



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**USO DE BASE DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA PARA EL
DISEÑO EFECTIVO DE MARKETING PROACTIVO**

Dina Magnolia Vásquez Antuche

Asesorado por el Ing. Sergio Arnaldo Méndez Aguilar

Guatemala, noviembre de 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**USO DE BASE DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA
PARA EL DISEÑO EFECTIVO DE MARKETING PROACTIVO**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR

DINA MAGNOLIA VÁSQUEZ ANTUCHE

ASESORADO POR EL ING. SERGIO ARNALDO MÉNDEZ AGUILAR

AL CONFERÍRSELE EL TÍTULO DE

INGENIERA EN CIENCIAS Y SISTEMAS

GUATEMALA, NOVIEMBRE DE 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Angel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Jurgen Andoni Ramírez Ramírez
VOCAL V	Br. Oscar Humberto Galicia Nuñez
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. César Augusto Fernández Cáceres
EXAMINADOR	Inga. Mirna Ivonne Aldana Larrazábal
EXAMINADOR	Ing. Luis Fernando Espino Barrios
SECRETARIA	Inga. Lesbia Magalí Herrera López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

USO DE BASE DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA PARA EL DISEÑO EFECTIVO DE MARKETING PROACTIVO

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha febrero de 2016.

Dina Magnolia Vásquez Antuche

Guatemala, 31 de octubre de 2016.

Ingeniero
Carlos Azurdía
Escuela de Ciencias y Sistemas
Facultad de Ingeniería

Respetable ingeniero Azurdía:

Por este medio le informo, como asesor del trabajo de graduación de la estudiante de la carrera de Ingeniería en Ciencias y Sistemas, Dina Magnolia Vásquez Antuche, carné 201020150, he procedido a revisar el trabajo de graduación titulado "Uso de bases de datos NoSQL en la manipulación de Big Data para el diseño efectivo de marketing proactivo" y de acuerdo a mi criterio todos los capítulos han sido concluidos.

Sin otro particular me suscribo de usted,

Atentamente,



Sergio Arnaldo Méndez Aguilar
Ingeniero en Ciencias y Sistemas
Colegiado No. 10958
Ingeniero Sergio Méndez
Asesor de trabajo de graduación
Colegiado: 10958



Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 16 de Noviembre de 2016

Ingeniero
Marlon Antonio Pérez Türk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación de la estudiante **DINA MAGNOLIA VÁSQUEZ ANTUCHE** con carné **201020150**, titulado: **“USO DE BASES DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA PARA EL DISEÑO EFECTIVO DE MARKETING PROACTIVO”**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



E
S
C
U
E
L
A

D
E

I
N
G
E
N
I
E
R
Í
A

E
N

C
I
E
N
C
I
A
S

Y

S
I
S
T
E
M
A
S

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

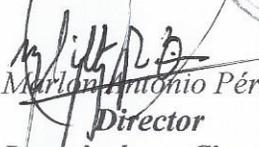


FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN
CIENCIAS Y SISTEMAS
TEL: 24188000 Ext. 1534

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación, **“USO DE BASE DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA PARA EL DISEÑO EFECTIVO DE MARKETING PROACTIVO”** realizado por la estudiante, **DINA MAGNOLIA VÁSQUEZ ANTUCHE**, aprueba el presente trabajo y solicita la autorización del mismo.*

“ID Y ENSEÑAD A TODOS”



Ing.  Marlon Antonio Pérez Türk
Director

Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 09 de noviembre de 2017

Universidad de San Carlos
de Guatemala

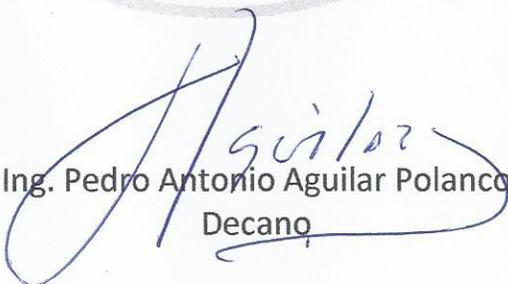


Facultad de Ingeniería
Decanato

DTG. 552.2017

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **USO DE BASE DE DATOS NoSQL EN LA MANIPULACIÓN DE BIG DATA PARA EL DISEÑO EFECTIVO DE MARKETING PROACTIVO**, presentado por la estudiante universitaria: **Dina Magnolia Vásquez Antuche**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:


Ing. Pedro Antonio Aguilar Polanco
Decano

Guatemala, noviembre de 2017

/gdech



ACTO QUE DEDICO A:

- Dios** Por colocar las oportunidades que me ayudaron a ser quien soy.
- Mi madre** Victoria Antuche, que siempre ha estado y está a mi lado incondicionalmente y me ama como soy.
- Mi novio** Kevyn Rodas. Por ser la persona que inspiró el tema de tesis y porque está en los momentos felices y difíciles apoyándome.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser una oportunidad para crecer como profesional y ser un lugar importante que marcó mi vida.
Facultad de Ingeniería	Por ser una importante influencia en mi carrera, un lugar donde obtuve conocimientos profesionales, valores y amigos.
Mis amigos	Que siempre estuvieron para apoyarme y darme ánimos cuando la situación se complicaba.
Mis hermanos	Que son mis consejeros y compañeros de aventuras, apoyándome y escuchándome.
Mi asesor	Ing. Sergio Méndez, quien me ayudo y me guío en la realización del trabajo de graduación.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN.....	XIII
OBJETIVOS.....	XV
INTRODUCCIÓN	XVII
1. ASPECTOS GENERALES.....	1
1.1. Historia del Big Data	1
1.2. Big data	2
1.3. Bases de datos NoSQL y SQL	3
1.3.1. Bases de datos NoSQL	3
1.3.2. Bases de datos relacionales	4
1.3.3. Ventajas del uso de bases de datos NoSQL	5
1.3.4. Diferencias entre las bases de datos NoSQL y bases de datos relacionales	5
1.4. Tipo de bases de datos NoSQL.....	6
1.5. Bases de datos NoSQL en la actualidad	7
1.6. Teorema de CAP.....	8
1.7. La Nube	10
1.8. Cadena de valor	11
1.8.1. Software en la cadena de valor	12
1.9. Marketing.....	13
1.9.1. Las 4P y las 4C del marketing	15
1.10. Marketing proactivo	17

1.11.	Profesional en Informática y su relación con el marketing	17
2.	BIG DATA, NOSQL Y PUBLICIDAD PROGRAMÁTICA CON ENFOQUE EN MARKETING	21
2.1.	Las 5V del Big Data.....	21
2.1.1.	Volumen	21
2.1.2.	Velocidad.....	22
2.1.3.	Variedad	22
2.1.4.	Veracidad	22
2.1.5.	Valor.....	22
2.2.	Origen de los datos en el Big Data.....	23
2.2.1.	Actividades utilizando el internet	23
2.2.2.	Uso de membresías en establecimientos comerciales	24
2.2.3.	Actividades financieras.....	24
2.2.4.	Uso del teléfono móvil en actividades diarias.....	25
2.2.5.	Consulta de fuentes de datos gratuitas	25
2.3.	Big Data Marketing.....	26
2.3.1.	Aplicación y uso de las 5V en el marketing	26
2.3.2.	Dificultades en la implementación del Big Data Marketing.....	27
2.4.	Adopción del Big Data al modelo de negocio de una empresa	28
2.4.1.	Fases de adopción del Big Data en las empresas ..	30
2.5.	Aplicaciones de uso de bases de datos NoSQL en el Big Data Marketing.....	32
2.6.	Beneficios en la adopción del Big Data en el ámbito de marketing	33
2.7.	Publicidad programática.....	35

3.	TENDENCIAS Y APLICACIONES DE SOFTWARE EN EL USO DE BIG DATA.....	37
3.1.	Big Data en la Nube	37
3.2.	Minería de datos.....	38
3.2.1.	Funcionamiento de la minería de datos.....	38
3.2.2.	Web Content Mining	41
3.2.3.	Minería de datos frente al Web Content Mining.....	42
3.3.	Machine Learning	43
3.3.1.	Deep Learning	43
3.4.	Uso de <i>clúster</i> en el Big Data	44
3.4.1.	Clúster y bases de datos NoSQL.....	45
3.5.	Business Intelligence.....	47
3.6.	Aplicaciones de software para el uso del Big Data.....	49
3.6.1.	Análisis de datos en tiempo real utilizando Big Data.....	51
3.6.2.	Software en análisis predictivo	52
4.	APLICACIÓN DE PRUEBAS A BASES DE DATOS NOSQL PARA MANEJO DE BIG DATA.....	55
4.1.	Definición de ambiente de pruebas	55
4.1.1.	Explicación técnica del ambiente de prueba.....	57
4.1.2.	Descripción del escenario de prueba.....	60
4.1.3.	Aspectos a evaluar	61
4.2.	Análisis de los aspectos que se evalúan	62
4.2.1.	Inserciones efectivas en disco	62
4.2.2.	Porcentaje de consumo de CPU y memoria RAM..	63
4.2.3.	Memoria RAM física consumida	65
4.2.4.	Tiempo de respuesta	67

4.2.5.	Factores que influyen en costos de implementación	68
4.3.	Elementos que intervienen en el costo de implementación.....	69
4.4.	Evaluación de resultados	72
CONCLUSIONES.....		75
RECOMENDACIONES		77
BIBLIOGRAFÍA.....		79

