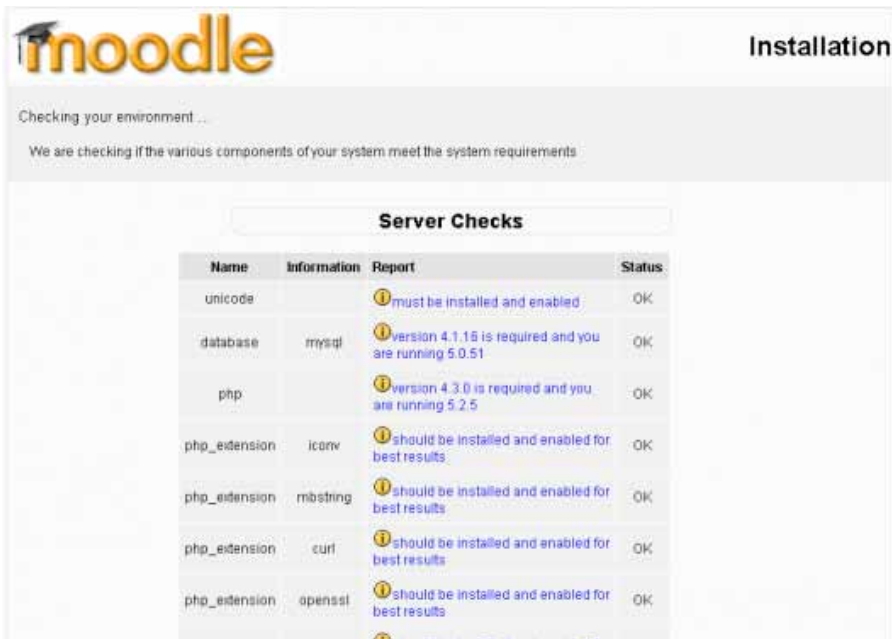
- Deberá ingresar al directorio <http://localhost/phpmyadmin> y crear la base de datos y el usuario que serán asociados a moodle. Para la instalación que se ha realizado se utiliza como base de datos moodle y como usuario *root*.



The screenshot shows the Moodle installation 'Installation' screen. At the top left is the Moodle logo, and at the top right is the word 'Installation'. Below the logo, there is a paragraph of text: 'Now you need to configure the database settings where most Moodle data will be stored. This database will be created automatically by the installer with the settings specified below.' Below this text are several key-value pairs: 'Type: fixed to "mysql" by the installer', 'Host: fixed to "localhost" by the installer', 'Name: database name, eg moodle', 'User: fixed to "root" by the installer', 'Password: your database password', and 'Tables Prefix: optional prefix to use for all table names'. Below these are input fields for 'Type' (mysql), 'Host Server' (localhost), 'Database' (moodle), 'User' (root), 'Password' (masked with asterisks), and 'Tables prefix' (mdl_). At the bottom are 'Previous' and 'Next' navigation buttons.

*Figura 34. Configuración de usuario y base de datos moodle
Fuente: Elaboración propia.*

- La siguiente ventana presenta una revisión del entorno del servidor en el que se está instalando moodle.



*Figura 35. Revisión del entorno del servidor
Fuente: Elaboración propia.*

- Muestra una ventana en la que se debe confirmar el paquete de idioma inicial.



*Figura 36. Confirma paquete de idioma
Fuente: Elaboración propia.*

- Aparece una ventana en la que se confirma que el archivo `config.php` ha sido creado.



Figura 37. Confirmación que config.php ha sido creado
Fuente: Elaboración propia.

- Se abre una ventana en la que se muestra el *copyright* y el acuerdo de Moodle.

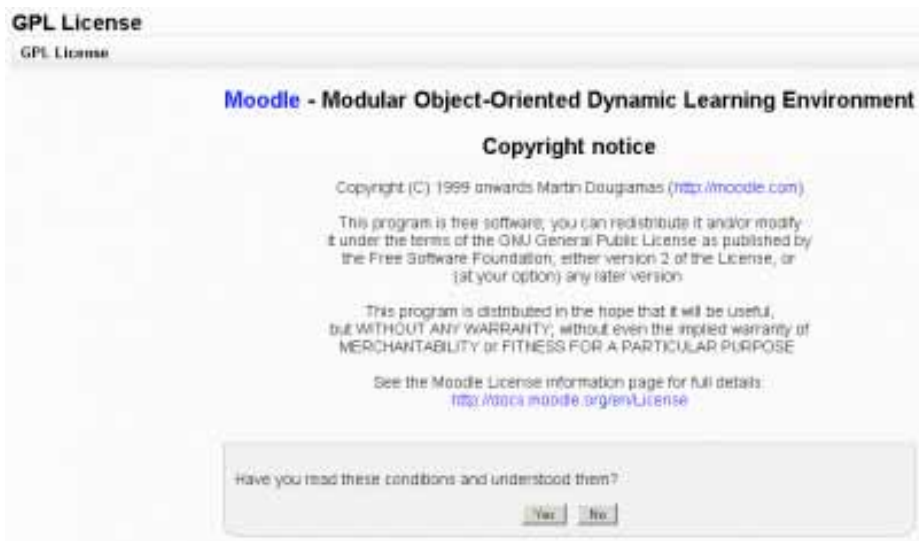


Figura 38. Copyright y acuerdo de Moodle
Fuente: Elaboración propia.

- En esta pantalla se deberá responder que sí se acepta el contrato de licencia. El instalador comienza a crear las tablas en la base de datos. Deberá esperar un momento.
- Al finalizar este proceso mostrará la versión de Moodle instalado; y usted podrá configurar la plataforma.

Moodle 2.6.3 (Build: 20140512)
Copyright © 1999 onwards, Martin Dougiamas
and many other contributors.
GNU Public License

Figura 39. Versión de Moodle
Fuente: Elaboración propia.

- Como siguiente paso se procede a configurar el perfil del usuario administrador.

The screenshot shows the Moodle user profile configuration page for an administrator. The page is titled "Perfil de usuario administrador" and contains various settings for the user's profile. The settings are organized into sections:

- Nombre de usuario:** admin
- Escoger un método de identificación:** Cuenta manual
- Cuenta de usuario suspendida:** No
- Nueva contraseña:** admin123456 (with a "Desenmascarar" button)
- Forzar cambio de contraseña:** No
- Nombre:** Super
- Apellido(s):** User
- Dirección de correo:** carlosferraz@gmail.com
- Mostrar correo:** Mostrar a todos mi dirección de correo
- Formato de correo:** Formato HTML
- Tipo de resumen de correo:** Sin anexos sin correo por cada mensaje del foro
- Subscripción automática al foro:** Si, cuando envía un mensaje suscribirse a este foro
- Rastreo del foro:** No, no registrar los mensajes que he visto
- Editor de texto:** Editor por defecto
- Ciudad:** (empty field)
- Seleccionar su país:** Seleccione su país
- Zona horaria:** Hora local del servidor
- Idioma preferido:** Español - Internacional (es)

Figura 40. Perfil de usuario administrador
Fuente: Elaboración propia.

- Posteriormente se configura la portada.



Figura 41. Perfil de usuario administrador
Fuente: Elaboración propia.

- Para visualizar la instalación de Moodle, se puede ingresar a la dirección que se configuró (<http://localhost/moodle>).



Figura 42. Instalación de Moodle finalizada
Fuente: Elaboración propia.

5.1.5 Instalación de Joomla

5.1.5.1 Proceso de instalación

- Para descargar los archivos de instalación de Joomla debe visitarse el sitio oficial del proyecto: <http://sourceforge.net/projects/moodle-joomla/files/>.

Looking for the latest version? [Download joomla_v1.0.0-for-joomla30-UNZIP-FIRST.zip \(914.7 kB\)](#)

Home RSS















Name ↕	Modified ↕	Size ↕	Downloads / Week ↕
joomla_v1.0.0-for-joomla30-UNZIP-...	2014-12-16	914.7 kB	16  
joomla-0.95-for-joomla30.zip	2014-02-24	973.8 kB	1  
joomla-0.95-for-joomla25.zip	2014-02-24	969.2 kB	4  
joomla-0.94-for-joomla30.zip	2013-11-22	976.6 kB	1  
joomla-0.94-for-joomla25.zip	2013-11-22	977.7 kB	1  
joomla-0.93-for-joomla30.zip	2013-10-29	970.4 kB	1  
joomla-0.93-for-joomla25.zip	2013-10-29	968.3 kB	1  

Figura 43. Sitio de descarga de Joomla
Fuente: Elaboración propia.

- Se descomprime el archivo descargado y luego se copian las carpetas `auth_joomla` y `enrol_joomla` que se encuentran dentro de la carpeta que descomprimió; y se copian dentro de las carpetas `auth` y `enrol` que están dentro del directorio de Moodle.





Nombre	Fecha de modifica...	Tipo	Tamaño
 <code>auth_joomla</code>	04/06/2015 09:55 ...	Carpeta de archivos	
 <code>enrol_joomla</code>	04/06/2015 09:55 ...	Carpeta de archivos	
 <code>auth_joomla.zip</code>	27/02/2015 11:56 a...	Carpeta comprimi...	77 KB
 <code>enrol_joomla.zip</code>	27/02/2015 11:56 a...	Carpeta comprimi...	20 KB

Figura 44. Sitio de descarga de Joomla
Fuente: Elaboración propia.

- Se abre Moodle como administrador desde el URL <http://localhost/moodle/admin/>.

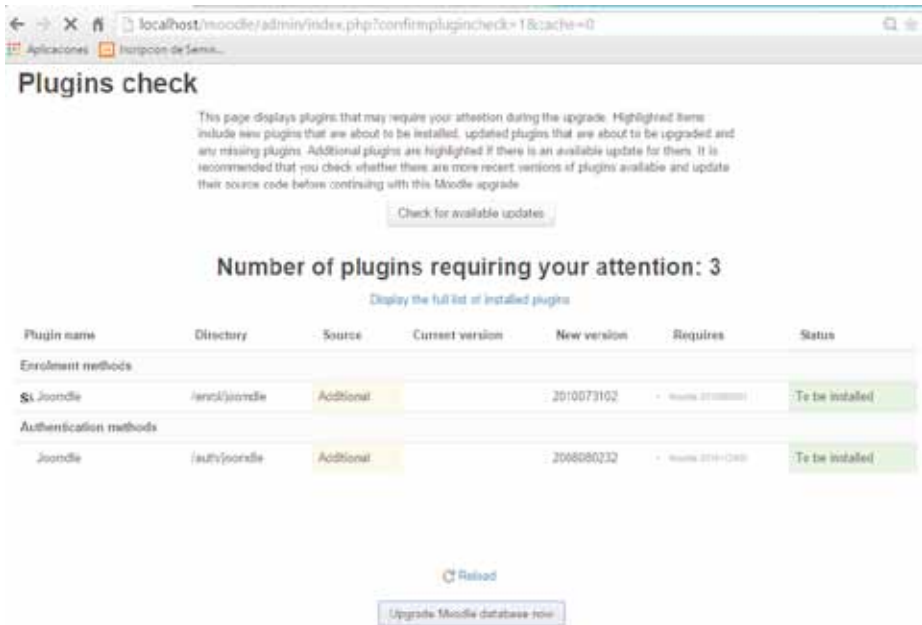


Figura 45. Actualizar base de datos Joomla
Fuente: Elaboración propia.

- Se procede a la activación de la extensión Joomla en Moodle.

Gestionar identificación

Plugins de identificación disponibles

Nombre	Usuarios	Habilitar	Arriba/Abajo	Configuración	Configuración del test	Desinstalar
Cuentas manuales	5	<input type="checkbox"/>		Configuración		
No hay sesión	0	<input type="checkbox"/>		Configuración		
Identificación basada en Email	0	<input type="checkbox"/>		Configuración		Desinstalar
Joomla!	4997	<input type="checkbox"/>		Configuración		
Usar un servidor CAS (SSO)	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar una base de datos externa	0	<input type="checkbox"/>		Configuración	Configuración	Desinstalar
Usar servidor FirstClass	0	<input type="checkbox"/>		Configuración		
Usar un servidor IMAP	0	<input type="checkbox"/>		Configuración		
Usar un servidor LDAP	0	<input type="checkbox"/>		Configuración		
Identificación de la Joomla!	0	<input type="checkbox"/>		Configuración		
Usar un servidor NNTP	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar un servidor POP3	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar un servidor SMTP	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar un servidor IMAP	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar un servidor POP3	0	<input type="checkbox"/>		Configuración		Desinstalar
Usar un servidor SMTP	0	<input type="checkbox"/>		Configuración		Desinstalar

Activar / Desactivar joomla haciendo clic en el ojo de la columna Habilitar.

Acceder a Extensiones / Identificación / Gestor de identificación para activar JOOMDLE

Figura 46. Actualizar base de datos Joomla
Fuente: Elaboración propia.

- Se habilitan los servicios web correspondientes.

Habilitar servicios web

Los servicios Web permiten otros sistemas para acceder a este Joomla y realizar operaciones. Para mayor seguridad esta característica debe ser desactivada a menos que realmente la esté utilizando.

[Quitar cookies](#)

Figura 47. Actualización de servicios web
Fuente: Elaboración propia.

- Se activan protocolos XML-RPC en Moodle; Joomla! usa XML-RPC para conectar con Moodle. Se debe seleccionar *Habilitar protocolos*, de la lista, y activar XML-RPC. Es posible que algunos servidores además requieran habilitar el protocolo SOAP.

Para la configuración debe seguir la ruta Administración del sitio -Extensiones-Servicios web-Administrar protocolos.



Figura 48. Protocolo XML-RPC

Fuente: Elaboración propia.

- Se comprueban los permisos de usuario para conectar con Joomla!, para ello se sigue la ruta Administración del sitio-Usuarios-Permisos-Definir roles.

Gestionar roles Permitir asignar roles Permitir cancelar roles Permitir cambios de rol			
Rol	Descripción	Nombre corto	Editar
Gestor	Los gestores pueden acceder a la clase y modificarla, pero la gestión no pertenece en los casos.	manager	↓ ↕ ↗ ✕
Creador de curso	Los creadores de cursos pueden crear nuevos cursos.	coursecreator	↑ ↓ ↕ ↗ ✕
Profesor	Los profesores pueden acceder a cualquier clase dentro de un curso, se asigna como un profesor y modifica a los estudiantes.	ecilinglesonier	↑ ↓ ↕ ↗ ✕
Profesor sin permiso de edición	Los profesores sin permiso de edición pueden enseñar en los cursos y modificar a los estudiantes, pero no pueden modificar los contenidos.	teacher	↑ ↓ ↕ ↗ ✕
Estudiante	Los estudiantes tienen por lo general acceso pero solo dentro de un curso.	student	↑ ↓ ↕ ↗ ✕
Incluido	Los invitados tienen privilegios mínimos y normalmente no están autorizados para escribir.	guest	↑ ↓ ↕ ↗
Usuario identificado	Todas las acciones de lectura.	user	↑ ↓ ↕ ↗
Usuario identificado en la página principal	Todas las acciones de lectura en el caso de la página principal.	homepage	↑ ↓ ↕ ↗

[Añadir un nuevo rol](#)

Figura 49. Permisos de usuario
Fuente: Elaboración propia.

Agregando nuevo rol

[Crear este rol](#) [Cancelar](#)

Nombre:

Nombre corto:

Descripción:

Fuente: Tamaño: Formato:

Web service

Plata:

Arquetipo de rol :

Tipos de contexto en que puede asignarse este rol:

- Sistema
- Usuario
- Categoría
- Curso
- Módulo
- Bloque

Figura 50. Configuración de permisos
Fuente: Elaboración propia.

- Se habilitan las capacidades del rol para el usuario, como sigue:
 - Usar protocolo XML-RPC webservice/xmlrpc. Use *Permitir*
 - Usar protocolo SOAP webservice/soap. Use *Permitir*
 - Ver debatesmod/fórum viewdiscussion. Use *Permitir*



Figura 51. Capacidades del rol para el usuario
Fuente: Elaboración propia.

- Se añade al nuevo rol creado *Servicios web* el usuario al que se le asignará Joomla!. Se sigue la ruta Administración del sitio-Usuarios-Permisos-Asignar roles globales.



Figura 52. Usuario asociado a Joomla!
Fuente: Elaboración propia.

- Se agrega un nuevo servicio para Joomla! Se sigue la ruta Administración del sitio-Extensiones-Servicios web-Servicios externos.

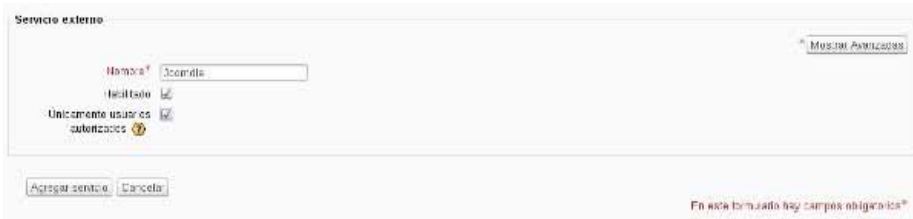


Figura 53. Usuario asociado a Joomla!
Fuente: Elaboración propia.

- Se agregan funciones al servicio externo de Joomla! en Moodle. Se sigue la ruta Administración del sitio-Extensiones-Servicios web-Servicios externos Funciones-Agregar funciones.

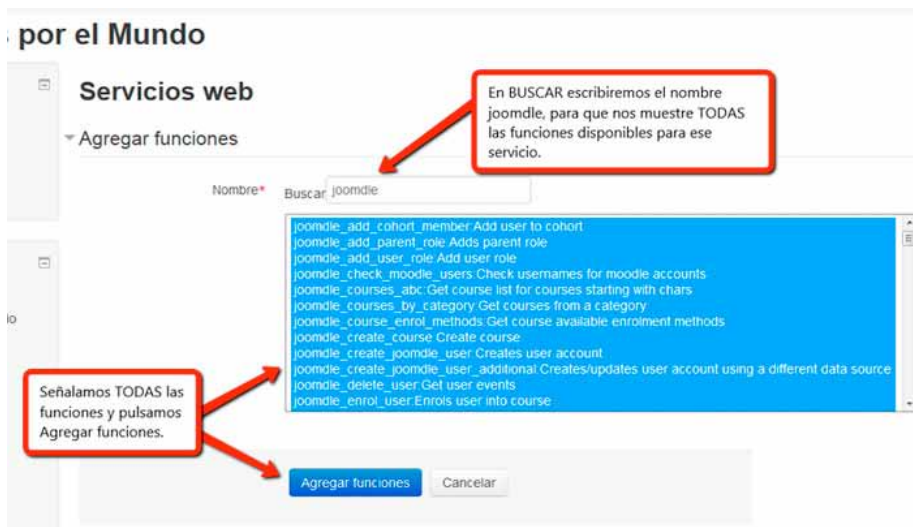


Figura 54. Agregar funciones al servicio externo Joomla!
Fuente: Elaboración propia.

- Se procede a crear el *token* de Moodle para Joomla!. Se sigue la siguiente ruta: Administración del sitio-Extensiones-Servicios web-Administrar *tokens*- Agregar.

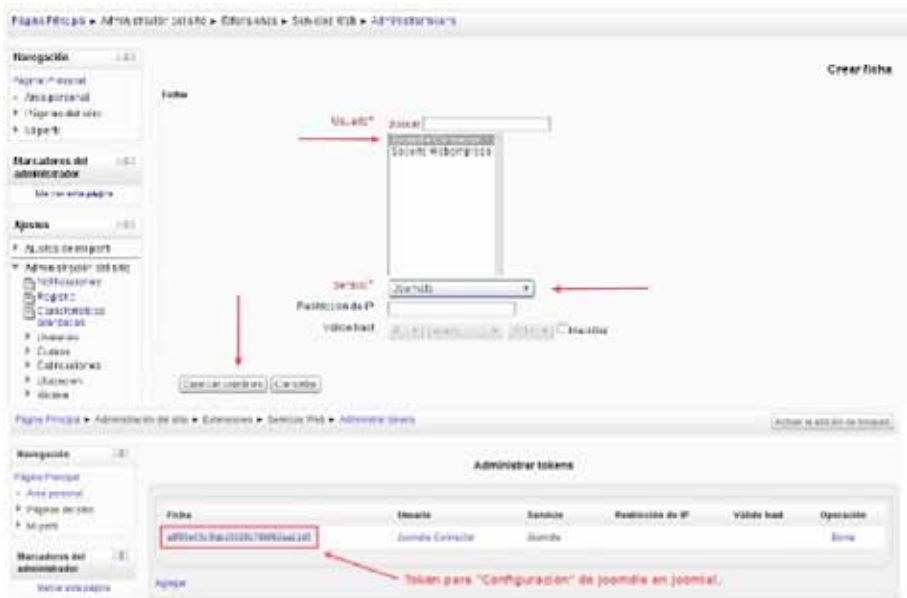


Figura 55. Agregar funciones al servicio externo Joomla
Fuente: Elaboración propia.

- Se gestiona identificación en Moodle y se asigna URL de Joomla. Para ello se sigue la ruta Administración del sitio-Extensiones-Identificación-Gestionar identificación; dar clic en configuración de Joomla! y luego se escribe el URL en el que está instalado Joomla.

AVST1 Español - Internacional (es) Admin User

Aula Virtual

Página Principal > Administración del sitio > Extensiones > Identificación > Gestionar identificación Activar la edición de bloques

NAVEGACIÓN

- Página Principal
- Área personal
- Páginas del sitio
- Mi perfil
- Cursos

MARCA S DEL ADMINISTRADOR

Marcar esta página

ADMINISTRACIÓN

- Ajustes de mi perfil
- Administración del sitio
 - Notificaciones
 - Registro
 - Características avanzadas
- Usuarios
- Cursos
- Calificaciones

Gestionar identificación

Plugins de identificación disponibles

Nombre	Usuarios	Habilitar	Arriba/Abajo	Configuración	Configuración del test	Desinstalar
Cuentas manuales	2			Configuración		
No hay sesión	0			Configuración		
Identificación basada en Email	0	on	↓	Configuración		Desinstalar
Josefa	0	on	↑	Configuración		Desinstalar
Usar un servidor CAS (SSO)	0	on		Configuración		Desinstalar
Usar una base de datos externa	0	on		Configuración	Configuración del test	Desinstalar
Usar servidor FirstClass	0	on		Configuración		Desinstalar
Usar un servidor (MAP)	0	on		Configuración		Desinstalar
Usar un servidor LDAP	0	on		Configuración		

Figura 56. Gestionar identificación de Moodle
Fuente: Elaboración propia.

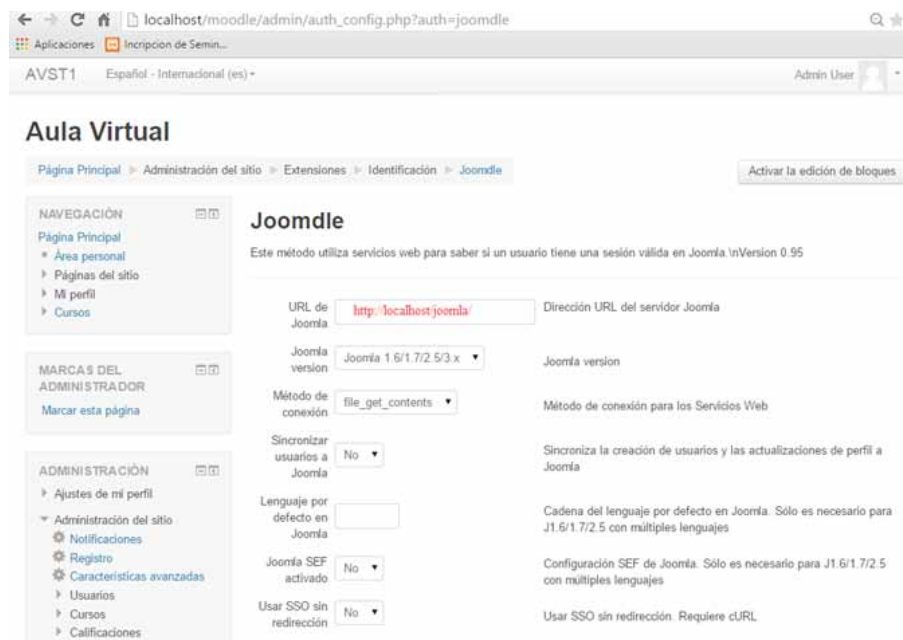


Figura 57. Gestionar identificación de Moodle
Fuente: Elaboración propia.

- Configurando Joomla! en Joomla!. Para ello se debe ingresar al URL del módulo de administración donde está instalado Joomla!. Para este caso, <http://localhost/administrator>. Especifica su usuario y *password*.



Figura 58. Credenciales de administrador
Fuente: Elaboración propia.

- Luego se procede a instalar el *plugin* de Joomla.



Figura 59. Instalando plugin Joomla
Fuente: Elaboración propia.

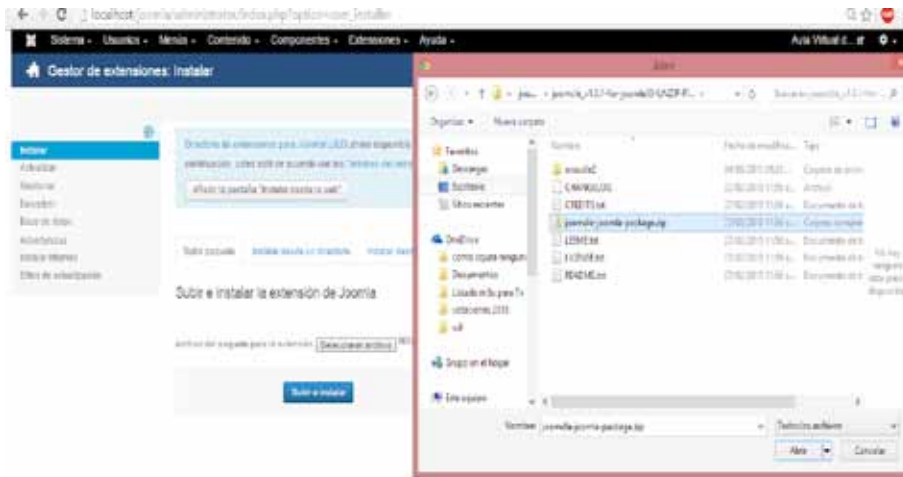


Figura 60. Instalando plugin Joomla
Fuente: Elaboración propia.

- Después de la instalación, se procede a configurar el componente Joomla instalado para ello; se busca en la barra de menú del panel de adminis-

tración la opción *Componentes* y da clic sobre Joomla! y luego aparece *Panel de control*.

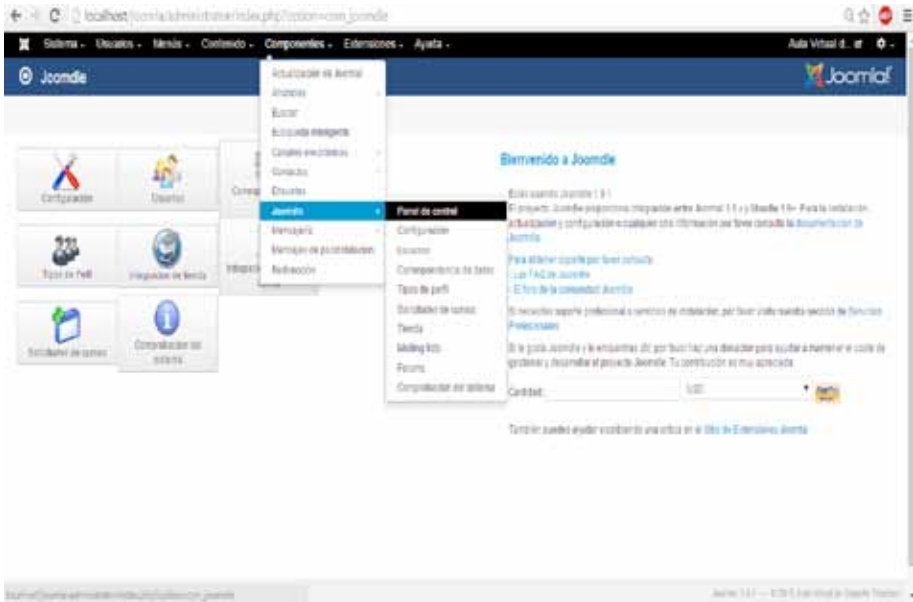


Figura 61. Configuración componente Joomla!
Fuente: Elaboración propia.

- En configuración general, se deberá ingresar el URL donde se tiene instalado Moodle, así como el *token* de autenticación.

Configuración general

URL de Moodle	<input type="text" value="http://10.0.0.4/moodle"/>
Version de Moodle	<input type="text" value="Moodle 2.0.x"/>
Token de autenticación de Moodle 2.0	<input type="text" value="b7c9ef7083cd25b8390a2326b3ee59ac"/>
Método de conexión	<input type="text" value="file_get_contents"/>
COM_JOOMDLE_INTERNAL_IP	<input type="text"/>
Usar SSO sin redirección	<input type="text" value="No"/>
Crear automáticamente usuarios Moodle	<input type="text" value="Si"/>
Borrar automáticamente usuarios Moodle	<input type="text" value="Si"/>
Auto login de usuarios al registrarse	<input type="text" value="No"/>

*Figura 62. Configuración componente Joomdle
Fuente: Elaboración propia.*

- Pueden realizarse más configuraciones; y dependerá de la necesidad del usuario administrador y de lo que él quiera mostrar al usuario final.
- Se finaliza el proceso haciendo la comprobación de que Joomla y Moodle se puedan comunicar. En caso de aparecer algún indicador en rojo, habrá que revisar de nuevo todo el proceso para identificar qué es lo que no se hizo bien.



Figura 63. Configuración componente Joomla

Fuente: Elaboración propia.

- Para mostrar el portal se configurará una entrada del menú principal como *Moodle Wrapper*. La aplicación se mostrará dentro de un *frame*, tal y como observa en la siguiente captura de pantalla.



*Figura 64. Portal Joomla
Fuente: Elaboración propia.*

Para visitar el portal web ingrese al sitio <http://www.geniusUtec.com>.

6. Conclusiones

Al finalizar la investigación se comprueba que la extensión Joomla permite integrar la plataforma de *E-learning* Moodle con la plataforma Joomla; con esto se logra obtener una mayor flexibilidad del portal corporativo, así como un portal multiplataforma que podrá visualizar en cualquier dispositivo o computadora sin perder la trascendencia y usabilidad del sitio web.

Al fusionar ambas plataformas, el usuario podrá disponer de Joomla y Moodle sin necesidad de volver a usar credenciales; ya registrado en una plataforma, por defecto estará registrado en la otra. Las sesiones en ambas plataformas

están sincronizadas, por lo que la información podrá actualizarse al instante, haciendo del portal un medio de comunicación muy efectivo.

Joomla proporciona agradables experiencias a los usuarios, y para el manejo de clases, cursos y diplomados en modalidad virtual es una excelente herramienta, ya que facilita al estudiante la habilidad de mostrar calificaciones y eventos de tareas que son específicos.

En cuanto a los indicadores del proyecto de investigación, estos han sido alcanzados, ya que se integró un portal en Joomla con la plataforma de *E-learning* Moodle. Esto permite que los usuarios utilicen solo una clave para ambas plataformas; y el portal es flexible.

El portal está implementado y se desarrolló para la Incubadora de Empresas de Alta Tecnología Genius. El URL para ingresar es <http://www.geniusUtec.com>, y se espera que pueda ser una propuesta que en un futuro se implemente en muchas otras instituciones.

7. Bibliografía

- Alejo, L.M. (21 de 10 de 2015). “Conectando Moodle con Joomla”. Obtenido de <http://www.webempresa.com/blog/item/787-conectando-moodle-con-joomla-mediante-joomdle.html>
- Bargas, B.C. (2011). *La educación a distancia vista desde la perspectiva bibliotecológica*. Mexico: UNAM.
- Bilib (21 de 10 de 2015). “Gestion de Contenidos Web”. Obtenido de http://www.bilib.es/uploads/media/estudio_sistemas_gestion_contenidos_web_cms.pdf
- Carleto, J.A. (16 de 10 de 2015). “Claroline, una herramienta libre”. Obtenido de <http://jornadaie.unvm.edu.ar/ponencia12.pdf>
- Clarenc, C.A. (2013). *19 Plataformas de E-learning*. Buenos Aires: Grupo Geipite.
- E-ABC (15 de 09 de 2015). “*E-AbcLearning*”. Obtenido de <http://www.e-abclearning.com/quesunaplataformadeelearning>
- EducAr (10 de 09 de 2015). “Ministerio de Educación Argentina”. Obtenido de http://coleccion.educ.ar/coleccion/CD26/datos/recursos/plataformas_virtuales.pdf
- Faver, H. (20 de 09 de 2015). “Educación del futuro”. Obtenido de <http://www.listindiario.com/la-vida/2015/08/25/385434/la-educacion-del-futuro>
- Inserver (22 de 10 de 2015). “Instalación básica de Moodle”. Obtenido de <http://es.slideshare.net/FelipeCasajsIbez/taller-de-moodle-27-instalacion-y-configuracion>
- Intef (21 de 10 de 2015). “Instituto Nacional de Tecnologías Educativas y de formación del profesorado”. Obtenido de http://www.ite.educacion.es/formacion/materiales/99/cd/mod_02/requerimientos_del_servidor.html
- Moodle (13 de 9 de 2015). “Requisitos de instalación Moodle”. Obtenido de https://docs.moodle.org/all/es/Requisitos_para_instalaci%C3%B3n_de_moodle_2.6
- Moreira, M.A. (20 de 09 de 2015). “*Tecedu Web*”. Obtenido de <http://tecedu.webs.ull.es/textos/eLearning.pdf>
- Ntic (15 de 09 de 2015). “La Nave Tic”. Obtenido de <http://www.lanavetic.com/las-mejores-plataformas-e-learning-de-software-libre/>

- Parra, C.L. (20 de 09 de 2015). “McGraw-Hill Interamericana de España”. Obtenido de https://www.mhe.es/secundaria/informatica/8448168933/archivos/Software/Ud07/ud_7_cms-IIjornadas.pdf
- TES, N.o. (14 de 10 de 2015). “Plataformas de *E-learning*”. Obtenido de <http://plataformas-e-learning.wikispaces.com/TIPOS+DE+PLATAFORMAS>
- UNAM, U. N. (18 de 10 de 2015). “Computo Académico, UNAM”. Obtenido de <http://www.ru.tic.unam.mx:8080/bitstream/DGTIC/81750/1/mod4.pdf>
- Wise (10 de 09 de 2015). “Cumbre mundial para la innovación en educación”. Obtenido de <http://www.wise-qatar.org/>

Big data, análisis de datos en la nube

Verónica Idalia Rosa,
Docente investigadora Utec. veronica.rosa@utec.edu.sv,
José Guillermo Rivera,
Docente investigador Utec. joseguillermo.rivera@gmail.com

Agradecimientos

Estamos muy agradecidos con la Vicerrectoría de Investigación; con la Dirección de Investigaciones de la Universidad Tecnológica de El Salvador (Utec) por el patrocinio de esta investigación y por la confianza y apoyo para que este trabajo pudiera realizarse. A los profesionales que muy gentilmente participaron en la investigación, contestando la encuesta que ayudó a determinar la necesidad existente de incursionar en la tendencia *Big Data* en El Salvador y específicamente en la Utec. A compañeros profesionales de la universidad, que con mucha disposición colaboraron en ciertos procesos necesarios para la instalación de programas fundamentales para la investigación. No fue un trabajo fácil debido a la novedad del tema y porque se carece de profesionales expertos en el área, pero al final se lograron satisfactoriamente los resultados.

