

ARTÍCULO DE INVESTIGACIÓN

DISEÑO DE MODELO TECNOLÓGICO PARA EL USO DE BIG DATA EN EL ANÁLISIS Y VISUALIZACIÓN DE INFORMACIÓN PARA LA PEQUEÑA Y MEDIANA EMPRESA

Guillermo Rivera Pleitez

Guillermo.rivera@uees.edu.sv

Yancy Staffany Ventura Aguilar

yancy.ventura@uees.edu.sv

Saúl Antonio Cornejo Hernández

saul.cornejo@uees.edu.sv

Facultad de Ingenierías, Universidad Evangélica de El Salvador

Recibido 23/05/17

Aceptado 5/09/17

RESUMEN

El objetivo de esta investigación fue diseñar un modelo tecnológico que permitiera la incorporación sistemática de las PYMES al comercio electrónico mediante la adecuación y actualización de su tecnología informática en uso. Se presenta la base documental técnica de los elementos involucrados en el área de Big Data que deberán considerarse para elaborar el modelo. Se realizó un estudio exploratorio descriptivo mediante la aplicación de encuestas a PYMES de servicios del departamento de San Salvador para identificar su infraestructura tecnológica y analizar cómo esta puede facilitar su incorporación al comercio electrónico. Se elaboró una guía metodológica en la que se reflejó el uso de herramientas Big Data para almacenar, procesar y analizar grandes cantidades de datos, a fin de obtener conclusiones que ayuden a cualquier empresa en la toma de decisiones, independientemente del rubro en el que operen. Con los dataset¹ públicos utilizados, que son datos no estructurados, se muestra a las PYMES la forma dinámica de obtener información útil de datos de clientes, preferencias de productos, regiones de mayor venta, presentación de productos que los clientes prefieren, categorías de productos y cualquier otro tipo de información que la empresa necesite que esté contenida en la base de datos. Con esta información, la empresa puede tomar decisiones, conocer preferencia de los clientes, clasificar por categorías los productos, proyectar ventas y conocer la tendencia del mercado.

Palabras clave: PYMES, Comercio electrónico, Big Data, Hadoop, Hive, dataset, macrodatos, El Salvador.

¹ Daset es una representación de datos residentes en memoria que proporciona un modelo de programación relacional coherente independientemente del origen de datos que contiene. <https://olgeferr.es>.

DESIGN OF TECHNOLOGICAL MODEL FOR THE USE OF BIG DATA IN THE ANALYSIS AND VISUALIZATION OF INFORMATION FOR SMALL AND MEDIUM-SIZED ENTERPRISE

ABSTRACT

The purpose of this research was to develop a technological model that would allow the systematic incorporation of SMEs in e-commerce through the suitability and update of their technology in use. The evidence-based technique is presented with the elements involved in the areas of Big Data that should be considered to develop the model. An exploratory and descriptive study was conducted through the application of surveys to SMEs services of the City of San Salvador to identify their technological infrastructure and to analyze how this can facilitate their incorporation into the e-commerce. A methodological guide was developed in which was reflected the use of tools of Big Data to store, process and analyze large amounts of data in order to obtain conclusions that help any company in the decision making, regardless of the industry in which they operate. With the use of this public dataset, which is unstructured data, is shown to the SMEs the dynamic form to get useful information from customer data, product preferences, greatest sales areas, products categories that customers prefer, categories of products and any other type of information that the company needs that is contained in the database.

Keywords: SMEs, e-commerce, Big Data, Hadoop, Hive, dataset, macro data, El Salvador

INTRODUCCIÓN

Big Data es un término que hace referencia a una cantidad de datos que supera la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. Los Big Data han sido muy usados en el medio de la informática y de las grandes empresas por su capacidad para visualizar la gran cantidad de información que se maneja hoy en día. El volumen de los datos masivos crece constantemente. En 2012, su tamaño se estimaba entre una docena de terabytes hasta varios petabytes de datos en un único conjunto de datos. 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 (Rosa y Rivera Pleitez, 2016).