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Sistemas NoSQL y el Teorema de CAP.....	10
2.	Análisis de las actividades de la cadena de valor	12
3.	Evolución de las 4P a las 4C de marketing	16
4.	Atribuciones de un Ingeniero en informática	19
5.	Componentes de la infraestructura de Big Data.....	30
6.	Fases de la adopción del Big Data.....	31
7.	Entorno de la publicidad programática	36
8.	Relación entre Data Mining y Big Data.....	41
9.	Beneficios del uso de un <i>clúster</i>	47
10.	Topología de red	59
11.	Cantidad de datos insertados en las bases de datos	62
12.	Porcentaje de consumo del CPU	64
13.	Porcentaje de consumo de memoria RAM.....	64
14.	Memoria RAM física consumida.....	66
15.	Consulta de datos en la base de datos Redis y MySQL	67
16.	Costos de implementación de una base de datos NoSQL	71
17.	Porcentaje de éxito en los aspectos evaluados	73

TABLAS

I.	Diferencias entre Big Data y Business Intelligence.....	49
II.	Características técnicas base de datos Redis	56
III.	Características técnicas base de datos MySQL	56
IV.	Especificación del escenario de prueba.....	60
V.	Estructura de tabla/lista en las bases de datos.....	60
VI.	Diferencia en la inserción de datos	63
VII.	Promedio de uso de CPU y memoria RAM.....	65
VIII.	Promedio de uso de memoria RAM	66
IX.	Promedio de datos insertados versus tiempo	68
X.	Resumen de pruebas.....	72

LISTA DE SÍMBOLOS

Símbolo	Significado
GB	Gigabyte
MB	Megabyte
%	Porcentaje
S	Segundo
ZB	Zettabyte

GLOSARIO

ACID	Características de una base de datos relacional que permiten que una serie de instrucciones sea considerada una transacción. Es el acrónimo de atomicidad, consistencia, aislamiento y durabilidad.
BSON	Es un formato para el intercambio de datos basado en el término JSON y significa <i>Binary JSON</i> . Diseñado para tener un almacenamiento y velocidad eficiente, el cual se utiliza en la base de datos MongoDB.
CPU	Es la sigla de Unidad Central de Procesamiento, la cual es la parte central de todo computador y cumple todas las funciones de procesamiento.
IP	Conocida como dirección IP, es un número que identifica de manera lógica a una interfaz de red de un dispositivo que utilice el protocolo IP.
JavaScript	Es un lenguaje de programación interpretado orientado a objetos. Se usa enfocado a las páginas web.
JSON	Es el acrónimo de <i>JavaScript Object Notation</i> , el cual es un formato ligero de intercambio de datos y usa la sintaxis de JavaScript.

Middleware	Se le conoce comúnmente al software que permite la comunicación y la gestión de información entre sistemas distribuidos.
Open Source	Conocido también como código abierto. Son programas informáticos que permiten acceso y modificaciones a su código fuente, el cual puede ser desarrollado de forma colaborativa y pública.
Proxy	Es un agente que tiene la autorización para actuar en nombre de otro dispositivo y que hace de intermediario de peticiones de recursos de un cliente a un servidor.
RAM	Memoria de acceso aleatorio que se utiliza como almacén para los programas y datos con los que se trabaja en el momento del uso de la computadora, ya que se cargan todas las instrucciones de la unidad central de procesamiento.
ROI	Sigla de retorno de la inversión, el cual es un valor que mide la relación entre las ganancias y los costos de actividades de marketing.
SQL	Es un acrónimo de <i>Structured query language</i> . Lenguaje estándar para manejar las bases de datos relacionales y con el cual se pueden realizar varias operaciones de administración y control.

Transacción	Es un conjunto de instrucciones, las cuales forman una unidad lógica de trabajo la cual debe realizarse de forma indivisible.
URL	Es la sigla de <i>Uniform resource locator</i> , está formada por una secuencia de caracteres de acuerdo con un estándar con el cual se asigna una dirección única a cada uno de los recursos disponible en una red.
XML	De sus siglas en inglés extensible markup language. Es un lenguaje para representar información estructurada en páginas web. Diseñado para cualquier lenguaje y alfabeto, basado en texto.
Zettabyte	Unidad de almacenamiento de información que equivale a un trillón de gigabytes, cuyo símbolo es ZB.

RESUMEN

Es una investigación acerca de los conceptos relacionados con el uso de las bases de datos NoSQL y el uso del Big Data Marketing, la cual describe aplicaciones de software adecuadas para su administración.

En el capítulo uno se definen los conceptos básicos que son necesarios para entrar en contexto sobre el Big Data, Marketing, las bases de datos NoSQL y los tipos que existen.

En el segundo capítulo se define el concepto de Big Data Marketing y la relación que tiene con las bases de datos NoSQL, se explica también como una empresa puede adoptar el uso de Big Data según su modelo de negocios, así como las aplicaciones que puede tener y los beneficios de su implementación.

En el tercer capítulo se describen las tendencias y aplicaciones de software que actualmente hay en el mercado.

Finalmente, en el cuarto capítulo, se describen las pruebas comparativas que se realizan entre bases de datos relacionales y bases de datos NoSQL, midiendo aspectos como consumo de recursos, tiempo de respuesta y costos de implementación.

OBJETIVOS

General

Demostrar que el software puede participar activamente en la cadena de valor de un producto, empleando el Big Data Marketing como herramienta, para dar un enfoque personalizado de un producto o servicio a un cliente.

Específicos

1. Dar a conocer los beneficios y ventajas en el uso de Big Data Marketing en el campo de la informática y en el plan de mercado de una empresa.
2. Enfatizar la importancia de la recolección de datos y su eficiente interpretación para una mejor implementación de un plan de marketing de una empresa.
3. Evidenciar la importancia del conocimiento de las herramientas de análisis de datos orientado al marketing que debe tener un profesional en informática.

INTRODUCCIÓN

En la actualidad las empresas han evolucionado debido a causas como la tecnología, el cambio cultural y las condiciones ambientales, entre otros factores y surgen nuevos datos que se deben recolectar, manejar y analizar para la toma de decisiones.

El concepto de Big Data toma parte importante en el análisis de los datos y en el aumento de la actividad comercial de una empresa, haciendo uso del marketing, para dar al cliente productos y servicios más personalizados en lugar de crear estrategias comerciales generalizadas y para buscar nuevas opciones en el manejo y almacenamiento de estos datos en bases de datos NoSQL y analizarlos en tiempo real, a través de aplicaciones de software especializado; para que los productos y servicios adquieran un nuevo valor comercial, único para cada cliente y se reconozcan tendencias por medio del análisis predictivo para facilitar la toma de decisiones efectiva en el campo del marketing proactivo.

1. ASPECTOS GENERALES

Este capítulo trata sobre los conceptos básicos que se involucran en el uso de bases de datos NoSQL enfocadas en conceptos de Big Data en el ámbito del marketing proactivo, para establecer la cadena de valor de un producto y su relación con el software.

1.1. Historia del Big Data

Las organizaciones comenzaron a implementar sistemas de almacenamiento centralizado, para poder guardar toda la información que poseían de sus usuarios, inventarios y otros datos que les fueran útiles para el desarrollo del negocio. Los sistemas centralizados fueron madurando con el tiempo, permitiendo que las organizaciones pudieran tomar decisiones en base a la información que tenían almacenada.

Desde el inicio de la década de los 60 según William Staunton se tiene noción de la cantidad de datos que crece alrededor del mundo y las empresas los obtenían a través de diversos medios como los son: encuestas, entrevistas y entre otros, para poder tener la información necesaria para implementar un mensaje más adecuado y así poder llegar mejor a su público objetivo.

El 26 de julio 1989, el autor y escritor Erick Larson fue el primero que empezó a utilizar el término “Big Data” en un artículo bajo el título “*How did they get your name? direct-mail firms have vast intelligence network tracking consumers*” en donde detalla en varios párrafos, sobre la recolección de grandes cantidades de datos de los consumidores y las implicaciones que se dan al utilizar

los datos de los clientes, con lo que posteriormente el término se comenzó a utilizar más frecuentemente.¹

En el año 2013 fue un momento decisivo para el término de Big Data, dado que las empresas Google, IBM, Facebook y Twitter comenzaron a absorber compañías que ya lo usaban. Posteriormente en el año 2014 se empezaron a relacionar los términos de Big Data con NoSQL, MapReduce y Maching Learning.

El crecimiento de la tecnología ha sido beneficioso para la sociedad. El uso de dispositivos electrónicos aumentó considerablemente, así como el acceso a internet. Algunos autores con estudios sobre minería de datos como Rakesh Agrawal, Gio Wiederhold, Robert Blum o Gregory Piatetsky-Shapiro consolidaron el término de minería de datos y empezaron a observar cierta relación de elementos y características que se aproximaba a lo que es actualmente el Big Data.

1.2. Big data

Se define Big Data como “un conjunto de tecnologías que nos permite resolver cualquier tipo de problemas a través del análisis de datos”.²

Otra definición es: “además de un conjunto de datos, es una tendencia que las empresas deben utilizar para tener un enfoque de ganar-ganar en relación a los clientes, para mejorar sus productos, sus servicios y a la empresa en sí misma”, así lo menciona el experto Juan David Vargas, publicista especializado en conocimiento del consumidor.³

¹ <http://barker.co.uk/bigdata>. Consulta: 30 de abril de 2016.