Resumen

El tema *Big Data* en El Salvador es novedoso y las empresas están con muchos deseos en incursionar en esta tecnología, aunque no saben cómo hacerlo, por lo que el objetivo de esta investigación fue elaborar una guía metodológica en la que se refleje el uso de herramientas *Big Data* para almacenar, procesar y analizar grandes cantidades de datos, con el fin de obtener conclusiones que puedan ayudar en la toma de decisiones de cualquier empresa, independientemente del rubro que manejen. Para esta investigación se hará uso de dos *dataset* que contienen información sobre registro de productos alimenticios y medicamentos y con ellos se trabajará, desde como almacenarlos y procesarlos con Hadoop, para hacer consultas con Hive y también haciendo análisis estadísticos con el programa R, finalizando con la creación de visualizaciones, ya sea con Google charts, Jqplot o D3.js.

Palabras Clave: Macrodatos, conjunto de datos, análisis estadístico

Abstract

The *Big Data* theme in El Salvador is novel, companies have a great desire to venture into this technology, although they don't know how, so that the objective of this research will be to elaborate a methodological guide, in which is reflected the using of *Big Data* tools to store, process and analyze large amounts of data, in order to draw conclusions that can help in making decisions of any company, regardless of the category they handle. For this research we'll make use of two dataset containing information on registration of foodstuffs and medicines, and will work with them, from storing and processing with Hadoop, Hive for consultations and doing statistical analysis with the R program, ending with creating displays with Google chart, Jqplot or D3.js.

Keywords: *Big Data*, dataset, statistical analysis

Introducción

Big Data ha sido muy usado en el medio de la informática y de las grandes empresas, ya que se puede visualizar la gran cantidad de información que fluye hoy en día. Es tan grande la cantidad de información que entra y sale, en empresas e instituciones, que representa un reto su manejo.

Big Data es un término que hace referencia a una cantidad de datos tal que supera la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. El volumen de los datos masivos crece constantemente. En 2012 se estimaba su tamaño de entre una docena de terabytes hasta varios petabytes de datos en un único conjunto de datos.

En el 2001 se realizó un informe de investigación en el que el analista Doug Laney del META Group (ahora Gartner)⁴, definía “el crecimiento constante de datos como una oportunidad y un reto para investigar en el volumen, la velocidad y la variedad”

Hoy en día, se continúan usando datos masivos y en mayor escala que hace 14 años, por lo tanto, para las empresas se hace necesario buscar herramientas que permitan dar soluciones a la demanda de grandes cantidades de datos para el procesamiento y análisis, tal es el caso de MapR, Cyttek Group, Cloudera, Hadoop, entre otros.

“*Big Data* es desde hace unos años el término de moda dentro del mundo de la informática. Dicho de otra manera, durante 2012 y parte de 2013 el 60% de los artículos de opinión de tecnología avanzada hablan de *Big Data* como la nueva estrategia indispensable para las empresas de cualquier sector, declarando, poco menos, que aquéllos que no se sumen a este nuevo movimiento se quedarán “obsoletos” en cuanto a la capacidad de reacción en sus decisiones, perdiendo competitividad y oportunidades de negocio contra su competencia.”⁵

4 <http://www.gartner.com/analyst/40872/Douglas-Laney>

5 Aitor Moreno. Responsable de inteligencia artificial de Ibermática

Debido a todo lo anterior, estamos en una realidad que no se puede cambiar y que se debe ir en la misma dirección con los avances de la ciencia y tecnología, por lo tanto, existe la necesidad de trabajar con una gran cantidad de datos, pero la gran mayoría de empresas no saben cómo hacerlo.

Esta investigación servirá como referencia para dar a conocer el uso de herramientas de *Big Data* en El Salvador.

En cuanto a la tecnología, se trata de ir a la vanguardia, sobre todo en temas como las telecomunicaciones, pero el concepto de *Big Data* es algo novedoso, aunque con muchas ganas de incursionar en el uso de las herramientas que esto conlleva, pues las empresas se preguntan cómo procesar y almacenar grandes volúmenes de datos y luego analizarlos.

Es tanta la información que se genera a diario en la web a través de las redes sociales, buscadores, almacenamiento de datos en la nube, etc. por lo que resulta abrumador y solo el hecho de saber cómo se consigue captar y analizar dicha información es sorprendente.

También se sabe que las redes sociales hoy en día aportan mucha información relevante que los usuarios comparten libre y públicamente en la web. Para los que están inmersos en este medio, no es desconocido que a muchas personas les encanta publicar los lugares en los que están en un momento dado, las marcas que prefieren, ya sea ropa, zapatos, accesorios, perfumes, comidas, restaurantes, etc.

Todo esto es aprovechable por las empresas, para detectar tendencias en el mercado y enfocar las acciones que se van a llevar a cabo, algo que ayuda a tomar mejores decisiones y a que los resultados sean mejores.

Por supuesto, las ventajas las obtendrán aquellas empresas que sepan cómo procesar y analizar esos datos, y es allí donde muchas se quedan estancada al seguir haciendo los procedimientos cotidianos, por falta de conocimiento sobre el uso de herramientas que facilitarían el procesado masivo de datos y en poco tiempo.

Por otro lado, están los *dataset* públicos, que son archivos que se encuentran alojados en la nube accesible a cualquiera, los cuales están en distintos formatos y es allí donde también surge el problema, cuando los datos ya no son estructurados como comúnmente se está acostumbrado a utilizarlos en las bases de datos relacionales tradicionales, pues estos se encuentran en formatos tales como json, csv, dat, arff, ncol, etc. En estos casos se hace necesario el uso de herramientas que permitan almacenar y procesar ese tipo de ficheros.

Es por ello que se considera importante una propuesta metodológica en la que se haga uso de herramientas propias de *Big Data* para el procesamiento masivo de información, análisis de los resultados y visualización de los datos. Para ello se trabajará con dos *dataset* públicos los cuales contienen una gran cantidad de datos sobre productos alimenticios, bebidas y medicamentos de fabricantes de muchos países del mundo, esto permitirá tener un control detallado de esos productos y hacer los análisis necesarios.

Los *dataset* tienen un formato CSV (*Comma Separated Value*) y uno de ellos consiste en el registro sanitario, en el que aparecen campos tales como número de registro, nombre del producto, fecha de vencimiento, estado, fabricante, dirección, teléfono.

El otro *dataset* contiene listado de medicamentos y los campos son número, principio activo, producto, registro, titular, resolución, fecha resolución, uso o tratamiento.

Lo que se pretende con esa información es ayudar a empresas gubernamentales y privadas que necesiten conocer los procedimientos necesarios y las herramientas que serán útiles para solventar el problema de trabajar con datos masivos y obtener resultados en menor tiempo.

Se ha tomado a bien hacer uso de dos *dataset* para aprovechar las herramientas *Big Data* y mostrar lo que se puede hacer con la información que contienen.

2. Objeto de Estudio

2.1 Planteamiento del problema

Debido a todo lo anterior, surge la necesidad de hacer una propuesta metodológica en la que se hará uso de varias herramientas de *Big Data* para almacenar, procesar, analizar y visualizar datos de dos *dataset* públicos, los cuales contienen registros de productos alimenticios y medicamentos de fabricantes de muchos países del mundo, por lo que esto podría ser de gran interés para instituciones gubernamentales como el Ministerio de Salud, Instituto Salvadoreño del Seguro Social (ISSS), Fondo Solidario para la Salud (FOSALUD), Farmacias públicas y privadas, así como para cualquier empresa que se encargue de la fabricación de alimentos, bebidas y medicamentos, ya que mediante las metodologías propuestas se podrán adaptar a las necesidades de cada una de ellas.

En El Salvador y en todos los países del mundo existen leyes que regulan el registro de los productos alimenticios y medicamentos, por lo que se hace necesario contar con el procesamiento y almacenamiento del registro sanitario de dichos productos, para que mediante un análisis a esos datos se pueda conocer la condición en la que se encuentran dichos productos y cuáles son los fabricantes en los que hay menos movimiento en el mercado, debido a ello han llegado a embodegarse y a vencerse.

Con las metodologías que se proponen, se dará a conocer lo que se puede hacer con un conjunto de datos utilizando algunas de las herramientas existentes de *Big Data* y al final las empresas podrán decidir si utilizan las nuevas tecnologías o continúan realizando los procedimientos como los hacen hasta el momento.

No hay estudios previos acerca de este tema en El Salvador, pero en otros países del mundo si hay investigaciones relevantes acerca del uso de *Big Data* y de cómo ha ayudado a mejorar los procesos que se realizan en muchas áreas de la vida cotidiana. En el marco teórico se hará mención de esas investigaciones y aplicaciones que se han realizado sobre el uso de *Big Data*.

En El Salvador no se encontraron *dataset* públicos disponibles en la web con la información que se necesita para implementar las metodologías y los que habían no eran útiles para el propósito de la investigación, por lo tanto, se han seleccionado dos *dataset* que son utilizados República Dominicana y el otro en Colombia, ya que estos tienen almacenado el registro de una gran cantidad de productos alimenticios, bebidas y medicamentos, con los cuales se puede realizar un estudio más profundo y dar a conocer lo que se puede hacer con conjuntos grandes de datos y de esa manera implementarlo en el país con la información que se disponga en las instituciones que estuvieran interesadas en las metodologías propuestas.

Sabiendo que independientemente del rubro al que se dediquen las empresas, siempre es necesario contar con un control de los productos que se comercializan. Por lo tanto, este es un tema de interés, pues mediante las metodologías propuestas se podrán hacer análisis de grandes volúmenes de datos.

Con toda la información que se tiene, se procesarán los datos haciendo uso de Hadoop y luego con la herramienta Hive se realizarán las consultas necesarias, también se utilizará el programa R para hacer análisis estadístico de los datos y por último se hará uso de herramientas de visualización tales como Google Chart, Jqplot y D3.js, para presentar de forma visual los resultados obtenidos. El objetivo principal que se pretende lograr con la investigación y la propuesta metodológica es proporcionar herramientas de *Big Data* para el manejo de grandes volúmenes de datos, para su respectivo almacenamiento, análisis y representación gráfica de la información generada.

2.2 Justificación

La justificación de la realización de dicho proyecto, se debe a la gran notoriedad que está teniendo esta tendencia y parte de las nuevas tecnologías.

Cualquier persona con o sin conocimientos tecnológicos se pregunta cómo se almacena toda la información que se genera en el mundo: como son las redes sociales Facebook, Twitter, Smartcities, Instagram o como Google es capaz de manejar todas las transacciones que se hacen a diario. Pero no solo se trata de

eso, ya que *Big Data* alcanza todos los ámbitos: bolsa, climatología, astronomía, marketing, etc. por lo que la cantidad de datos que se genera actualmente es abrumadora; y solo el hecho de saber cómo se consigue captar y analizar dicha información parece una justificación bastante razonable para buscar herramientas que proporcionen soluciones atractivas.

Por otra parte, el almacenamiento de la información cada día se incrementa, por tal razón se ha decidido implementar nuevas tecnologías que cumplan con los requisitos de las grandes empresas, ya que almacenan cantidades enormes de información y requieren de mecanismos que les permitan realizar sus procesos de forma rápida y eficiente.

En la actualidad, la tecnología del *Big Data* está tomando cada vez más realce dentro del mundo de los negocios y las estrategias, el conocimiento de esta tecnología puede ser aprovechado por cualquier empresa, con el fin de ofrecer una mejor forma de brindar sus productos y servicios.

La explotación de la tecnología del *Big Data* permite que las empresas conozcan más de cerca a sus clientes, prestarle un mejor servicio, mejorar la calidad de sus productos, generar oportunidad para ingresar a nuevos mercados, completar su portafolio de clientes, entre otras tareas que generen beneficios al negocio.

Por lo tanto, la investigación sobre la tecnología de *Big Data* y el uso de herramientas que faciliten el procesamiento, análisis y visualización de los datos está basada en los siguientes indicadores: 1. Explorar los conocimientos que se tienen sobre *Big Data*, 2. Uso de las distintas herramientas para el procesamiento de los datos, 3. Conocimiento de herramientas de visualización de grandes cantidades de datos y 4. Conocer las preferencias y los elementos necesarios que se pueden utilizar para mejorar los procesos en las empresas.

2.3 Objetivos

2.3.1 General

- Diseñar una guía metodológica para explicar el análisis, procesamiento y presentación de los datos en la nube y su función en el manejo de la información.

2.3.2 Específicos

- Explorar y analizar *dataset* públicos en la nube, que contengan una gran cantidad de registros para el procesamiento y almacenamiento de la información.
- Aplicar herramientas tecnológicas que permitan el procesamiento y análisis de los datos, tales como R y Hadoop con su almacen de datos Hive.
- Analizar y presentar los resultados obtenidos utilizando herramientas de visualización tales como Google Chart, Jqplot y D3.js

2.4 Alcances

- Uno de los alcances como profesionales es aprender el uso de nuevas herramientas tecnológicas para poderlas implementar en nuestras áreas de trabajo, ya que día a día los datos van aumentando y por lo tanto se necesitarán nuevas tecnologías que sean capaces de almacenar grandes cantidades de información.
- Hacer uso de las herramientas necesarias de *Big Data* que permitan procesar y analizar grandes volúmenes de datos, para adquirir un mejor conocimiento y habilidad en la implementación de nuevas tecnologías, ya que a medida que el tiempo avanza, los sistemas de almacenamiento van creciendo cada vez más y para ello debemos de estar preparados.

2.5 Delimitación

La investigación se basa principalmente en conocer la implementación que tiene la tecnología *Big Data* y la forma en cómo se almacena la información que día a día va creciendo, por lo que se necesita que las personas que están a cargo conozcan los procesos que se deben de seguir para trabajar de una forma más efectiva.

Esta tecnología es usada en las grandes empresas que a diario generan enormes cantidades de información y es de suma importancia que las empresas puedan invertir en la implementación de nuevos equipos que cumplan con las características necesarias para que se genere una mayor seguridad en la transferencia de la información.

En esta investigación se trabajará únicamente con *dataset* públicos que están en la nube y se utilizarán las herramientas tecnológicas necesarias para el procesamiento de los datos, para su posterior análisis y presentación de la información.

3. Marco Teórico

3.1 ¿Qué es *Big Data*?

Big Data es un término aplicado a conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. Se considera un conjunto de datos que crecen rápidamente y que no pueden ser manipulados por las herramientas de gestión de bases de datos tradicionales (Aguilar, 2013)

El ser humano se ha visto en la necesidad de crear nuevas formas de comunicación y almacenar dicha información de manera constante siendo está de rápido crecimiento. Esta contribución a la acumulación masiva de datos se puede encontrar en diversas industrias, las compañías mantienen grandes cantidades de datos transaccionales, reuniendo información acerca de sus clientes, proveedores y operaciones, de la misma manera sucede con el sector público.

De acuerdo con un estudio realizado por Cisco, entre el 2011 y el 2016 la cantidad de tráfico de datos móviles crecerá a una tasa anual de 78% y el número de dispositivos móviles conectados a Internet excederá el número de habitantes del planeta. Las naciones unidas proyectan que la población mundial alcanzará los 7.5 billones para el 2016 de tal modo que habrá cerca de 18.9 billones de dispositivos conectados a la red a escala mundial, esto conllevaría a que el tráfico global de datos móviles alcance 10.8 Exabytes mensuales o 130 anuales.

Este volumen de tráfico previsto para 2016 equivale a 33 billones de DVD anuales o 813 cuatrillones de mensajes de texto. Pero no solamente son los seres humanos quienes contribuyen a este crecimiento enorme de información, existe también la comunicación denominada máquina a máquina (M2M, machine-to-machine) cuyo valor en la creación de grandes cantidades de datos también es muy importante (TestingSoft, 2014)

En la siguiente infografía se observa una representación de *Big Data*.



Figura 1 Infografía Big Data

Fuente: <http://n-economia.com/notasalerta/trasformacion-digital-big-data-infografia/>

3.2 Tipos de datos

- a. **Datos estructurados (Structured Data).** Datos que tienen bien definidos su longitud y su formato, como las fechas, los números o las cadenas de caracteres. Se almacenan en tablas. Un ejemplo son las bases de datos relacionales y las hojas de cálculo.
- b. **Datos no estructurados (Unstructured Data).** Datos en el formato tal y como fueron recolectados, carecen de un formato específico. No se pueden almacenar dentro de una tabla, ya que no se puede desgranar su información a tipos básicos de datos. Algunos ejemplos son los PDF, documentos multimedia, e-mails o documentos de texto.
- c. **Datos semiestructurados (Semistructured Data).** Datos que no se limitan a campos determinados, pero que contiene marcadores para separar los diferentes elementos. Es una información poco regular como para ser gestionada de una forma estándar. Estos datos poseen sus propios metadatos semiestructurados que describen los objetos y las relaciones entre ellos, y pueden acabar siendo aceptados por convención. Algunos ejemplos son HTML, XML, JSON, CSV. También se tiene la transformación que a continuación se explica.

Transformación

Una vez encontradas las fuentes de los datos necesarios, muy posiblemente dispongamos de un sinfín de tablas de origen sin estar relacionadas. El siguiente objetivo consiste en hacer que los datos se recojan en un mismo lugar y darles un formato.

Aquí entran en juego las plataformas ETL (Extract, Transform and Load). Su propósito es extraer los datos de las diferentes fuentes y sistemas, para después hacer transformaciones (conversiones de datos, limpieza de datos, cambios de formato...) y finalmente cargar los datos en la base de datos o Data Warehouse especificada. Un ejemplo de plataforma ETL es el Pentaho Data Integration, más concretamente su aplicación Spoon.

Almacenamiento NoSQL

El acrónimo NoSQL se refiere a Not Only SQL y son sistemas de almacenamiento que no cumplen con el esquema entidad-relación. Proveen un sistema de almacenamiento mucho más flexible y concurrente y permiten manipular grandes cantidades de información de manera mucho más rápida que las bases de datos relacionales.

Distinguimos los siguientes cuatro grandes grupos de bases de datos NoSQL.

Almacenamiento Clave-Valor (Key-Value). Los datos se almacenan de forma similar a los maps o diccionarios de datos, donde se accede al dato a partir de una clave única. Los valores (datos) son aislados e independientes entre ellos, y no son interpretados por el sistema. Pueden ser variables simples como enteros o caracteres, u objetos.

Por otro lado, este sistema de almacenamiento carece de una estructura de datos clara y establecida, por lo que no requiere un formateo de los datos muy estricto. Son útiles para operaciones simples basadas en las claves. Un ejemplo es el aumento de velocidad de carga de un sitio web que puede utilizar diferentes perfiles de usuario, teniendo mapeados los archivos que hay que incluir según el id de usuario y que han sido calculados con anterioridad.

Almacenamiento Documental. Las bases de datos documentales guardan un gran parecido con las bases de datos Clave-Valor, diferenciándose en el dato que guardan. Si en la anterior no requería una estructura de datos concreta, en este caso guardamos datos semiestructurados. Estos datos pasan a llamarse documentos, y pueden estar formateados en XML, JSON, Binary, JSON o el que acepte la misma base de datos. Todos los documentos tienen una clave única con la que puede ser accedido e identificado explícitamente.

Almacenamiento en Grafo. Las bases de datos en grafo rompen con la idea de tablas y se basan en la teoría de grafos, donde se establece que la información son los nodos y las relaciones entre la información son las aristas, algo similar a lo que sucede en el modelo relacional. Su mayor uso se contempla en casos de relacionar grandes cantidades de datos que pueden ser muy variables.

Almacenamiento Orientado a Columnas. Por último, el almacenamiento Column-Oriented es parecido al Documental. Su modelo de datos es definido como “un mapa de datos multidimensional poco denso, distribuido y persistente”. Se orienta a almacenar datos con tendencia a escalar horizontalmente, por lo que permite guardar diferentes atributos y objetos bajo una misma Clave.

3.2.1 ¿Qué tipo de datos se deben explorar en *Big Data*?

Ya se ha mencionado que es tanta la información que se maneja hoy en día y hay mucha en la web, por lo tanto, se tiene que tener claro lo que se desea analizar y el problema que se quiera resolver.

Como bien se sabe, existe una gran variedad de datos y en distintos formatos, su clasificación se puede observar a continuación, aunque esta puede variar de acuerdo con los avances tecnológicos.

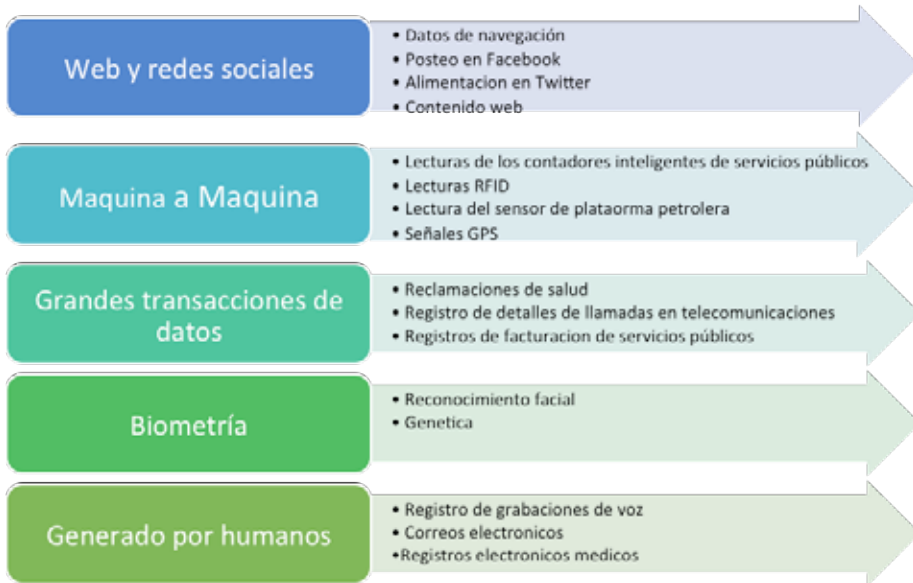


Figura 2 Clasificación de los datos Big Data

Fuente: Sitio web de IBM

1. Web y Redes Sociales (Web and Social Media). Incluye contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, blogs, etc. (IBM, developerworks, 2012)

2. Maquina a Maquina (Machine-to-Machine M2M). M2M se refiere a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.) los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa (IBM, developerworks, 2012)

3. Grandes Transacciones de Datos (Big Transaction Data). Incluye registros de facturación en telecomunicaciones, registros detallados de las llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados (IBM, developerworks, 2012)

4. Biometría (Biometrics). Información biométrica en la que se incluye huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación (IBM, developerworks, 2012)

5. Generado por Humanos (Human Generated). Las personas generamos diversas cantidades de datos como la información que guarda un call center al establecer una llamada telefónica, notas de voz, correos electrónicos, documentos electrónicos, estudios médicos, etc. (IBM, developerworks, 2012)

3.3 Componentes de una plataforma *Big Data*

Las empresas a escala mundial han atacado esta problemática desde diferentes ángulos. Todas esas montañas de información generan un costo al no descubrir el valor asociado. Actualmente, quien tiene el liderazgo en términos de popularidad para analizar enormes cantidades de información es la plataforma de código abierto Hadoop.

Hadoop

Es utilizado en la actualidad por numerosas compañías para satisfacer sus necesidades de procesamiento de *Big Data*. Algunas de las grandes compañías que emplean Hadoop son Yahoo!, para realizar los cálculos requeridos por su motor de búsqueda, y Facebook, que presume de tener el clúster más grande de Hadoop con más de 100 PB de datos.

Hadoop está inspirado en el proyecto de Google File System(GFS) y en el paradigma de programación *MapReduce*, el cual consiste en dividir en dos tareas (mapper-reducer) para manipular los datos distribuidos a nodos de un clúster logrando un alto paralelismo en el procesamiento. Hadoop está compuesto de tres piezas: *Hadoop Distributed File System* (HDFS), *Hadoop MapReduce* y *Hadoop Common*.

- **Hadoop Distributed File System(HDFS)**

Los datos en el clúster de Hadoop son divididos en pequeñas piezas llamadas *bloques* y son distribuidas a través del clúster; de esta manera, las funciones map y reduce pueden ser ejecutadas en pequeños subconjuntos y esto provee de la escalabilidad necesaria para el procesamiento de grandes volúmenes de datos.

HDFS es un sistema de ficheros que está especialmente diseñado para funcionar bien cuando se almacenan archivos grandes, que posteriormente se leerán de forma secuencial (Ghemawat S. G., 2003)

La siguiente figura ejemplifica como los bloques de datos son escritos hacia HDFS. Se observa que cada bloque es almacenado tres veces y al menos un bloque se almacena en un diferente rack para lograr redundancia.

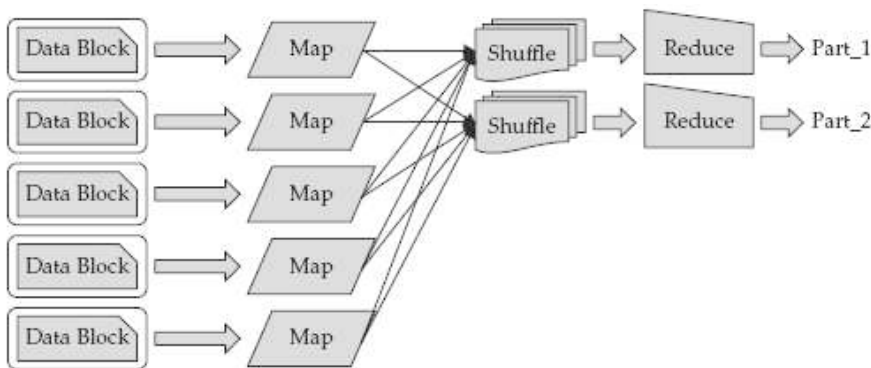


Figura 3 Ejemplo de HDFS

Fuente: Sitio web IBM

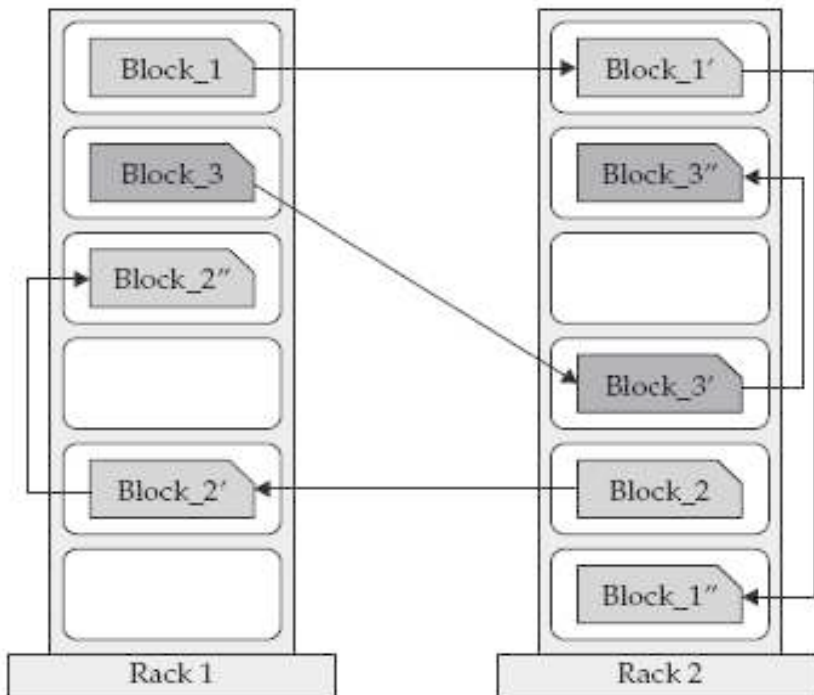
- **Hadoop MapReduce**

El motor MapReduce es un sistema que gestiona los mecanismos para ejecutar tareas *MapReduce* de forma distribuida entre los dife-

rentes nodos del clúster Hadoop. De nuevo, la forma en la que los datos se distribuyen en diferentes subtareas y cómo estas se asignan a cada máquina resulta transparente para el desarrollador.

Además, el ecosistema de Hadoop se compone de otros proyectos que, sin ser vitales para su funcionamiento, permiten realizar determinadas tareas de un modo más sencillo o más eficiente (Dean, MapReduce: Simplified Data Processing on Large Clusters. , 2004)

La siguiente figura ejemplifica un proceso sencillo de MapReduce.



*Figura 4 Ejemplo de MapReduce
Fuente: Sitio web IBM*

- **Hadoop Common**

Hadoop Common Components son un conjunto de librerías que soportan varios subproyectos de Hadoop.

3.3.1 Principales distribuciones de Hadoop

En principio, todos los componentes del proyecto Hadoop, así como los demás proyectos relacionados, se pueden descargar del sitio web de Apache, donde se puede encontrar también documentación para llevar a cabo su instalación.

No obstante, muchas compañías han decidido lanzar sus propias distribuciones de Hadoop. Las características de cada una de estas distribuciones dependen del fabricante, pero en general tienen las siguientes características:

- ✓ Ofrecen un ecosistema completo e interoperable. Como se comentó anteriormente, son muchos los proyectos que surgen en torno a Hadoop y al evolucionar cada uno de ellos de forma independiente, en ocasiones hay que ser cuidadoso de escoger una combinación de versiones que funcione correctamente. Las distribuciones de Hadoop ofrecen un ecosistema completo que ha sido testeado para garantizar que todos sus componentes funcionen correctamente.
- ✓ Ofrecen soporte (más allá del ofrecido por la comunidad de Hadoop), si bien este soporte no tiene por qué ser gratuito.

Además de las distribuciones (que pueden ser gratuitas o de pago), muchos fabricantes ponen a disposición de los usuarios lo que llaman un *sandbox*, en el que ofrecen una máquina virtual con Hadoop preinstalado (no apto para entornos en producción), que complementan con tutoriales u otros recursos útiles para los desarrolladores que no tengan experiencia.

a. Hortonworks

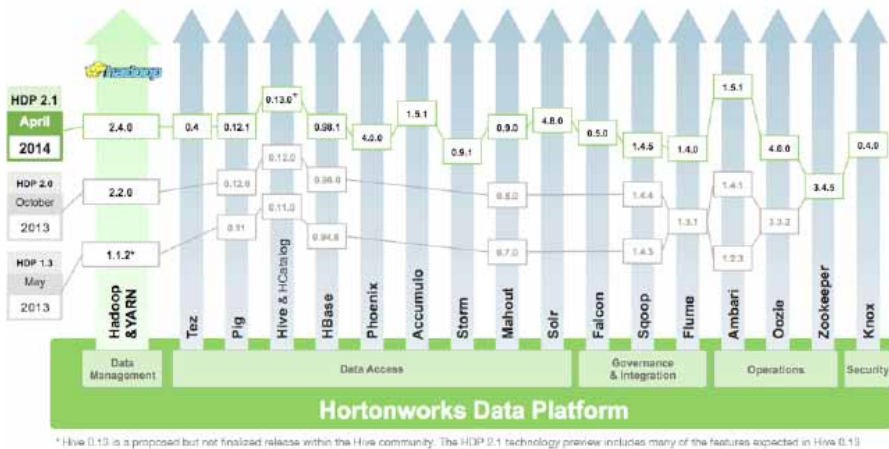


Una de las distribuciones más extendidas de Hadoop es Hortonworks Data Platform, que se presenta a sí misma como la distribución 100% opensource de Hadoop y una de las que más ha contribuido al desarrollo de código del proyecto Hadoop.⁶

Hortonworks incorpora numerosos proyectos que se integran con Hadoop para aumentar el abanico de posibilidades que ofrecer a los desarrolladores y a los usuarios.

La siguiente figura muestra los diferentes componentes incluidos en Hortonworks, así como las versiones que se han escogido.

⁶ <http://hortonworks.com/>



* Hive 0.13 is a proposed but not finalized release within the Hive community. The HDP 2.1 technology preview includes many of the features expected in Hive 0.13

Figura 5 Componentes de Hortonworks Data Platform

Fuente: <http://hortonworks.com/>

Una de las particularidades de Hortonworks es que la distribución de Hadoop que ofrece está disponible tanto para sistemas GNU/Linux como para Microsoft Windows, siendo la única distribución a día de hoy que soporta este último sistema operativo.

Además, Hortonworks ofrece una sandbox consistente en una máquina virtual con Hadoop preinstalado, complementado con el proyecto Hue (www.gethue.com), que ofrece una interfaz web sencilla y manejable para realizar algunas operaciones con Hadoop.

Además, Hortonworks ofrece la posibilidad de instalar y gestionar Hadoop a través de Ambari (ambari.apache.org), lo que simplifica sustancialmente la tarea de desplegar Hadoop en un clúster.

b. Cloudera



Cloudera ofrece CDH (Cloudera Data Hub) como distribución de Hadoop, que también se presenta como una distribución 100% open-source al igual que Hortonworks. Según indica la propia empresa, esta distribución cuenta con un volumen de descargas que supera a la suma de todas las demás distribuciones juntas.⁷

Basándose en esta distribución, Cloudera ofrece diferentes soluciones, tales como Cloudera Express (que es gratuita de forma ilimitada), o diferentes paquetes de Cloudera Enterprise, que ofrece funcionalidades complementarias y soporte adicional. Además de estas soluciones, también pone a disposición de los usuarios Cloudera Live, que permite probar Hadoop online utilizando la interfaz Hue sin necesidad de instalarlo ni de descargar ninguna máquina virtual, por lo que resulta una alternativa interesante a Hortonworks Sandbox para usuarios que están empezando con Hadoop y no tienen recursos para desplegar una distribución. Finalmente, Cloudera es reconocido por ofrecer numerosos cursos de formación y también certificaciones en Hadoop, si bien estos cursos no son gratuitos.

c. MapR



⁷ <http://www.cloudera.com/>

MapR ofrece M3 como distribución básica de Hadoop, disponible de forma gratuita para su descarga. Al igual que las otras distribuciones discutidas anteriormente, combina Hadoop con otros proyectos de Apache que los complementan y que ofrecen funcionalidad extra, además de una consola propia para la gestión del clúster.⁸

Por encima de M3, MapR ofrece las distribuciones de M5 y M7, que añaden más funcionalidad para entornos en producción (tales como protección de datos, alta disponibilidad, etc.) y soporte extendido.