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). Debido a esto, la realidad actual demanda trabajar con una gran cantidad de datos, pero la mayoría de empresas no saben cómo hacerlo.

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

El propósito de este estudio es diagnosticar la infraestructura tecnológica de las PYMES para elaborar un modelo tecnológico que les permita implementar herramientas Big Data. Las herramientas se probaron en dataset públicos con datos relacionados con el manejo operativo de las PYMES.

METODOLOGÍA

Tomando en cuenta que la investigación es de tipo innovadora con un enfoque experimental, la propuesta fue diseñada por medio de la implementación de la tecnología Open Source para cumplir con los objetivos, de tal forma que se elaboró un modelo metodológico que facilite la implementación de las herramientas Big Data en las PYMES.

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

Durante los meses de mayo a octubre de 2016 se realizaron las pruebas experimentales con dataset públicos almacenados en la nube y se diseñó y creó un modelo metodológico capaz de administrar grandes volúmenes de información, como los procesos de almacenamiento, administración de los datos e interoperabilidad, así como las técnicas y herramientas para llevar a cabo los análisis de datos y la visualización de información para la toma de decisiones.

Los datos fueron obtenidos de un instrumento de evaluación de diagnósticos inicial administrado a un grupo muestral de 72 PYMES para conocer de primera mano la situación actual de las empresas en cuanto a infraestructura, conocimientos y aplicación de herramientas Big Data. Durante la primera semana de enero de 2016 se envió a cada empresa la encuesta con herramienta. Luego de pasar las encuestas se efectuó, entre la segunda y la tercera semana de enero de 2016, la tabulación de resultados; posteriormente, estos resultados serían objeto de análisis para establecer conclusiones que describieran información importante para medir el grado de utilización y aprovechamiento de la tecnología Big Data por parte de las PYMES.

Resultados de la infraestructura tecnológica con la que cuentan las empresas entrevistadas:

Como se observa en la Figura 1, la porción mayor indica que un porcentaje significativo de empresas encuestadas cuentan con al menos un tipo de infraestructura dentro de su empresa, sin embargo, el porcentaje es variante entre los diferentes tipos de infraestructura tecnológica.

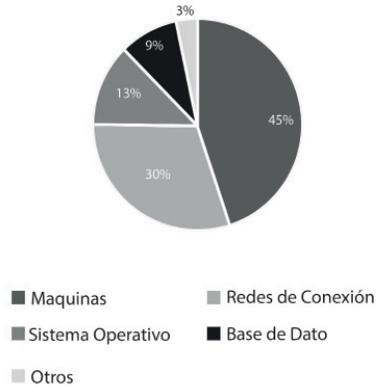


Figura 1. Infraestructura tecnológica que posee la empresa.

Fuente: Elaboración propia.

Con respecto a los medios digitales de comunicación con los que cuentan las empresas, como las redes sociales, comunicación interactiva y medios electrónicos, se tuvieron en mayoría porcentajes similares y en menor porcentaje de otros medios (Figura 2).



Figura 2. Medios digitales de comunicación en las empresas.

Fuente: Elaboración propia.

El siguiente gráfico nos permite saber las tecnologías o técnicas utilizadas por las empresas

encuestadas para dar a conocer productos nuevos o promociones a sus clientes (Figura 3).

En la actualidad, las redes sociales son comúnmente utilizadas a nivel mundial para promocionar y ofertar productos o servicios, como se muestra en el gráfico, logrando el primer lugar con un 52%, seguido de medios electrónicos, con un 21% sobre el resto de herramientas.