² <https://www.youtube.com/watch?v=4Ay3AL-NeVs>. Consulta: 2 de julio 2016.

³ *Ibíd.*

Por lo que se puede decir que Big Data es el término con el cual se designa a un gran volumen de información de cualquier tipo, que se obtiene de diferentes medios como lo son las redes sociales, medios escritos, blogs, páginas de internet y sensores, entre otros; con lo que se busca conocer al cliente lo mejor posible en relación a la empresa, para mejorar la calidad de los productos de la misma. Big data no es un tema de tecnología, sino que se trata de las posibilidades y beneficios que ofrece.

1.3. Bases de datos NoSQL y SQL

En los siguientes apartados se abarcan conceptos que se relacionan con las bases de datos SQL o relacionales y bases de datos NoSQL y se mencionan las características, ventajas y diferencias de cada una de ellas.

1.3.1. Bases de datos NoSQL

Se refiere comúnmente a “No Only SQL”, es decir un marco de trabajo completamente diferente al que se usa en las bases de datos relacionales; ya que con este marco de trabajo se obtiene un alto rendimiento, un procesamiento ágil de la información en una escala masiva; además, no cumplen con el esquema entidad-relación y para el almacenamiento hacen uso de formatos como clave-valor y documentos, entre otros.⁴

Las características principales de las bases de datos NoSQL, son las siguientes:

⁴ <http://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>. Fecha: 29 de mayo de 2016. Consulta: 15 de mayo de 2016.

- No son bases de datos estructuradas, por lo que presentan una gran eficiencia.
- Se centran en el concepto de bases de datos distribuidas, donde los datos pueden ser almacenados en distintas localidades usando múltiples servidores, lo que provee facilidad en la búsqueda de información.
- No manejan las reglas ACID, por lo que puede presentar cierta inconsistencia en la información y tiene transacciones restringidas.
- Fácilmente escalables, con lo que se reduce el costo de inversión en hardware.

1.3.2. Bases de datos relacionales

Las bases de datos relacionales también se conocen como bases de datos SQL, tienen varias utilidades relacionadas con el mantenimiento de la integridad de los datos y a la organización de los mismos. Cuentan con varias características que las definen:

- Son bases de datos que tienen una estructura definida, de forma que un dato es almacenado una única vez en una estructura de filas y columnas conocida como tabla y por lo tanto al consultar los datos es más lento el proceso.
- Diseñadas para tener configuraciones en un único servidor, con lo cual los datos se acumulan y mientras mayor es el volumen las búsquedas toman más tiempo.

- Utiliza un conjunto de reglas ACID, con lo que los datos no presentan inconsistencias y permiten usar transacciones para las operaciones que se realicen.
- Maneja una escalabilidad vertical, con lo que se tiene que invertir más en hardware que pueda potenciar el rendimiento de los servidores usados en bases de datos relacionales.

1.3.3. Ventajas del uso de bases de datos NoSQL

Debido a la naturaleza de las bases de datos NoSQL se pueden observar las siguientes ventajas en el manejo de grandes cantidades de datos:

- Manejo distribuido de datos en servidores de bajo costo.
- Manejo de escalabilidad horizontal, por lo que no se necesita hardware especializado, para añadir más equipos si se necesitan.
- Gestión de gran cantidad de datos al mismo tiempo con gran rapidez que no se genera cuellos de botella al acceder a ellos.

1.3.4. Diferencias entre las bases de datos NoSQL y bases de datos relacionales

Al comparar el manejo de la información entre las bases de datos relacionales y NoSQL, se pueden observar algunas diferencias que sobresalen en el uso de unas y otras bases de datos:

- En las bases de datos NoSQL no se utiliza el lenguaje SQL como base para el manejo de los datos, por lo que tampoco se usan estructuras como tablas para almacenar los datos.
- Las bases de datos relacionales utilizan como lenguaje estándar SQL y usan estructuras como tablas y relaciones para almacenar los datos.
- La normalización no aplica en las bases de datos NoSQL, dado que usan conceptos como embebido de información o anidamiento.
- Las bases de datos relacionales no soportan el uso de los tipos de datos semi-estructurados y no estructurados.
- No se usa el modelo entidad-relación en las bases de datos NoSQL y tampoco se tienen estructuras del tipo maestro-esclavo.

1.4. Tipo de bases de datos NoSQL

Existen varios tipos de bases de datos NoSQL que difieren en la forma de almacenar los datos, pero tienen el mismo objetivo: manejar grandes cantidades de datos, por lo que se mencionan las siguientes:

1. Bases de datos clave-valor: modelo más sencillo, en el cual cada elemento se identifica con una clave o llave única, lo que permite que la información se recupere de una forma muy rápida. Este modelo es rápido y eficiente para escribir y leer los datos.
2. Bases de datos documentales: como su nombre lo indica la información se almacena en un documento que tiene una estructura

de tipo XML o bien JSON y también se utiliza una clave única para cada registro almacenado.

3. Bases de datos en grafo: la información se presenta como nodos los cuales se relacionan con otros nodos por medio de aristas. Cada nodo debe tener una columna y cada relación dos columnas.
4. Bases de datos orientados a objetos: la información se representa mediante objetos, de la misma forma en la que se representan los lenguajes orientados a objetos.
5. Bases de datos Columnar: se introduce un patrón de jerarquía para ordenar los datos y almacenarlos. Las filas pueden variar, ya que pueden contener un número de campos diferente.

1.5. Bases de datos NoSQL en la actualidad

Una de las tendencias en la actualidad es el almacenamiento masivo utilizando bases de datos NoSQL, debido a que el volumen de datos crece constantemente y grandes empresas como Google, Facebook, Amazon y Oracle incursionan en ese campo, ya sea creando su propia base de datos NoSQL o bien consumiendo el servicio. Se presentan algunas de las bases de datos más populares:

- Apache Cassandra: base de datos creada por Apache. Es del tipo clave-valor. Es una aplicación realizada en el lenguaje Java, por lo que puede ser utilizada en cualquier tipo de sistema operativo que cuente con la máquina virtual de java.

- Redis: una base de datos clave-valor, las operaciones que realiza son atómicas y son persistentes, además es un software de código abierto. Es compatible con sistemas Unix, Linux y sus derivados, Solaris, OS/X, pero no existe un soporte oficial para Windows.
- MongoDB: modelo de base de datos orientado a documentos. Cada entrada puede tener un esquema de datos diferente al resto de registros almacenados. Es muy rápido en la ejecución de operaciones, utiliza un sistema propio de documento con el nombre de BSON que es una evolución de JSON, ya que puede almacenar datos binarios.
- Neo4j: Es una base de datos *Open source* orientada a grafos. Cada nodo o relación puede tener una estructura de datos diferente, lo que favorece que tenga mejor escalabilidad. La diferencia que tiene Neo4j es el recorrido de las relaciones ya que el tiempo de una consulta depende del área de búsqueda del grafo.
- CouchDB: es una base de tipo documental que usa formato JSON, soporta configuraciones maestro-maestro con detección automática de conflictos y utiliza una semántica basada en las reglas ACID.

1.6. Teorema de CAP

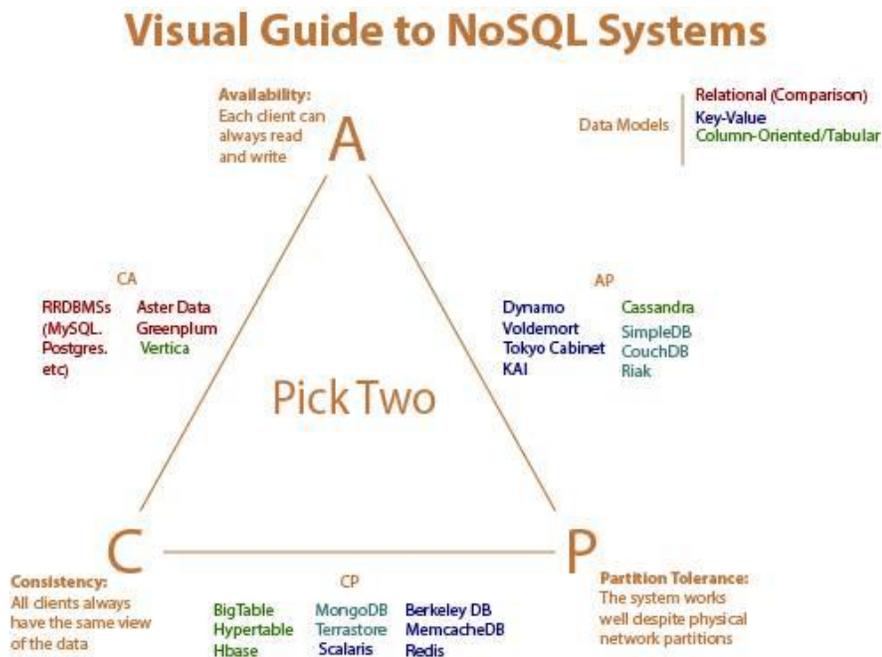
Fue propuesto en el año 2000 por Eric Brewer, como una conjetura solamente y posteriormente fue probado formalmente y se le reconoce como un teorema a partir del año 2002. El teorema menciona que es imposible que un sistema de datos distribuido pueda garantizar al mismo tiempo la consistencia, disponibilidad y tolerancia de particiones.