Finalmente, MapR también ofrece la posibilidad de descargar un sandbox para comenzar con Hadoop, funcionando sobre una máquina virtual que el usuario puede ejecutar en su ordenador.

3.4 Investigación y ejemplos de aplicación de *Big Data*

En este apartado se hace referencia a investigaciones que se han hecho anteriormente, así como de la historia general sobre el auge de *Big Data* en la actualidad y como beneficiarse con las herramientas que existen en esta área.

Existen investigaciones sobre la aplicación de *Big Data*, entre ellas están las siguientes:

- ✓ Investigación de *Big Data* en los entornos de defensa y seguridad (Carrillo Ruiz, y otros, 2013)
- ✓ El *Big Data* puede ayudar en el diagnóstico y tratamiento del cáncer (Bernardo, 2013)
- ✓ *Big Data*, recurso para ciudades inteligentes (Souto, 2015) el nuevo recurso natural para las ciudades inteligentes, el cual aprovecha múltiples fuentes de datos; este analiza los datos a través de la tecnología analítica y permite a los líderes servir mejor a los ciudadanos y negocios en un mundo cambiante. Este recurso aprovecha algoritmos predictivos para resolver los problemas proactivamente.

⁸ <http://www.mapr.com/>

- ✓ *Big Data* ayuda al transporte inteligente en Dublín (Irlanda). Mediante el uso de herramientas de *Big Data*, se puede conocer el estado actual de toda la red de buses de un vistazo, y en forma detallada en áreas donde hay problemas para identificar la causa de la congestión antes de que se extienda por otras rutas. La población de Dublín en el 2013 era de 1660000. La información archivada ha servido para analizar a posteriori y entender lo que pasó y así tomar medidas para optimizar el tráfico (Zikopoulos, 2015)
- ✓ *Big Data* ayuda a cerveceras a delinear su marketing de acuerdo con el país. La brasilera Vortio (2013) analizó todas las conversaciones en las redes sociales sobre la palabra clave cerveza en diferentes idiomas y países. Resultado: a diferentes culturas, diferentes comportamientos frente a la cerveza. En dos semanas de análisis, parece que los americanos (EEUU) beben para relajarse. Los italianos beben por causa de problemas de relación con su pareja; los alemanes para abastecer el tanque; los franceses para apreciar el gusto; los argentinos porque es saludable; los brasileños beben cuando salen de fiesta. Para esta investigación se analizaron y midieron percepciones, opiniones, etc.
- ✓ *Big Data* ayudó a Obama a ganar las elecciones el 2012 (Scherer, 2012) El equipo de dirección de campaña creó una mega base de datos con información de votantes y simpatizantes a partir de múltiples fuentes desde las elecciones de 2008. Analizaron la información a fin de identificar los gustos y preferencias de sus seguidores. Crearon diferentes aplicaciones para lograr lo siguiente.
 1. Mejores decisiones con mayor volumen de datos.
 2. Traducir datos en bruto para realizar análisis predictivo.
 3. Establecer las preferencias de los votantes.
 4. Solucionar problemas de volumen.
- ✓ *Big Data* ayuda a las tiendas Macy's de EEUU a incrementar sus

ventas (De Juana, 2015). Hasta el 2010, Macy's seguía utilizando hojas de cálculo Excel para analizar grandes volúmenes de datos de clientes. Ahora, con *Big Data* analiza decenas de millones de Terabytes (1012) de información cada día, y ha pasado de 22 horas a 19 minutos para rehacer el precio de sus artículos. Han conseguido lo siguiente:

1. Mismas decisiones en menos tiempo.
 2. Un incremento del 10% en sus ventas.
- ✓ *Big Data* ayuda a General Electric a mejorar sus productos (Fondos-Fidelity, 2012). En el 2011, GE invirtió mil millones de dólares en un centro de investigación para mejorar sus diferentes productos. Allí analizan un gran volumen de datos procedentes de multitud de sensores y otros dispositivos digitales.

Las investigaciones citadas anteriormente solo son una pequeña muestra de las aplicaciones de *Big Data* en algunas áreas, pero en términos generales los principales sectores que usan esta tecnología son los siguientes:

- **Marketing**, conocido también como Business Intelligence o Inteligencia de Negocios y es muy utilizado en las empresas de muchos países del mundo, debido a que mediante estrategias bien definidas logran crear y administrar conocimiento sobre el medio, a través de análisis de los datos existentes dentro de la empresa.
- **Redes Sociales**, estas son las que más beneficios han obtenido con el auge *Big Data*, pues mediante técnicas de recolección de datos se puede llegar a saber las preferencias de las personas, lo cual es bien utilizado por las empresas para ser más competitivas.
- **Meteorología**, manejan datos de gran tamaño y bases de datos de los matemáticos desde hace mucho tiempo, por lo tanto, las herramientas *Big Data* han ayudado con ese problema.
- **Ingeniería**, esta área es muy amplia y es por ello que se ve beneficia-

da con las tecnologías *Big Data*, como el transporte, sector energético, telecomunicaciones, etc.

- **Medios de comunicación**, debido a que éstos están inmersos en todos los aspectos en el manejo de la información masiva.

3.4.1 Ventajas de utilizar herramientas de *Big Data*

El uso de las tecnologías *Big Data*⁹ en la actualidad, ha venido a ayudar en gran manera en el manejo de grandes volúmenes de datos, eso se ha podido comprobar en las investigaciones que se han mencionado anteriormente.

El uso de estas tecnologías permite el tratamiento y análisis de grandes repositorios de datos que de otra manera no fuera posible si se quisiera lograr con las herramientas de bases de datos tradicionales, ya que estos se vuelven insuficientes en todos los aspectos.

Se ha mencionado que hoy en día existe una mayor cantidad de datos que se almacenan, los cuales proceden de páginas web, aplicaciones de imágenes y videos, redes sociales, dispositivos móviles, apps o sensores, por lo tanto, es necesario contar con herramientas potentes que permitan, almacenar, procesar y analizar esos datos para fines diversos, los cuales pueden ser negocios, salud, comercialización de productos, etc.

3.4.2 Desventajas de utilizar herramientas de *Big Data*

Se mencionan mucho las ventajas que trae consigo el uso de las herramientas *Big Data* y se podría pensar que no hay desventajas que mencionar, pero realmente si las hay, sobre todo en países en desarrollado en los que existe mucho desconocimiento en esta área, pero que tienen enormes deseos por incursionar en las nuevas tecnologías.

Las desventajas que se pueden mencionar por la experiencia propia en El Salvador son las siguientes:

⁹ <http://www.alten.es/wp-content/uploads/2014/07/CONCLUSIONES-PROYECTO-BIG-DATA.pdf>

- Falta de profesionales expertos.
- Resistencia al cambio por miedo al fracaso.
- Falta de inversiones destinadas a implementar soluciones *Big Data*.
- Dificultad de integración en los procesos internos de las empresas.
- Falta de interés por innovar y por capacitar al personal de las empresas.

Por estas desventajas, las empresas en El Salvador y probablemente en otros países del mundo, no se deciden a incursionar en la tecnología *Big Data*, pues la disposición de profesionales en esta área es bien limitada o nula, aparte que hay que invertir de alguna manera en equipo y capacitaciones y por otro lado hay que cambiar todos los procesos internos para migrar a las nuevas tecnologías.

Muchas veces la resistencia al cambio y el miedo al fracaso truncan las posibilidades de mejorar y ser mucho más productivos profesionalmente. Pero las empresas que han asumido ese reto se han dado cuenta de los beneficios que han obtenido, debido a que los procesos se hacen en menos tiempo y pueden almacenar grandes volúmenes de datos.

En El Salvador, muchas empresas trabajan con las bases de datos relacionales tradicionales, como SQL, Oracle, SyBase, MySQL, pero en esos casos se trabaja con datos que están estructurados. En el peor de los casos, utilizan Excel o Access para el almacenamiento de los datos, por lo que el tiempo de respuesta es mucho mayor y la cantidad de datos que se pueden almacenar no es en grandes volúmenes.

Se sabe que los datos en la nube carecen de estructura y están almacenados en formatos que no se pueden trabajar con las bases de datos relacionales, es por ello que se necesita de herramientas que permitan procesar y analizar grandes volúmenes de datos y en el menor tiempo posible.

A pesar de las desventajas que se mencionan, las empresas se van dando cuenta que cada día se encuentran con mayor cantidad de datos, los cuales provienen de fuentes diversas y el problema es que muchas veces no son datos

estructurados y es allí donde surge la necesidad de hacer uso de otras herramientas que no sean las convencionales.

Las herramientas *Big Data* proporcionan la habilidad de procesar grandes volúmenes de datos en solo segundos, lo cual permite elevar la competitividad y además da lugar a nuevos enfoques e ideas y por ende se puede cambiar la forma fundamental en la que se gestiona la empresa.

3.5 Teorías

Big Data es un sistema genérico que debe tratar una gran cantidad de datos al que le hace falta integrar muchas herramientas para que sea lo que dice su nombre, dependiendo de la cantidad de datos, su tipo, relación entre ellos, modelos y algoritmos que han de ejecutar.

En esencia, se trata de un conjunto de tecnologías y arquitecturas diseñadas para conseguir un mejor rendimiento de grandes volúmenes de información. Como ocurre con cualquier modelo de negocio, el factor clave para obtener beneficios de *Big Data* no depende de la capacidad tecnológica sino de la capacidad humana para realizar la correcta interpretación de la información que permita obtener valor de su análisis.

3.5.1 Potencial *Big Data*

Big Data no es sólo una herramienta o una tecnología si no un conductor de una disciplina de toma de decisiones mejorada basada en análisis predictivos, que marca el comienzo de una era de cambio cultural y mejora del rendimiento. La experiencia del usuario será clave no sólo en la venta de servicios, sino también en los productos.

Con *Big Data* la venta de productos o servicios podrá diferenciarse haciendo que su consumo suponga una experiencia personalizada para los gustos y preferencias de cada cliente. *Big Data* permitirá llevar a cabo la gestión de emociones a la hora de enriquecer el consumo de los productos y servicios.

Big Data no es una actividad aislada. Para el éxito se necesita más que nunca el conocimiento del negocio que permita hacer las preguntas correctas y establecer las correlaciones oportunas. Negocio y Tecnología de la Información deben de ir de la mano desde el primer momento.

Sin duda alguna, uno de los retos de *Big Data* es incorporar a su capacidad analítica información de contexto que permita adaptar y comprender el resultado del análisis en base a las condiciones del entorno. Para ello, el verdadero conocimiento será aquel que incorpore los atributos que contextualicen el análisis.

La contextualización del dato trata de responder e incorporar al análisis información relativa a cuándo se obtuvo la fuente origen, cómo se obtuvo, de dónde procede, cuál es su naturaleza.

Existe una gran complejidad para realizar un análisis cuando el número de variables es muy alto, mucha de la información puede no ser útil o considerarse falsa. *Big Data* puede derivar que se encuentren correlaciones falsas o falsos positivos. Para intentar solventar esta problemática en el rigor del análisis de los datos, existen ciertas premisas que ayudan a evitar errores, las más importantes son las siguientes.

- Es primordial comenzar con pequeños pilotos para ganar experiencia y conocimiento de las nuevas tecnologías.
- Es recomendable trabajar con expertos para evitar cometer grandes errores.
- Construir un modelo que permita conocer a futuro y corrija los errores permitiendo la optimización de los procesos de negocio.

Un ejemplo claro de esto lo podemos encontrar en Google, respecto a las predicciones sobre la epidemia de gripe en América del Norte, Google hizo un estudio para conocer cómo es que se había propagado la epidemia, dos años más tarde el estudio ya no era válido debido a que los datos habían cambiado y el sistema no había sido realimentado de forma asistida.

Big Data es mucho más que volumen de información, muchos tipos de variables, muchos tipos de observaciones, muchos resultados. Lo realmente importante es lo que se muestra a continuación.

- Cómo están extraídos los datos.
- De dónde provienen.
- Su fiabilidad.
- De todos los datos, cuáles son los que son relevantes para integrar en el sistema.
- La relevancia forma parte de la respuesta.
- Esto limita el alcance del sistema.
- Influye sobre la definición misma del sistema.

Por otro lado, para la meteorología, el tema ya está abordado completamente por la riqueza en cuanto a la gran cantidad de datos continuos con los que se alimenta el sistema. La evolución de orígenes de datos es constantemente reevaluada con el fin de poder integrarlas de forma adecuada en el sistema.

3.6 Contexto

Como se ha venido hablando anteriormente, *Big Data* está siendo utilizado en muchas aplicaciones e incluso en nuestra vida cotidiana. La economía familiar, la educación, el entretenimiento, la política o incluso nuestra salud se ven implicada e influenciada.

Según el libro del ingeniero informático Joyanes Aguilar, *Big Data: Análisis de grandes volúmenes de datos en organizaciones* (Joyanes Aguilar, 2013), podemos darnos cuenta hasta qué punto revoluciona el *Big Data* nuestra vida cotidiana.

Hoy en día, en el ámbito económico, los grandes almacenes se benefician de estrategias basadas en la comparación analítica de datos para proporcionar ofertas a los clientes. En Estados Unidos, el conocido Black Friday usa tecnologías *Big Data* para ofrecer productos a diferentes clientes dependiendo de la hora del día.

Otro aspecto en el que encontramos la intervención del *Big Data* es en la educación, una de las áreas más importantes de nuestras vidas. En últimas investigaciones realizadas, se ha descubierto el uso de data analytics para métodos pedagógicos, o para itinerarios personalizados para cada alumno. Sin duda es uno de los mayores beneficios que podemos encontrar, teniendo en cuenta la decadencia que dicha área sufre en los últimos años. Con esta nueva tecnología, se adapta personalmente cada alumno a sus propios intereses, su talento o aptitudes, mejorando el rendimiento escolar.

Otro de los aspectos más importantes en la vida cotidiana es la salud, en la cual el *Big Data* se implica también de lleno. Gracias a las nuevas tecnologías, es posible monitorizar nuestra actividad física para obtener pautas personalizadas o consejos de rendimiento a largo plazo, así como diagnosticar enfermedades.

Encontramos también, la influencia del *Big Data* en el entretenimiento. Interviene proporcionando a los usuarios recomendaciones personales, teniendo en cuenta los servicios que ya han consumido. Una de las tiendas virtuales más importantes en la actualidad, Amazon, ofrece ya el servicio de predecir cuáles serán nuestras propias transacciones, haciendo más fácil así la elección de nuestros productos. Esto lo logra haciendo uso de técnicas de Inteligencia Artificial, específicamente haciendo uso de sistemas de recomendación.

Se ha visto, por lo tanto, que el *Big Data* se encuentra en la vida cotidiana en casi todos los aspectos. Facilita los servicios personalizados, predice elecciones y monitoriza actividades para un mayor control sobre cada persona. Sin duda, el *Big Data* puede llegar a cambiar la vida tal y como se concibe.

Y por último como dato curioso y eso se pudo apreciar en el Marco Teórico, el presidente Barack Obama también incluyó la utilización de técnicas *Big Data* para su campaña electoral.

Es asombroso todo lo que se ha investigado y como *Big Data* está involucrado en muchos aspectos de nuestra vida, por lo tanto, no nos podemos quedar atrás en la aplicación de esta tecnología y es por ello que surge la necesidad de trabajar en la elaboración de metodologías para el uso de herramientas de *Big Data* y darlas a conocer a empresas en El Salvador que

puedan estar interesadas en incursionar en esta área y poder mejorar sus procesos de almacenamiento y el análisis de los datos para llegar a ser más productivos.

La metodología propuesta consiste en la elaboración de los pasos necesarios para la utilización de herramientas *Big Data* tales como Hadoop y Hive para el almacenamiento, procesamiento y análisis de los datos más completo o para facilitar algunas tareas y realizar consultas posteriores, utilización del programa R para el análisis estadístico de los datos y luego hacer uso de herramientas de visualización ya sea Google Chart, jqplot o D3.js, según sea necesario.

A continuación, se presenta una breve descripción de las herramientas que se van a utilizar en la metodología propuesta.

3.7 Descripción de las herramientas que han de utilizarse en la investigación

3.7.1 Qué es Hadoop

En la actualidad, Hadoop¹⁰ es un proyecto de *software* libre, con licencia Apache, cuya finalidad es prestar una plataforma para la gestión de grandes cantidades de datos. Los principales componentes que constituyen Hadoop son el sistema de archivos HDFS y el motor MapReduce.

Hadoop Distributed File System (HDFS) es un sistema de archivos distribuido inspirado en el GFS de Google, que permite distribuir los datos entre distintos nodos de un clúster (llamados *datanodes*), gestionando la distribución y la redundancia de forma transparente para el desarrollador que vaya a hacer uso de esos datos.

El motor MapReduce es un sistema que gestiona los mecanismos para ejecutar tareas MapReduce de forma distribuida entre los diferentes nodos del clúster Hadoop. De nuevo, la forma en la que los datos se distribuyen en diferentes subtareas y cómo estas se asignan a cada máquina resulta transparente para el desarrollador, además, el ecosistema de Hadoop se compone de otros proyec-

¹⁰ Información tomada del libro *Hadoop. The Definitive Guide (3rd Edition)*.

tos que, sin ser vitales para su funcionamiento, permiten realizar determinadas tareas de un modo más sencillo o más eficiente.

Existen varias distribuciones de Hadoop, tales como Hortonworks, Cloudera, MapR, pero para este trabajo de investigación se utilizará el primero. Anteriormente se habló de esta distribución de una forma más específica y de las ventajas que tiene para utilizarla.

Para llevar a cabo el despliegue de Hadoop es necesario comenzar realizando los siguientes pasos:

1. Decidir la arquitectura física del sistema. Este paso es probablemente el más complicado en un principio, pues requiere tener una visión global del uso que se le dará al sistema. Esta decisión incluye conocer el número de nodos del clúster, los servicios que ejecutará cada uno de ellos, la distribución física de los equipos, etc. Una de las principales ventajas que ofrece Hadoop es que es relativamente sencillo adaptar la infraestructura física a las necesidades que surjan en un futuro (por ejemplo, añadiendo nuevos nodos).

2. Decidir la distribución que se va a desplegar, en caso de que se haga. Como se ha comentado anteriormente, la distribución que se utilizará será Hortonworks, aunque todas las demás ofrecen combinaciones de diversos proyectos del ecosistema Hadoop cuya interoperabilidad ha sido testeada, además de funcionalidades extra y soporte.

Para la elaboración de la propuesta metodológica se empleará una arquitectura basada en dos máquinas virtualizadas (de las cuales una hará de máster y la otra de esclavo), esto por la carencia de recursos físicos para llevar a cabo el despliegue.

En cuanto a la distribución, se sugiere instalar Hortonworks, puesto que dispone de un instalador y un configurador del clúster basado en Apache Ambari, lo que simplifica enormemente el proceso.

Por qué Hadoop.

En la actualidad, la cantidad de datos que se generan a cada segundo es inmensa. En 2012, IBM publicó una infografía¹¹ en la que, basándose en fuentes tales como estudios de IDC o EMC, resumía el estado actual en lo referente a la cantidad de datos que inundaba la web.

Algunas de las cifras más significativas son las *siguientes*:

- Se mandan 294.000 millones de emails diariamente.
- Se suben 100 terabytes de datos a Facebook diariamente.
- Se generan 5 exabytes (millones de terabytes) cada dos días.
- Existen en el universo digital 2.7 zettabytes (miles de millones de Terabytes) de datos.

Esta inmensidad de datos se puede explicar fundamentalmente en base a los siguientes tres orígenes:

- La interacción entre humanos que emplean un sistema informático que registra información mientras se produce dicha interacción, son los casos de los emails, de los foros de internet y de las redes sociales, en los que los datos los generamos los humanos y estos son almacenados o procesados por máquinas.
- La interacción entre un humano y una máquina. Este caso se da cuando navegamos por Internet y los servidores web generan logs con información sobre el proceso de navegación, o cuando compramos en una plataforma de comercio electrónico o empleamos la banca online y un sistema registra nuestras transacciones.
- La interacción entre máquinas (M2M), en las que son varias máquinas las que intercambian información entre ellas y la registran. Algunos ejemplos son sistemas de monitorización, en los que un sistema de

¹¹ <http://www.ibmbigdatahub.com/infographic/flood-big-data>

sensores proporciona la información que recibe a otras máquinas para que realicen algún procesado sobre ella.

Probablemente, más importante que la cantidad de datos que se genera en la actualidad es entender que este ritmo crecerá en el futuro, puesto que cada vez son más las personas que tienen acceso a Internet y la variedad de dispositivos que se conectan a la red. En esta misma infografía, IBM revela que en 2020 se generarán 35 zettabytes de datos anualmente.

Este ritmo de generación de datos introduce numerosos desafíos en lo que concierne al modo en que se almacenan y, especialmente, la forma en la que deben ser procesados. Resulta evidente que los sistemas tradicionales son incapaces de manejar esta información de una forma eficiente, puesto que no están preparados para soportar la explosión de datos de los últimos años.

Las principales compañías que vinieron notando esta necesidad de sistemas para el almacenamiento y procesado más eficiente de grandes cantidades de datos fueron los buscadores de Internet.

La labor de un buscador de Internet es, en teoría, relativamente sencilla. Lo primero que debe hacer es rastrear la web, siguiendo los hipervínculos de cada página, para ir construyendo un grafo (una estructura de datos que relaciona diferentes nodos, en este caso páginas web entre sí por medio de enlaces). A continuación, debe elaborar un índice invertido en el que se pueda localizar fácilmente una web dados unos términos de búsqueda.

Además, los buscadores suelen emplear una función de ranking por medio de la que asignan a cada página web un peso en función de su relevancia, que se puede calcular en base a la cantidad de páginas que enlazan con la página asignada y al mismo tiempo, a la relevancia de cada una de las páginas.

La principal dificultad que encuentran los buscadores es que la cantidad de páginas web disponibles es inmensa, lo que dificulta llevar a cabo todo el proceso de indexación en un tiempo razonable; como para mantener los resultados de búsqueda actualizados.

En el año 2003 (Ghemawat, Gobioff, & Leung, 2003), Google publica su famoso artículo¹² en el que describe Google File System, un sistema de ficheros escalable y distribuido que pretende subsanar la dificultad de tener que almacenar grandes cantidades de datos de forma confiable y proporcionando un alto rendimiento para aplicaciones que realizan un uso intensivo de estos datos.

Al año siguiente, Google publica otro artículo (Dean, MapReduce: Simplified Data Processing on Large, 2004) en el que describe el paradigma de programación MapReduce, cuya finalidad es llevar a cabo un procesado distribuido de grandes cantidades de datos de forma eficiente.

MapReduce

MapReduce es un desarrollo que responde a la necesidad de Google de procesar grandes cantidades de datos de manera eficiente, de forma paralela. Además, es un paso intuitivo tras el desarrollo de Google File System (GFS); puesto que ahora hay grandes cantidades de datos almacenadas de forma distribuida entre varios equipos, resulta oportuno realizar un procesamiento también distribuido de estos datos.

La idea detrás de MapReduce es sencilla: una aplicación MapReduce cuenta con una rutina `map()` y otra rutina `reduce()`, que son las que dan nombre a este modelo de programación.

La rutina `map()` recibe una tupla clave-valor ($\langle k, v \rangle$) y devuelve un conjunto de tuplas clave-valor ($\langle km, vm \rangle$).

Posteriormente, todas las claves devueltas por las rutinas `map()` ejecutadas se ordenan y se agrupan por clave, resultando un conjunto de tuplas que contienen una clave y una lista de valores. Por ejemplo, si las rutinas `map()` habían devuelto las tuplas

¹² <http://static.googleusercontent.com/media/research.google.com/es//archive/gfs-sosp2003.pdf>

$\langle k_{mi}, v_{mi,1} \rangle, \langle k_{mi}, v_{mi,2} \rangle, \dots, \langle k_{mi}, v_{mi,n} \rangle$, tras esta fase todas estas tuplas se agruparán en una tupla $\langle k_{mi}, [v_{mi,1}, v_{mi,2}, \dots, v_{mi,n}] \rangle$.

Por último, la rutina `reduce ()` recibe como entradas estas tuplas agrupadas y devuelven, para cada una de ellas, un conjunto de valores ($\langle vr \rangle$).

Probablemente trabajar bajo este paradigma de programación, para muchos no sea fácil de comprender y aplicarlo, pero por esa razón se trabajará con una herramienta que viene incluida dentro de Hadoop, la cual es mucho más fácil, sobre todo para aquellos que en alguna ocasión han trabajado con base de datos relacionales como SQL. Esta herramienta es conocida como Hive, la cual incorpora el procesado MapReduce.

Hive

Apache Hive¹³ es un proyecto que forma parte del ecosistema Hadoop y, por ello, viene incluido en muchas distribuciones de Hadoop, y también en la distribución Hortonworks.

El propósito de Hive es, en cierto modo, emular un sistema de bases de datos relacional encima de Hadoop.

Así, el usuario podrá crear tablas e insertar datos (o crearlas a partir de ficheros existentes en HDFS), para posteriormente consultarlas empleando un lenguaje de modelado y de consulta muy similar a SQL.

Es importante entender que esta lógica funciona bien cuando trabajamos con datos que son estructurados, puesto que el concepto de tablas en el modelo relacional estructura los datos en columnas (campos) y en filas (registros).

Hive es una herramienta adecuada para usuarios que estén familiarizados con las bases de datos relacionales. Permite crear tablas y hacer consultas sobre

13 Información tomada del libro Programming Hive de la dirección web <https://books.google.es/books?id=n8FffirFaQIC&printsec=frontcover&hl=es#v=onepage&q&f=false>

ellas empleando un lenguaje similar a SQL, si bien estas consultas se traducirán automáticamente a rutinas MapReduce.

3.7.2 Qué es R

Se puede definir R¹⁴ desde dos perspectivas:

- R es un entorno de software.
- R es un lenguaje de programación.

Fundamentalmente R puede ser definido como un entorno software para el análisis matemático y estadístico de datos, en cierto sentido similar a herramientas tales como Microsoft Excel. A través del entorno de R vamos a ser capaces de manipular datos (por ejemplo, cargarlos desde ficheros, editarlos, volverlos a almacenar...), realizar análisis sobre esos datos y presentar los resultados gráficamente para facilitar su interpretación.

El entorno software viene acompañado de un lenguaje de programación que pone a nuestra disposición las funcionalidades típicas de un lenguaje de propósito general (manejo de variables, tipos y estructuras de datos, operadores, mecanismos de control del flujo de ejecución, funciones, etc.) combinadas con librerías y herramientas específicas para facilitar el análisis de datos. Utilizando este lenguaje es relativamente sencillo implementar nuestras propias funciones y scripts para automatizar el procesamiento de ciertos datos.

En la práctica, estas dos perspectivas están muy relacionadas. Así, por ejemplo, para interactuar con el entorno de R se utilizarán expresiones escritas en el lenguaje R.

¿Por qué utilizar R?

Actualmente existe una amplia gama de herramientas que pudiéramos pensar en utilizar a la hora de llevar a cabo análisis de datos (por ejemplo, Microsoft Excel, S-PLUS, una versión comercial del lenguaje S, SAS, SPSS

¹⁴ Fuente: <https://www.r-project.org/about.html>

de IBM, etc.). Así pues, una de las cuestiones que las empresas pudieran plantear en esta metodología es por qué elegir R como herramienta de análisis de datos.

Algunas de las razones que se podría emplear a la hora de justificar la decisión incluye lo siguiente:

- R es software de código libre con licencia GNU GPL (General Public License).
- Mientras que las principales herramientas de análisis tienen un costo (algunas con precios bastante elevados), R es completamente gratuito.
- Existen versiones para los sistemas operativos más comunes: Windows, Mac OS X y Linux.
- Posee una comunidad de usuarios amplia y muy activa, por lo que va a resultar relativamente sencillo encontrar documentación o ayuda en foros si resulta necesario.
- El entorno es fácilmente extensible, mediante el desarrollo de paquetes. Debido a esto, evoluciona rápidamente: nuevos algoritmos y técnicas de análisis se incorporan con regularidad.
- R y sus extensiones nos ofrece una gran variedad de herramientas de análisis y visualización de datos. Actualmente existen más de 5000 paquetes disponibles para ser instalados en el entorno.

3.7.3 Herramientas de visualización

Google Chart¹⁵ es una aplicación de Google para realizar estadísticas web, de fácil uso para desarrolladores de software web, usado en muchos campos como Google Analytics, se puede usar con diferentes formatos, Json, Javascript y plugins que se pueden integrar con varios lenguajes de programación.

Esta herramienta permite realizar gráficos atractivos y existe una gran variedad de galerías disponibles en el sitio de Google para utilizarlas y adaptarlas a las necesidades de análisis de cada persona.

¹⁵ <https://developers.google.com/chart>

Jqplot¹⁶ es un framework para el trazado de gráficos y plugins jQuery Javascript, jqPlot. Produce hermosas líneas, barras y gráficos circulares con muchas características; las siguientes son algunas de estas:

- Numerosas opciones de estilo gráfico.
- Fecha eje con formato personalizable.
- Hasta nueve ejes *Y*.
- Texto eje girado.
- Cálculo automático de la línea de tendencia.
- La información sobre herramientas y punto de datos resaltada.
- Valores predeterminados razonables para facilitar su uso.

D3.js¹⁷ o simplemente D3 de documentos basados en datos. Es una biblioteca JavaScript para producir visualizaciones de datos dinámicos e interactivos en los navegadores web. Hace uso ampliamente de SVG, HTML5 y estándares CSS. En contraste con muchas otras bibliotecas, D3.js permite un gran control sobre el resultado visual final. Para poder hacer uso de esta herramienta, es necesario conocer de JavaScript, por consiguiente, hay que aprender ese lenguaje de programación.

4. Metodología de la Investigación

4.1 Metodología

La investigación se realizó con el objetivo de conocer sobre el auge de *Big Data*, y como se podrían utilizar algunas herramientas que incluye *Big Data* para el procesamiento, análisis y visualización de los datos. Se hizo un estudio exploratorio de tipo descriptivo por medio de una encuesta en la que se seleccionó únicamente a profesionales en el área de Informática, debido a que se necesitaba una opinión bastante real en cuanto a conocimientos con respecto a las nuevas tecnologías.

¹⁶ <http://www.jqplot.com/index.php>

¹⁷ <https://d3js.org/>

4.2 Participantes

El proceso de recolección de datos se hizo con profesionales en el área de Informática que labora en el departamento de San Salvador. La muestra fue aleatoria, en la cual participaron 40 personas de ambos géneros y de distintas edades.

4.3 Instrumento

En la encuesta se busca explorar los conocimientos con los que cuentan actualmente los profesionales en el área de Informática sobre *Big Data*. Dicha encuesta fue elaborada con la herramienta Google Forms, la cual cuenta con 12 preguntas cerradas y a cada participante se les proporciono el link para que la pudieran contestar en línea, en el momento que dispusieran. La encuesta nos ayudó a determinar la importancia de la elaboración de una propuesta metodológica en el uso de herramientas de *Big Data*.

Puede verse la encuesta en el anexo 8 y los resultados en el anexo 9.

4.4 Procedimiento

Lo primero que se realizó fue enviarles el link de la encuesta a compañeros colegas que trabajan en el área de Informática. Posteriormente se les solicitó a los catedráticos de Programación, Desarrollo de Sistemas y Redes de la Utec, para que les enviaran a sus docentes el link. Este proceso se realizó en la segunda semana del mes de noviembre de 2015.

Luego de pasar las encuestas, Google Forms también da la facilidad de enviar las respuestas en una hoja de Google Spreadsheets en el drive de la cuenta de correo electrónico en la cual se elaboró la encuesta, a parte que proporciona los gráficos respectivos, que son elaborados con Google Chart. Se aprovechó esta herramienta porque es una forma de hacer uso de nuevas tecnologías en la web. Este proceso se realizó entre la cuarta semana de noviembre de 2015, en la que se obtienen resultados que permiten realizar un análisis y se establecen conclusiones que describen información importante para medir el grado de conocimiento sobre las herramientas de *Big Data*.

5. Resultados

Con la encuesta se busca medir el grado de conocimiento que hay en los profesionales acerca de *Big Data* y debido a ello, la necesidad de incursionar en esa área.

Como se mencionó anteriormente, la encuesta tiene 12 preguntas cerradas. A continuación, se presentan los datos obtenidos durante el mes de noviembre de 2015.

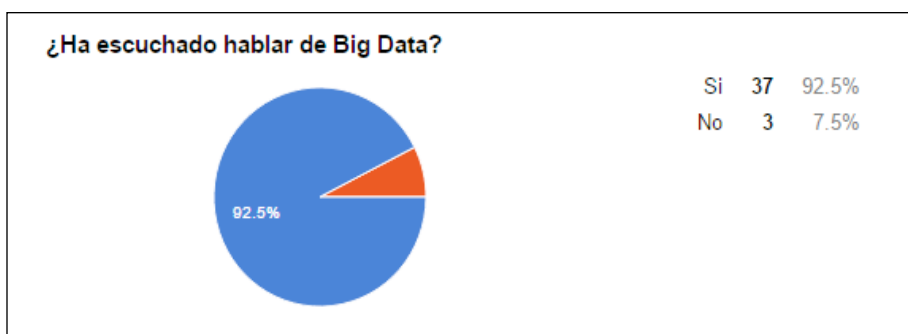


Figura 6

Fuente: Resultado de Google Forms

Se puede observar que el 92,5% de los encuestados si ha escuchado acerca de *Big Data*, lo que significa que, de alguna manera hay un conocimiento previo con respecto al tema y esto puede llegar a facilitar la implementación de herramientas basados en esta tecnología en determinado momento.