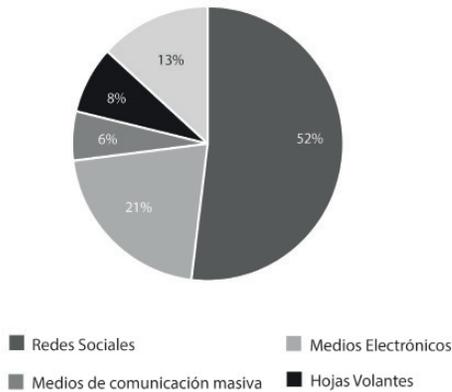


Figura 3. Herramientas de internet para dar a conocer los nuevos productos o promociones a los clientes existentes o potenciales.

Fuente: Elaboración propia.

En su mayoría, las empresas tomadas como muestras para este estudio consideran que es de gran beneficio realizar análisis de datos para el desarrollo y crecimiento de su entidad. A continuación, una gráfica que ejemplifica lo anterior con datos reales (Figura 4).

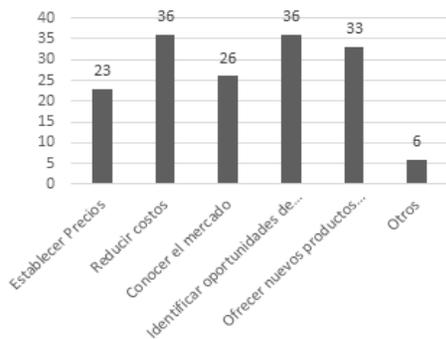


Figura 4. Factores que beneficiarían a la empresa si realizara análisis de datos.

Fuente: Elaboración propia.

El comercio electrónico consiste en realizar electrónicamente transacciones comerciales. Está basado en el tratamiento y transmisión electrónica de datos, incluidos texto, imágenes y vídeo. Por su parte, Guisado Moreno desde una perspectiva jusprivatista señala que se entiende por comercio electrónico «aquel que abarca las transacciones comerciales electrónicas, compraventa de bienes y prestación de servicios realizados entre empresarios, o bien entre empresarios y consumidores, a través de los soportes electrónicos proporcionados por las nuevas tecnologías de la información y la comunicación, básicamente Internet, así como también las negociaciones previas y posteriores estrechas y directamente relacionadas con aquellos contratos (ofertas contractuales, contraofertas, pago electrónico)» (Gerencia, 2014).

En el comercio electrónico B2C (Business to Consumer), el negocio va dirigido de las empresas al consumidor (Lozoya, 2013).

Las ventajas más destacables del comercio electrónico B2C son:

- Compras más cómodas y más rápida.
- Ofertas y precios siempre actualizados.
- Centro de atención al cliente integrado en la web.

Los inconvenientes, como sucede en toda transacción, también existen. El consumidor debe prestar especial atención a la seguridad de compras (Lozoya, 2013).

Planteamiento de la propuesta tecnológica

Modelo de incorporación de las PYMES a la tecnología Big Data.

El modelo se construyó partiendo de los resultados de la investigación. Los elementos que conforman el modelo son:

- a. Recursos humanos
 - Capacitación
 - Soporte
- b. Tecnología
- c. Seguridad
- d. Marco legal
- e. Factibilidad económica
- f. Redes sociales
- g. Comercio electrónico

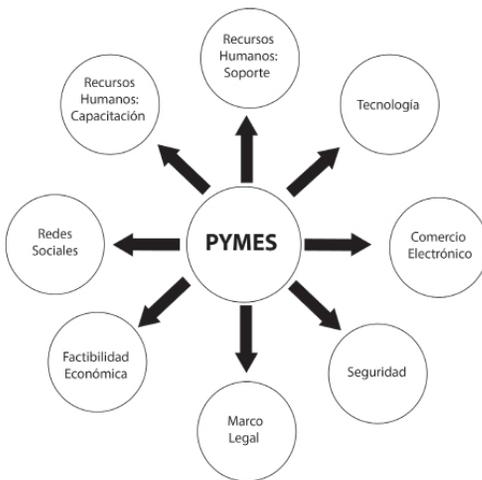


Figura 5: Modelo de incorporación de las PYMES a la Tecnología Big Data.