Las bases NoSQL pueden tener dos condiciones, de las tres que menciona el teorema, las cuales son las siguientes:

- Disponibilidad y tolerancia de particiones, pero sin contar con una consistencia completa en sus datos, dado que algunas de las bases de datos tienen una consistencia parcial.
- Consistencia y tolerancia de particiones, ya que utilizan la replicación de datos como método usando nodos pero no se cuenta con la disponibilidad de estos.
- Consistencia y disponibilidad, ya que no se tiene la tolerancia de particiones pero se puede resolver replicando los datos.

Estas situaciones ayudan a clasificar a las bases de datos NoSQL según el uso que se le quiera dar para poder aplicarlo a las estrategias de marketing, según las necesidades de las empresas. En la figura 4 se muestra en forma gráfica la relación de los sistemas NoSQL y el teorema de CAP.

Figura 1. **Sistemas NoSQL y el Teorema de CAP**



Fuente: <http://i.stack.imgur.com/3NAOs.png>. Consulta: mayo de 2016.

1.7. La Nube

La nube es un concepto más abstracto, una metáfora. Considerada como un paradigma que ofrece servicios de computación por medio del Internet.

La computación en la Nube o bien *Cloud Computing* se define como un servicio de computación a través del Internet. En otras palabras se puede decir que es la administración de recursos mediante aplicaciones desde centros de datos localizados en algún lugar remoto, a través de Internet y que funcionan por medio de una forma de pago del cliente.

Se cuentan con tres modelos de despliegue para los servicios de computación en la Nube:

- Nube pública: la cual es propiedad de empresas que ofrecen acceso a otras organizaciones o bien a individuos. Los usuarios no tienen que adquirir hardware o infraestructura de apoyo ya que estos recursos los aportan los proveedores.
- Nube privada: es propiedad de una compañía que controla los servicios y recursos de forma personalizada y posee un mejor manejo de recursos.
- Nube híbrida: son aquellas que tienen una base en una nube privada y se combinan con el uso de una nube pública.

El uso de la Nube en una empresa depende de las necesidades de la misma ya que se puede optar por cualquiera de los tipos de Nube mencionadas anteriormente para poder administrar su información.

1.8. Cadena de valor

Se conoce como cadena de valor a un concepto teórico que describe el modo en que se desarrollan las acciones y actividades de una empresa, como se muestra en la figura 2. Con base en la definición de cadena, es posible hallar en ella diferentes eslabones que intervienen en un proceso económico: se inicia con la materia prima y llega hasta la distribución del producto terminado. En cada eslabón se añade valor, que en términos competitivos está entendido como la cantidad de consumidores que están dispuestos a abonar por un determinado producto o servicio.

Figura 2. **Análisis de las actividades de la cadena de valor**



Fuente: <http://www.webyempresas.com/la-cadena-de-valor-de-michael-porter/>. Consulta: marzo de 2016.

1.8.1. **Software en la cadena de valor**

La mayoría de las empresas usa métricas para evaluar y medir que tan efectiva es la inversión que hacen en determinada situación en el curso normal de los negocios. Esta métrica se puede aplicar al software, la cual se conoce como ROI, que aplicado al software se utiliza para tomar las decisiones acerca de las aplicaciones de software que utilizarán en sus negocios para poder validar la inversión que se hacen sobre sus productos de software, calculando el ROI antes y después de las implementaciones de software, con el fin de medir que tan exitosa fue dicha implementación, aunque el enfoque que proporciona la métrica ROI en términos de ganancias en el desarrollo del software es difícil de tratar directamente.

El software es aplicado directamente como parte del proceso de producción de un determinado producto o servicio, el resultado es visible y se consideran procesos basados en proyectos. Por el contrario si el software es

aplicado a actividades de soporte de servicio, o bien de información para la toma de decisiones, se dicen que son procesos no basados en proyectos, ya que no dan un resultado tangible en el momento, más bien dan un resultado a largo plazo.

Finalmente, se establecen aspectos en el panorama del análisis de la cadena de valor que influyen de manera contundente en relación al software, los cuales se mencionan a continuación:

- Grado de integración: es un término con el que se define a todas aquellas actividades que se realizan en la propia casa o empresa y no en otras compañías independientes.
- Panorama industrial: es el conjunto de sectores industriales que están relacionados entre sí y que son en los que compete una empresa gracias a una estrategia perfectamente delimitada y coordinada con el claro objetivo de conseguir las metas que se han marcado.

1.9. Marketing

Se define *marketing* como: “las actividades en las que una empresa centra su atención para facilitar el intercambio mutuo de beneficios”, según William J. Stanton, Michael J. Etzel y Bruce J. Walker en su libro Fundamentos del Marketing.

Marketing es planear, elaborar, observar y examinar. Para que una empresa exista, se necesita de su materia prima y esta no es el producto o

servicio en sí, sino sus consumidores o clientes.⁵ En otras palabras, marketing es una herramienta utilizada por cualquier persona, empresa u organización para que todas las partes involucradas tengan un beneficio ya sea un producto o servicio de parte de la empresa y efectivo de parte del cliente o consumidor.

Para que el *marketing* exista, debe haber una necesidad de las personas por satisfacer. Por lo que se pueden destacar las siguientes características:

- Comunicación bilateral entre un cliente y la empresa, ya sea de forma directa o indirecta, por medio de la publicidad.
- Atención personalizada a cada cliente, enviando un mensaje distinto a cada uno de ellos.
- La publicidad como actividad primaria para la captura de los clientes por medio de información recopilada.
- Promoción de las marcas, de forma que los clientes realizan las compras sin presencia de las mismas.
- Recolección de la información para formular una estrategia exitosa.

El surgimiento del *marketing mix* se enfoca en aspectos internos para crear una estrategia de mercado y es donde las 4P posteriormente evolucionan a las 4C.

⁵ <http://ciberconta.unizar.es/LECCION/marketing/100.HTM>. Consulta: 9 de marzo de 2016.

1.9.1. Las 4P y las 4C del marketing

Con el avance de la tecnología, el *marketing* lo hace constantemente, y el motivo de este avance es la adaptación de las estrategias y actividades a las necesidades de cada uno de los clientes, para que todo negocio o empresa pueda sobrevivir en el mundo comercial.

En la actualidad han evolucionado los conceptos antes desarrollados por Jerome McCarthy sobre las 4P, según se muestra en la figura 2, las cuales se describen a continuación:

- Producto: lo que se ofrece al mercado.
- Precio: costo y margen de beneficio.
- Plaza o posicionamiento: canales de distribución existentes.
- Promoción: con qué imagen o mensaje se va a dar a conocer el producto.

A principio de la década de los 90, Robert Lauterborn realizó una investigación sobre un nuevo modelo de consumo, el cual se enfocaba en satisfacer las necesidades de los clientes, con lo que surgen las 4C del *marketing*, las cuales se describen a continuación:

- Consumidor: se enfoca más en las necesidades del cliente y menos en promover un producto o dar un servicio sin conocer lo que el cliente requiere, y se busca mantener un contacto activo con los clientes para mejorar los productos.

- Costo: se enfoca en el precio de venta que percibe el consumidor y también se evalúa el tiempo en que se adquiere el producto, o bien el tiempo en que se consume el servicio; se enfoca también en que la necesidad de adquirir cierto producto sea decisiva.
- Conveniencia: se establece una ruta de distribución cercana al cliente para analizar en qué lugares prefiere adquirir los productos o servicios.
- Comunicación: se establece una comunicación básica y fundamental con los clientes para mejorar las relaciones entre la empresa, sus proveedores y sus consumidores. Además, se pretende transmitir al cliente el mayor número de mensajes con el menor costo posible.

Figura 3. **Evolución de las 4P a las 4C de marketing**



Fuente: <https://www.customfitonline.com/news/2012/10/19/4-cs-versus-the-4-ps-of-marketing/>.

Consulta: abril de 2016.

1.10. Marketing proactivo

Es una actividad que tiene por objetivo incrementar el rendimiento que proviene de los clientes por medio de relaciones interactivas con un valor agregado a largo plazo, en donde se fortalecen y se construyen más relaciones con los clientes, las que con el paso del tiempo se vuelven esenciales y por lo tanto el cliente se compromete con la marca, el producto, servicio o la empresa.⁶ El *marketing* proactivo ayuda a conocer al cliente, su identidad y las interacciones que realiza con la empresa ayuda; a identificar al cliente y saber cuál es su demanda en un determinado producto o servicio, así como sus preferencias.

Cada cliente es diferente dependiendo del valor que la empresa le dé y del valor que el cliente perciba del producto o servicio que consume. Cuando se identifican este tipo de clientes con valor, se crea una interacción única para conocerlos mejor y de una forma personalizada. Finalmente se busca desarrollar una relación de aprendizaje con el fin de personalizar los productos y servicios para cada uno de los clientes y registrar el conocimiento adquirido en bases de datos.