Figura 7

Fuente: Resultado de Google Forms

Los resultados nos dicen que son pocos los que saben acerca de los datos en la nube (45%), pero también hay un buen porcentaje (40%) de los que dicen saber mucho acerca del tema; solamente un 10% sabe bastante y un 5% no sabe nada al respecto. Esto nos hace ver la necesidad existente de dar a conocer que son los datos en la nube de una forma más amplia.

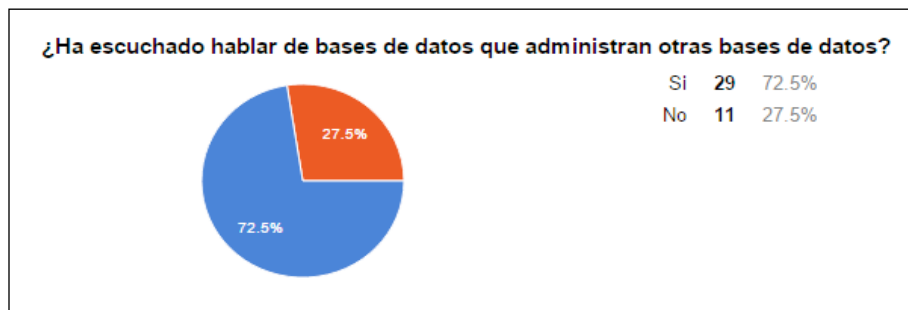


Figura 8

Fuente: Resultado de Google Forms

El 72,5% manifiesta haber escuchado hablar de las bases de datos que administran otras bases de datos, esto se debe a que son profesionales que se desenvuelven en el área de Informática, por lo tanto, el término es bastante utilizado en el medio.



Figura 9

Fuente: Resultado de Google Forms

Más de la mitad de los encuestados han escuchado hablar poco sobre *hadoop*: un 35% nada, solamente el 7,5% dice que bastante y un 5% que mucho. Esto significa que hay una gran necesidad de dar a conocer esta herramienta, ya que es una de las más importantes de *Big Data*, para el procesamiento, almacenamiento y análisis de grandes volúmenes de datos.

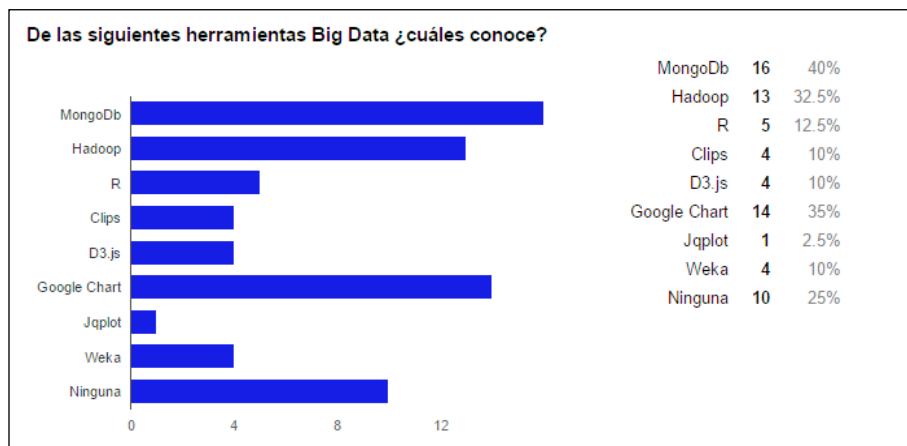


Figura 10

Fuente: Resultado de Google Forms

Las herramientas que más conocen los encuestados son MongoDB y Google Chart, aunque muy contradictoriamente han mencionado que Hadoop, pero en la pregunta anterior fue todo lo contrario. Puede observarse también que la cuarta parte de ellos no conoce ninguna de las herramientas mencionadas, todo

esto conlleva a ver la necesidad de presentar y dar a conocer cada una de esas herramientas para incursionar en las nuevas tecnologías de *Big Data*.



Figura 11

Fuente: Resultado de Google Forms

Es notoria la necesidad de conocer las herramientas de análisis de datos en la nube, debido a que un 72,5% manifiesta estar completamente seguro de querer conocer herramientas de análisis de datos en la nube, aunque todavía hay cierto temor en querer hacerlo esto lo hace ver un 22,5%

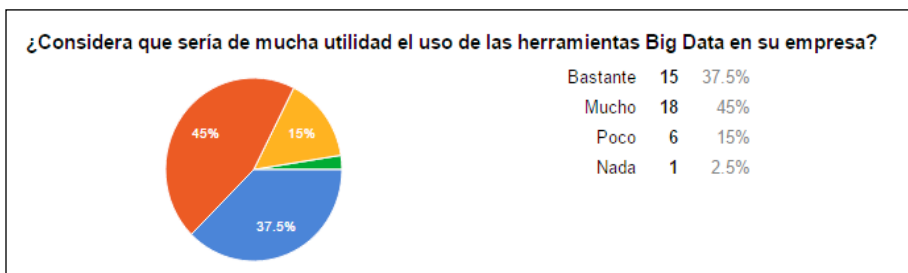


Figura 12

Fuente: Resultado de Google Forms

El 45% de los encuestados considera que sería de mucha utilidad el uso de herramientas *Big Data* en sus empresas, y un 37,5% considera que bastante, por lo tanto, se puede apreciar la necesidad existente en el uso de dichas herramientas, ya que esto facilitaría en gran medida el procesamiento y análisis de los datos.

¿Sabía usted que *Big Data* ha ayudado a resolver problemas de marketing, ventas, elecciones presidenciales, diagnóstico de enfermedades, relaciones interpersonales, meteorología y muchas más?



Figura 13

Fuente: Resultado de Google Forms

Debido a que la gran mayoría de los encuestados ha oído hablar de *Big Data*, entonces han podido darse cuenta de las ventajas y de las áreas que han sido beneficiadas, es por ello que un 35% sabe bastante, el 15% mucho, pero también hay un buen porcentaje (32,5%) que sabe poco al respecto y un 17,5% no sabe nada.

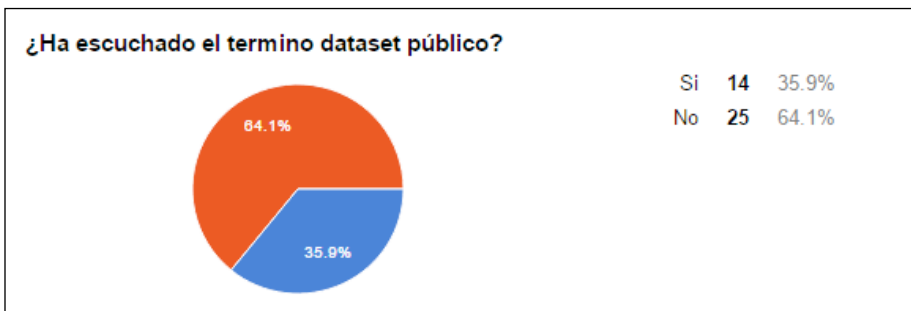


Figura 14

Fuente: Resultado de Google Forms

Acerca del *dataset* se puede apreciar que existe desconocimiento, ya que el 64,1% dice que no ha escuchado hablar de ese término, lo cual nos confirma que hay necesidad de dar a conocer en que consiste, porque *Big Data* trabaja mucho con *dataset* públicos.

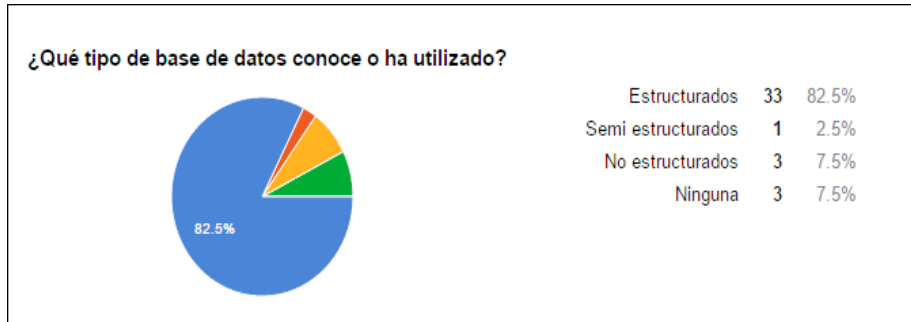


Figura 15
Fuente: Resultado de Google Forms

Los resultados obtenidos nos confirman que en El Salvador las bases de datos con las cuales se trabaja son las estructuradas, esto lo manifiesta el 82,5%, lo cual es comprensible, pues al desconocer herramientas *Big Data* no se puede trabajar con datos semi estructurados y mucho menos con datos no estructurados.

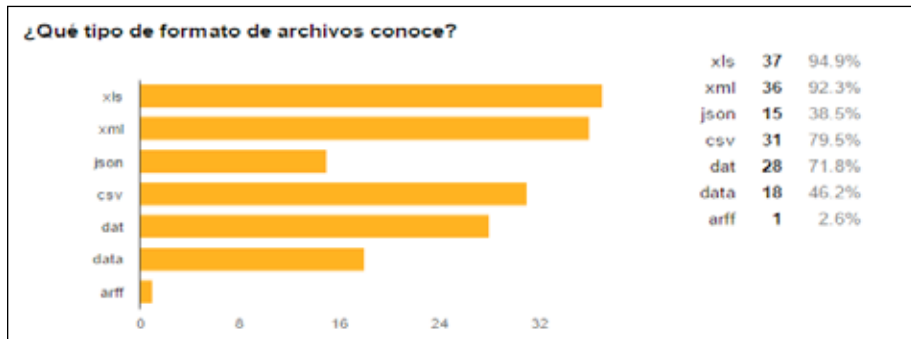


Figura 16
Fuente: Resultado de Google Forms

El tipo de formato más conocido por los encuestados es el xls debido a que la gran mayoría ha trabajado o trabaja con hojas de cálculo, el otro formato es el xml, esto se debe a que muchos profesionales trabajan en el diseño de páginas web. También es bien conocido para los encuestados el formato csv y el dat.

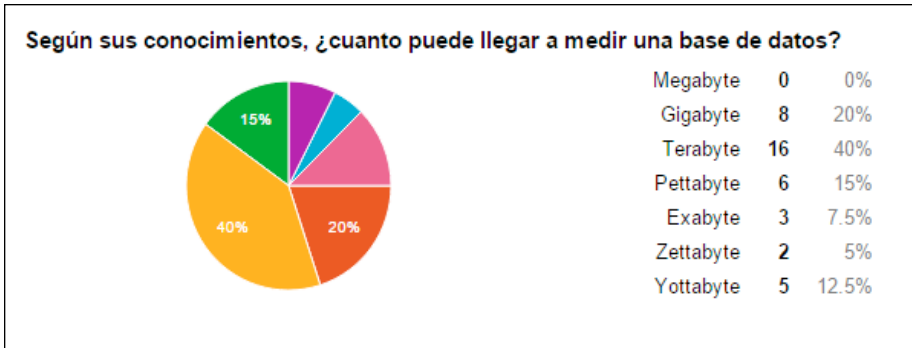


Figura 17

Fuente: Resultado de Google Forms

Los resultados son comprensibles, esto es porque en nuestro medio las bases de datos no son muy extensas, por lo que el 40% de los encuestados manifiesta que el tamaño puede ser en Terabytes y un 20% dice que Gigabyte. Aunque la realidad no es así y de esto es lo que se encarga *Big Data*.

6. Conclusiones

Después de presentar los resultados, se concluye lo siguiente:

Debido a que los datos arrojan que hay poco conocimiento de elementos importantes de *Big Data* en El Salvador, pero que a la vez hay muchos deseos de querer incursionar y utilizar las herramientas que faciliten el procesamiento y análisis de grandes volúmenes de datos, por lo tanto, en esta investigación se ha decidido hacer uso de metodologías que permitan describir los pasos necesarios para utilizar algunas herramientas de *Big Data* y los requisitos que deben considerarse para poder usarlas.

En términos generales la metodología consistirá en lo siguiente:

- Almacenar y procesar un *dataset* público el cual contiene información sobre registro de medicamentos, haciendo uso de Hadoop.

- Con la herramienta Hive que viene en la distribución Hortonworks de Hadoop, se harán las consultas necesarias, ya que esta herramienta es similar a las instrucciones que se utilizan en SQL, por lo que, para los que están acostumbrados a trabajar con base de datos relacionales, les será fácil entender la lógica de cómo trabaja Hive y en nuestro país SQL es el software más utilizado para bases de datos, y esa es la razón por la que se seleccionó esta herramienta.
- Para el análisis de datos estadístico, se va a utilizar el programa R con otro *dataset*. Las razones del porque se seleccionó este programa fueron explicadas en el Marco Teórico. Este programa nos devolverá información sobre datos importantes de los productos que están almacenados y a la vez permitirá que realicemos conclusiones en base a los resultados.
- Después de haber hecho un análisis estadístico y las consultas pertinentes de los datos, se procederá a realizar los gráficos necesarios para una mejor comprensión de los resultados y en base a ello sacar conclusiones y poder tomar decisiones. Las herramientas que se pueden utilizar son Google Chart, Jqplot o D3.js

7. Propuesta: “Metodología analítica y visualización para el control de productos alimenticios y medicamentos”

7.1 Desarrollo y Metodologías utilizadas

En este apartado se tratará de explicar los procesos fundamentales que será necesario implementar para lograr el objetivo planteado. Estos deben ser de forma ordenada iniciando desde la elección del *dataset* público y luego con el uso de las herramientas necesarias para el almacenamiento, procesamiento, análisis y visualización de los datos.

Para este caso, como se ha mencionado anteriormente, se hará uso de dos *dataset*, los cuales contienen registros sanitarios de alimentos, medicamentos

y bebidas. Estos ficheros serán manipulados de tal manera que puedan ser almacenados, uno de ellos con Hadoop y luego con la herramienta Hive hacer las consultas necesarias y el otro *dataset* será utilizado en el programa R, ya que se harán análisis estadísticos de los datos más relevantes del fichero seleccionado.

Con la información obtenida, se procederá a elaborar los gráficos, usando cualquiera de las herramientas de visualización mencionadas anteriormente, los cuales reflejarán lo más importante de los datos y así obtener conclusiones que servirán en la toma de decisiones.

Ficheros utilizados

Los ficheros están en formato csv, los cuales han sido descargados de la web. Uno de los ficheros contiene 33028 registros y 9 campos: Registro sanitario, Nombre del producto, Fecha de vencimiento, Fábrica, Dirección de la Fábrica, Teléfono de la Fábrica, Representante, Teléfono del representante.

El otro fichero tiene 757 registros y 8 campos: Numero, Principio Activo, Producto, Registro, Titular, Resolución, Fecha resolución, Uso o Tratamiento.

7.1.2 Almacenando y procesando datos con Hadoop

Como se dijo antes, Hadoop es un proyecto de software libre, con licencia Apache, cuya finalidad es prestar una plataforma para la gestión de grandes cantidades de datos. Los principales componentes que constituyen Hadoop son el sistema de archivos HDFS y el motor MapReduce.

Por lo tanto, se procederá a instalar CentOS, ya que es la opción recomendada para desplegar posteriormente Hortonworks y para este proyecto se hará uso de la herramienta Hive para realizar las consultas respectivas de uno de los *dataset*, el cual tiene registro de medicamentos, estas podrían ser las siguientes:

- Mostrar el principio activo del medicamento, nombre del producto y el laboratorio que lo fabrica.

- Mostrar los medicamentos y la fecha de resolución
- Mostrar los medicamentos y para que tratamiento son usados
- Mostrar medicamentos que son usados para cualquier tipo de enfermedad que interese consultar.
 - Mostrar los medicamentos con sus números de registro.
 - Mostrar la cantidad de medicinas que hay para un tratamiento específico.
 - Mostrar la cantidad de Titulares o fabricantes de los medicamentos.

7.1.3 Analizando datos con R

Dentro de los objetivos se encuentra la utilización del programa R para el análisis estadístico de los datos, por lo tanto, hay que importar el *dataset* al programa R y comenzar a realizar los análisis estadísticos respectivos.

Debido a que este proyecto consiste en el análisis y visualización para el control de productos alimenticios y medicamentos, se hará uso del segundo *dataset*, el cual contiene el registro sanitario de alimentos, bebidas y medicamentos, por lo tanto, interesa conocer lo siguiente:

- Cantidad de productos vencidos
- Cantidad de productos vigentes
- Cantidad de fabricantes distintos
- Cantidad de productos distintos
- Cuantos productos distintos hay por fabricante
- Conocer los datos de un producto específico, etc.
- Mostrar la cantidad de productos de acuerdo con la fecha de vencimiento
- Mostrar resúmenes de productos y de fábricas

7.1.4 Visualizando datos

Cada día se generan billones de datos e incluso de forma individual estamos inundados de correos electrónicos, fotografías, videos, música, libros electró-

nicos y aparte de ello se almacena más información y documentos en la nube, como, por ejemplo, Google Drive o mediante cualquier disco virtual lo cual hace que exista demasiada información y de la que no es posible conocer el poder que puede tener.

Así mismo, si se añade el concepto de *Big Data*, «grandes conjuntos de datos (*datasets*)», se podrá comprender que la principal dificultad no es la captura y almacenamiento de los datos, sino más bien el análisis y su posterior representación visual estática o interactiva.

Los datos por si solos no tienen sentido, a menos que estos se conviertan en información útil y comprensible para luego acceder al conocimiento.

Se sabe que los datos son la materia prima de la información y esta la del conocimiento y, por supuesto al tener conocimiento se pueden tomar decisiones.

Para las empresas el tiempo en procesar todos los datos generados es valioso, pero se sabe que no se puede lograr sin las herramientas adecuadas que permitan transformar la multitud de datos. Una vez se haya logrado transformar los datos en información, es necesario presentarlos de manera atractiva y que sea fácil de comprender, por lo tanto, la magia de la visualización de información radica en la captura y síntesis previa de la información mediante el uso de diversas técnicas visuales como diagramas, gráficas, esquemas, nubes palabras, conexiones, grafos, para que pueda ser transformada y con ello facilitar su comprensión.

Es por ello que, después de haber hecho uso de herramientas para el almacenamiento, procesamiento y análisis de los datos, es necesario hacer uso de herramientas de visualización que permitan comprender los resultados, pero en un formato gráfico, debido a que para la mente humana es mucho más fácil entender imágenes que únicamente texto o números.

De esto se encarga la visualización de datos, que no es otra cosa que el diseño de la comprensión de manera atractiva. Los datos deben ser comprendidos de manera efectiva. El objetivo de toda buena visualización tiene que centrar la atención del interesado en la información en aquello que realmente es relevante e importante.

Una vez se disponga de ese nuevo conocimiento, este permitirá hacer análisis y sacar conclusiones que serán importantes para la toma de decisiones.



Figura 18 Diagrama de estructura de los datos hacia la sabiduría

Fuente: Hey, J.: *The Data, information, knowledge, Wisdom Chain: The Metaphorical Link*

Lo que se pretende con el diagrama anterior es pasar los datos a la sabiduría, porque de eso dependerá tomar buenas decisiones.

Las herramientas de visualización que se pueden utilizar para este proyecto son Google chart y Jqplot o D3, cada una de ellas tienen sus propias características tal y como se mencionó en apartados anteriores, por lo que, dependiendo del uso que quiera darse en la representación gráfica de los datos, así será la selección de cualquiera de ellas o se pueden utilizar las tres si se desea.

Para la implementación de la metodología propuesta es necesario tomar en cuenta ciertos requisitos técnicos, los cuales se detallan a continuación.

7.2. Identificación de requisitos

En cualquier país del mundo existen reglamentos para el control sanitario de productos alimenticios y de bebidas, a los cuales deben someterse todas las empresas cuyo rubro sea el de ese tipo.

En El Salvador, la institución que regula el registro sanitario de productos alimenticios y bebidas es el Ministerio de Salud, el cual en la sección 12 y Art. 95 del Código de Salud, dice lo siguiente: **Alimentos y Bebidas**. El Ministerio llevará un registro de alimentos y bebidas, en consecuencia se prohíbe la importación, exportación, comercio, fabricación, elaboración, almacenamiento, transporte venta o cualquiera otra operación de suministros al público, de alimentos o bebidas empacadas o envasadas cuya inscripción en dicho registro no se hubiere efectuad¹⁸

Por lo tanto, todas las empresas que se dediquen a ese rubro, deben tener registrados sus productos y de allí que la cantidad de datos se vuelve extensa para poder procesarla y analizarla. Lastimosamente en El Salvador, no está disponible un *dataset* público con esa información, por esa razón los *dataset* que se seleccionaron son de otros países, pero servirán para hacer la metodología con herramientas de *Big Data* y de esa manera demostrar lo que se puede hacer cuando se tienen grandes volúmenes de datos.

Con respecto al *dataset* de medicamentos, estos datos son muy utilizados por Laboratorios, farmacias y entidades de salud que trabajan con esa información y por lo tanto, la cantidad de datos que se almacenan es bastante y es necesario utilizar herramientas que faciliten su análisis.

18 <https://www.salud.gob.sv/servicios/descargas/documentos/Documentaci%C3%B3n-Institucional/C%C3%B3digo-de-Salud-Leyes-y-Reglamento/>

Los requisitos técnicos y las tecnologías a utilizar son los siguientes:

Herramienta tecnológica	Requisito de hardware	Requisito de software
Hadoop	<ul style="list-style-type: none"> ✓ 5.^a generación del procesador Intel® Core™ i3, i5 o i7 ✓ Memoria de 16 GB expandible a 32 GB ✓ Disco duro de 1 a 2 TB 	<ul style="list-style-type: none"> ✓ Windows y Linux ✓ Linux de 64 bits ✓ Oracle VM VirtualBox ✓ CentOS 6.5 ✓ Distribución Hortonworks
Programa R	No requiere de características especiales.	<ul style="list-style-type: none"> ✓ Windows, Mac o Linux ✓ R o R Studio
Google chart, Jqplot, D3.js	No requiere de características especiales.	<ul style="list-style-type: none"> ✓ Cualquier sistema operativo ✓ Cualquier editor de texto: bloc de notas, sublime text o notepad ++ ✓ Navegador web ✓ Servidor web

Hadoop es la única herramienta que requiere de muchos recursos de hardware, sobre todo de memoria RAM y de disco duro, es por ello que lo más recomendable es que esté instalado en un servidor, pero por las limitantes encontradas y no disponer de suficientes recursos en el equipo utilizado se optó por crear máquinas virtuales para simular un master y un esclavo y hacer el despliegue en un entorno similar a uno de producción.

Los demás programas que han de utilizarse son gratuitos y con lecturas de manuales en la web se pueden llegar a utilizar sin ningún problema.

7.3 Desplegando Hadoop



1. Como se va a utilizar Hadoop para el procesamiento de los datos y posteriormente realizar las consultas necesarias con Hive, será necesario desplegarlo, para ello es recomendable tomar en cuenta lo que se mencionó en el apartado 3.7.1 con respecto a la decisión de la arquitectura física y la distribución.
2. Como se ha mencionado en la **Identificación de los Requisitos**, Hadoop requiere de muchos recursos de hardware, los cuales son una limitante para esta investigación, por lo tanto, se ha optado por crear máquinas virtuales y simular un entorno de producción real.
3. Se comenzará haciendo uso de una de las herramientas de virtualización como Oracle VM VirtualBox (www.virtualbox.org), la cual está disponible para los principales sistemas operativos.
4. Después de haber instalado VirtualBox, se procede a crear las máquinas virtuales, las cuales tendrán iguales características.
5. Se recomienda seleccionar como sistema operativo Linux de 64 bits y asignarle 4GB de RAM y 20GB de disco duro. Estos valores pueden cambiar si se dispone de más recursos de hardware en el equipo donde se hará el despliegue, incluso no habría necesidad de crear máquinas virtuales pues se haría con todos los recursos del equipo, aunque para esta investigación esos valores serán suficientes, y lo más probable es que la máquina no consuma todos estos recursos.

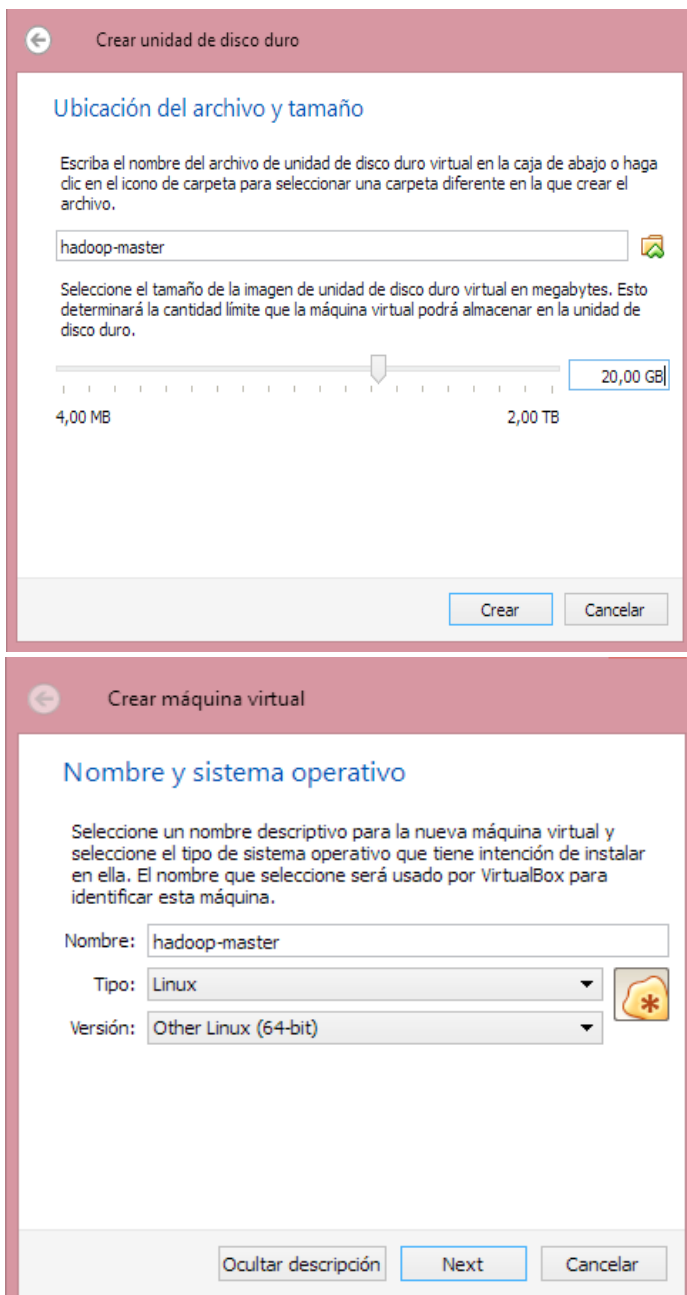
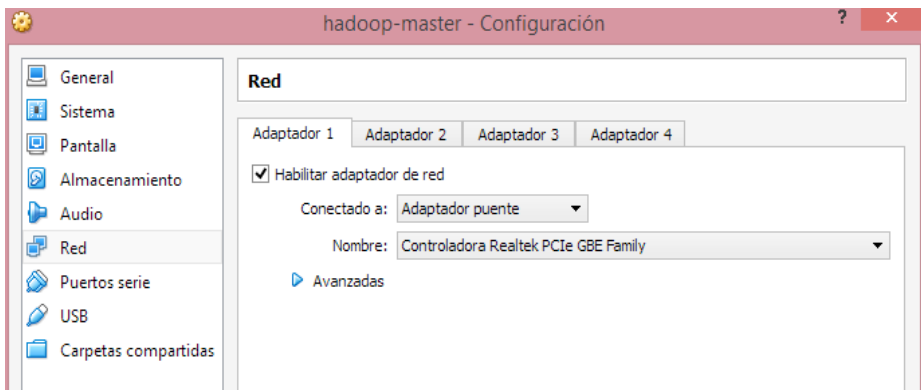


Figura 19 Asignación de recursos a las máquinas virtuales
Fuente: propia

Además, se recomienda cambiar la configuración de red para que el adaptador esté conectado como adaptador puente. De este modo, la máquina funciona como si estuviera conectada directamente al router, lo que quiere decir que, si se usa DHCP, asignará a cada máquina una IP distinta, si no se usa, se tendrá que configurar una IP manualmente como si se tratara de una máquina física.



*Figura 20 Configuración de la red en las máquinas virtuales
Fuente: propia*

6. Instalación del sistema operativo. Una vez creadas las máquinas virtuales, se instalará en ellas el sistema operativo CentOS, que es la opción recomendada para desplegar posteriormente Hortonworks. Para ello, se debe descargar la imagen de DVD de la última versión disponible (para esta investigación se utilizó la versión 6.5, aunque después se actualizó con la versión 6.7) desde la página web oficial (www.centos.org).

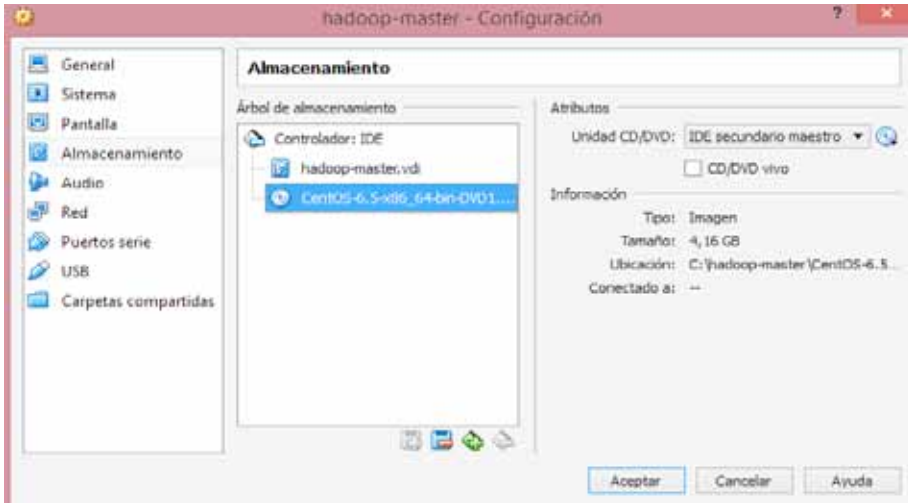


Figura 21 Selección del Sistema Operativo
Fuente: propia

7. Para instalar CentOS, solo es de seguir los pasos en todo momento. El proceso es sencillo y solo tarda unos cuantos minutos.



Figura 22 Instalación de CentOS
Fuente: propia

- Se escribe un nombre al host, aunque este sea ficticio (para el master se ha seleccionado `hadoop-master.Utec.net` y para el slave, `hadoop-slave.Utec.net`) y en configuración de red, debido a que se está utilizando DHCP, únicamente hay que activar el chequecito en la casilla Conectar automáticamente.

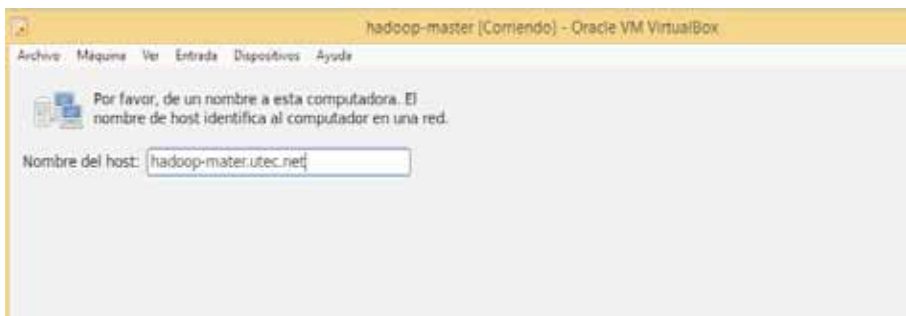


Figura 23 Escribiendo un nombre de host para el master

Fuente: propia



Figura 24 Escribiendo un nombre de host para el slave

Fuente: propia

- Después de seleccionar la localización y el idioma, se asigna un password al usuario root.

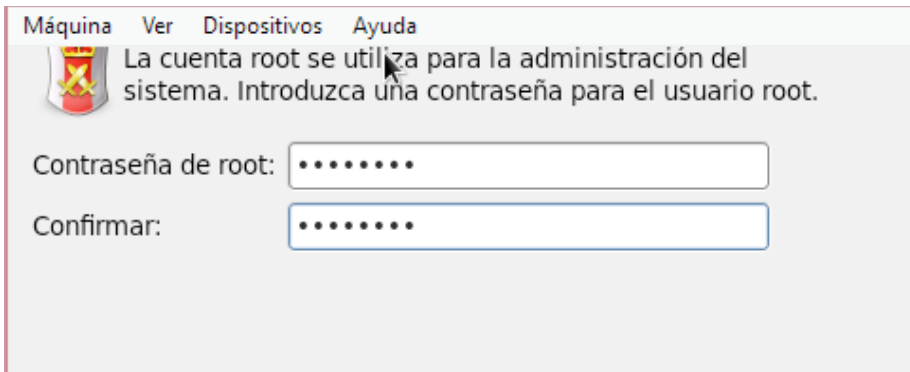


Figura 25 Asignando contraseña al usuario root
Fuente: propia

10. Al momento de preguntar el tipo de instalación, hay que asegurarse de no escoger la mínima (*minimal*), porque carece de algunos comandos necesarios para proceder con el despliegue de Hadoop, y por lo tanto será necesario instalar aparte. Además, cuando pida el nombre del equipo, se debe poner un nombre de red completo (con el dominio incluido), aunque este sea ficticio.