Fuente: Elaboración propia

Resultados del modelo metológico tecnológico para el uso de Big Data

Se presenta una serie de pasos que deben seguirse para la utilización de las herramientas de Big Data, comenzando con la instalación de los programas, pasando por el proceso de almacenamiento y procesamiento del Dataset seleccionado, para luego finalizar con la visualización de los resultados.

Debido a la gran notoriedad que está teniendo esta tendencia y parte de las nuevas tecnologías, cualquier persona, con conocimientos tecnológicos o sin ellos, se pregunta cómo se almacena toda la información que se genera en el mundo por medio de las redes sociales (Facebook, Twitter, Smartcities, Instagram) o cómo Google es capaz de manejar todas las transacciones que se hacen a diario. Pero no solo se queda aquí, ya que Big Data alcanza todos los ámbitos: bolsa, climatología, astronomía, marketing, 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 incrementa cada día, por lo que 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 aprovechada por cualquier empresa con el fin de ofrecer una mejor forma de brindar sus productos y servicios.

El aprovechamiento de tecnología del Big Data permite que las empresas conozcan más de cerca a sus clientes para prestarles 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: explorar los conocimientos que

se tienen sobre Big Data, el uso de las distintas herramientas para el procesamiento de los datos, los conocimientos de herramientas de visualización de grandes cantidades de datos, conocer las preferencias y los elementos necesarios que se pueden utilizar para mejorar los procesos en las empresas. Se realizaron pruebas con dataset públicos con datos relacionados con la operatividad de las PYMES utilizando herramientas Big Data.

DISCUSIÓN

El uso intensivo de los datos ha pasado a ser el petróleo de muchas compañías. El nuevo enfoque es almacenar cualquier dato, por irrelevante que pueda parecer, para su posterior análisis.

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

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).

Las PYMES están creciendo en grandes volúmenes de datos y deben prepararse para enfrentar el cambio tecnológico. Este crecimiento podría ser inhibido si las capacidades de seguridad de la tecnología no son combinadas con políticas y procesos diseñados para proteger la privacidad, tanto de las compañías como de los particulares (Ecointeligencia, 2013).

Con los resultados obtenidos en las gráficas se muestra la utilidad de las herramientas de visualización en la toma de decisiones de las empresas, no importando si los volúmenes de datos son extremadamente grandes y con datos sin formato. A las empresas les interesa ver los resultados con gráficos atractivos y exactos en

de datos de clientes, preferencias de productos, regiones de mayor venta, presentación de productos que los clientes prefieren, categorías de productos y otra información que la empresa necesite que esté contenida en la base de datos.

Con esta información, la PYMES puede tomar decisiones, conocer preferencia de los clientes, clasificar por categorías los productos, proyectar ventas y conocer la tendencia del mercado.

Las PYMES serán más competitivas en el mercado utilizando herramientas de última tecnología que les permitan competir aun con la gran industria.

Por lo tanto, el presente trabajo de investigación deja un modelo metodológico para que las PYMES puedan acceder al uso de herramientas tecnológicas Big Data para facilitar el almacenamiento, análisis y visualización de grandes volúmenes de información para la toma de decisiones.

CONCLUSIONES

- Se elaboró un procedimiento tecnológico que permitiera reflejar un buen uso de las herramientas Big Data, ya que son novedosas y no se conocen las herramientas que se utilizan para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos. Es por ello que, con las metodologías propuestas, se ha tratado de dar a conocer el funcionamiento de alguna de ellas, para motivar a las empresas a involucrarse en las nuevas tecnologías.
- Se identificaron los dataset públicos en la nube que contenían gran cantidad de registros, permitiéndonos procesar y analizar la información almacenada.
- Se mostraron herramientas para dar al uso de Hadoop para el almacenamiento y procesamiento de los datos para luego realizar consultas al estilo SQL por medio de la herramienta Hive. Esta herramienta involucra implícitamente el proceso Map Reduce, lo cual facilita aún más su uso.