1.11. Profesional en Informática y su relación con el marketing

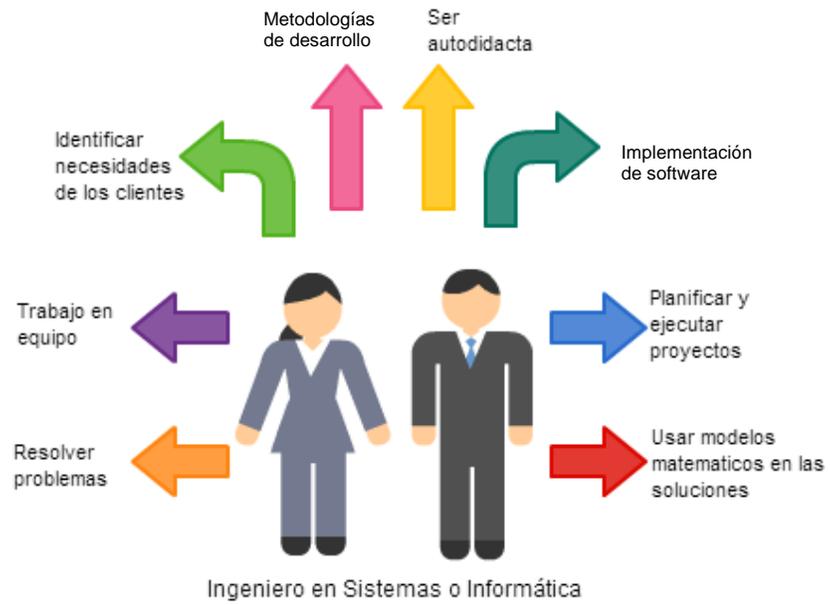
Un profesional en informática, cuenta con un perfil en el momento de egresar de la carrera, este perfil se enfoca en el conocimiento y selección de metodologías de desarrollo de software para un adecuado análisis, el diseño e implementación de un determinado sistema, la identificación de las necesidades de un cliente, conocimientos de software, hardware, redes, capacidad de trabajar en equipo y auto-aprendizaje, entre otras más, según se muestra en la figura 3.

⁶ http://www.deguate.com/artman/publish/gestion_merca/Marketing_Proactivo_2981.shtml#.VzyvLh97Dc. Consulta: 15 de junio de 2016.

Con las actuales tecnologías se protege la calidad de la información que se recibe y se obtiene de los clientes, ya que es muy importante en el *marketing*, como un activo relevante. Con esto se busca que el cliente tenga una mejor experiencia y más contacto con la empresa. Por lo que con esta información se realizan análisis predictivos, aplicando conocimientos estadísticos, matemáticos y de software especializado, logrado así una abstracción para la toma de decisiones basadas en hechos y no en intuiciones.

Al proporcionar una visión técnica y administrativa, un profesional en informática tendría la posibilidad de perfilarse como un facilitador de negocios, aplicando el conocimiento en las tecnologías de información para que esto dé un impulso competitivo y favorable a una empresa, siendo este su valor agregado; teniendo como punto focal al cliente y sus necesidades, dado que las tecnologías de la información tienen un valor intangible y por lo tanto, las acciones que se tomen deben tener un valor tangible que beneficie a los clientes y con ello también a la empresa.

Figura 4. **Atribuciones de un Ingeniero en informática**



Fuente: elaboración propia.

2. BIG DATA, NOSQL Y PUBLICIDAD PROGRAMÁTICA CON ENFOQUE EN MARKETING

Este capítulo describe la estructura de las relaciones entre los temas que influyen en las decisiones que se toman en el ámbito del *marketing*, así como también cuáles son los datos e información que se deben considerar para poder realizarlo de una manera efectiva y eficaz, así como la forma de almacenamiento de la información, la cual es en este caso el uso de bases de datos NoSQL.

2.1. Las 5V del Big Data

Los datos son sumamente importantes para realizar una buena estrategia de *marketing* en una empresa, por lo que el uso del Big Data se ha vuelto una forma muy útil de obtenerlos y de analizarlos; tiene cinco aspectos importantes que definen el término Big Data, los cuales se describen a continuación:

2.1.1. Volumen

El volumen de los datos varía en función del sector o la localización geográfica. La escala de datos que son producidos por las personas, al usar sus dispositivos electrónicos o móviles, cuando entran a las redes sociales o bien usan otros medios. Los datos que se obtienen por cada visita a un sitio de Internet, el uso de aplicaciones para móviles y otros medios son prácticamente ilimitados.

2.1.2. Velocidad

Es la velocidad con la que se crea la información. Es uno de los principales fenómenos del surgimiento de Big Data. La mayor parte de información se produce en tiempo real, por lo que debe ser lo suficientemente ágil y rápido para poder capturar dicha información y ser flexibles para reaccionar de una manera más eficiente al manejar la inmediatez de la información y poder analizar la secuencia de los datos.

2.1.3. Variedad

Se refiere a las diferentes formas de los datos. Hay distintos datos que se generan diariamente, estos están en diferentes formatos en los que se representan, teniendo en común la heterogeneidad de representación, lo que dificulta su interrelación. La mayoría de datos son generados por el consumo de Internet.

2.1.4. Veracidad

Es la incertidumbre de los datos, es decir el establecimiento de qué información es exacta y actualizada. Por este fenómeno se realiza un análisis para evaluar la fiabilidad y autenticidad de los datos que se recolectan.

2.1.5. Valor

Es el aspecto más relevante, dado que se debe tener algún beneficio de toda la información que se recibe y se almacena, posterior a su análisis, para generar una ventaja competitiva en cada ámbito que se necesite según el rol de cada empresa.

2.2. Origen de los datos en el Big Data

En la vida diaria, en el ámbito de las empresas, existen distintos orígenes de tipos de datos que alimentan de información a los sistemas, por lo que estos datos que se obtienen sirven para la toma de decisiones, ya que permiten establecer patrones acerca del comportamiento de los consumidores de cierto producto, servicio o marca, preparando predicciones para lo que posteriormente pueda suceder.

Usar el Big Data para las empresas es muy efectivo, pero para tener los datos necesarios y correctos se necesitan fuentes de donde se puedan obtener esa información. A continuación, se mencionan las principales fuentes de origen de los datos:

2.2.1. Actividades utilizando el internet

Es una de las primeras fuentes, ya que se puede obtener cualquier tipo de información sobre los consumidores, como por ejemplo al comprar en línea, al obtener datos como su correo electrónico, el lugar en donde ingresa a Internet, su historial de compras, o bien las búsquedas que realiza en los navegadores de Internet y las actividades que realiza en las redes sociales.

También se obtiene información del consumidor, por lo que puede mostrársele determinados contenidos. Por ejemplo, el precio al comprar un producto puede variar dependiendo del país de donde se ingrese a Internet: si compró un cierto producto, puede mostrársele más productos similares o relacionados, mientras realiza búsquedas en el navegador de internet.

2.2.2. Uso de membresías en establecimientos comerciales

Son las tarjetas exclusivas que generalmente se adquiere en los negocios o empresas, y que al ser un cliente acreedor de esa tarjeta puede tener beneficios y ciertas exclusividades. Se utilizan en localidades físicas de negocios, cuando los clientes compran o bien realizan transacciones en las cuales al presentar la tarjeta obtienen ciertas facilidades en la actividad a realizar.

Tiempo atrás, este tipo de tarjeta aseguraba que los clientes fueran fieles a un determinado negocio, pero actualmente, además de asegurar la fidelidad, se puede obtener datos sobre las preferencias de compra de los consumidores y sus hábitos de consumo, por lo cual las empresas pueden realizar un análisis y adelantarse a las necesidades de sus consumidores.

2.2.3. Actividades financieras

Se usa la tarjeta de crédito la cual está ligada a una institución bancaria, por lo que un banco podrá extraer mucha información del dueño de la tarjeta: saber dónde, qué, cuándo y cómo compra los artículos para sus necesidades. Además de que la institución bancaria obtiene la información, la puede compartir con las instituciones o negocios en donde compra una persona.

Comprar con tarjeta de crédito, permite ofrecer promociones exclusivas a los clientes, o bien conceder o denegar un crédito, basándose en su historial de pagos y compras.

2.2.4. Uso del teléfono móvil en actividades diarias

Es una de las fuentes que más da información acerca de un consumidor, y la que puede aportar más a las empresas con datos recientes, debido a que acompaña al consumidor la mayoría de tiempo en su vida diaria, ya sea por el uso de datos de geolocalización, o para realizar una publicación en una red social. Cada aplicación instalada en un teléfono móvil, envía datos distintos de cada uno de los aspectos de la vida de una persona a las empresas que manejan dichas aplicaciones.

2.2.5. Consulta de fuentes de datos gratuitas

Son fuentes (se les conoce como *Data Sets*) que se especializan en recopilar información de cierta localidad geográfica, o de ciertas páginas de Internet, para el uso de empresas que requieran cierta información, la cual se clasifica por categorías de temas. En su mayoría, se encuentran fuentes gratuitas, aunque existen opciones pagadas. Algunas de las que se pueden encontrar en Internet, son:

- Portal de datos abiertos de la Unión Europea.⁷
- Portal de datos abiertos del Gobierno de los Estados Unidos.⁸
- Amazon, a través de Amazon Web Services.⁹
- Facebook Graph API.¹⁰
- Indicadores de Gapminder.¹¹
- Buscador de data set de datos numéricos Quandl.¹²

⁷<http://open-data.europa.eu/es/data/>. Consulta: 10 de marzo de 2016.

⁸<http://www.data.gov/>. Consulta: 12 de marzo de 2016.

⁹<http://aws.amazon.com/es/datasets/>. Consulta: 12 de marzo de 2016.

¹⁰<https://developers.facebook.com/docs/graph-api>. Consulta: 12 de marzo de 2016.