Por lo tanto, se selecciona la opción Basic Server

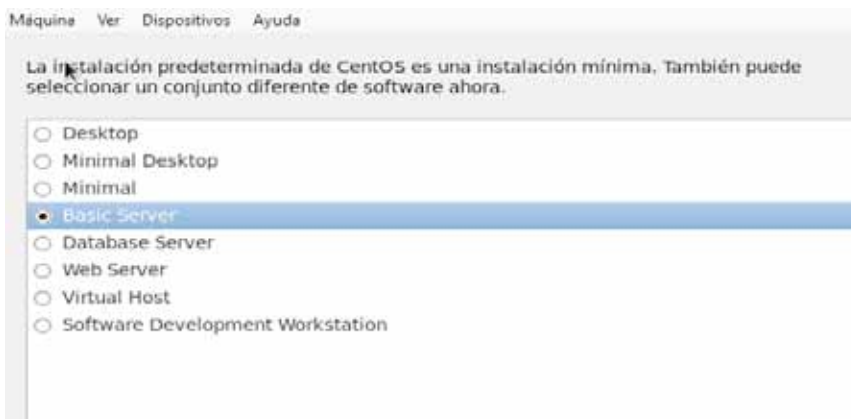
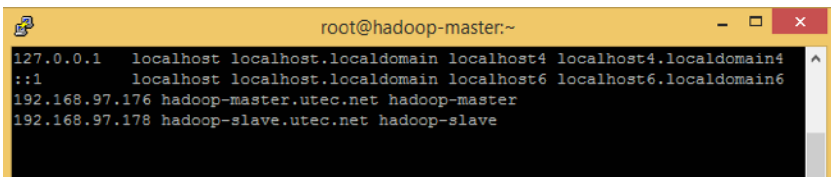


Figura 26 Seleccionando instalación Basic Server
Fuente: propia

11. Se deben editar los ficheros de hosts para simular entradas DNS utilizando el comando `vi /etc/hosts` y se escriben las IP generadas de cada una de las máquinas, con su respectivo host y se guardan los cambios. Este proceso se debe hacer en ambas máquinas virtuales.

Para nuestro caso quedaría de la siguiente manera:

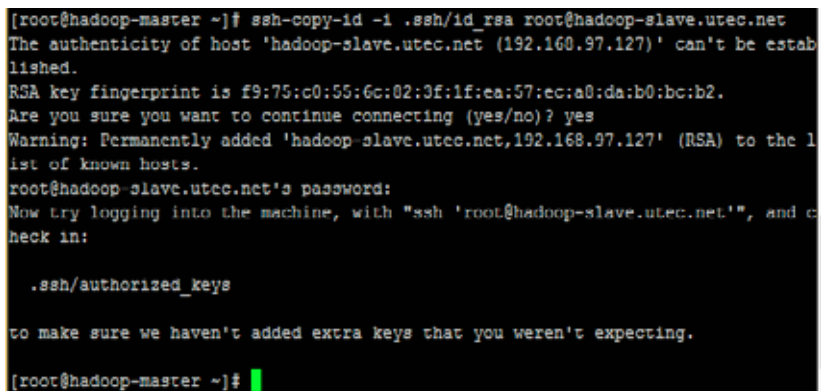


```
root@hadoop-master:~
127.0.0.1    localhost localhost.localdomain localhost4 localhost4.localdomain4
::1        localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.97.176 hadoop-master.utec.net hadoop-master
192.168.97.178 hadoop-slave.utec.net hadoop-slave
```

Figura 27 Escribiendo las IP generadas
Fuente: propia

12. Posteriormente se debe crear una clave haciendo uso del comando `ssh-keygen` y utilizar las opciones por defecto.

Luego se copian las claves en cada uno de los equipos.



```
[root@hadoop-master ~]# ssh-copy-id -i .ssh/id_rsa root@hadoop-slave.utec.net
The authenticity of host 'hadoop-slave.utec.net (192.168.97.127)' can't be established.
RSA key fingerprint is f9:75:c0:55:6c:02:3f:1f:ea:57:ec:a0:da:b0:bc:b2.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'hadoop-slave.utec.net,192.168.97.127' (RSA) to the list of known hosts.
root@hadoop-slave.utec.net's password:
Now try logging into the machine, with "ssh 'root@hadoop-slave.utec.net'", and check in:

    .ssh/authorized_keys

to make sure we haven't added extra keys that you weren't expecting.
[root@hadoop-master ~]#
```

Figura 28 Copiando las claves en cada uno de los equipos
Fuente: propia

- 13. Se ejecuta el servicio de tiempo en red en todos los nodos usando el comando `service ntpd start`.

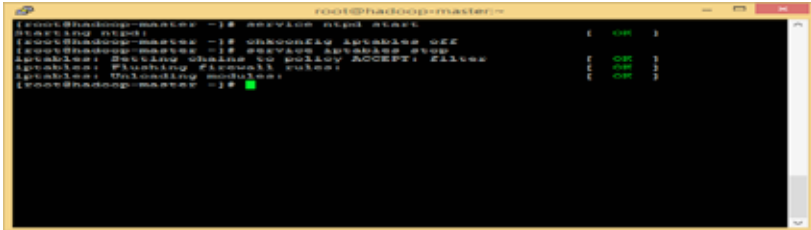


Figura 29 Ejecutando el servicio en tiempo de red
Fuente: propia

- 14. Luego hay que desactivar el firewall de Linux y el sistema de seguridad en cada nodo, porque puede ocasionar algunas incompatibilidades con Ambari, esto se hace con el comando `chkconfig iptables off` y `service iptables stop` y luego deshabilitar el módulo de seguridad con el comando `setenforce 0`.

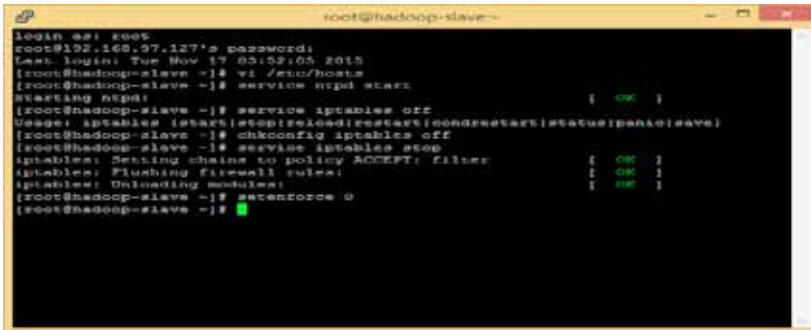
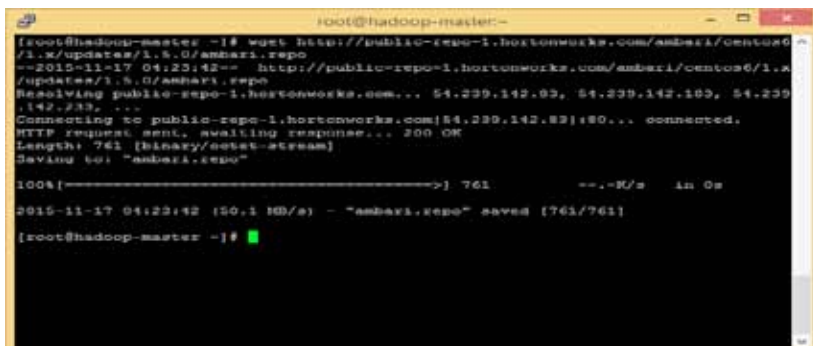


Figura 30 Desactivando el Firewall de Linux
Fuente: propia

- De aquí en adelante únicamente se trabaja con el nodo master
- Se descarga el repositorio que contiene los paquetes de ambari de la siguiente manera:



```
root@hadoop-master:~# wget http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.5.0/ambari.repo
--2015-11-17 04:23:42-- http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.5.0/ambari.repo
Resolving public-repo-1.hortonworks.com... 54.239.142.89, 54.239.142.189, 54.239.142.239, ...
Connecting to public-repo-1.hortonworks.com|54.239.142.89|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 761 [Binary/octet-stream]
Saving to: "ambari.repo"

100%[=====>] 761  --.-K/s  in 0s

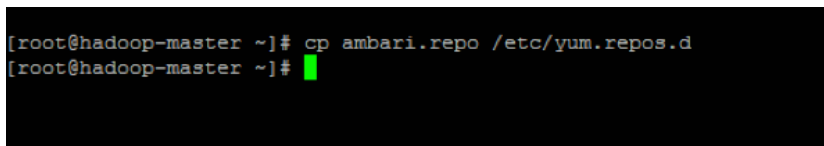
2015-11-17 04:23:42 (50.1 MB/s) - "ambari.repo" saved [761/761]

[root@hadoop-master ~]#
```

Figura 31 Descargando ambari

Fuente: propia

- Después de descargado se copia el fichero al directorio de repositorios del sistema de paquetes yum.



```
[root@hadoop-master ~]# cp ambari.repo /etc/yum/repos.d
[root@hadoop-master ~]#
```

Figura 32 Copiando el fichero en el directorio de repositorios del sistema de paquetes yum

Fuente: propia

- Posteriormente se comienza a instalar el servidor de ambari

```

root@hadoop-master:~
[root@hadoop-master ~]# yum install ambari-server
Loaded plugins: fastestmirror, security
HDP-UTILS-1.1.0.16                | 2.9 kB    00:00
HDP-UTILS-1.1.0.16/primary_db    | 35 kB    00:00
Updates-ambari-1.5.0             | 2.9 kB    00:00
Updates-ambari-1.5.0/primary_db  | 4.7 kB    00:00
ambari-1.x                       | 1.3 kB    00:00
ambari-1.x/primary               | 1.9 kB    00:00
ambari-1.x                       |          5/5
base                              | 3.7 kB    00:00
base/primary_db                  | 4.6 MB    00:45
extras                           | 3.4 kB    00:00
extras/primary_db                | 33 kB    00:00
updates                          | 3.4 kB    00:00
updates/primary_db               | 2.6 MB    00:20
Setting up Install Process
Resolving Dependencies
--> Running transaction check
--> Package ambari-server.noarch 0:1.5.0.14-1 will be installed
--> Processing Dependency: postgresql-server >= 8.1 for package: ambari-server-1.5.0.14-1.noarch
--> Running transaction check
--> Package postgresql-server.x86_64 0:8.4.20-3.el6_6 will be installed
--> Processing Dependency: postgresql-libs(x86-64) = 8.4.20-3.el6_6 for package:

```

Figura 33 Instalando el servidor ambari
Fuente: propia

19. Una vez instalado ambari, podrá acceder a la interfaz web entrando en la URL del equipo donde se ha instalado y en el puerto 8080. En este caso, esta URL sería <http://hadoop-master.Utec.net:8080> siempre y cuando se haya agregado las IP con sus nombres al archivo host de la maquina física o se escribe directamente la IP, que en nuestro caso sería 192.168.97.126:8080. El usuario y contraseña para iniciar sesión es admin/admin.

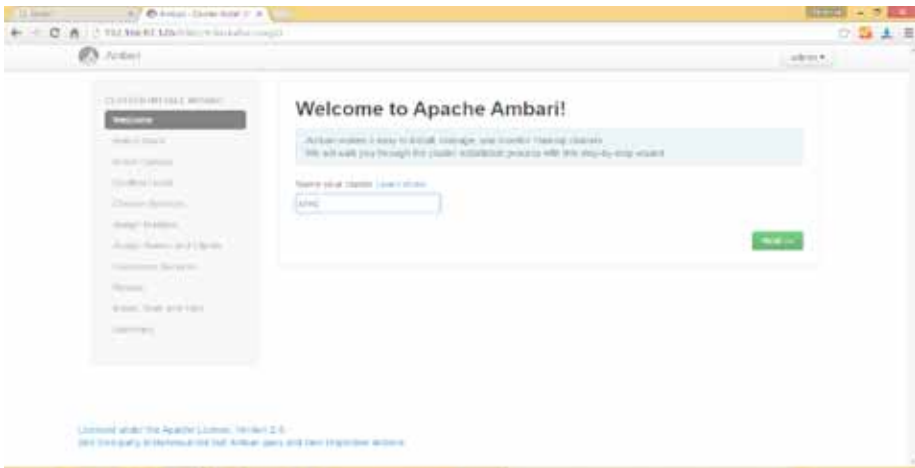


Figura 34 Ingresando un nombre de clúster en ambari
Fuente: propia

El proceso de instalación es casi automático, pero requiere de algunas configuraciones:

- ✓ En la pantalla **Select Stack** se debe seleccionar la versión HDP 2.0, debido a que es la versión estable más reciente.
- ✓ En la pantalla **Install Options** se debe incluir los nombres de todos los equipos que formarán parte del clúster, así como cargar la clave SSH que se ha generado previamente. Ambari detecta los equipos y permite continuar.

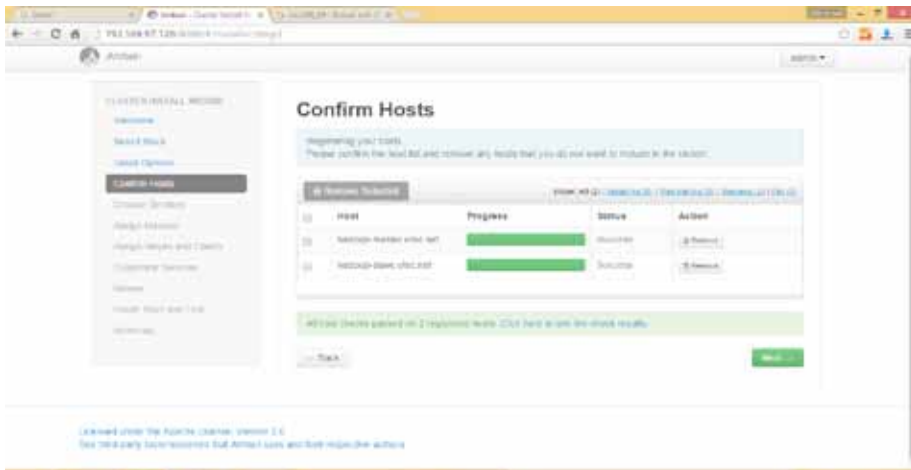


Figura 35 Confirmando los hosts
Fuente: propia

- ✓ En la pantalla **Choose Services** se debe asegurar de que todos los servicios están seleccionados.

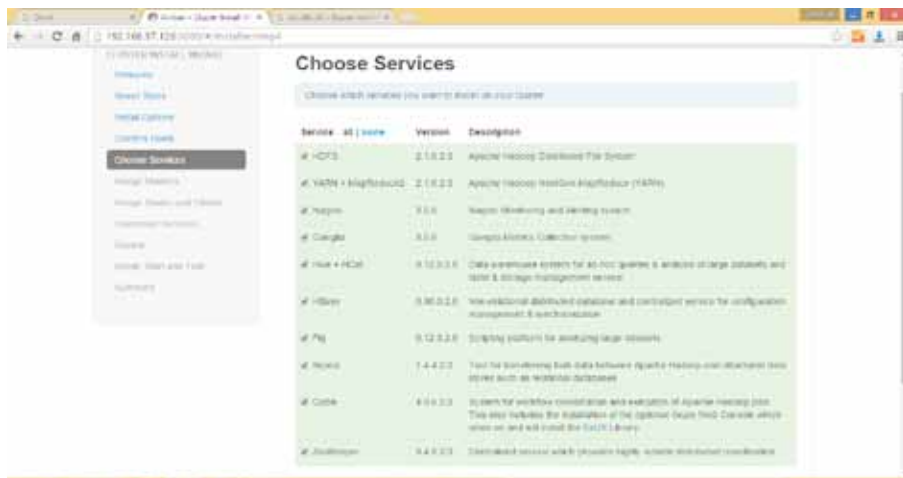


Figura 36 Seleccionando todos los servicios
Fuente: propia

- ✓ En la pantalla **Assign Masters** se asignan todos los servicios a uno de los nodos, que hará de máster (para este caso sería `hadoop-master.Utec.net`).

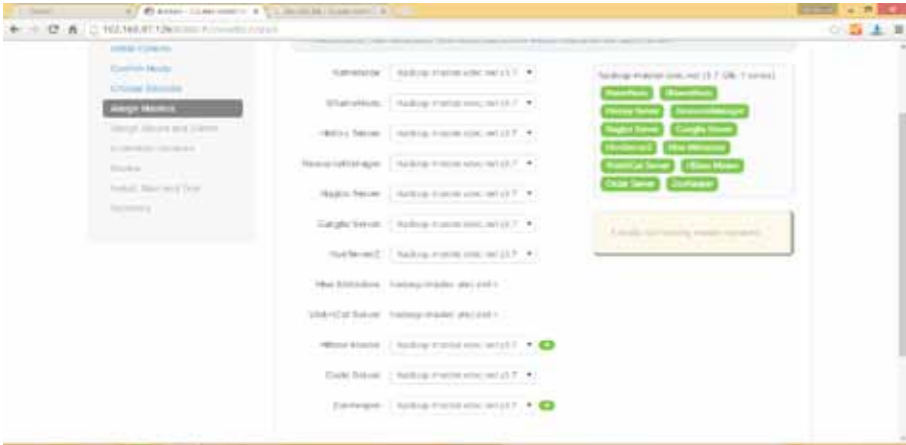


Figura 37 Asignando el master a todos los servicios

Fuente: propia

- ✓ En la pantalla **Assign Slaves and Clients** se seleccionan todos los equipos, tanto el que ya hace de máster como todos los demás en el clúster.
- ✓ En la pantalla **Customize Services** no es necesario cambiar la configuración por defecto, aunque se solicita para algunos servicios introducir una contraseña, y en el caso de Nagios (el servicio de monitorización para Hadoop) hay que escribir un correo electrónico donde mandar las alertas.

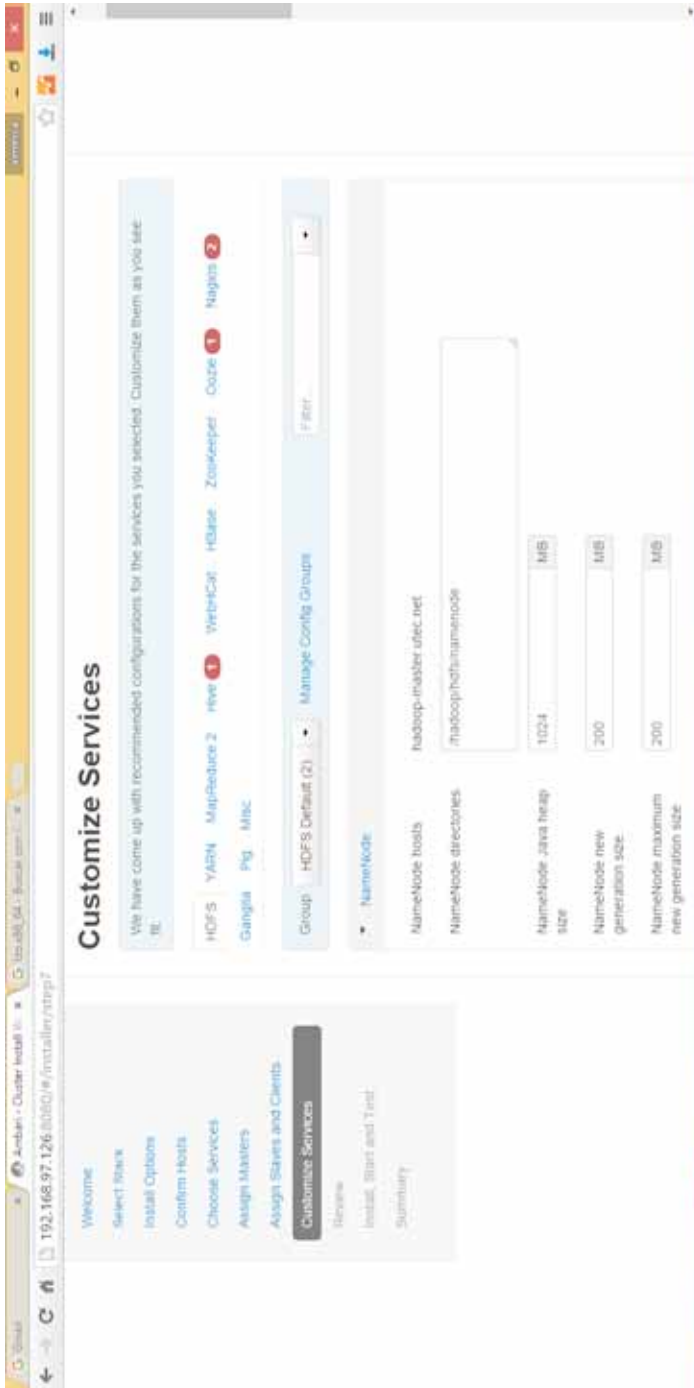


Figura 38 Personalizando los servicios
Fuente: propia

- ✓ Una vez completada esta configuración se comienza el proceso de despliegue de Hadoop en el clúster. Este proceso toma bastante tiempo para completarse.



Figura 39 Pantalla de revisión de la configuración realizada

Fuente: propia

- ✓ Para verificar que el despliegue de Hadoop es adecuado se debe asegurar de que el proceso ha acabado correctamente y que ambari indica que todos los servicios están en ejecución.

✓

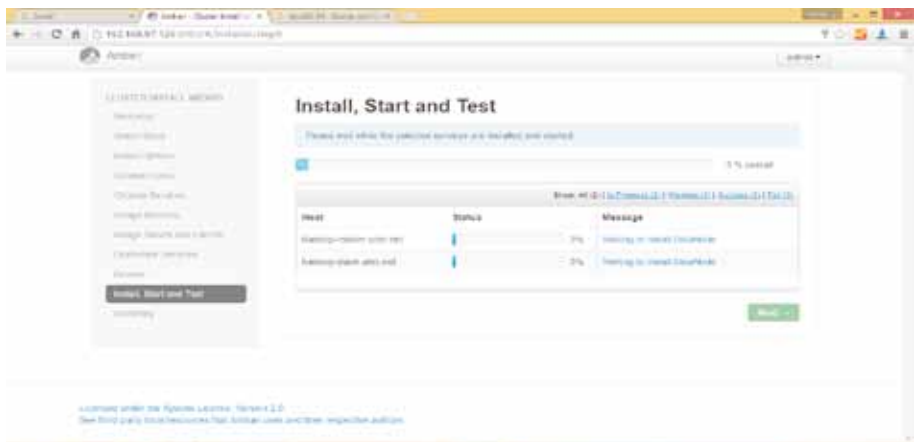


Figura 40 Pantalla de inicio e instalación y prueba

Fuente: propia

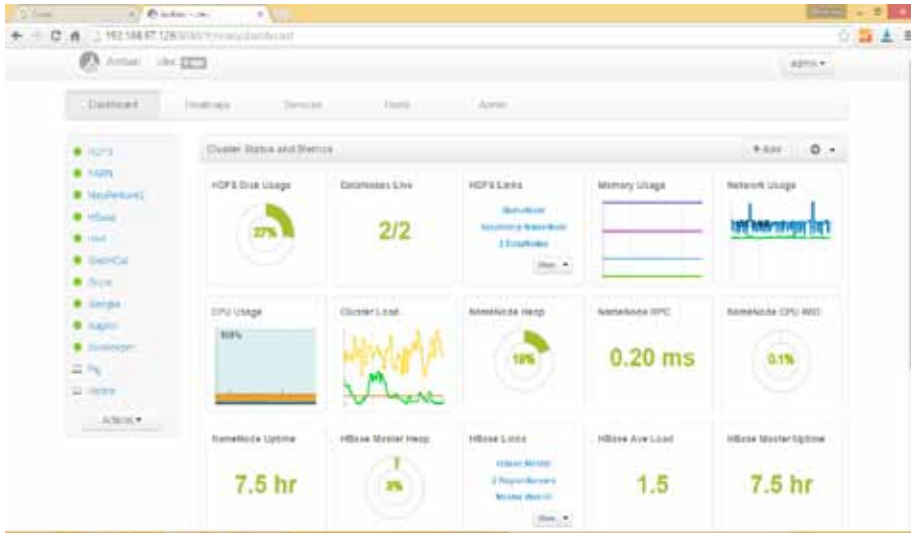


Figura 41 Despliegue de Hadoop finalizado
Fuente: propia

7.3.1 Trabajando con Hadoop

Una vez se haya hecho el despliegue correctamente de Hadoop, se accederá por medio de conexión ssh, haciendo uso del programa Putty, en el que se escribirá la IP de la máquina master.

Una vez estemos en Hadoop, se procede a descargar el fichero correspondiente al registro de medicamentos desde la web. El fichero tiene 757 registros y 8 campos y se encuentra en la siguiente URL: http://www.partidopirata.cl/wp-content/uploads/2015/09/Listado-de-productos-Bioequivalentes_agosto_2015.csv

El comando que se debe utilizar para descargar el fichero es el siguiente:
wget http://www.partidopirata.cl/wp-content/uploads/2015/09/Listado-de-productos-Bioequivalentes_agosto_2015.csv

En ese momento comienza a realizar la descarga del fichero

```
[root@hadoop-master ~]# wget http://www.partidopirata.cl/wp-content/uploads/2015/09/Listado-de-productos-Bioequivalentes_agosto_2015.csv
--2015-11-22 17:24:02-- http://www.partidopirata.cl/wp-content/uploads/2015/09/Listado-de-productos-Bioequivalentes_agosto_2015.csv
Resolving www.partidopirata.cl... 178.62.148.7
Connecting to www.partidopirata.cl|178.62.148.7|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 127257 (124K) [text/csv]
Saving to: "Listado-de-productos-Bioequivalentes_agosto_2015.csv"

100%[=====>] 127,257      124K/s   in 1.0s

2015-11-22 17:24:04 (124 KB/s) - "Listado-de-productos-Bioequivalentes_agosto_2015.csv" saved [127257/127257]

[root@hadoop-master ~]#
```

Figura 42 Descarga del fichero en Hadoop
Fuente: propia

- Si no existe una carpeta dentro del fichero HDFS para el usuario, habrá que crear una y es probable que sea necesario ejecutar este paso como root para evitar problemas de permisos posteriores.

El comando es el siguiente:

```
sudo -u hdfs hadoop fs -mkdir /user/<uname>
sudo -u hdfs hadoop fs -mkdir /user/root
```

- Habrà que dar permisos al usuario para manipular este directorio. Este paso, aunque no es obligatorio, es conveniente ya que por defecto el único usuario con permisos para operar sobre HDFS es hdfs, razón por la que antes se ha tenido que ejecutar el comando `sudo -u hdfs`. Para ello, se ejecutará el siguiente comando:

```
sudo -u hdfs hadoop fs -chown <<uname>> /user/<uname>
sudo -u hdfs hadoop fs -chown root /user/root
```

- Ahora se copiará el fichero del *dataset* que se ha extraído a HDFS:

```
hadoop fs -copyFromLocal Listado-de-productos-Bioequivalentes_agosto_2015.csv /user/root
```

5. Para asegurarse que el archivo se encuentra en la carpeta, hay que mostrar las últimas líneas del contenido del fichero, haciendo uso del comando tail

```
hadoop fs -tail /user/root/Listado-de-productos-Bioequivalentes_
agosto_2015.csv
```

```
[root@hadoop-master ~]# hadoop fs -tail /user/root/Listado-de-productos-Bioequiv
alentes_agosto_2015.csv
IÓN RW N° 14381", "20□08□2015", "Diurético"
"752", "ETINILESTRADIOL / DROSPIRENONA", "LADEE 20 COMPRIMIDOS RECUBIERTOS", "FD216
59", "EXELTIS CHILE S.P.A.", "RESOLUCIÓN RW N° 14382", "20□08□2015", "Anticonceptivo
"
"753", "ETINILESTRADIOL / DROSPIRENONA", "LADEE COMPRIMIDOS RECUBIERTOS", "FD21650"
, "EXELTIS CHILE S.P.A.", "RESOLUCIÓN RW N° 14514", "21□08□2015", "Anticonceptivo"
"754", "AZITROMICINA", "AZITROMICINA COMPRIMIDOS RECUBIERTOS 500 mg", "FD22082", "BL
UEPHARMA CHILE S.P.A.", "RESOLUCIÓN RW N° 14522", "21□08□2015", "Antibiótico"
"755", "CETIRIZINA", "ALERTOP COMPRIMIDOS RECUBIERTOS 10 mg", "FD7193", "LABORATORIO
CHILE S.A.", "RESOLUCIÓN RW N° 14524", "21□08□2015", "Antialérgico"
"756", "CETIRIZINA", "CETIRIZINA DICLORHIDRATO COMPRIMIDOS RECUBIERTOS
10 mg", "FD13749", "LABORATORIO CHILE S.A.", "RESOLUCIÓN RW N° 14525", "21□08□2015",
"Antialérgico"
"757", "LORATADINA", "LORATADINA COMPRIMIDOS 10 mg", "FD8455", "LABORATORIO CHILE S.
A.", "RESOLUCIÓN RW N° 14532", "21□08□2015", "Antialérgico"
[root@hadoop-master ~]#
```

Figura 43 Mostrando las últimas líneas del fichero descargado

Fuente: propia

6. Habrá que mover el fichero copiado en HDFS a la carpeta creada, para que pueda ser utilizado posteriormente en Hive:

```
hadoop fs -mv Listado-de-productos-Bioequivalentes_agosto_2015.
csv user/root/root
```


7.3.2 Uso de Hive



1. Se procede a hacer uso de Hive, la cual es una de las herramientas de Hadoop, que anteriormente se ha mencionado las razones del porque se seleccionó.
2. En la consola de hadoop, se digita el comando hive para trabajar con esta herramienta.
3. Estando en hive, se crea la base de datos haciendo uso del comando **CREATE DATABASE** y el nombre de la base.

```
hive> CREATE DATABASE Listado_Productos;  
OK  
Time taken: 1.811 seconds
```

*Figura 44 Creado la base de datos en Hive
Fuente: propia*

4. A continuación, se crea la tabla, con el comando **CREATE EXTERNAL TABLE**

La tabla debe llevar todos los campos que contiene el fichero, con su respectivo tipo de datos.

```
hive> CREATE EXTERNAL TABLE Medicinas(num INT, principio STRING, producto STRING
, registro STRING, titular STRING, resolucion STRING, fecha_resol STRING, tratam
iento STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\,' LINES TERMINATED BY
'\n' LOCATION '/user/root/';
OK
Time taken: 0.074 seconds
```

Figura 45 Creado la tabla dentro de la base de datos

Fuente: propia

5. Se comprueba que los datos se hayan cargado en la tabla, haciendo uso del comando SELECT y que muestre al menos 10 registros.

```
hive> SELECT * FROM Medicinas LIMIT 10;
OK
NULL
NULL
NULL
NULL
Listado de productos Bioequivalentes,
actualizado al 21 de Agosto de 2015
NULL
NULL
NULL
Principio Activo      Producto      Registro      Titular ResoluciOn
Fecha resoluciOn Uso / Tratamiento
1 ANASTROZOL ANASTROZOL COMPRIMIDOS RECUBIERTOS 1 mg F-16801 SYNTHON CHILE
LTDA. RESOLUCION NO 1336 16-03-2009 CDncer de mama
2 ANASTROZOL MADELEN COMPRIMIDOS RECUBIERTOS 1 mg F-16807 SYNTHON CHILE
LTDA. RESOLUCION NO 1336 16-03-2009 CDncer de mama
3 ESCITALOPRAM NEOPRESOL COMPRIMIDOS RECUBIERTOS 10 mg F-15836 LABORATORIO BAG
O DE CHILE S.A. RESOLUCION NO 4204 20-08-2009 Antidepressivo
4 ABACAVIR MESABEX COMPRIMIDOS RECUBIERTOS 300 mg F-17883 LABORATORIOS LA
FI LTDA. RESOLUCION NO 77 12-01-2010 tratamiento de VIH/SIDA
Time taken: 0.588 seconds, Fetched: 10 row(s)
```

Figura 46 Mostrando 10 registros de la tabla

Fuente: propia

6. Realizando consultas en Hive dentro de la tabla Medicinas
 - ✓ Se desea un listado que contenga el principio activo del medicamento, el nombre del producto y el fabricante o titular.

```
hive> SELECT principio,producto,titular FROM Medicinas;
OK
```

Listado de productos Bioequivalentes,
actualizado al 21 de Agosto de 2015

Principio Activo	Producto	Titular
ANASTROZOL	ANASTROZOL COMPRIMIDOS RECUBIERTOS 1 mg	SYNTHON CHILE LTDA.
ANASTROZOL	MADELEN COMPRIMIDOS RECUBIERTOS 1 mg	SYNTHON CHILE LTDA.
ESCITALOPRAM	NEOPRESOL COMPRIMIDOS RECUBIERTOS 10 mg	LABORATORIO BAGO DE CHILE S.A.
ABACAVIR	MESABEX COMPRIMIDOS RECUBIERTOS 300 mg	LABORATORIOS LAFI LTDA.
QUETIAPINA	GOFYL COMPRIMIDOS RECUBIERTOS 25 mg	AXON PHARMA S.p.A.
QUETIAPINA	GOFYL COMPRIMIDOS RECUBIERTOS 200 mg	AXON PHARMA SPA.
QUETIAPINA	GOFYL COMPRIMIDOS RECUBIERTOS 100 mg	AXON PHARMA SPA.
CICLOBENZAPRINA	TENSIOMAX COMPRIMIDOS RECUBIERTOS 10 mg	LABORATORIO BAGO DE CHILE S.A.
CLORFENAMINA	CLORFENAMINA MALEATO COMPRIMIDOS RECUBIERTOS 4 mg	LABORATORIOS ANDREMACO S.A.
BICALUTAMIDA	BIOLEV COMPRIMIDOS RECUBIERTOS 150 mg	SYNTHON CHILE LTDA.
DESLOMATADINA	NEOLARMAX COMPRIMIDOS RECUBIERTOS 5 mg	LABORATORIOS ANDREMACO S.A.