- Se analizaron y se presentaron los resultados obtenidos con las herramientas de procesamiento y análisis. Se ha hecho uso de tres de ellas: Google Chart, Jqplot y D3.js; por supuesto, no es necesario trabajar con todas, sino que dependerá de las necesidades a representar y de las ventajas que ofrece cada herramienta, así como de la habilidad que se tenga en la programación con Java Script; sin embargo, como se ha mencionado, a la hora de trabajar con las herramientas de visualización existen ejemplos y galerías en la Web sobre cada una de ellas, las cuales se pueden reutilizar y adaptar a las necesidades de cada empresa u organización.
- Se diseñó la propuesta de un modelo apropiado para la pequeña y mediana empresa (PYMES), aplicando las herramientas Big Data para el análisis de datos.

RECOMENDACIONES

- El uso de la información proveniente de la PYMES analizadas con herramientas de Big Data y aplicando el modelo propuesto permitirá a la empresa conocer patrones conductuales de los potenciales clientes que le faciliten a las PYMES llegar a tomar decisiones efectivas respecto a los productos y servicios que le puedan llegar a ofrecer a sus clientes, según sus necesidades.
- Diagnosticar las necesidades de las PYMES para obtener los datos idóneos para el análisis de los dataset.
- Implementar herramientas actualizadas de administración de datos Big Data.
- Aplicar correctamente las herramientas de visualización de los datos.
- Se recomienda aplicar la propuesta tecnológica en una PYME para utilizar la información de su Base de Datos, para lo cual es necesario establecer un convenio entre la PYME y la UEES, a fin de garantizar la confidencialidad de la información.
- Una vez teniendo el convenio con la PYME, se podría tener la base necesaria para una siguiente investigación, en la cual se podría aplicar los sistemas de seguridad de Big Data.

AGRADECIMIENTOS

Los autores agradecen primeramente a Dios Todopoderoso por darnos la salud, sabiduría y destrezas necesarias para culminar con éxito este proyecto de investigación; a las autoridades de la Universidad Evangélica de El Salvador (UEES); al grupo de docentes y estudiantes de la Facultad de Ingenierías por su apoyo a la investigación.

REFERENCIAS BIBLIOGRÁFICAS

- Aguilar, L. J. (2013). Big Data, Análisis de los grandes volúmenes de datos. Edi. Alfaomega, 2013. Mexico.
- Bustillo, I. (2013). Business Intelligence Obtenido de Pentaho: <http://ignaciobustillo.focalrock.com/blog/blog/63-ique-es-pentaho>
- Gerencia, v. (2014). Funciones gerenciales. Recuperado el 2 de diciembre de 2015, de http://www.degerencia.com/tema/comercio_electronico
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google File System. In Proceedings of the 19th ACM Symposium on Operating Systems Principles. Recuperado el 16 de diciembre de 2015. Obtenido de: <http://static.googleusercontent.com/media/research.google.com/es//archive/gfs-sosp2003.pdf>
- International Business Machines Corp. (2012). Tipos de datos de Big Data, Web y redes sociales. Recuperado el 2014, de <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- Lozoya, J. (2013). Clase de comercio electrónico, B2E. Recuperado 25 de noviembre de 2014. Obtenido de: <http://suite101.net/articulo/clases-de-comercio-electronico-b2b-b2c-b2ab2e-c2c-c2g-b2g-a26589>
- Martínez, J. E. (2013). Desafío y Oportunidades de las Pymes Salvadoreñas. San Salvador: N/A.
- Rosa, V. I., Rivera, J. G., (2016) Big Data, Análisis de Datos en La Nube. El Salvador: Universidad Tecnológica de El Salvador.
- TestingSoft. (2014). Tipos de datos a explorar. Recuperado el 2014, de <http://www.evaluandosoftware.com/nota-3684-Que-es-el-Big-Data.html>.