¹¹<http://www.gapminder.org/data/>. Consulta: 12 de marzo de 2016.

¹²<https://www.quandl.com/>. Consulta: 12 de marzo de 2016.

- DBpedia.¹³

2.3. Big Data Marketing

Se conoce como Big Data Marketing al proceso de análisis de los datos almacenados en una estructura de Big Data, el cual permite guardar toda la información obtenida mediante herramientas especializadas, medir de forma cuantitativa la información, establecer los objetivos de rendimiento y medir procesos, conocer de forma detallada el comportamiento de los clientes, predecir las tendencias del mercado ofreciendo una personalización a gran escala a cada cliente, e influir en sus decisiones mediante estrategias de mercadeo.

Uno de los beneficios del Big Data Marketing, según Enrique Infante, director *marketing* de contenido de UNIR, es que “permite a las empresas conocer mejor a sus clientes, segmentarlos por su comportamiento y no por su perfil, y personalizar la oferta, la publicidad y las promociones comerciales para adaptarlas a cada usuario en el momento y soporte más conveniente”.

2.3.1. Aplicación y uso de las 5V en el marketing

Por la naturaleza del Big Data se pueden observar varias aplicaciones que tiene en el *marketing* y como esto se relaciona con la aplicación de las 5V, las cuales se mencionan a continuación.

- Big Data ofrece una mejor segmentación de los consumidores, y la información que se obtiene de ellos se puede conseguir de una forma más rápida y eficaz, con lo cual se logra más detalle.

¹³<http://wiki.dbpedia.org/About>. Consulta: 12 de marzo de 2016.

- Se realizarán campañas de *marketing* más precisas y concretas. Y se enfocan en un público objetivo más determinado.
- Además, si se realiza un análisis preciso y exacto de los datos que proporciona el Big Data pueden identificarse características que revelan tendencias y necesidades que otros competidores en el mismo mercado no han identificado.
- El Big Data marca patrones y comportamiento de los consumidores, con lo que se pueden construir predicciones que se puedan necesitar en el futuro.

2.3.2. Dificultades en la implementación del Big Data Marketing

Existen diferentes dificultades en el momento de implementar el uso del Big Data en una empresa, según una encuesta realizada en un evento efectuado por la Asociación de Marketing Directo en el año 2013, las cuales se listan a continuación:

1. El hardware utilizado para el análisis y procesamiento del gran volumen de datos debe ser especializado.
2. La calidad de los datos, dado que son datos variados y sin estructura. Se pueden guardar datos que no tengan una implicación directa con el negocio, por lo cual se debe tener un estricto filtrado de los mismos.

3. El presupuesto limitado para tener todo lo que requiere la empresa para implementar y usar de forma correcta el almacenamiento y análisis de datos.
4. La implementación de una nueva tecnología y la falta de profesionales con experiencia en el área.

2.4. Adopción del Big Data al modelo de negocio de una empresa

Big Data no es un concepto nuevo, dado que algunas compañías llevan varios años almacenando datos en grandes cantidades, como las empresas de telecomunicaciones o bien las de extracción de petróleo. Cada organización identifica y analiza su modelo de negocio y los requisitos, para luego adaptar su infraestructura con el fin de tener una mejor oportunidad, utilizando el concepto de Big Data.

A través del análisis y enfoque exclusivo sobre las necesidades, preferencias y demandas del cliente, las empresas encuentran formas nuevas de interactuar con sus consumidores y se forma una relación bilateral entre el cliente y la empresa. Pero existen retos que se deben superar en el momento de adoptar el Big Data, entre los cuales se pueden mencionar los siguientes:

- Incapacidad de conectar e integrar los datos que se tienen almacenados.
- Capacidad de almacenamiento, por lo que se debe verificar si la infraestructura es correcta para realizar esta tarea.

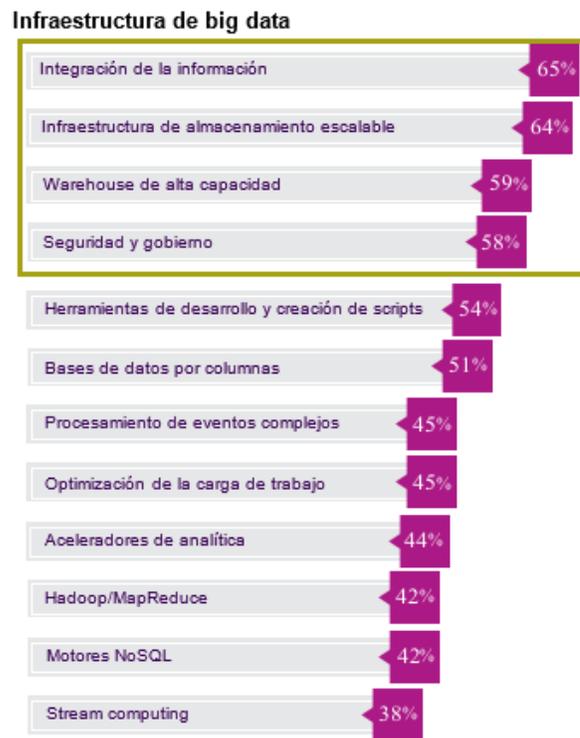
- La legislación en la protección de los datos, por lo cual debe informarse al cliente a que datos se accederá en el momento de cierta acción.

Según el informe basado en el “Big Data @ Work Survey”¹⁴ en el año 2012, con una muestra de 1,144 profesionales encuestados, se define la infraestructura que debe tener, o que se considera necesaria para el Big Data, la cual se muestra en la figura 5.

La encuesta reveló que las principales fuentes de datos para las empresas, son aquellas que provienen de fuentes internas, ya que es la información que más se maneja y se tiene una mejor comprensión sobre ella. Los datos internos crecen rápidamente por lo que no se pueden gestionar con herramientas tradicionales, debido a lo cual aunque se tengan muchos datos, la mayoría no se analiza por medios tradicionales, ya que con el avance de la tecnología se logran utilizar funciones de simulación de datos y modelos de optimización y de análisis.

¹⁴http://www05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf. Consulta: 13 de febrero de 2016.

Figura 5. **Componentes de la infraestructura de Big Data**



Fuente: *Analytics: el uso de Big Data en el mundo real*. http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf.
Consulta: marzo de 2016.

2.4.1. **Fases de adopción del Big Data en las empresas**

Se debe realizar un estudio para conocer la infraestructura y otros elementos que una empresa necesita para implementar el uso del Big Data. Según una encuesta realizada en el “Big Data @ Work Survey”¹⁵, se definen 4 etapas para que una empresa adopte y evolucione en el proceso de Big Data.

¹⁵http://www05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf. Consulta: 10 de marzo de 2016.

En la figura siguiente se muestran y se explican brevemente las fases que se deben seguir para que las empresas puedan aplicar Big Data.

Figura 6. **Fases de la adopción del Big Data**



Fuente: *Analytics: el uso de Big Data en el mundo real*. http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf.

Consulta: marzo de 2016.

- Educar y crear una base de conocimiento. En esta fase se realiza la concientización y el desarrollo en el conocimiento del uso de los datos en grandes cantidades, con base en las ventajas y resultados que cada empresa quiere obtener respecto a sus competidores.
- Explotar, definir el caso del negocio y hoja de ruta. Se trata de analizar cómo se puede beneficiar la empresa con el uso del Big Data y que ruta es la más indicada para comenzar a utilizarlo, conociendo los recursos y la tecnología disponible, para que el plan que se desarrolle se acople a las estrategias de negocio de la empresa.
- Interactuar y adoptar el Big Data. En esta fase las empresas comprueban el valor que les agrega la utilización del Big Data en sus relaciones de negocios. Además, se analizan, prueban y

comprenden las habilidades y la tecnología para poder aprovechar y buscar nuevas fuentes de datos que beneficien a su negocio.

- Ejecutar e implementar Big Data. En este punto las empresas implementan las funciones y recursos que ofrece el Big Data aplicada a su modelo de negocio y obtienen el máximo valor de sus activos con la información que se tiene y se recibe de los clientes.

2.5. Aplicaciones de uso de bases de datos NoSQL en el Big Data Marketing

El Big Data es una oportunidad para el negocio, para agregar valor importante al producto o servicio que el consumidor necesite. Las bases de datos NoSQL pueden apoyar de varias formas, por lo que se puede tener todo tipo de información y analizarla en tiempo real. Las aplicaciones de uso se mencionan a continuación:

- Almacenar gran cantidad de información, tanto en tiempo real como fuera de línea.
- Atender a más usuarios, más consumidores y tener una visión más clara de lo que les interesa a los clientes y de cómo tomar mejores decisiones basadas en el análisis de los datos almacenados.
- La forma de los datos y de la información cambia constantemente, por lo que se requiere un sistema que soporte este tipo de cambios, debido a lo cual el usar una base de datos NoSQL permite un almacenamiento rápido y el tiempo de consulta se reduce.

- Permitir un análisis de lotes, lo que incluye trabajar con grandes cargas de trabajo.
- Asegurar la disponibilidad en todo momento, ya que las bases de datos NoSQL tienen la característica de manejar los datos de forma distribuida.
- Analizar servicios financieros, tendencias del clima, algoritmos de negociación, análisis de perfil de ciudadanos, datos geoespaciales, identificar comportamientos únicos en los consumidores, tendencias de compra y analizar datos de pequeños negocios que deseen expandir sus operaciones.