Figura 47 Listado de medicamentos y el fabricante

Fuente: propia

- ✓ Se desea conocer el listado en el que muestra el nombre de la medicina y para qué tratamiento es utilizado

```
hive> SELECT producto, tratamiento FROM Medicinas WHERE tratamiento="Trastornos mentales";
OK
```

GOFYL COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales
GOFYL COMPRIMIDOS RECUBIERTOS 200 mg	Trastornos mentales
GOFYL COMPRIMIDOS RECUBIERTOS 100 mg	Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 2 mg	Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 1 mg	Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 3 mg	Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 100 mg.	Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 200 mg	Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 100 mg	Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 200 mg	Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 300 mg	Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBIERTOS DE LIBERACION PROLONGADA 50 mg	Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBRIMIENTO DE LIBERACION PROLONGADA 200 mg	Trastornos mentales
QUETIUM XR COMPRIMIDOS DE LIBERACION PROLONGADA 400 mg	Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBIERTOS DE LIBERACION PROLONGADA 300 mg	Trastornos mentales
QUETIAPINA COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales
RISPERIDONA COMPRIMIDOS DE DISOLUCION BUCAL 1 mg	Trastornos mentales
QUETIAPINA COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales
EUFREMIN COMPRIMIDOS RECUBIERTOS 100 mg	Trastornos mentales
EUFREMIN COMPRIMIDOS RECUBIERTOS 25 mg	Trastornos mentales

Figura 48 Listado de medicamentos y el tratamiento

Fuente: propia

- ✓ Muestra los medicamentos que son utilizados para Trastornos mentales

```
hive> SELECT producto, tratamiento FROM Medicinas WHERE tratamiento="Trastornos mentales";
OK
GOFYL COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
GOFYL COMPRIMIDOS RECUBIERTOS 200 mg Trastornos mentales
GOFYL COMPRIMIDOS RECUBIERTOS 100 mg Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 2 mg Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 1 mg Trastornos mentales
RISPYL COMPRIMIDOS RECUBIERTOS 3 mg Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 100 mg Trastornos mentales
QUETKARE COMPRIMIDOS RECUBIERTOS 200 mg Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 100 mg Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 200 mg Trastornos mentales
QUETIUM COMPRIMIDOS RECUBIERTOS 300 mg Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBIERTOS DE LIBERACION PROLONGADA 50 mg Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBIERTOS DE LIBERACION PROLONGADA 200 mg Trastornos mentales
QUETIUM XR COMPRIMIDOS DE LIBERACION PROLONGADA 400 mg Trastornos mentales
QUETIUM XR COMPRIMIDOS RECUBIERTOS DE LIBERACION PROLONGADA 300 mg Trastornos mentales
QUETIAPINA COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
RISPERIDONA COMPRIMIDOS DE DISOLUCION BUCAL 1 mg Trastornos mentales
QUETIAPINA COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
EFFENEM COMPRIMIDOS RECUBIERTOS 100 mg Trastornos mentales
EFFENEM COMPRIMIDOS RECUBIERTOS 25 mg Trastornos mentales
```

Figura 49 Listado de medicamentos utilizados para trastornos mentales

Fuente: propia

- ✓ Se quiere conocer la cantidad de productos existentes para el tratamiento de Antidepresivo.

```
EMERGEN COMPRIMIDOS 100 mg Antidepresivo 1
ERLINIZ COMPRIMIDOS RECUBIERTOS 10 mg (ESCITALOPRAM) Antidepresivo 1
ERLINIZ COMPRIMIDOS RECUBIERTOS 20 mg (Escitalopram) Antidepresivo 1
ESC COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
ESC COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
ESCITALOPRAM COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
ESCITALOPRAM COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
ESCITALOPRAM COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 2
ESCITALOPRAM COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
ESCITAVITAE COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
ESCITAVITAE COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
ETALOKARE COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
FLUOXETINA COMPRIMIDOS 20 mg Antidepresivo 3
IPRAN COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
IPRAN COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
LIVERTOX COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
LOWFIN COMPRIMIDOS RECUBIERTOS 50 mg Antidepresivo 1
LUNDYA COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
LUNDYA COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
NEOFRESOL COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
NEOZENTIS COMPRIMIDOS RECUBIERTOS 10 mg Antidepresivo 1
NERVIX COMPRIMIDOS RECUBIERTOS 75 mg Antidepresivo 1
PAMAX COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
PARLUXETIL COMPRIMIDOS RECUBIERTOS 20 mg Antidepresivo 1
```

```
hive> SELECT producto, tratamiento, COUNT(*) FROM Medicinas WHERE tratamiento="Antidepresivo" G
ROUP BY producto, tratamiento;
```

Figura 50 Cantidad de productos existentes antidepresivos

Fuente: propia

Estas son una muestra de las consultas que se pueden realizar con el fichero, pero se pueden hacer más y de cualquier tipo, todo dependerá de las necesidades de cada usuario. Los comandos que se usan son bastante similares y hasta cierto punto iguales a los utilizados en SQL.

7.4 Utilización de R



R es el programa que se va a utilizar para el análisis estadístico de los datos, por lo tanto, hay que seguir algunos pasos para la instalación y luego se procede a utilizarlo.

7.4.1 Instalación de R

En primer lugar, hay que descargar el instalador específico de la plataforma utilizada en el equipo. Para este caso es Windows, pero hay versiones para Mac OS o Linux.

Los pasos que se deben seguir son.

1. Descargar el instalador del sitio web:

<https://www.r-project.org/>

Aparecerá la siguiente pantalla:



*Figura 51 Instalación de R
Fuente: Sitio web de R*

2. Si se observa al lado izquierdo de la pantalla, aparece un enlace con el texto CRAN, acrónimo de «Comprehensive R Archive Network». CRAN es una red de servidores web y FTP distribuidos por todo el mundo que actúan como réplicas (mirrors) para la distribución del software y la documentación de R. Habrá que dar un clic a ese enlace y observará que aparece una página web conteniendo la lista de servidores de la red CRAN desde los que podrá descargar el software.

La lista está organizada geográficamente para que los usuarios escojan el servidor que se encuentre más próximo a su ubicación, porque es

probable que también se encuentre más cerca en Internet y, por tanto, los tiempos de descarga sean menores.

CRAN Mirrors	
The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. More statistics on the status of the mirrors can be found here: www.r-project.org/mirrors/cran/	
<p>Chile</p> <p>http://cran.usach.cl/ http://cran.usach.cl/</p> <p>Algeria</p> <p>http://cran.univ-bz.dz/</p> <p>Argentina</p> <p>http://cran.usc.edu.ar/cran/</p> <p>Australia</p> <p>http://cran.usc.edu.au/ http://cran.usc.edu.au/cran/</p> <p>Austria</p> <p>http://cran.usc.edu.at/ http://cran.usc.edu.at/cran/</p> <p>Bahamas</p> <p>http://cran.usc.edu.bs/ http://cran.usc.edu.bs/cran/</p> <p>Brazil</p> <p>http://cran.usc.edu.br/ http://cran.usc.edu.br/cran/ http://cran.usc.edu.br/cran/ http://cran.usc.edu.br/cran/ http://cran.usc.edu.br/cran/</p> <p>Canada</p> <p>http://cran.usc.edu.ca/ http://cran.usc.edu.ca/cran/ http://cran.usc.edu.ca/cran/</p>	<p>CRAN Mirrors</p> <p>Zürcher, Institute for Data Science in Medicine Zürich, Institute for Data Science in Medicine</p> <p>University of Science and Technology, Havana, Dominican Republic</p> <p>Venezuela, Universidad de La Plata</p> <p>CMDO University of Melbourne</p> <p>Wirtschaftsuniversität Wien Wirtschaftsuniversität Wien</p> <p>R.U. Leuven, Astronomical Ghent University Library</p> <p>Centre for Comp. Biol. at Universidade Estadual de Santa Cruz Universidade Federal de Pernambuco Oswaldo Cruz Foundation, Rio de Janeiro University of Sao Paulo, Sao Paulo University of Sao Paulo, Piracicaba</p> <p>Simon Fraser University, Toronto Dalhousie University, Halifax University of Toronto</p>

Figura 52 Lista de red de servidores CRAN

Fuente: Sitio web de R

Para este caso se selecciona el servidor de El Salvador

<p>El Salvador</p> <p>http://cran.salud.gob.sv/</p>	<p>Ministry of Health (Ministerio de Salud)</p>
---	---

- Una vez se haya seleccionado el servidor de descarga, verá que aparece una nueva página, en cuya parte superior aparecen enlaces específicos para cada sistema operativo. Verá también que desde esta página se tiene acceso al código fuente de la herramienta, en caso de que se quiera realizar la instalación a partir de él.



Figura 53 Pantalla para seleccionar el Sistema Operativo
Fuente: Sitio web de R

- Se selecciona el enlace **Download R para Windows** y lo llevará a otra página en la que le pedirá que seleccione qué es exactamente lo que se quiere descargar, si el sistema base, los paquetes de extensión (contributed packages) u otras herramientas. Para este caso, se selecciona el sistema base, por lo que lo llevará a la página donde se encuentra el instalador y procederá a descargarlo y a esperar para poder ejecutarlo.

R-3.2.2 for Windows (32/64 bit)

[Download R 3.2.2 for Windows](#) (62 megabytes, 32/64 bit)
[Installation and other instructions](#)
[New features in this version](#)

- Una vez instalado el programa, aparecerá el acceso directo en el escritorio siempre y cuando lo haya especificado al momento de la instalación.



6. Se procederá a ejecutar el programa y se abrirá una ventana que le ofrece una consola (R console), la cual funciona de manera similar a la línea de comandos del sistema operativo MS-DOS, en la que se digitan los comandos (denominados expresiones en R) y se obtendrá a través de ella los resultados (incluyendo mensajes de error si algo va mal).

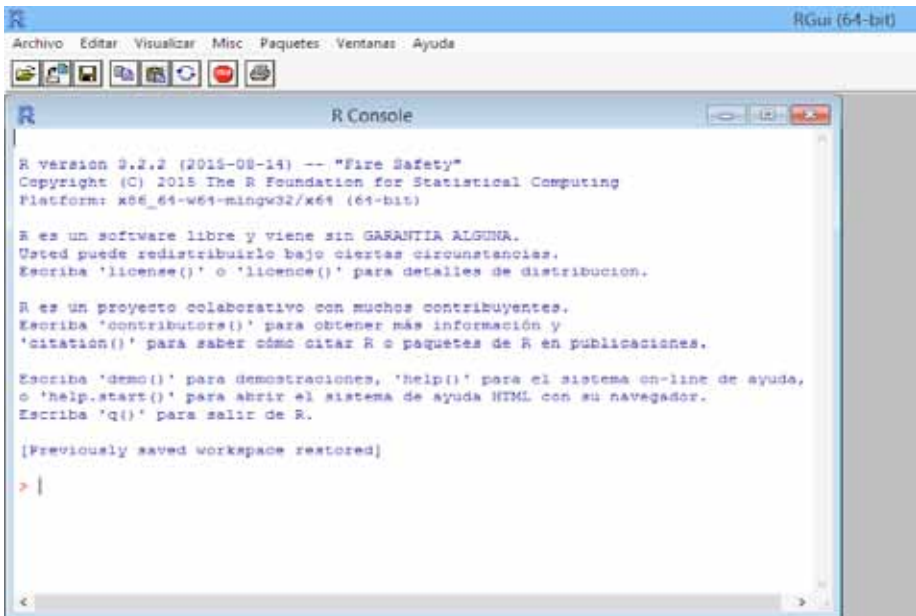


Figura 54 Interfaz de R (consola)

Fuente: propia

7.4.2 Utilización de R para el análisis estadístico de los datos

Como ya se tiene instalado R, se procede a hacer los análisis estadísticos necesarios con el *dataset* seleccionado. Para este caso se hará uso del *dataset* que contiene el registro sanitario de productos alimenticios, bebidas y medicamentos, por lo que se utilizarán algunos comandos que devolverán los resultados esperados.

1. Primeramente, se debe importar el *dataset* a la consola de R con el siguiente comando:

```
> datos<-read.csv("c:/datos/registro_sanitario_alimentos.csv",sep=",",head=T)
```

En este caso lo que se ha hecho es que se ha cargado el *dataset* en una variable a la que se ha nombrado datos. En R el operador de asignación es <- y no el signo = como es acostumbrado en algunos lenguajes de programación.

2. Al digitar el comando `head(datos)` mostrará los datos del *dataset* de la siguiente manera:

```
> head(datos)
Registro.Santiario  Nombre.del.Producto Estado      Fecha.Vence
1                   152  GUANDULES BANIA Vigente 25/09/2016 00:00
2                   153  SALMON HERONIC Vencido 17/05/2003 00:00
3                   154  SALMON SEAJJOY Vencido 17/05/2003 00:00
4                   155  FANTA GRAPE PARTE 1B Vencido 25/12/2010 00:00
5                   156  FANTA GRAPE PARTE 2 Vencido 25/12/2010 00:00
6                   157  HARINA DE MAIZ GRANO DE Vencido 17/05/2003 00:00

      Fábrica
1          3 ARROYOS, S.A.
2 ALEJANDRO CABRERA SANCHEZ
3 BELAERMINIO MOREL
4 CARIBBEAN REFRESCOS, INC.
5 CARIBBEAN REFRESCOS, INC.
6      JOSE A. FLAQUER.

      Dirección.de.la.Fábrica
1 CALLE CHACO 350, TRES ARROYOS, PROVINCIA DE BUENOS AIRES, ARGENTINA
2          SANTO DOMINGO R.D.
3          SABANA DE LA MAR
4          CIDRA, PUERTO RICO
5          CIDRA, PUERTO RICO
6          LICKY AL MEDIO

Tel.Fábrica      Representante Tel.Representante X X.1 X.2 X.3
1          ALCA DIE, SRL.
2          ALEJANDRO CABRERA SANCHEZ
3          LUIS E. MOLASCO
4          5038556 REFRESCOS NACIONALES, C. X A.      5038556
5          5038556 REFRESCOS NACIONALES, C. X A.      5038556
6          JOSE A. FLAQUER, IA
```

Figura 55 Mostrando los datos del dataset

Fuente: propia

3. Si se quiere saber cuántas filas y columnas tiene el fichero, se hace con el siguiente comando:

```
> dim (datos)
[1] 32890    9
```

Se puede observar que el fichero tiene 32890 filas y 9 columnas. Esto es debido a que se tuvo que hacer una limpieza del *dataset* original, ya que había filas con datos nulos y eso no era representativo a la hora de hacer análisis, por lo tanto, se eliminaron.

4. Como el *dataset* almacena registro sanitario de productos alimenticios, medicamentos y bebidas, se quiere saber la cantidad de cada uno de los productos que están registrados, y la cantidad de fábricas existentes, para ello se utilizará la función **summary**, la cual es usada para producir resultados resúmenes de las diversas funciones de ajuste del modelo

```
> summary (datos$Nombre.del.Producto)
ZUKO (PARA 1.5 LITROS)      16
  GALLETAS PANCHITOS""    13
    ALPINA BABY           10
      MI FRUTA PLUS POPULAR NARANJA EN POLVO  9
        PROTEINA DE SOYA  8
          GELATINA OLI CON TUTTI FRUTTI BALDOM"" 7
            QUESO MEDIA CREMA 7
              VINO TINTO DENOMINACION DE ORIGEN PENEDES 7
                ITALIAN FOOD 6
                  LIVEAN BEBIDA INSTANTANEA EN POLVO CON SABOR A 6
                    MINUTE MAID CITRUS PUNCH 6
                      QUESO DE LECHE DENASTADA 6
```

Figura 56 Listado de productos y las cantidades registradas

Fuente: propia

Los datos anteriores son una muestra de los resultados, debido a que el *dataset* contiene muchos registros, pero se observa el nombre del producto y la cantidad que hay registrada.

A continuación, se observa los nombres y las cantidades de las fábricas.

```
> summary (datos$Fábrica)
      BALTIMORE DOMINICANA, S. A. (BALDOM)
                                159
      GENERAL MILLS, INC.
                                136
      GALLETAS GULLON, S. A.
                                134
      VINICOLA DEL NORTE, S.A.
                                133
      RADIA SPICES, INC.
                                124
      INDUVECA, S. A.
                                123
      SIGMA ALIMENTOS DOMINICANA, S.A.
                                123
      ANGEL CAMACHO ALIMENTACION S.L
                                120
      KRAFT FOODS GROUP EXPORT, LLC.
                                114
      CORPORA TRESMONTES, S. A.
                                113
      KRAFT FOODS COSTA RICA, S. A.
                                107
      HELADOS BON, S. A.
                                105
```

Figura 57 Listado de Fábricas y las cantidades registradas
Fuente: propia

- Se desea conocer la cantidad de productos distintos que hay registrados:

```
> length (levels (datos$Nombre.del.Producto))
[1] 31661
```

Se observa que hay 31661 productos distintos.

Además, se quiere conocer la cantidad de Fábricas distintas que hay:

```
> length (levels (datos$Fábrica))
[1] 7283
```

Son 7283 fábricas distintas que tienen sus productos registrados.

- Se puede conocer todos los productos que están registrados y todos los fabricantes, haciendo uso de la siguiente instrucción:

```
> datos $Fábrica|
```

```
[32873] DYMATIZE ENTERPRICES, LLC.
[32874] DYMATIZE ENTERPRICES, LLC.
[32875] DYMATIZE ENTERPRICES, LLC.
[32876] MIGASA ENVASADO S.L.U. PARA CASH BAREA
[32877] CEXAC PARA CASH BAREA
[32878] GALLETAS GULLON, S. A.
[32879] BARCELO & CO., SRL.
[32880] CONSORCIO TECNO-DEAH, S.A.
[32881] PROCESADORA DE ALIMENTOS ARCOR DE ARGENTINA
[32882] BALTIMORE DOMINICANA, S. A. (BALDOM)
[32883] COLONIAL INDUSTRIAL C.POR A.
[32884] AGUA TINA
[32885] RALSTON FOODS, INC.
[32886] GOYA SANTO DOMINGO, S. A.
[32887] CONSERVAS EL CIDACO
[32888] ABBOTT LABORATORIES
[32889] WEMADO B.V.
[32890] KRAFT FOODS COSTA RICA, S. A.
7283 Levels: KARILU "" ... ZWOLSCHKE ZUIVELHANDEL
```

Figura 58 Listado de fábricas registradas

Fuente: propia

En la imagen se observa una parte de los fabricantes que hay, debido a que son muchos, pero al final muestra la cantidad total, que anteriormente se había conocido.

A continuación, el listado de productos existentes

```
> datos $Nombre.del.Producto|
```

```
[32873] DYMATIZE NUTRITION AMINO PRO SABOR PONCHE DE FRUTAS
[32874] DYMATIZE AMINO PRO SABOR NARANJA
[32875] DYMATIZE AMINO PRO LIMA LIMON (SUPLEMENTO)
[32876] ACEITE GIRASOL BARESOL
[32877] ACEITE OLIVA VIRGEN EXTRA OLIBAREA SL.
[32878] GULLON DIET FIBRA MUESLI
[32879] MAMAJUANA REPUBLIK (MIXTA)
[32880] LICOR ANISSETE CHEVALIER 25° %VOL.ALC.
[32881] GALLETITAS SABOR CHOCOLATE CON RELLENO SABOR
[32882] MAYONESA BALDOM""
[32883] REQUESON (QUARK) CON NATURAL "ELITE""
[32884] AGUA TINA
[32885] AVENA INSTANTANEA PAQUETE CON 4 VARIEDADES FOOD
[32886] ACEITE DE OLIVA EXTRA VIRGEN GOYA
[32887] FIMIENTO PIQUILLO
[32888] PEDIASURURE PLUS ( SABOR CHOCOLATE)
[32889] MINICUBOS CEBOLLA SAZONADOR KNORR""
[32890] TANG CHINOLA-MARACUYA
31661 Levels: 'AJONJOLI FELOMAR' ... ZUMO GASIFICADO SIN ALCOHOL DE PIÑA MAY"
```

Figura 59 Listado de productos registrados

Fuente: propia

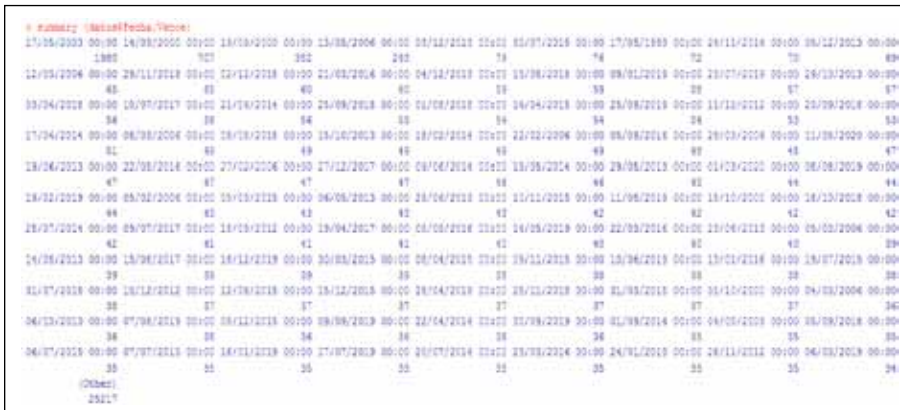
De igual manera que el listado de fabricantes, se muestran todos los productos registrados y al final indica cuantos son en total, dato que anteriormente se conoció.

- Se desea conocer cuántos productos han vencido y cuántos están vigentes
- Se desea conocer cuántos productos están registrados por fechas

```
> summary (datos$Estado)
Vencido Vigente
 19990   12900
```

Figura 60 Cantidad de productos registrados por fecha

Fuente: propia



- Se quiere ver el listado de productos y el estado (vencido o vigente), para ello se debe crear una tabla(dataframe) con los productos y el estado, por lo tanto, el procedimiento es el siguiente:

```
> productos<-datos$Nombre.del.Producto
> estado<-datos$Estado
> productos_estado<-data.frame(productos,estado)
.
```

Se observa que se han creados dos variables, una en la que se almacenan los productos y la otra el estado, posteriormente con esas variables se crea el dataframe.

Para visualizar los datos del dataframe, se hace lo siguiente:

```
> productos_estado|
```

```
FIBER ONE SOFT BAKED COOKIES OATMEAL RAISIN Vigente
DYMATIZE NUTRITION AMINO PRO SABOR PONCHE DE FRUITAS Vigente
DYMATIZE AMINO PRO SABOR NARANJA Vigente
DYMATIZE AMINO PRO LIMA LIMON (SUPLEMENTO Vigente
ACEITE GIRASOL BARESOL Vigente
ACEITE OLIVA VIRGEN EXTRA OLIBAREA 5L. Vigente
GULLON DIET FIBRA MUESLI Vigente
MAMAJUANA REPUBLIK (MIXTA) Vigente
LICOR ANISSETE CHEVALIER 25° %VOL.ALC. Vigente
GALLETITAS SABOR CHOCOLATE CON RELLENO SABOR Vencido
MAYONESA BALDOM"" Vencido
REQUESON (QUARK) CON NATURAL "ELITE"" Vencido
AGUA TINA Vencido
AVENA INSTANTANEA PAQUETE CON 4 VARIEDADES FOOD Vencido
ACEITE DE OLIVA EXTRA VIRGEN GOYA Vencido
PIMIENTO PIQUILLO Vencido
PEDIASURURE PLUS ( SABOR CHOCOLATE) Vencido
MINICUBOS CEBOLLA SAZONADOR KNORR"" Vigente
TANG CHINOLA-MARACUYA Vigente
```

Figura 61 Listado de productos y el estado en el que se encuentran
Fuente: propia

Para hacer un resumen de la tabla creada se haría de la siguiente manera:

```
> summary (productos_estado)
                productos          estado
ZUKO (PARA 1.5 LITROS)      :    16  Vencido:19990
GALLETAS PANCHITOS""      :    13  Vigente:12900
ALPINA BABY                 :    10
MI FRUTA PLUS POPULAR NARANJA EN POLVO :     9
PROTEINA DE SOYA           :     8
GELATINA OLI CON TUTTI FRUTTI BALDOM"" :     7
(Other)                     : 32827
```

Figura 62 Muestra el resumen de los productos
Fuente: propia

Se puede observar que hace un resumen de la cantidad de productos que hay y cuantos están vencidos y vigentes.

10. La misma tabla se puede utilizar para seleccionar las filas que contengan un producto determinado y el estado que queremos conocer. Por ejemplo, se ha seleccionado el producto “DOS PINOS DELIGURT YOGURT LIQUIDO SEMIDESCREMADO” y que su estado sea Vigente.

```
> productos_estado[productos=="DOS PINOS DELIGURT YOGURT LIQUIDO SEMIDESCREMADO"&&estado=="Vigente",]
      productos estado
32427 DOS PINOS DELIGURT YOGURT LIQUIDO SEMIDESCREMADO Vigente
32428 DOS PINOS DELIGURT YOGURT LIQUIDO SEMIDESCREMADO Vigente
```

Figura 63 Muestra el resumen de los productos
Fuente: propia

Se observan dos filas, es decir, hay dos productos con el estado solicitado.

11. A la tabla anterior se le ha agregado otra columna y es la fecha de vencimiento, por lo que quedaría de la siguiente manera el dataframe:

```
> fecha<-datos$Fecha.Vence
> productos_estado<-data.frame(productos,estado,fecha)
> productos_estado|
```

FIBER ONE SOFT BAKED COOKIES OATMEAL RAISIN	Vigente	17/05/2020	00:00
DYMATIZE NUTRITION AMINO PRO SABOR FONCHE DE FRUTAS	Vigente	11/05/2020	00:00
DYMATIZE AMINO PRO SABOR NARANJA	Vigente	11/05/2020	00:00
DYMATIZE AMINO PRO LIMA LIMON (SUPLEMENTO)	Vigente	11/05/2020	00:00
ACEITE GIRASOL BAREBOL	Vigente	12/05/2020	00:00
ACEITE OLIVA VIRGEN EXTRA OLIBAREA 5L.	Vigente	12/05/2020	00:00
GULLON DIET FIBRA MUESLI	Vigente	02/04/2019	00:00
MAMAJUANA REPUBLIK (MIXTA)	Vigente	19/02/2019	00:00
LICOR ANISSETE CHEVALIER 25° 4VOL.ALC.	Vigente	20/03/2019	00:00
GALLETITAS SABOR CHOCOLATE CON RELLENO SABOR	Vencido	29/04/2014	00:00
MAYONESA BALDON™	Vencido	22/06/2013	00:00
REQUESON (QUARK) CON NATURAL "ELITE"	Vencido	26/01/2013	00:00
AGUA TINA	Vencido	08/02/2013	00:00
AVENA INSTANTANEA PAQUETE CON 4 VARIETADES FOOD	Vencido	31/12/2012	00:00
ACEITE DE OLIVA EXTRA VIRGEN GOYA	Vencido	17/12/2012	00:00
PIMIENTO PIQUILLO	Vencido	29/05/2013	00:00
PEDIASURE PLUS (SABOR CHOCOLATE)	Vencido	30/05/2013	00:00
MINICUBOS CEBOLLA SAZONADOR "KHORR"	Vigente	27/09/2016	00:00
TANG CHINGOLA-MARACUYA	Vigente	08/05/2017	00:00

Figura 64 Listado de productos, con su estado y su fecha de vencimiento
Fuente: propia

Mostrando el resumen de la tabla en la que se ha agregado la fecha de vencimiento.

```
> summary (productos_estado)
```

	productos	estado	fecha
ZUCO (PARA 1.5 LITROS)	: 16	Vencido:19999	17/05/2003 00:00: 1980
GALLETAS PANCHITOS**	: 13	Vigente:12900	14/09/2000 00:00: 708
ALPINA BABY	: 10		19/09/2000 00:00: 382
LIC. EMMA MARIA DE LOURDES LEE.	: 9		13/05/2006 00:00: 293
MI FRUTA PLUS POPULAR NARANJA EN POLVO:	9		03/12/2013 00:00: 79
PROTEINA DE SOYA	: 8		30/07/2015 00:00: 76
(Other)	:32834		(Other) :29381

Figura 65 Resumen de la tabla generada

Fuente: propia

- Generando una tabla que contenga el producto y el fabricante, para luego hacer un resumen de esos datos.

```
> fabrica<-datos$Fábrica
> productos_fabrica<-data.frame(productos,fabrica)
> productos<-datos$Nombre.del.Producto
```

```
> summary (productos_fabrica)
```

	productos	fabrica
ZUCO (PARA 1.5 LITROS)	: 16	BALTIMORE DOMINICANA, S. A. (BALDOM): 169
GALLETAS PANCHITOS**	: 13	GENERAL MILLS, INC. : 136
ALPINA BABY	: 10	GALLETAS GULLON, S. A. : 134
MI FRUTA PLUS POPULAR NARANJA EN POLVO :	9	VINICOLA DEL NORTE, S.A. : 133
PROTEINA DE SOYA	: 8	BADIA SPICES, INC. : 124
GELATINA OLI CON TUTTI FRUTTI BALDOM**:	7	INDUVECA, S. A. : 123
(Other)	:32827	(Other) :32061

Figura 66 Muestra el resumen de fábricas

Fuente: propia

Se observa que, en esta tabla se muestra el resumen de la cantidad de los productos y la cantidad de Fábricas existentes.

7.5 Visualización de la Información

La visualización de la información es muy importante, debido a que a través de los gráficos se puede entender mejor los resultados obtenidos al hacer uso de cualquier herramienta de procesamiento y análisis estadístico de los datos.

Anteriormente se usó Hive en Hadoop para consultas al estilo SQL, pero también se hicieron análisis estadísticos con R. Partiendo de ello se elaboran gráficos utilizando las siguientes herramientas: Google Chart, Jqplot y D3.js

En esta investigación se hace uso de las tres herramientas para demostrar lo que se puede hacer con cada una de ellas y como muestran los datos. Al final es opcional con cual trabajar y dependerá de la habilidad que se tenga en la programación con Java Script, html y css, solo que debe tomarse en cuenta las ventajas de cada una de ellas y de las necesidades que se tenga en un momento dado en la representación de los datos.

Los *dataset* que se han utilizado para esta investigación han sido revisados y depurados con Google Spreadsheets, debido a que contenían columnas con campos vacíos los cuales no eran útiles en la representación de los datos, por lo tanto, esas filas han sido eliminadas para que la información sea lo más real posible.

Para ello, el *dataset* de listado de medicamentos originalmente tenía 666 registros, pero al depurarlo quedó con 627

Para el caso del *dataset* de registro de productos alimenticios, originalmente tenía 33033 registros, pero quedó finalmente con 32890.

Los códigos utilizados para la elaboración de los gráficos se han basado en galerías existentes en la web, debido a que hay ejemplos desarrollados y que pueden ser adecuados a las necesidades de cada usuario, por ejemplo, para Google Chart existe una diversidad de galerías que pueden utilizarse, estas se encuentran en la siguiente URL:

<https://google-developers.appspot.com/chart/interactive/docs/gallery>

Para hacer uso de D3.js al igual que en Google Chart y Jqplot, es necesario tener conocimiento de java script. Sin embargo, también existen galerías y ejemplos que pueden ayudar a crear los gráficos de forma fácil y adaptar los datos como se considere necesario. Estas galerías pueden encontrarse en <https://github.com/mbostock/d3/wiki/Gallery>

Para Jqplot, también existen ejemplos en: <http://www.jqplot.com/tests/> y al igual que en D3.js, se hace uso de librerías que en muchos casos hay que descargarlas para utilizar algunas opciones y efectos en los gráficos.

Todos los gráficos elaborados son dinámicos, por lo que, para poder ver las animaciones, es necesario usar un servidor web, por ejemplo, Xampp. También se puede usar plunker y ejecutarlos desde el sitio <http://plnkr.co/edit/?p=catalogue>.

En algunos gráficos se ha hecho uso de archivos csv, es decir; se han importado los datos en el código. Este archivo csv ha sido elaborado con los datos más importantes que se obtuvieron en los análisis realizados con las otras herramientas, los códigos utilizados podrán verse en los anexos de este documento, a continuación, se presentan las visualizaciones generadas.

Gráfica 1

En la siguiente gráfica se muestran los principales titulares o fabricantes de medicamentos que se tienen registrados en el *dataset*, estos son los que tienen mayores cantidades de productos bajo su distribución. La gráfica ha sido elaborada en D3.

El grafico animado se puede ver en:

<http://embed.plnkr.co/2nYJgPkfwfsXhnaoxwSO/preview>

<http://plnkr.co/edit/1kdc1RsnHaWtpNATc7hh?p=preview>

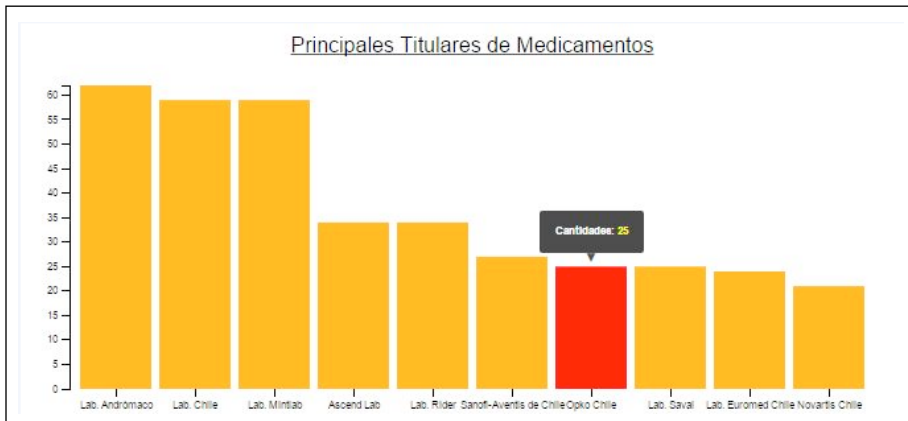


Figura 67 Grafica que muestra los principales titulares de medicamentos
Fuente: propia

Como se mencionó, los *dataset* han sido depurados en Google Spreadsheet y según los datos obtenidos en los análisis realizados se creó una hoja en donde aparecen los datos más importantes para generar pequeños archivos csv que son utilizados dentro de algunos códigos.

En la siguiente dirección se puede ver el archivo en una hoja de Google Spreadsheet

<https://docs.google.com/spreadsheets/d/1KuU2Ns5pF1a03tobCHk9nM9jXj2HQQDmr8lszjyzWA/edit#gid=1175116494>

La hoja que en Google Spreadsheet tiene por nombre *datos importantes* y son los datos que han sido obtenidos en el análisis con R y con Hive.

En este caso se observa que, de 54 titulares o fabricantes de medicamentos, 10 de ellos son los que más productos de medicamentos tienen registrados. Siendo Laboratorios Andrómaco el que tiene la mayor cantidad 62 productos.

El código de este gráfico puede verse en el Anexo 1.

Gráfica 2

En la siguiente gráfica se muestran los medicamentos que tienen mayores cantidades registradas. La gráfica ha sido elaborada en Google Chart

El gráfico animado se puede ver en el siguiente enlace <http://embed.plnkr.co/sThGhTevFgwpUSI6Gzkp/preview>

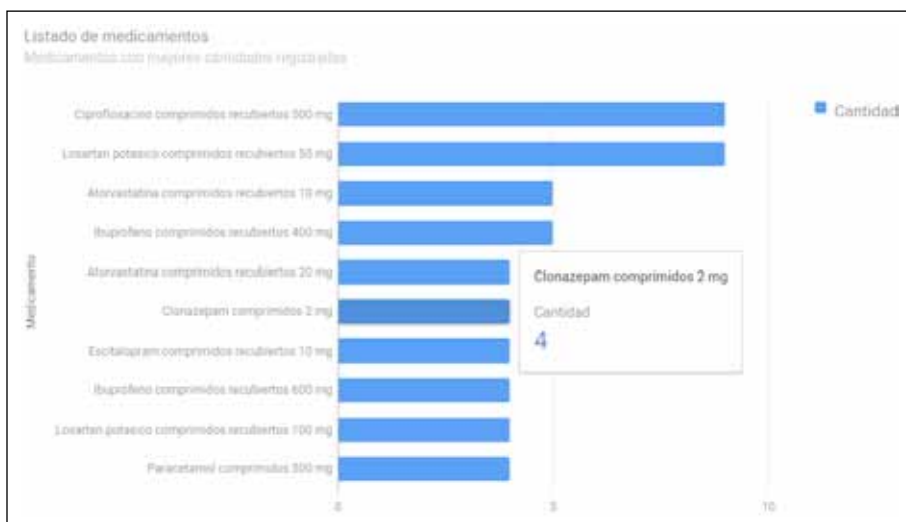


Figura 68 Gráfica que muestra los medicamentos con mayor existencia
Fuente: propia

Se puede observar que los medicamentos Ciprofloxacino y Losartán potásico son los que tienen mayores cantidades registradas y al consultar el *dataset*, estos medicamentos son usados en tratamiento como antibiótico y para hipertensión arterial respectivamente, por lo que se puede concluir que son los que más demanda tienen por los pacientes que requieren de este medicamento, por lo tanto, siempre deben tenerse registrados o en existencia.

El código utilizado para la elaboración del gráfico puede verse en el Anexo 2.

Gráfica 3

En la siguiente gráfica se hace una comparación de la cantidad de productos o medicamentos con la cantidad de fabricantes o titulares. Por supuesto que hay muchos más productos que fabricantes, debido a que la mayoría de los titulares elaboran diferentes productos en el mercado. El gráfico ha sido elaborado en Jqplot.

Ver el gráfico animado en <http://embed.plnkr.co/dB6Tt5VN4LgXLBm4Gvc4/preview>

<http://plnkr.co/edit/g5IUuKnA3gPBHp1u8oY9?p=preview>

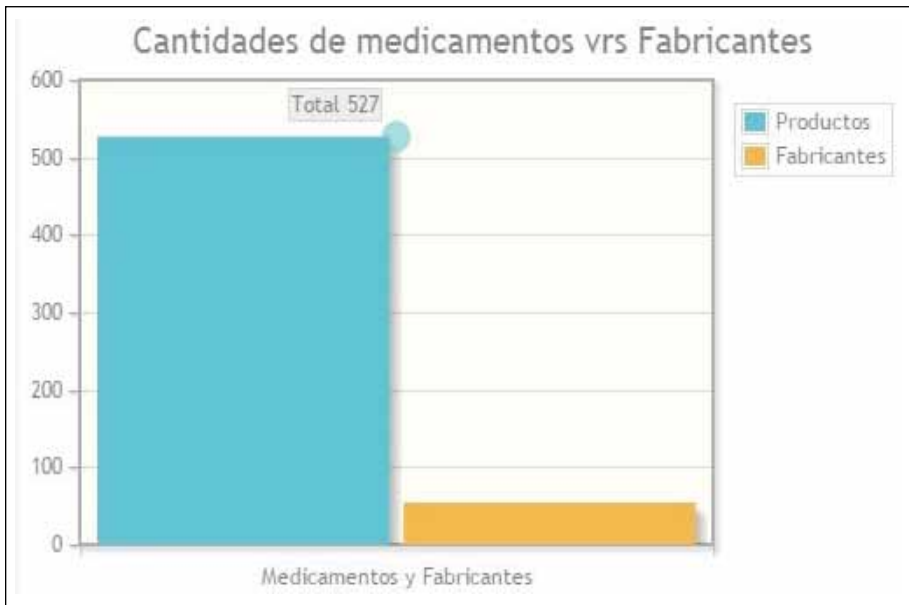


Figura 69 Gráfica que muestra una comparativa entre medicamentos y fabricantes
Fuente: propia

El código utilizado para este gráfico, puede revisarse en el Anexo 3.

Gráfica 4

A continuación, se muestran las gráficas del *dataset* de registro sanitario de productos alimenticios y medicamentos.

Los datos se depuraron en una hoja de Google Spreadsheet, en la que se colocaron los datos necesarios para la gráfica.

La gráfica ha sido elaborada en Google Chart y muestra los productos que tienen mayores cantidades registradas.

De la cantidad total de productos se han seleccionado únicamente los 10 más representativos en cantidades, los cuales han sido obtenidos del análisis estadístico en R.



Figura 70 Gráfica que muestra los productos con mayores cantidades registradas

Fuente: propia

El gráfico animado se puede ver en

<http://embed.plnkr.co/7IiWJAqgg67KGPoNdOdI/preview>

Se logra ver en la gráfica que los productos que tienen más cantidades registradas son las bebidas en polvo, como Zuko, Mi fruta plus y Livean bebida. Pero también las golosinas tales como galletas y gelatinas.

Se puede revisar el código utilizado para la elaboración del gráfico en el Anexo 4.

Gráfica 5

En la siguiente grafica se muestran las principales fábricas que tienen más productos registrados y es por ello que aparecen muchas veces en el *dataset*.

La gráfica animada ha sido elaborada en D3 y puede ser vista en el siguiente enlace <http://plnkr.co/edit/U8p9G55AX8jv6Rs0NcGn?p=preview>

Se observa que la Fábrica Baltimore (Baltimore Dominicana S.A.) es la que tiene mayor cantidad de productos registrados, luego le sigue General Mills, Inc y Galletas Guillon S.A.

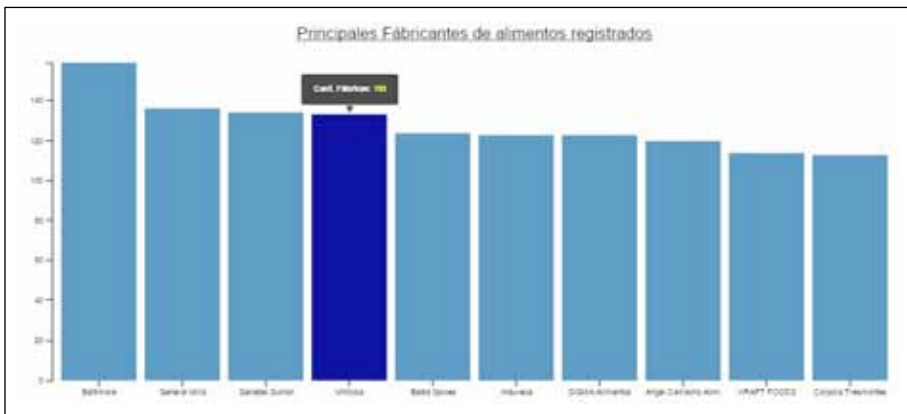


Figura 71 Grafica que muestra los fabricantes con mayores cantidades registradas

Fuente: propia

Puede verse el código utilizado para la elaboración del gráfico en el **Anexo 5**.

Gráfica 6

El siguiente gráfico es un pie chart (gráfico de pastel) en 3D, elaborado en Google Chart, en el que se refleja el porcentaje de productos vigentes y vencidos.



Figura 72 Gráfica que muestra el porcentaje de productos vigentes y vencidos
Fuente: propia

El gráfico animado puede verse en el siguiente enlace

<http://embed.plnkr.co/wJNt3gGHm9ip2BkNwW2s/preview>

Se observa que más del 60% de los productos que están registrados están vencidos y solo un poco más del 39% están vigentes. Esto significa que deben revisarse los productos vigentes para estar pendientes con las fechas de vencimiento y tener un mejor control.

Se puede ver el código utilizado para la elaboración del gráfico en el Anexo 6.

Gráfica 7

La siguiente gráfica muestra la cantidad de productos que está próximo a venderse a finales del año 2015 y a principios del 2016, la gráfica ha sido elaborada en Google Chart.

El código utilizado para la elaboración del gráfico puede verse en el Anexo 7

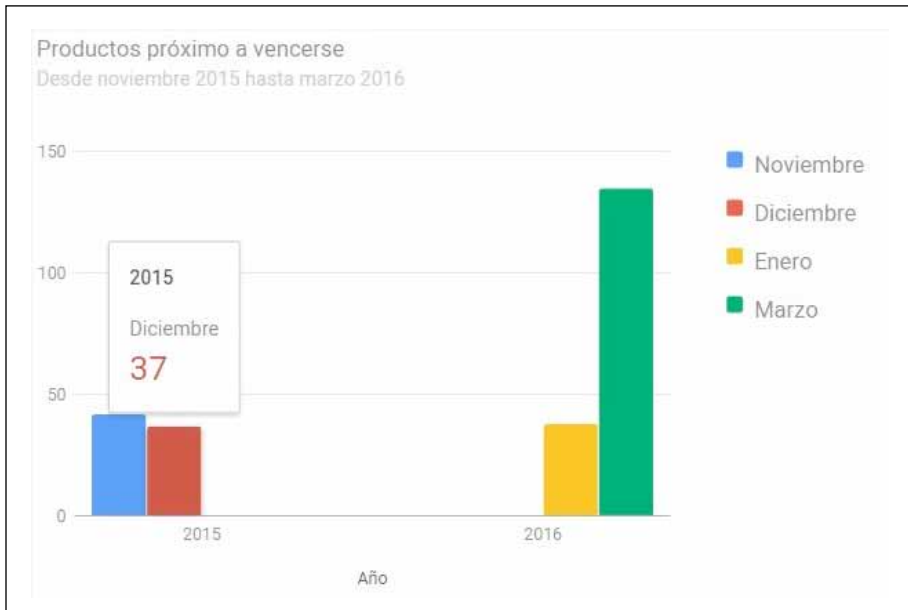


Figura 73 Gráfica que muestra la cantidad de productos próximos a venderse
Fuente: propia

El enlace para ver la gráfica en forma dinámica es el siguiente:

<http://embed.plnkr.co/L0eHthQ1uM2f5E15PTIW/preview>

Se puede observar que una buena cantidad de productos están próximos a venderse en lo que resta del año 2015 y los otros en el primer trimestre del 2016.

8. Bibliografía

- Aguilar, L. J. (2013). *Big Data, Analisis de los grandes volúmenes de datos*. Mexico: Alfaomega.
- Bernardo, A. (8 de mayo de 2013). *Think Big*. Obtenido de <http://blog-thinkbig.com/big-data-cancer/>
- Carrillo Ruiz, J. A., Marco de Lucas, J. E., Dueñas López, J. C., Cases Vega, F., Fernández, J. C., Pereda Laredo, L. F., & González Muñoz de Morales, G. (marzo de 2013). <http://www.ieee.es/>. Obtenido de http://www.ieee.es/Galerias/fichero/docs_investig/DIEEINV03-2013_Big_Data_Entornos_DefensaSeguridad_CarrilloRuiz.pdf
- De Juana, R. (24 de febrero de 2015). *MuyComputerPro*. Obtenido de <http://www.muycomputerpro.com/2015/02/24/ejemplos-reales-uso-inteligente-big-data>
- Dean, J. a. (2004). *MapReduce: Simplified Data Processing on Large*. Obtenido de <http://static.googleusercontent.com/media/research.google.com/es//archive/mapred>
- Dean, J. a. (2004). *MapReduce: Simplified Data Processing on Large Clusters*. .
- FondosFidelity. (2012). Obtenido de https://www.fondosfidelity.es/static/pdfs/informes-fondos/Fidelity_ArgInvSXXI_BigData_Sept12_ES.pdf
- Ghemawat, S. G. (2003). *The Google File System*.
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (octubre de 2003). *The Google File System*. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles*. Obtenido de <http://static.googleusercontent.com/media/research.google.com/es//archive/gfs-sosp2003.pdf>
- IBM, *developerworks*. (2012). Obtenido de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- IBM, *developerworks*. (2012). Obtenido de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- IBM, *developerworks*. (2012). Obtenido de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

IBM, *developerworks*. (2012). Obtenido de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

IBM, *developerworks*. (2012). Obtenido de <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

Joyanes Aguilar, L. (2013). *Big Data: Análisis de grandes volúmenes de datos en las organizaciones*. Mexico: AlfaOmega.

Scherer, M. (9 de noviembre de 2012). *DataPrix*. Obtenido de <http://www.dataprix.com/noticias-it/tendencias-tecnologicas/big-data/big-data-ayudaron-obama-ganar-las-elecciones>

Souto, S. (4 de mayo de 2015). *Hechos de Hoy*. Obtenido de <http://www.hechosdehoy.com/big-data-pilar-fundamental-para-el-desarrollo-de-las-ciudades-inteligentes-43285.htm>

TestingSoft. (2014). *Tipos de datos a explorar. Recuperado el 2014, de http://www.evaluandosoftware.com/nota-3684-Que-es-el-Big-Data.html.*

Zikopoulos, P. (2015). *BeSmart*. Obtenido de <http://www.besmart.com-company/blog/big-data-transforma-ciudades/>

URL utilizadas

Dataset sobre registro sanitario de alimentos

http://datos.gob.do/vi_VN/dataset/registro-sanitario-de-alimentos-y-bebidas/resource/464e4ed5-92ef-48ee-9dd3-27016b43ff28

Dataset sobre medicamentos

http://www.partidopirata.cl/wp-content/uploads/2015/09/Listado-de-productos-Bioequivalentes_agosto_2015.csv

Documentos regulatorio Ministerio de Salud

<http://usam.salud.gob.sv/index.php/temas/alimentos>

http://usam.salud.gob.sv/archivos/pdf/alimentos/tarifa_por_servicios.pdf

Proyecto Big Data

<http://www.alten.es/wp-content/uploads/2014/07/CONCLUSIONES-PROYECTO-BIG-DATA.pdf>

Big Data en la vida diaria

<https://thebigbunkerdata.wordpress.com/2015/04/08/big-data-en-nuestra-vida-cotidiana/>

Infografía *Big Data*

<http://www.ibmbigdatahub.com/infographic/flood-big-data>

Herramienta de visualización D3.js

<http://gcoch.github.io/D3-tutorial/>

<https://d3js.org/>

Herramienta Google Chart

<https://developers.google.com/chart/?hl=es>

https://es.wikipedia.org/wiki/Google_Chart

Herramienta jqplot

<https://www.jqplot.com/>

9. ANEXOS

Anexo 1

Código en D3.js para la elaboración del gráfico de barras, que muestra los principales titulares de medicamentos.

```

1  <!DOCTYPE html>
2  <meta charset="utf-8">
3  <style>
4
5  body {
6    font: 8px sans-serif;
7  }
8
9  .axis path,
10 .axis line {
11   fill: none;
12   stroke: #000;
13   shape-rendering: crispEdges;
14 }
15
16 .bar {
17   fill: orange;
18 }
19
20 .bar:hover {
21   fill: red ;
22 }
23
24 .x.axis path {
25   display: none;
26 }
27
28 .d3-tip {
29   line-height: 1;
30   font-weight: bold;
31   padding: 12px;
32   background: rgba(0, 0, 0, 0.8);
33   color: #fff;
34   border-radius: 2px;
35 }
36
37 /* Creates a small triangle extender for the tooltip */
38 .d3-tip:after {
39   box-sizing: border-box;
40   display: inline;
41   font-size: 10px;
42   width: 100%;
43   line-height: 1;
44   color: rgba(0, 0, 0, 0.8);

```

```

45   content: "\25BC";
46   position: absolute;
47   text-align: center;
48 }
49
50 /* Style northward tooltips differently */
51 .d3-tip.n:after {
52   margin: -1px 0 0 0;
53   top: 100%;
54   left: 0;
55 }
56 </style>
57 <body>
58 <script src="http://d3js.org/d3.v3.min.js"></script>
59 <script src="http://labratrevenge.com/d3-tip/javascripts/d3.tip.v0.6.3.js"></script>
60 <script>
61 // Gráfico que muestra los principales titulares de medicamentos.
62 var margin = {top: 40, right: 20, bottom: 30, left: 60},
63     width = 900 - margin.left - margin.right,
64     height = 300 - margin.top - margin.bottom;
65
66 var x = d3.scale.ordinal()
67     .rangeRoundBands([0, width], .1);
68
69 var y = d3.scale.linear()
70     .range([height, 0]);
71
72 var xAxis = d3.svg.axis()
73     .scale(x)
74     .orient("bottom");
75
76 var yAxis = d3.svg.axis()
77     .scale(y)
78     .orient("left");
79
80
81 var tip = d3.tip()
82     .attr('class', 'd3-tip')
83     .offset([-10, 0])
84     .html(function(d) {
85         return "<strong>Cantidades:</strong> <span style='color:yellow'> " + d.Cantidad + "</span>";
86     });
87
88 var svg = d3.select("body").append("svg")
89     .attr("width", width + margin.left + margin.right)
90     .attr("height", height + margin.top + margin.bottom)
91     .append("g")
92     .attr("transform", "translate(" + margin.left + ", " + margin.top + ")");
93
94 svg.call(tip);
95
96 d3.csv("data.csv", type, function(error, data) {
97     x.domain(data.map(function(d) { return d.Titular; }));
98     y.domain([0, d3.max(data, function(d) { return d.Cantidad; })]);
99
100    svg.append("g")
101        .attr("class", "x axis")
102        .attr("transform", "translate(0," + height + ")")
103        .call(xAxis);
104

```

```

105     svg.append("g")
106         .attr("class", "y axis")
107         .call(yAxis)
108         .append("text")
109             .attr("transform", "rotate(-90)")
110             .attr("y", 6)
111             .attr("dy", ".71em")
112             .style("text-anchor", "end")
113
114     svg.append("text")
115         .attr("class", "title")
116         .attr("x", (width / 2))
117         .attr("y", -25)
118         .attr("text-anchor", "middle")
119         .style("font-size", "16px")
120         .style("text-decoration", "underline")
121         .text("Principales Titulares de Medicamentos");
122
123     svg.selectAll(".bar")
124         .data(data)
125         .enter().append("rect")
126             .attr("class", "bar")
127             .attr("x", function(d) { return x(d.Titular); })
128             .attr("width", x.rangeBand())
129             .attr("y", function(d) { return y(d.Cantidad); })
130             .attr("height", function(d) { return height - y(d.Cantidad); })
131             .on('mouseover', tip.show)
132             .on('mouseout', tip.hide)
133
134     });
135
136     function type(d) {
137         d.Cantidad = +d.Cantidad;
138         return d;
139     }
140 </script>

```

Archivo csv utilizado (data.csv)

Titular	Cantidad
Laboratorio Andrómaco	62
Laboratorio Chile	59
Laboratorio Mintlab	59
Ascend Laboratories	34
Laboratorio Rider	34
Sanofi-Aventis de Chile	27
Opko Chile	25
Laboratorio Saval	25
Laboratorio Euromed Chile	24
Novartis Chile	21

Anexo 2

Código en Google Chart para la elaboración del gráfico de barras horizontales, que muestra el listado de productos de medicamentos con mayor cantidad registrada.

```

1 </html>
2 <head>
3 <script type="text/javascript" src="https://www.google.com/jsapi"></script>
4 <script type="text/javascript">
5   google.load("visualization", "1.1", {packages:["bar"]});
6   google.setOnLoadCallback(drawChart);
7   function drawChart() {
8     var data = google.visualization.arrayToDataTable([
9       ['Medicamento', 'Cantidad', { role: 'style' }],
10      ['Clorofeniramina comprimidos recubiertos 500 mg', 9, 'color: gray'],
11      ['Losartan potásico comprimidos recubiertos 50 mg', 9, 'color: #76A7FA'],
12      ['Atorvastatina comprimidos recubiertos 10 mg', 5, 'opacity: 0.2'],
13      ['Ibuprofeno comprimidos recubiertos 400 mg', 5, 'stroke-color: #783593; stroke-width: 4; fill-color: #C3A5CF'],
14      ['Atorvastatina comprimidos recubiertos 20 mg', 4, 'stroke-color: #871B47; stroke-opacity: 0.6; stroke-width: 0; fill-color: #8C5679; fill-opacity: 0.2'],
15      ['Clonacepan comprimidos 2 mg', 4, 'color: gray'],
16      ['Escitalopram comprimidos recubiertos 10 mg', 4, 'color: #76A7FA'],
17      ['Ibuprofeno comprimidos recubiertos 600 mg', 4, 'opacity: 0.2'],
18      ['Losartan potásico comprimidos recubiertos 100 mg', 4, 'stroke-color: #783593; stroke-width: 4; fill-color: #C3A5CF'],
19      ['Paracetamol comprimidos 500 mg', 4, 'color: #76A7FA'],
20    ]);
21
22     var options = {
23       chart: {
24         title: 'Listado de medicamentos',
25         subtitle: 'Medicamentos con mayores cantidades registradas',
26       },
27       bars: 'horizontal' // Required for Material Bar Charts.
28     };
29
30     var chart = new google.charts.Bar(document.getElementById('bar_chart_material'));
31
32     chart.draw(data, options);
33   }
34 </script>
35 </head>
36 <body>
37 <div id="bar_chart_material" style="width: 500px; height: 500px;"></div>
38 </body>
39 </html>

```

Anexo 3

Código en Jqplot para la elaboración del gráfico de barras verticales, que muestra la cantidad de medicamentos versus fabricantes.

```

1 <!DOCTYPE html>
2 <html>
3 <head>
4 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/jquery.min.js"></script>
5 <script language="javascript" type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/jquery.jqplot.min.js"></script>
6 <link rel="stylesheet" type="text/css" href="http://rd.unr.net/imagenes/jquery/jqplot/jquery.jqplot.css" />
7 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.jqplot.barRenderer.min.js"></script>
8 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.categoryAxisRenderer.min.js"></script>
9 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.pointLabels.min.js"></script>
10 <!-- Agrega estas tres librerías más para mostrar el Tooltip-->
11 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.highlighter.min.js"></script>
12 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.cursor.min.js"></script>
13 <script type="text/javascript" src="http://rd.unr.net/imagenes/jquery/jqplot/plugins/jquery.dateAxisRenderer.min.js"></script>
14 </script>
15
16 $(document).ready(function(){
17     var a1 = [227];
18     var a2 = [34];
19     // Can specify a custom tick array.
20     // Ticks should match up one for each y value (category) in the series.
21     var ticks = ["Medicamentos y Fabricantes"];
22
23     var plot1 = $.jqplot("chart1", [a1, a2], {
24         // The "seriesDefaults" option is an optional object that will
25         // be applied to all series in the chart.
26         title: "Cantidad de medicamentos por Fabricantes", // Vers el título
27
28         // para el tooltip que muestra el pasar el mouse
29         show: true,
30         sizeOfJuts: 7.5,
31         multiSeries: 'y',
32         formatting: 'total %d'
33     });
34     // para el título que muestra el pasar el mouse
35     show: false
36     },
37     seriesDefaults: {
38         renderer: $.jqplot.BarRenderer,
39         rendererOptions: {
40             barDirection: "vertical"
41         }
42     },
43     // Custom labels for the series are specified with the "label"
44     // option on the series option. Here a series option object
45     // is specified for each series.
46     series: [
47         {label: "Medicamentos"},
48         {label: "Fabricantes"},
49     ]
50     // Show the legend and put it outside the grid, but inside the
51     // plot container, shrinking the grid to accommodate the legend.
52     // A value of "outside" would not shrink the grid and allow
53     // the legend to overflow the container.
54     legend: {
55         show: true,
56         placement: "outsideGrid"
57     },
58     axes: {
59         // Set a category axis on the y axis and use our custom ticks.
60         yaxis: {
61             renderer: $.jqplot.CategoryAxisRenderer,
62             ticks: ticks,
63             tickOptions: {
64                 formatting: '%d'
65             }
66         },
67         // Set the x axis just a little so bars can get close to, but
68         // not touch, the grid boundaries. 1.2 is the default spacing.
69         xaxis: {
70             pad: 1.0,
71             tickOptions: {formatting: '%d'}
72         }
73     }
74 });
75 </script>
76 </head>
77 <body>
78 <div id="chart1">
79 </div>
80 </body>
81 </html>

```

Anexo 4

Código en Google Chart, para la elaboración del gráfico de barras horizontales, que muestra los productos con mayores cantidades registradas.

```

1- <html>
2-   <head>
3-     <script type="text/javascript" src="https://www.google.com/jsapi"></script>
4-     <script type="text/javascript">
5-       google.load("visualization", "1.1", {packages:["bar"]});
6-       google.setOnLoadCallback(drawStuff);
7-
8-       function drawStuff() {
9-         var data = new google.visualization.arrayToDataTable([
10-          ['Productos Registrados', 'Cantidades'],
11-          ['ZUKO (1.5 Litros)', 16],
12-          ['Galletas Panchitos', 13],
13-          ['Alpina Baby', 10],
14-          ['Mi fruta plus popular naranja en polvo', 9],
15-          ['Proteina de soya', 8],
16-          ['Gelatina Oli con tutti frutti', 7],
17-          ['Queso media crema', 7],
18-          ['Vino tinto denominacion de origen penedes', 7],
19-          ['Italian Food', 6],
20-          ['Livean bebida instantanea en polvo', 6],
21-        ]]);
22-
23-         var options = {
24-           title: 'Chess opening moves',
25-           width: 900,
26-           legend: { position: 'none' },
27-           chart: { title: 'Productos mas populares registrados',
28-                   subtitle: 'Cantidades de productos' },
29-           bars: 'horizontal', // Required for Material Bar Charts.
30-           axes: {
31-             x: {
32-               0: { side: 'top', label: 'Cantidades' } // Top x-axis.
33-             }
34-           },
35-           bar: { groupWidth: "90%" }
36-         };
37-
38-         var chart = new google.charts.Bar(document.getElementById('top_x_div'));
39-         chart.draw(data, options);
40-       };
41-     </script>
42-   </head>
43-   <body>
44-     <div id="top_x_div" style="width: 900px; height: 500px;"></div>
45-   </body>
46- </html>

```

Anexo 5

Código en D3.js para la elaboración del gráfico de barras verticales, que muestra las principales fábricas que tienen sus productos registrados.

```

1  <!DOCTYPE html>
2  <meta charset="utf-8">
3  <style>
4
5  body {
6    font: 8px sans-serif;
7  }
8
9  .axis path,
10 .axis line {
11   fill: none;
12   stroke: #000;
13   shape-rendering: crispEdges;
14 }
15
16 .bar {
17   fill: steelblue;
18 }
19
20 .bar:hover {
21   fill: darkblue ;
22 }
23
24 .x.axis path {
25   display: none;
26 }
27
28 .d3-tip {
29   line-height: 1;
30   font-weight: bold;
31   padding: 12px;
32   background: rgba(0, 0, 0, 0.8);
33   color: #fff;
34   border-radius: 2px;
35 }
36
37 /* Creates a small triangle extender for the tooltip */
38 .d3-tip:after {
39   box-sizing: border-box;
40   display: inline;
41   font-size: 12px;
42   width: 100%;
43   line-height: 1;
44   color: rgba(0, 0, 0, 0.8);
45
46   content: "\25BC";
47   position: absolute;
48   text-align: center;
49 }
50 /* Style northward tooltips differently */
51 .d3-tip.n:after {
52   margin: -1px 0 0 0;
53   top: 100%;
54   left: 0;
55 }
56 </style>
57 <body>
58 <script src="http://d3js.org/d3.v3.min.js"></script>
59 <script src="http://labratrevenge.com/d3-tip/javascripts/d3.tip.v0.6.3.js"></script>
60 </script>

```

```

61
62 var margin = {top: 40, right: 20, bottom: 30, left: 60},
63   width = 900 - margin.left - margin.right,
64   height = 400 - margin.top - margin.bottom;
65
66 var x = d3.scale.ordinal()
67   .rangeRoundBands([0, width], .1);
68
69 var y = d3.scale.linear()
70   .range([height, 0]);
71
72 var xAxis = d3.svg.axis()
73   .scale(x)
74   .orient("bottom");
75
76 var yAxis = d3.svg.axis()
77   .scale(y)
78   .orient("left");
79
80
81 var tip = d3.tip()
82   .attr('class', 'd3-tip')
83   .offset([-10, 0])
84   .html(function(d) {
85     return "<strong>Cantidad:</strong> <span style='color:yellow'>" + d.Cantidad + "</span>";
86   });
87
88 var svg = d3.select("body").append("svg")
89   .attr("width", width + margin.left + margin.right)
90   .attr("height", height + margin.top + margin.bottom)
91   .append("g")
92   .attr("transform", "translate(" + margin.left + "," + margin.top + ")");
93
94 svg.call(tip);
95
96 d3.csv("datos_fab.csv", type, function(error, data) {
97   x.domain(data.map(function(d) { return d.Fabricas; }));
98   y.domain([0, d3.max(data, function(d) { return d.Cantidad; })]);
99
100  svg.append("g")
101    .attr("class", "x axis")
102    .attr("transform", "translate(0," + height + ")")
103    .call(xAxis);
104
105  svg.append("g")
106    .attr("class", "y axis")
107    .call(yAxis)
108    .append("text")
109    .attr("transform", "rotate(-90)")
110    .attr("y", 6)
111    .attr("dy", ".71em")
112    .style("text-anchor", "end")

```

```

113 |
114 |     svg.append("text")
115 |         .attr("class", "title")
116 |         .attr("x", (width / 2))
117 |         .attr("y", -25)
118 |         .attr("text-anchor", "middle")
119 |         .style("font-size", "16px")
120 |         .style("text-decoration", "underline")
121 |         .text("Principales Fabricantes de alimentos registrados");
122 |
123 |     svg.selectAll(".bar")
124 |         .data(data)
125 |         .enter().append("rect")
126 |         .attr("class", "bar")
127 |         .attr("x", function(d) { return x(d.Fabricas); })
128 |         .attr("width", x.rangeBand())
129 |         .attr("y", function(d) { return y(d.Cantidad); })
130 |         .attr("height", function(d) { return height - y(d.Cantidad); })
131 |         .on('mouseover', tip.show)
132 |         .on('mouseout', tip.hide)
133 |
134 |     });
135 |
136 | function type(d) {
137 |     d.Cantidad = +d.Cantidad;
138 |     return d;
139 | }
140 | </script>

```

Archivo csv utilizado (datos_fab.csv)

	Fabricas.Cantidad
1	Baltimore,159
2	General Mills,136
3	Galletas Guillon,134
4	Vinicola,133
5	Badia Spices,124
6	Induveca,123
7	SIGMA Alimentos,123
8	Angel Camacho Alim.,120
9	KRAFT FOODS,114
10	Corpora Tresmontes,113

Anexo 6

Código para el gráfico de pastel en 3D elaborado en Google chart, en el que muestra el porcentaje de productos vigentes y vencidos.

```
1- <html>
2-   <head>
3-     <script type="text/javascript" src="https://www.google.com/jsapi"></script>
4-     <script type="text/javascript">
5-       google.load("visualization", "1", {packages:["corechart"]});
6-       google.setOnLoadCallback(drawChart);
7-       function drawChart() {
8-         var data = google.visualization.arrayToDataTable([
9-           ['Estado', 'Cantidades'],
10-          ['Vigentes', 12900],
11-          ['Vencidos', 19990]
12-        ]]);
13-
14-        var options = {
15-          title: 'Porcentaje del estado de los productos registrados',
16-          is3D: true,
17-        };
18-
19-        var chart = new google.visualization.PieChart(document.getElementById('piechart_3d'));
20-        chart.draw(data, options);
21-      }
22-    </script>
23-  </head>
24-  <body>
25-    <div id="piechart_3d" style="width: 900px; height: 500px;"></div>
26-  </body>
27- </html>
```

Anexo 7

Código para el gráfico de barras verticales en Google chart, en el que muestra la cantidad de productos próximo a vencerse.

```

1- <html>
2- <head>
3   <script type="text/javascript" src="https://www.google.com/jsapi"></script>
4- <script type="text/javascript">
5     google.load("visualization", "1.1", {packages:["bar"]});
6     google.setOnLoadCallback(drawChart);
7     function drawChart() {
8-      var data = google.visualization.arrayToDataTable([
9         ['Año', 'Noviembre', 'Diciembre', 'Enero', 'Marzo'],
10        ['2015', 42,37, 0,0],
11        ['2016', 0, 0, 38,135],
12      ]]);
13
14-     var options = {
15-       chart: {
16         title: 'Productos próximo a vencerse',
17         subtitle: 'Desde noviembre 2015 hasta marzo 2016',
18       }
19     };
20
21     var chart = new google.charts.Bar(document.getElementById('columnchart_material'));
22
23     chart.draw(data, options);
24   }
25 </script>
26 </head>
27- <body>
28   <div id="columnchart_material" style="width: 600px; height: 400px;"></div>
29 </body>
30 </html>

```


Anexo 8

Encuesta de investigación “Big Data, análisis de datos en la nube”

La siguiente encuesta es para explorar los conocimientos referentes a *Big Data*, lo cual será de mucha ayuda para la investigación institucional de la escuela de Informática y Ciencias Aplicadas de la Universidad Tecnológica.

1. ¿Ha escuchado hablar de *Big Data*?

- Sí
- No

2. ¿Qué tanto sabe de datos en la nube?

- Bastante
- Mucho
- Poco
- Nada

3. ¿Ha escuchado hablar de bases de datos que administran otras bases de datos?

- Sí
- No

4. ¿Qué tanto ha escuchado hablar de hadoop?

- Bastante
- Mucho
- Poco
- Nada

5. De las siguientes herramientas *Big Data* ¿cuáles conoce?

- MongoDB
- Hadoop
- R
- Clips
- D3.js
-

- Google Chart
- Jqplot
- Weka
- Ninguna

6. ¿Estaría interesado en conocer herramientas para el análisis de datos en la nube?

- Completamente seguro
- No estoy seguro
- De ninguna manera
- Probablemente

7. ¿Considera que sería de mucha utilidad el uso de las herramientas *Big Data* en su empresa?

- Bastante
- Mucho
- Poco
- Nada

8. ¿Sabía usted que *Big Data* ha ayudado a resolver problemas de marketing, ventas, elecciones presidenciales, diagnóstico de enfermedades, relaciones humanas, meteorología y muchas más?

- Bastante
- Mucho
- Poco
- Nada

9. ¿Ha escuchado el termino *dataset* público?

- Si
- No

10. ¿Qué tipo de base de datos conoce o ha utilizado?

- Estructurados
- Semi estructurados
- No estructurados
- Ninguna

11. ¿Qué tipo de formato de archivos conoce?