2.6. Beneficios en la adopción del Big Data en el ámbito de marketing

El Big Data surge en el área de tecnología y es aplicable en el área de *marketing*, y combinando ambos campos se logra obtener variados beneficios para el cliente y para la empresa que aplica el concepto de Big Data Marketing. A continuación se enumeran los más importantes:

- Mejora la gestión de inventarios, ya que permite obtener datos en tiempo real sobre la gestión de los productos en venta y abastecimiento en el menor tiempo posible, optimizando así la cadena de suministro y reduciendo costos.
- Maneja el análisis de comportamiento de los clientes, a través de un seguimiento y análisis de sus actos. Analiza también cuales campañas publicitarias tienen más interacción con el público.

- Aumenta la velocidad de producción de mercancías y puede realizar simulaciones de la producción de un artículo específico, lo que permite optimizar procesos.
- Controla factores de riesgo y la toma de decisiones empresariales en tiempo real.
- Reduce los costos de la infraestructura de almacenamiento, dado que se utilizan servicios en la Nube, con menor costo si se manejan servidores físicos.
- Utiliza índices y métricas para avalar los niveles de satisfacción de los clientes y diseñar planes de fidelidad que atiendan las necesidades del cliente y mejoren su experiencia.
- Identifica vulnerabilidades en la arquitectura del software usando una postura predictiva, previo a la recolección de datos de la empresa.
- Elabora estrategias de *marketing* de manera personalizada para cada cliente.
- Concurrencia en el manejo de distintos tipos de datos, dado que acepta datos no estructurados y semi-estructurados y se logra unir toda la información en menor tiempo para el análisis.
- Ayuda a la segmentación del target, en grupos específicos y en forma dinámica.

Identificar nuevas fuentes de ingreso, desarrollar nuevos productos o servicios utilizando el Big Data ofrece resultados tangibles en la búsqueda de nuevos ingresos para las empresas.

2.7. Publicidad programática

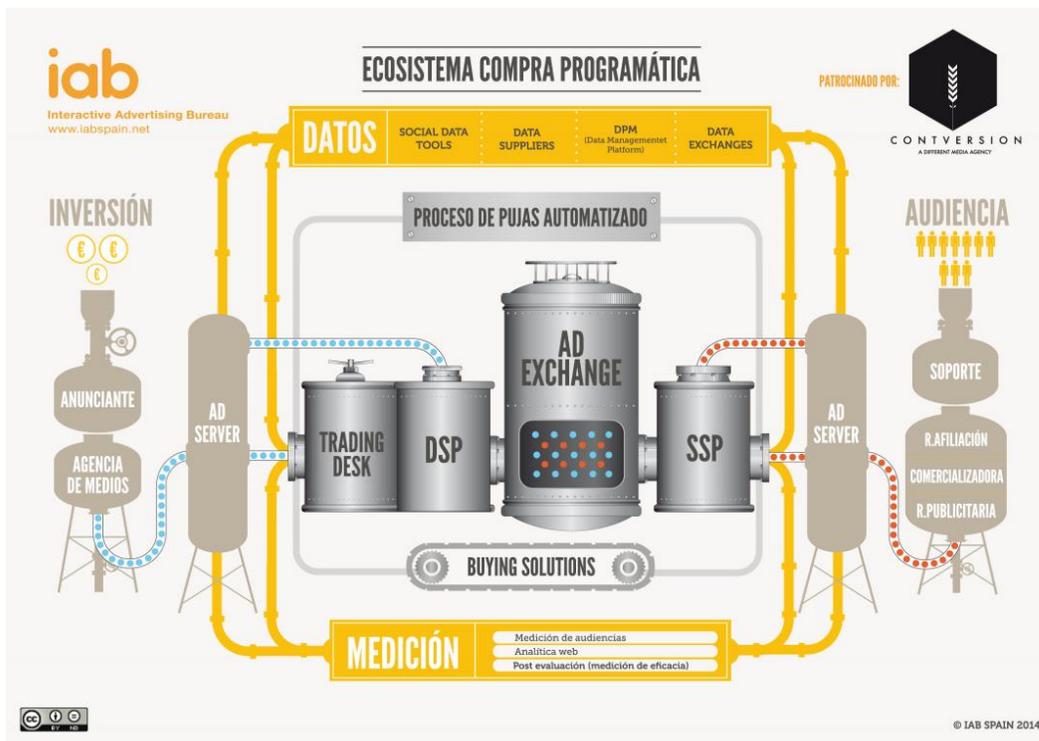
La publicidad ha evolucionado con el uso de Big Data, cuyo objetivo es obtener grupos de audiencias de clientes y no comprar espacios publicitarios. Este tipo de publicidad tiene como finalidad ayudar a las marcas a mostrar el producto correcto a la persona correcta.

La publicidad programática consiste en que las empresas compran publicidad mediante un proceso de pujas en tiempo real. En el proceso intervienen los *Ad Exchanges*, que son casas de subastas; las *Demand side Plataforms* que son proveedores de tecnología que realizan la optimización mediante el Big Data, los cuales manejan gran cantidad de información de la empresa, el target, entre otros; los *Data Partners* que son los que proporcionan los datos e información; los *Trading Desk*, que son el equipo que trabaja directamente con las casas de subasta y finalmente los *Sell Side Plataforms* que maximizan el rendimiento de los espacios publicitarios.

Todo este procedimiento se realiza mediante un algoritmo, el cual se basa en búsquedas en Internet que hacen los usuarios: se conoce como “*Search marketing*”. En el momento en que un anuncio se visualice en una página, un servidor especializado guarda la información de la búsqueda, y mediante un algoritmo realiza un cálculo de probabilidad de que se haga clic en el anuncio o no, entre otras variables, según los datos recolectados del usuario en todas sus búsquedas realizadas.

Actualmente la tendencia de las empresas está siendo utilizar la publicidad programática, según aseguró en una entrevista a Forbes, el director de estrategia digital de Kellogg's, Bob Arnold, quien asegura que algunas marcas han visto multiplicar por seis el ROI de sus campañas. Juan Sevillano, director general de Rocket Fuel España, añade al respecto que "Mediante el Big Data podemos llegar a tener un conocimiento tan exhaustivo de los intereses de cada uno de los internautas de manera que sepamos cuál es el mejor anuncio para mostrar a cada usuario en cada momento, e incluso elegir el canal más adecuado" y agrega que este proceso también ha incrementado el ROI en su empresa.

Figura 7. Entorno de la publicidad programática



Fuente: *La Publicidad Programática, explicada de forma visual.*

http://www.reasonwhy.es/actualidad/digital/la-publicidad-programatica-explicada-de-forma-visual_2014-03-10. Consulta: mayo de 2016.

3. TENDENCIAS Y APLICACIONES DE SOFTWARE EN EL USO DE BIG DATA

Este capítulo abarca las diferentes tendencias en las cuales se aplica el Big Data para el manejo de la información; y se presentan herramientas y aplicaciones de software en las que se usan procesos que aplican los conceptos de Big Data Marketing, se da también una breve descripción de su uso y de cómo intervienen las bases de datos NoSQL.

3.1. Big Data en la Nube

Para el caso del Big Data se pueden utilizar cualquiera de los tipos de nubes que se describen en capítulos anteriores, pero por motivos de seguridad acerca de los datos que se manejan se requiere el control directo sobre la Nube. Aunque uno de los inconvenientes más frecuentes es el costo de la infraestructura y el manejo de los problemas técnicos que surjan en el manejo de la Nube y por esto la mayoría de las empresas opta por contratar proveedores que ofrecen el servicio de Nubes públicas.

La mezcla de seguridad y la reducción de costos utilizan la opción de Nubes híbridas, y además las Nubes privadas soportan de una mejor manera una mayor carga de trabajo, mientras que las Nubes públicas acomodan de mejor forma las demandas sobre el uso de los datos. Sin embargo, es importante mencionar que el Big Data no depende de la Nube, es más bien un medio de almacenamiento y facilita el análisis de los datos que allí se almacenan.

3.2. Minería de datos

Se le conoce como exploración de datos o *data mining*, la cual consiste en el análisis de datos para encontrar patrones y resumir grandes volúmenes de datos en nuevos datos comprensibles y más útiles que se puedan usar en el análisis y toma de decisiones proactivas.¹⁶

Big data y minería de datos son dos términos que tienen relación entre sí. Uno de los medios que se utiliza en la minería de datos para obtener los mismos, es el Big Data, debido a su gran tamaño, en donde se escogen grupos menores de información para que los administradores tomen decisiones acerca de un tema en específico, por lo que podría decirse que el Big Data es la fuente y el *data mining* es el manejador de los datos.

El uso más común que tiene la minería de datos es la utilidad en las bases de datos de conocimientos, o bien en el descubrimiento de conocimiento en las bases de información, con lo que toma un gran volumen de información y extrae la que resulta válida, novedosa, útil y comprensible para el usuario. También se pueden descubrir modelos matemáticos para los datos los cuales se pueden aplicar en la segmentación de clientes y la predicción de comportamientos futuros, entre otros.

3.2.1. Funcionamiento de la minería de datos

La minería de datos proporciona el vínculo entre el software utilizado para la extracción de los datos, con el cual se analizan las relaciones y los patrones que se encuentran almacenados en las bases de datos de conocimientos.

¹⁶<http://www.tuataratech.com/2015/06/mineria-de-datos-data-mining-vs-grandes.html>.
Consulta: 1 de agosto de 2016.