- xls
- xml
- json
- csv
- dat
- data
- arff

12. Según sus conocimientos, ¿cuánto puede llegar a medir una base de datos?

- Megabyte
- Gigabyte
- Terabyte
- Pettabyte
- Exabyte
- Zettabyte
- Yottabyte

Anexo 9

Resultados de la encuesta

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
1	Si	Mucho	No	Poco	D3.js, Jqplot	Completamente seguro	Mucho	Poco	No	Estructurados	xls, xml, dat	Gigabyte
2	No	Mucho	Si	Poco	R, D3.js	No estoy seguro	Bastante	Mucho	No	No estructurados	xml, dat	Terabyte
3	Si	Poco	No	Nada	Ninguna	Completamente seguro	Mucho	Nada	No	Estructurados	xls, xml, dat, data	Yottabyte
4	Si	Bastante	Si	Bastante	MongoDb, Hadoop, R, Clips, D3.js	Completamente seguro	Bastante	Bastante	Si	Estructurados	xls, xml, json	Zettabyte
5	Si	Poco	Si	Poco	MongoDb, D3.js, Google Chart	Completamente seguro	Bastante	Bastante	Si	Estructurados	xls, xml, json, csv, dat, data, arff	Terabyte
6	No	Nada	No	Nada	Ninguna	Completamente seguro	Mucho	Nada	No	Estructurados	xls, xml, csv, dat, data	Terabyte
7	Si	Poco	Si	Poco	Hadoop	Completamente seguro	Bastante	Bastante	No	Estructurados	xls, csv	Pettabyte
8	Si	Mucho	Si	Nada	MongoDb	Completamente seguro	Bastante	Bastante	No	Ninguna	xls, xml, csv, dat, data	Gigabyte
9	Si	Poco	No	Nada	MongoDb, Google Chart	Probablemente	Poco	Nada	No	Estructurados	xls, xml, json, csv, data	Gigabyte
10	Si	Poco	No	Poco	MongoDb, Hadoop	Completamente seguro	Mucho	Bastante	No	Estructurados	xls, xml, json, csv	Yottabyte
11	Si	Mucho	Si	Nada	MongoDb	Completamente seguro	Mucho	Mucho	No	Estructurados		Exabyte
12	Si	Poco	Si	Nada	Ninguna	Completamente seguro	Poco	Poco	No	Estructurados	xls, xml, json, csv, dat, data	Gigabyte

13	Si	Mucho	No	Poco	MongoDb, Hadoop	Completamente seguro	Mucho	Nada	No	Estructurados	xls, xml, json, csv, dat	Terabyte
14	Si	Mucho	Si	Poco	Google Chart	No estoy seguro	Mucho	Poco	No	Ninguna	xls, xml, dat	Petabyte
15	Si	Poco	Si	Poco	Hadoop	Completamente seguro	Bastante	Poco	No	Estructurados	xls, xml, csv, dat, data	Yottabyte
16	Si	Poco	Si	Nada	Ninguna	Probablemente	Bastante	Poco	No	Estructurados	xls, xml, csv, dat	Terabyte
17	Si	Mucho	Si	Mucho	Hadoop, Clips, Google Chart	Completamente seguro	Mucho	Mucho	Si	Estructurados	xls, xml, json, csv, dat	Gigabyte
18	Si	Mucho	Si	Poco	MongoDb, Hadoop, Google Chart, Weka	Completamente seguro	Mucho	Mucho	Si	Estructurados	xls, xml, json, csv, dat	Terabyte
19	Si	Poco	Si	Nada	Google Chart	Completamente seguro	Bastante	Bastante	No	Estructurados	xls, xml, json, csv, dat, data	Yottabyte
20	Si	Poco	No	Nada	Clips, Google Chart	Completamente seguro	Poco	Poco	No	Ninguna	xls, xml, csv	Terabyte
21	Si	Mucho	No	Nada	Ninguna	Probablemente	Mucho	Poco	No	No estructurados	xls, xml, csv	Terabyte
22	Si	Poco	Si	Poco	Hadoop, Google Chart	Probablemente	Mucho	Poco	Si	Estructurados	xls, xml, csv, dat, data	Terabyte
23	Si	Poco	Si	Poco	Ninguna	Probablemente	Mucho	Poco	No	Semi estructurados	xls, xml, csv, dat, data	Terabyte
24	Si	Mucho	Si	Poco	MongoDb, R, Weka	Probablemente	Mucho	Bastante	Si	Estructurados	xls, xml, json, csv, dat	Terabyte
25	Si	Mucho	Si	Poco	MongoDb,	Probablemente	Mucho	Bastante	Si	Estructurados	xls, xml, json, csv, dat	Terabyte

26	Si	Poco	Si	Poco	R, Weka	Completamente seguro	Bastante	Bastante	Bastante	Bastante	Bastante	No	Estructurados	csv, dat xls, xml, csv, data	Terabyte				
27	Si	Poco	Si	Poco	Google Chart	Completamente seguro	Nada	Mucho	Mucho	Mucho	Mucho	Si	Estructurados	xls, xml, csv	Terabyte				
28	Si	Poco	Si	Nada	Ninguna	Completamente seguro	Nada	Nada	Mucho	Mucho	Mucho	No	No estructurados	xls, dat	Gigabyte				
29	Si	Poco	No	Nada	MongoDb	Probablemente	Nada	Poco	Poco	Poco	Poco	No	Estructurados	xls, xml, json, csv	Exabyte				
30	Si	Mucho	Si	Poco	MongoDb	Completamente seguro	Poco	Bastante	Bastante	Bastante	Mucho	Si	Estructurados	xls, xml, json, csv, dat, data	Terabyte				
31	No	Poco	No	Nada	Ninguna	Probablemente	Nada	Nada	Nada	Nada	Nada	No	Estructurados	dat	Gigabyte				
32	Si	Mucho	Si	Poco	MongoDb, Google Chart	Completamente seguro	Poco	Bastante	Bastante	Bastante	Bastante	Si	Estructurados	xls, xml, csv, dat, data	Petabyte				
33	Si	Mucho	Si	Poco	Hadoop, Google Chart	Completamente seguro	Poco	Mucho	Mucho	Mucho	Poco		Estructurados	xls, xml, dat, data	Yottabyte				
34	Si	Bastante	Si	Mucho	MongoDb	Completamente seguro	Mucho	Bastante	Bastante	Bastante	Bastante	Si	Estructurados	xls, xml, csv, dat, data	Zettabyte				
35	Si	Bastante	Si	Bastante	MongoDb, Hadoop, Google Chart, Weka	Completamente seguro	Bastante	Bastante	Bastante	Bastante	Bastante	Si	Estructurados	xls, xml, json, csv, dat, data	Petabyte				
36	Si	Bastante	Si	Bastante	Hadoop	Completamente seguro	Bastante	Bastante	Bastante	Bastante	Bastante	No	Estructurados	xls, xml, csv, dat, data	Petabyte				
37	Si	Mucho	Si	Poco	MongoDb, Hadoop	Completamente seguro	Poco	Mucho	Mucho	Mucho	Bastante	Si	Estructurados	xls, xml, csv, dat, data	Exabyte				
38	Si	Mucho	Si	Poco	Hadoop, Google Chart	Completamente seguro	Poco	Mucho	Mucho	Mucho	Poco	No	Estructurados	xls, xml, csv	Petabyte				

39	Si	Nada	Si	Nada	Ninguna	Completamente seguro	Poco	Poco	Si	Estructurados	xls, xml, json, csv, dat	Gigabyte
40	Si	Poco	No	Poco	R, Clips, Google Chart	Completamente seguro	Bastante	Nada	No	Estructurados	xls, xml, csv	Terabyte

RECOPIACIÓN DE LAS INVESTIGACIONES PUBLICADAS 2008-2015

PRODUCTO	COLECCIÓN
Curso de Derecho Penal Salvadoreño. Parte General Volumen III. Teoría de la ley penal. Miguel Alberto Trejo Escobar	Colección Jurídica
Etnografía de salvadoreños migrantes en Brentwood y Hempstead Nueva York Jorge Arturo Colorado Berríos	Facultad de Ciencias Sociales
Proyecto de registro y reconocimiento de sitios arqueológicos históricos de El Salvador (PAHES-UTEC) José Heriberto Erquicia Cruz	Facultad de Ciencias Sociales Escuela de Antropología
En defensa de la Patria. Historia del Conflicto Armado en El Salvador 1980-1992 General Humberto Corado Figueroa	
Las controversiales fichas de fincas salvadoreñas. Antecedentes, origen y final. José Luis Cabrera Arévalo	Vicerrectoría de Investigación y Proyección Social. Escuela de Antropología
Recopilación Investigativa. Tomo I	Vicerrectoría de Investigación y Proyección Social
Recopilación Investigativa. Tomo II	Vicerrectoría de Investigación y Proyección Social
Recopilación Investigativa. Tomo III	Vicerrectoría de Investigación y Proyección Social
Apuntes sobre Mercadeo moderno José A. Exprúa	
El Estado Constitucional Dr. Mario Antonio Solano Ramírez	Colección Jurídica
Las agrupaciones ilícitas como delincuencia organizada Leonardo Ramírez Murcia	Colección Jurídica
La mujer dormida. Novela corta Eduardo Badía Serra	Colección Literaria
Koot. Revista de museología No. 1	Museo Universitario de Antropología
De la ilusión al desencanto. Reforma económica en El Salvador 1989-2009 Juan Héctor Vidal	Colección Ciencias Sociales
Casa Blanca Chalchuapa, El Salvador. Excavación en la trinchera 4N Nobuyuki Ito	Vicerrectoría de Investigación y Proyección Social. Escuela de Antropología

Recopilación Investigativa 2009. Tomo 1	Vicerrectoría de Investigación y Proyección Social
Recopilación Investigativa 2009. Tomo 2	Vicerrectoría de Investigación y Proyección Social
Recopilación Investigativa 2009. Tomo 3	Vicerrectoría de Investigación y Proyección Social
El nuevo proceso civil y mercantil salvadoreño	Colección Jurídica
Koot. Revista de museología No. 2	Museo Universitario de Antropología
Discursos en el tiempo para graduados y otros temas educativos José Adolfo Araujo Romagoza	
Recopilación Investigativa 2010	Vicerrectoría de Investigación
Recopilación Investigativa 2010	Vicerrectoría de Investigación
Recopilación Investigativa 2010	Vicerrectoría de Investigación
Decisiones Dr. Jorge Bustamente	
Compendio Gramatical. José Braulio Galdámez	
Foro económico: El Salvador 2011	Colección Ciencias Sociales
La violencia social delincencial asociada a la salud mental en los salvadoreños. Investigación Dr. José Ricardo Gutiérrez Quintanilla	Vicerrectoría de Investigación
Recopilación de investigación 2011 Tomo I	Vicerrectoría de Investigación
Recopilación de investigación 2011 Tomo II	Vicerrectoría de Investigación
Recopilación de investigación 2011 Tomo III	Vicerrectoría de Investigación
Programa psicopreventivo de educación para la vida efectividad en adolescentes UTEC-PGR Ana Sandra Aguilar de Mendoza- Milton Alexander Portillo	Vicerrectoría de Investigación
El lenguaje delincencial en El Salvador. Braulio Galdámez	
Medicina tradicional entre los indígenas de Izalco, Sonsonate, El Salvador Beatriz Castillo	Colección Ciencias Sociales
Contenido y proyección del anteproyecto de constitución política de 1950. Dr. Alfredo Martínez Moreno	Colección Jurídica
Revista Koot No 3 Dr. Ramón Rivas	Museo Universitario de Antropología

<p>Causas de la participación del clero salvadoreño en el movimiento emancipador del 5 de noviembre de 1811 en El Salvador y la postura de las autoridades eclesiales del Vaticano ante dicha participación. Claudia Rivera Navarrete</p>	<p>Colección Investigaciones</p>
<p>Estudio Histórico proceso de independencia: 1811-1823 Tomo II Dr. José Melgar Brizuela</p>	<p>Colección Investigaciones</p>
<p>El Salvador insurgente 1811-1821 Centroamérica. Tomo III César A. Ramírez A.</p>	<p>Colección Investigaciones</p>
<p>Antropología en El Salvador. Recorrido histórico y descriptivo Dr. Ramón Rivas</p>	<p>Colección Ciencias Sociales</p>
<p>Representatividad y pueblo en las revueltas de principios del siglo XIX en las colonias hispanoamericanas Héctor Raúl Grenni Montiel.</p>	<p>Colección Investigaciones</p>
<p>Guía básica para la exportación de la Flor de Loroco, desde El Salvador hacia España, a través de las escuelas de hostelería del País Vasco. Álvaro Fernández Pérez</p>	<p>Colección Investigaciones</p>
<p>La regulación jurídico-penal de la trata de personas especial referencia a El Salvador y España Hazel Jasmin Bolaños Vásquez</p>	<p>Colección Investigaciones</p>
<p>Infancia y adolescencia en la prensa escrita, radio y televisión salvadoreña Camila Calles Minero Morena Azucena Mayorga</p>	<p>Colección Investigaciones</p>
<p>Participación científica de las mujeres en El Salvador Primera aproximación Camila Calles Minero</p>	<p>Colección Investigaciones</p>
<p>Mejores prácticas en preparación de alimentos en la micro y la pequeña empresa José Remberto Miranda Mejía</p>	<p>Colección Investigaciones No. 29</p>
<p>Evaluación de factores psicosociales de riesgo y de protección de violencia juvenil en El Salvador José Ricardo Gutiérrez Quintanilla</p>	<p>Colección Investigaciones No. 30</p>
<p>Historias, patrimonios e identidades en el municipio de Huizúcar, La Libertad, El Salvador José Heriberto Erquicia Cruz Martha Marielba Herrera Reina Ariana Ninel Pleitez Quiñónez</p>	<p>Colección Investigaciones No. 31</p>
<p>Condiciones socioeconómicas de preparación para la PAES de los estudiantes de Educación Media Saúl Campos Morán Paola María Navarrete</p>	<p>Colección Investigaciones No. 32</p>

<p>Inventario de las capacidades turísticas del municipio de Chilitupán, departamento de La Libertad Lisette Cristalina Canales de Ramírez Carlos Jonatán Chávez Mejía Mario Antonio Aguilar Flores</p>	<p>Colección Investigaciones No. 33</p>
<p>Delitos relacionados con la pornografía en personas menores de 18 años. Especial referencia a las tecnologías de información y la comunicación como medio comisivo Hazel Jasmin Bolaños Vásquez Miguel Ángel Boldova Pasamar Carlos Fuertes Iglesias</p>	<p>Colección Investigaciones No. 34</p>
<p>Condiciones culturales de los estudiantes de educación media para el aprendizaje del idioma inglés Saúl Campos Morán Paola María Navarrete Julio Anibal Blanco</p>	<p>Colección Investigaciones No. 35</p>
<p>Valoración económica del recurso hídrico como un servicio ambiental de las zonas de recarga en las subcuencas del río Acelhuate José Ricardo Calles</p>	<p>Colección Investigaciones No. 36</p>
<p>Migración forzada y violencia criminal: Una aproximación teórico-práctica en el contexto actual Elsa Ramos</p>	<p>Colección Investigaciones No. 37</p>
<p>La prevención del maltrato en la escuela. Experiencia de un programa entre alumnos de educación media Ana Sandra Aguilar de Mendoza José Manuel Andreu Rodríguez María Elena Peña Fernández</p>	<p>Colección Investigaciones No. 38</p>
<p>Percepción del derecho a la alimentación en El Salvador. Perspectiva desde la biotecnología Carolina Lucero Morán Jeremías Ezequiel Yanes Densy Samuel Trejo Quintana</p>	<p>Colección Investigaciones No. 39</p>
<p>Publicidad y violencia de género en El Salvador Camila Calles Minero Francisca Guerrero Morena L. Azucena Hazel Bolaños</p>	<p>Colección Investigaciones No. 40</p>
<p>El domo el güegüecho y la evolución volcánica. San Pedro Perulapán (departamento de Cuscatlán), El Salvador. Primer informe Walter Hernández Guillermo E. Alvarado Brian Jicha Luis Mixco</p>	<p>Colección Investigaciones No. 41</p>

<p>Imaginario colectivo, movimientos juveniles y cultura ciudadana juvenil en El Salvador Saúl Campos Morán Paola María Navarrete Carlos Felipe Osegueda</p>	<p>Colección Investigaciones No. 42</p>
<p>Evaluación del sistema integrado de escuela inclusiva de tiempo pleno implementado por el Ministerio de Educación de El Salvador (Estudio en las comunidades educativas del municipio de Zaragoza del departamento de La Libertad) Mercedes Carolina Pinto Benítez Julio Anibal Blanco Escobar Guillermo Alberto Cortez Arévalo Wilfredo Alfonso Marroquín Jiménez Luis Horaldo Romero Martínez</p>	<p>Colección Investigaciones No. 43</p>
<p>Estudio de buenas prácticas en clínica de psicología. Caso Utec Edgardo Chacón Andrade Sandra Beatriz de Hasbún Claudia Membreño Chacón</p>	<p>Colección Investigaciones No. 44</p>
<p>Aplicación de una función dosis-respuesta para determinar los costos sociales de la contaminación hídrica en la microcuenca del río Las Cañas, San Salvador, El Salvador José Ricardo Calles Hernández</p>	<p>Colección Investigaciones No. 45</p>
<p>Aplicación de buenas prácticas de negocio (pequeña y mediana empresa de los municipios de San Salvador, Santa Tecla y Soyapango en El Salvador) Vilma Elena Flores de Ávila Blanca Ruth Gálvez Rivas Rosa Patricia Vásquez de Alfaro</p>	<p>Colección Investigaciones No. 46</p>
<p>Modelo de reactivación y desarrollo para cascos urbanos Coralía Rosalía Muñoz Márquez</p>	<p>Colección Investigaciones No. 48</p>
<p>Historia, patrimonio e identidades en el municipio de Comasagua, La Libertad, El Salvador José Heriberto Erquicia Cruz Martha Marielba Herrera Reina</p>	<p>Colección Investigaciones No. 49</p>
<p>El derecho humano al agua en El Salvador y su impacto en el sistema hídrico Sandra Elizabeth Majano Carolina Lucero Morán Dagoberto Arévalo Herrera</p>	<p>Colección Investigaciones No. 50</p>

<p>El contexto familiar asociado al comportamiento agresivo en adolescentes de San Salvador José Ricardo Gutiérrez Quintanilla Delmi García Díaz María Elisabet Campos Tomasino,</p>	<p>Colección Investigaciones No. 52</p>
<p>Análisis del tratamiento actual de las lámparas fluorescentes, nivel de contaminantes y disposición final José Remberto Miranda Mejía Samuel Martínez Gómez John Figerald Kenedy Hernández Miranda</p>	<p>Colección Investigaciones No. 53</p>
<p>Niñas, niños, adolescentes y mujeres en la ruta del migrante</p>	<p>Colección Investigaciones No. 54</p>
<p>Práctica de prevención del abuso sexual a través del funcionamiento familiar Ana Sandra Aguilar de Mendoza Ivett Idayary Camacho Lazo José Manuel Andreu María Elena de la Peña Fernández</p>	<p>Colección Investigaciones No. 55</p>
<p>Desplazamiento interno forzado y su relación con la migración internacional Elsa Ramos</p>	<p>Colección Investigaciones No. 56</p>
<p>Monografía cultural y socioeconómica del cantón Los Planes de Renderos Saúl Enrique Campos Morán Paola María Navarrete Gálvez Carlos Felipe Osegueda Osegueda Julio Anibal Blanco Escobar Melissa Regina Campos Solórzano</p>	<p>Colección Investigaciones No. 57</p>
<p>Modelo de vivienda urbana sostenible Coralía Rosalía Muñoz Márquez</p>	<p>Colección Investigaciones No. 58</p>

RECOPIACIÓN DE COLECCIONES “CUADERNILLOS” 2008-2014

TITULO	COLECCIÓN
El método en la investigación. Breve historia del derecho internacional humanitario desde el mundo antiguo hasta el tratado de Utrecht Colección de Derecho No. 1 Dr. Jaime López Nuila Lic. Aldonov Frankeko Álvarez Ferrufino	Colección de Derecho
Modo de proceder en el recurso de casación en materias: civiles, mercantiles y de familia Colección de Derecho No. 2 Dr. Guillermo Machón Rivera	Colección de Derecho
La administración de justicia y la elección de los magistrados de la corte suprema de justicia luego de los acuerdos de paz Colección de Derecho No. 3 Lic. Rene Edgardo Vargas Valdez	Colección de Derecho
La Proyección Social una propuesta práctica Colección Cuaderno No. 1 Lic. Carlos Reynaldo López Nuila	Rectoría Adjunta
Hacia una nueva cultura jurídica en materia procesal civil y mercantil Colección de Derecho No. 4 Lic. Juan Carlos Ramírez Cienfuegos	Colección de Derecho
La educación: ¿derecho natural o garantía fundamental? Dr. Jaime López Nuila	Colección de Derecho
Realidad Nacional 1 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez	Colección Ciencias Sociales
Realidad Nacional 2 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez	Colección Ciencias Sociales
Realidad Nacional 3 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez	Colección Ciencias Sociales
Realidad Nacional 4 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez	Colección Ciencias Sociales
Realidad Nacional 5 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez	Colección Ciencias Sociales

<p>Realidad Nacional 6 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez</p>	<p>Colección Ciencias Sociales</p>
<p>Realidad Nacional 7 Lic. Rene Edgardo Vargas Valdez Lic. Aldonov Frankeko Álvarez</p>	<p>Colección Ciencias Sociales</p>
<p>Obstáculos para una investigación social orientada al desarrollo Colección de Investigaciones Dr. José Padrón Guillen</p>	<p>Vicerrectoría de Investigación</p>
<p>Estructura familia y conducta antisocial de los estudiantes en Educación Media Colección de Investigaciones No. 2 Luis Fernando Orantes Salazar</p>	<p>Vicerrectoría de Investigación</p>
<p>Prevalencia de alteraciones afectivas: depresión y ansiedad en la población salvadoreña Colección de Investigaciones No. 3 José Ricardo Gutiérrez</p>	<p>Vicerrectoría de Investigación</p>
<p>Violación de derechos ante la discriminación de género. Enfoque social Colección de Investigaciones No. 4 Elsa Ramos</p>	<p>Vicerrectoría de Investigación</p>
<p>Diseño de un modelo de vivienda bioclimática y sostenible. Fase I Colección de Investigaciones No. 5 Ana Cristina Vidal Vidales</p>	<p>Vicerrectoría de Investigación</p>
<p>Importancia de Iso indicadores y la medición del quehacer científico Colección de Investigaciones No. 6 Noris López de Castaneda</p>	<p>Vicerrectoría de Investigación</p>
<p>Situación de la educación superior en El Salvador Colección de Investigaciones No. 1 Lic. Carlos Reynaldo López Nuila</p>	<p>Vicerrectoría de Investigación</p>
<p>La violencia social delincuencial asociada a la salud mental. Colección de Investigaciones No. 7. Lic. Ricardo Gutiérrez Quintanilla</p>	<p>Vicerrectoría de Investigación</p>
<p>Estado de adaptación integral del estudiante de educación media de El Salvador Colección de Investigaciones No. 8 Luis Fernando Orantes</p>	<p>Vicerrectoría de Investigación</p>
<p>Aproximación etnográfica al culto popular del Hermano Macario en Izalco, Sonsonate, El Salvador. Colección de Investigaciones No. 9 José Heriberto Erquicia Cruz</p>	<p>Vicerrectoría de Investigación</p>

<p>La televisión como generadora de pautas de conducta en los jóvenes salvadoreños Colección de Investigaciones No. 10 Edith Ruth Vaquerano de Portillo Domingo Orlando Alfaro Alfaro</p>	Vicerrectoría de Investigación
<p>Violencia en las franjas infantiles de la televisión salvadoreña y canales infantiles de cable Colección de Investigaciones No. 11 Camila Calles Minero Morena Azucena Mayorga Tania Pineda</p>	Vicerrectoría de Investigación
<p>Factores que influyen en los estudiantes y que contribuyeron a determinar los resultados de la PAES 2011 Colección de Investigaciones No. 12 Saúl Campos Blanca Ruth Orantes</p>	Vicerrectoría de Investigación
<p>Responsabilidad legal en el manejo y disposición de desechos sólidos en hospitales de El Salvador Colección de Investigaciones No. 13 Carolina Lucero Morán</p>	Vicerrectoría de Investigación
<p>Obrajes de añil coloniales de los departamentos de San Vicente y La Paz, El Salvador Colección de Investigaciones No. 14 José Heriberto Erquicia Cruz</p>	Vicerrectoría de Investigación
<p>San Benito de Palermo: elementos afrodecendientes en la religiosidad popular en El Salvador. Colección de Investigaciones No. 16 José Heriberto Erquicia Cruz y Martha Marielba Herrera Reina</p>	Vicerrectoría de Investigación
<p>Formación ciudadana en jóvenes y su impacto en el proceso democrático de El Salvador Colección de Investigaciones No. 17 Saúl Campos</p>	Vicerrectoría de Investigación
<p>Turismo como estrategia de desarrollo local. Caso San Esteban Catarina. Colección de Investigaciones No. 18 Carolina Elizabeth Cerna, Larissa Guadalupe Martín y José Manuel Bonilla Alvarado</p>	Vicerrectoría de Investigación
<p>Conformación de clúster de turismo como prueba piloto en el municipio de Nahuizalco. Colección de Investigaciones No. 19 Blanca Ruth Galvez García, Rosa Patricia Vásquez de Alfaro, Juan Carlos Cerna Aguiñada y Oscar Armando Melgar.</p>	Vicerrectoría de Investigación
<p>Mujer y remesas: administración de las remesas. Colección de Investigaciones No. 15 Elsa Ramos</p>	Vicerrectoría de Investigación

<p>Estrategias pedagógicas implementadas para estudiantes de educación media Colección de Investigaciones No. 21 Ana Sandra Aguilar de Mendoza</p>	Vicerrectoría de Investigación
<p>Participación política y ciudadana de la mujer en El Salvador Colección de Investigaciones No. 20 Saúl Campos Morán</p>	Vicerrectoría de Investigación
<p>Estrategia de implantación de clúster de turismo en Nahuizalco (Propuesta de recorrido de las cuatro riquezas del municipio, como eje de desarrollo de la actividad turística) Colección de Investigaciones No. 22 Blanca Ruth Gálvez Rivas Rosa Patricia Vásquez de Alfaro Óscar Armando Melgar Nájera</p>	Vicerrectoría de Investigación
<p>Fomento del emprendedurismo a través de la capacitación y asesoría empresarial como apoyo al fortalecimiento del sector de la Mipyme del municipio de Nahuizalco en el departamento de Sonsonate. Diagnóstico de gestión. Colección de Investigaciones No. 23 Vilma Elena Flores de Ávila</p>	Vicerrectoría de Investigación
<p>Proyecto migraciones nahua-pipiles del postclásico en la Cordillera del Bálsamo Colección de Investigaciones No. 24 Marlon V. Escamilla William R. Fowler</p>	Vicerrectoría de Investigación
<p>Transnacionalización de la sociedad salvadoreña, producto de las migraciones Colección de Investigaciones No. 25 Elsa Ramos</p>	Vicerrectoría de Investigación
<p>Imaginario y discursos de la herencia afrodescendiente en San Alejo, La Unión, El Salvador Colección de Investigaciones No. 26 José Heriberto Erquicia Cruz Martha Marielba Herrera Reina Wolfgang Effenberger López</p>	Vicerrectoría de Investigación
<p>Metodología para la recuperación de espacios públicos Colección de Investigaciones No. 27 Ana Cristina Vidal Vidales Julio César Martínez Rivera</p>	Vicerrectoría de Investigación

Resumen de Hoja de Vida

Investigador principal

Datos Personales

- Nombre: Omar Otoniel Flores Cortez
- E-Mail: omar.flores@utec.edu.sv



Estudios Realizados

- Maestría: Master en docencia universitaria, Universidad Tecnológica de El Salvador
- Posgrado: Robótica y Programación, Universidad de La Rioja, España
- Posgrado: Entornos Virtuales de Aprendizaje, Universidad Tecnológica de El Salvador
- Educación Superior: Ingeniera Eléctrica, Universidad de El Salvador

Competencias

- Publicaciones: libro de texto Aprende Arduino – 2013 ISBN: 978-99961-0-346-9
- Áreas de conocimiento: docente en educación superior, investigador, programación de sistemas embebidos, robótica y domótica educativa.

Coinvestigador

Datos personales

- Nombre: German Antonio Rosa Castellanos
- E-Mail: germanr.castellanos@gmail.com
- Telefono: 2275-8992 /



Estudios realizados

- Postgrado: Entornos Virtuales de Aprendizaje, Universidad Tecnológica de El Salvador
- Educación superior: INGENIERO EN SISTEMAS Y COMPUTACION., Universidad Tecnológica de El Salvador

Competencias

- Áreas de conocimiento: - Modelado e impresión 3D.- Robótica Educativa.- Programación de PLC LOGO. - Programación de alto y bajo nivel (Leguaje C++, Ensamblador, Labview, Programación de Microcontroladores PIC).

Investigador principal

Datos personales

- Nombre: Carlos Antonio Aguirre
- E-mails: carlos99ayala@gmail.com, carlos.aguirre@utec.edu.sv



Estudios realizados

- Maestría: Máster en Computación con énfasis en Sistemas de Información, Instituto Tecnológico de Costa Rica
- Postgrado: Investigación e Innovación Científica, Universidad Centroamericana *José Simeón Cañas*
- Postgrado: Entornos Virtuales de Aprendizaje, Virtual Educa Argentina
- Educación superior: Licenciado en Administración de Empresas con especialidad en Computación, Universidad Tecnológica de El Salvador

Competencias

- Áreas de conocimiento: Administrador de proyectos, desarrollo de aplicaciones para la web, investigador, docente universitario

Coinvestigador

Datos personales

- Nombre: Marvin Elenilson Hernández Montoya
- E-Mail: marvin.hernandez@utec.edu.sv



Estudios realizados

- Maestría: Maestría en Educación Superior, Universidad Tecnológica de El Salvador
- Postgrado: Investigación e Innovación Científica, Universidad Centroamericana *José Simeón Cañas*
- Postgrado: Entornos Virtuales de Aprendizaje, Virtual Educa Argentina
- Educación superior: Licenciado en Administración de Empresas con especialidad en Computación, Universidad Tecnológica de El Salvador
- Postgrado: Investigación, Universidad de Alicante, España

Competencias

- Áreas de conocimiento: Administrador de proyectos, desarrollo de aplicaciones para la web, investigador, docente universitario

Investigador Principal

Datos Personales

Coinvestigador

- Nombre: Verónica Idalia Rosa Urrutia
- E-Mail: veronica.rosa@utec.edu.sv



Estudios Realizados

- Maestría: Master en docencia universitaria, Universidad Tecnológica de El Salvador
- Maestría: Master en Visual Analytics y *Big Data*
- Pre-especialidad: Comunicaciones y Seguridad de Redes de Datos Bajo Ambientes Windows y Linux
- Educación Superior: Ingeniera en Sistemas y Computación, Universidad Tecnológica de El Salvador

Competencias

- Áreas de conocimiento: Programador, analista, docente en educación superior, investigador.
-

Coinvestigador

Datos Personales

- Nombre: José Guillermo Rivera Pleitéz
- E-Mail: joseguillermo.rivera@gmail.com



Estudios Realizados

- Maestría: Master en Administración de Recursos Humanos, Universidad Evangélica
- Maestría: Master en Docencia Universitaria, Universidad Evangélica
- Diplomado en Actualización de Metodología en Investigación Científica, Universidad Francisco Gavidia
- Educación Superior: Ingeniero en Sistemas y Computación, Universidad Tecnológica de El Salvador

Competencias

- Áreas de conocimiento: Programador, analista, docente en educación superior, investigador

Internet de las cosas: Diseño e implementación de prototipo electrónico para el monitoreo vía internet de sistemas de generación fotovoltaico

Basados en la premisa *internet de las cosas* (IoT, por sus siglas en inglés), a través de esta investigación se diseñó un sistema electrónico que permita a un panel solar publicar en una página web, accesible desde cualquier punto de internet, aspectos de funcionamiento y generación eléctrica de voltaje. Se implementó un sistema embebido electrónico basado en el microcontrolador ATmega32u4, usando una plataforma de desarrollo Arduino modelo YUN; junto con esta se utilizó un sensor que mide el voltaje generado por el panel solar. Es esta información la que se publica automáticamente en un servicio web, o nube, para ser accesible por medio de un panel publicado en un sitio web, para ser accedido desde cualquier punto de internet. El resultado de esta investigación fue el prototipo electrónico junto con la programación en la nube, que permite que el panel solar publique automáticamente su voltaje generado.

Implementación de un entorno de aprendizaje virtual integrando herramientas de *E-learning* y CMS

La educación es un pilar fundamental para las sociedades en vías de desarrollo, ya que por medio del conocimiento se obtienen los escenarios reales, que se convierten en la fuerza impulsora para la generación de innovación que, a su vez, tiene su origen en la investigación, por ello se estudió la plataforma de *E-learning Moodle* y el CMS *Joomla*, integrando la potencialidad de ambos a través de una extensión llamada *Joomdle*, con el objetivo de mostrar una nueva plataforma con mayor potencial, para el manejo de cursos virtuales, en un entorno más dinámico y amigable al usuario.

Big data, análisis de datos en la nube

El tema *Big Data* en El Salvador es novedoso, pero las empresas están con muchos deseos en incursionar en esta tecnología, aunque no saben cómo hacerlo, por lo que el objetivo de esta investigación fue elaborar una guía metodológica en la que se refleje el uso de herramientas Big Data para almacenar, procesar y analizar grandes cantidades de datos, con el fin de obtener conclusiones que puedan ayudar en la toma de decisiones de cualquier empresa, independientemente del rubro que manejen. Para esta investigación se usaron dos *dataset* que contienen información sobre registro de productos alimenticios y medicamentos; y con ellos se trabajó, desde cómo almacenarlos y procesarlos con *Hadoop*, para hacer consultas con *Hive* y también haciendo análisis estadísticos con el programa R, finalizando con la creación de visualizaciones, ya sea con *Google charts*, *Jqplot* o *D3.js*.

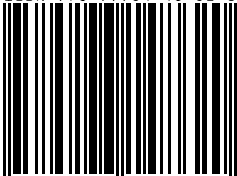
Vicerrectoría de Investigación y Proyección Social

Edificio Dr. José Adolfo Araujo Romagoza,

Calle Arce y 19.^a Avenida Sur No. 1045, 2.^o nivel, San Salvador, El Salvador, C. A.

Tel.: 2275-1011 • vicerrectoriadeinvestigacion@utec.edu.sv

ISBN 978-99961-48-62-0



9 789996 148620 >