Existen cuatro tipos de relaciones que se buscan en el proceso de minería de datos, los cuales son:

- Clases: son datos almacenados previamente para poder encontrar otros en grupos nuevos.
- *Clústeres*: los datos y sus elementos se agrupan de acuerdo a ciertas asociaciones lógicas, o bien de acuerdo a preferencias detectadas en los consumidores.
- Asociaciones: conjunto de datos distintos que tienen alguna relación.
- Patrones secuenciales: los datos se extraen de las acciones anticipadas de tendencias.

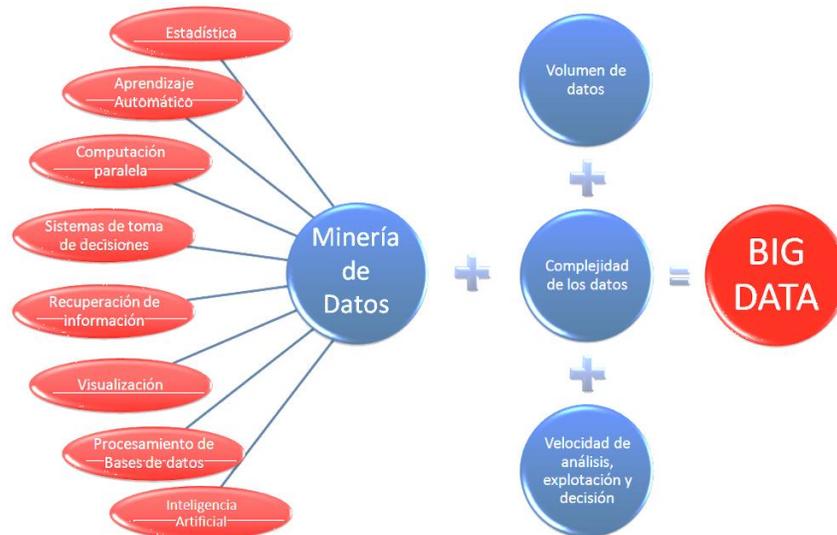
Entre los usos específicos de la minería de datos se incluyen los siguientes:

- Segmentación del mercado: identificar las características comunes de los clientes que compran los mismos productos de una compañía.
- Pérdida de clientes: predecir qué clientes son propensos a dejar su empresa e ir con un competidor.
- Detección del fraude: identificar cuáles son las transacciones probablemente fraudulentas.

- Marketing directo: identificar perspectivas que se incluyen en una misma lista para obtener la tasa de respuesta más alta.
- Marketing interactivo: predecir lo que probablemente es más interesante para cada individuo que accede a un sitio Web.
- Análisis de la cesta: entender qué productos o servicios son comúnmente comprados juntos.
- Análisis de tendencias: revelar la diferencia entre un cliente típico y uno permanente.

El uso a largo plazo de la minería de datos que puede verse en la figura 8, la convierte en una herramienta poderosa en la creación de estrategias innovadoras para una toma de decisiones efectiva.

Figura 8. Relación entre Data Mining y Big Data



Fuente: https://www.incibe.es/extfrontinteco/img/Image/BLOG/infografia_1000.png. Consulta: mayo de 2016.

3.2.2. Web Content Mining

Es una técnica de análisis que usa las técnicas de la minería de datos, pero exclusivamente en el ámbito web, y obtiene información de las búsquedas que se realizan a los sitios web en sus diferentes formatos de contenido como: texto libre, multimedia HTML y otros, en forma automática.¹⁷ También se define como la extracción o recuperación de datos que no se realiza de forma trivial, cuya información es útil y previamente era desconocida.¹⁸

¹⁷<http://www.scaleunlimited.com/about/web-mining/>. Consulta: 9 de julio de 2016.

¹⁸<http://www.bridgedworld.com/es/data-mining-y-web-mining>. Consulta: 10 de julio de 2016.

Utiliza reglas de clasificación y agrupación, asociación y sucesos frecuentes, las cuales permiten clasificar a los usuarios e identificar los patrones de comportamiento según como interactúen con la página web, con lo que se pueden identificar patrones de comportamiento según las acciones que se reiteren con respecto a determinada acción realizada por el usuario, y así darle un mejor servicio o producto de acuerdo a su perfil almacenado en la página web.

3.2.3. Minería de datos frente al Web Content Mining

Se presentan algunas diferencias entre la minería de datos enfocada exclusivamente a páginas web y la minería de datos en el ámbito relacional.

- Escala en el ámbito web la cantidad de información es: mayor en comparación con una base de datos relacional con miles de datos, dado que en una página web se pueden tener diferentes contenidos no estructurados y enlaces con otras páginas web.
- Analiza datos en las páginas web: se tiene un acceso público y pocas veces se requieren derechos de acceso, por lo que se usa software para rastrear el contenido de los usuarios.
- Estructura la minería de datos: se obtiene información de una base de datos que tiene cierta estructura en los datos, mientras que en el *web mining*, se obtienen miles de datos semiestructurados y no estructurados de las páginas web, aun cuando esta información proviene de una base de datos.

3.3. Machine Learning

Se le conoce como aprendizaje automático o bien aprendizaje de máquinas. Es una rama de la inteligencia artificial que consiste en un método de análisis de datos que realiza de forma automática un modelo analítico y usa algoritmos que aprenden en forma iterativa a partir de los datos de interés que se reciben de los usuarios.¹⁹

Otro término que se aplica es el que lo define como el estudio de algoritmos que calculan estadísticas que servirán para un posterior aprendizaje a partir de datos ejemplos, aplicado en áreas intensivas como el reconocimiento de voz, búsqueda de datos específicos y análisis de redes sociales, entre otros.²⁰

La clave del éxito del aprendizaje automático es tener gran volumen de datos y que sean recolectados y analizados sin intervención del usuario, es decir explotar las oportunidades ocultas que proporciona el Big Data.

3.3.1. Deep Learning

Se conoce también como aprendizaje profundo y es una serie de algoritmos que aplican el *Machine Learning* y se refiere a las redes neuronales artificiales. Es una tendencia en crecimiento dado sus aplicaciones favorables en conjuntos de datos con un gran volumen y tienen un objetivo complejo.²¹

El lenguaje profundo puede inferir a partir de una parte de un texto, aun así si no se encuentra lo que se busca de forma explícita, lo que ayuda a la toma

¹⁹http://www.sas.com/en_id/insights/analytics/machine-learning.html. Consulta: 10 de agosto de 2016.

²⁰<https://www.quora.com/What-is-machine-learning-4>. Consulta: 11 de agosto de 2016.

²¹<https://www.quora.com/What-is-deep-learning>. Consulta: 11 de agosto de 2016.

de decisiones a realizar un análisis predictivo, y se enfoca en el análisis y clasificación de patrones en donde se hace uso de las redes neuronales, modelos estadísticos y algoritmos de aprendizaje supervisado.

El uso de *Deep Learning* puede tener complicaciones, dado que su implementación puede resultar sumamente costosa y entre más grande sea el volumen de datos el tiempo de aprendizaje se eleva y se complica.

La automatización de la extracción de datos sin supervisión en *Deep Learning*, permite su aplicación a diferentes tipos de datos como imagen, textura, sonido y otros. El uso de modelos lineales para tratar de forma eficaz con conocimientos obtenidos a través de datos complejos que sirve para la toma de decisiones.

3.4. Uso de *clúster* en el Big Data

Un *clúster* es un grupo de dispositivos con similares elementos, que pueden ser computadores, que están interconectados y funcionan como una unidad de trabajo.²²

El gran volumen de datos debe almacenarse para no evitar pérdidas, por lo que en esta condición se presenta la situación de cómo y dónde guardar todos los datos, debido a esto se utilizan grupos de dispositivos, en este caso servidores, para evitar cuellos de botella. Regularmente estos *clúster* están localizados en la Nube, dado que las empresas tratan de reducir costos en infraestructura y también se logra una optimización en el movimiento de los datos.

²²http://suseos.weebly.com/uploads/1/8/7/0/18707470/23-_cluster.pdf. Consulta: 22 de agosto de 2016.

La principal características de un *clúster* para Big Data, es que los dispositivos usados, son de bajo costo, y cuentan con una escalabilidad horizontal para poder manejar de una forma distribuida la información. Se usan algoritmos especializados para poder distribuir los datos, así como en la forma de búsqueda de información.

En el proceso de almacenamiento mediante el uso de *clúster*, se obtiene la información en tiempo real, ya que según los datos que se analicen se pueden obtener aproximadamente 300,000 actualizaciones por minuto, esto se logra en base a la especialización de los *clúster*, ya que se basan en temas específicos, palabras claves, historial de cierto navegador o bien cambios de estado, si fuera un red social.

3.4.1. Clúster y bases de datos NoSQL

Existen actualmente varias herramientas que proveen el manejo de *clúster* para arquitecturas distribuidas que emplean bases de datos NoSQL, algunas de ellas son las siguientes:

- Hadoop: software especializado para manejar un *clúster* especialmente diseñado para almacenar grandes cantidades de datos no estructurados utilizando un ambiente distribuido, con máquinas de procesamiento de bajo costo, que además es altamente escalable.
- Redis: es una aplicación distribuida, de alto rendimiento y escalabilidad lineal, dado que los datos son fragmentados automáticamente en varios nodos y usa una replicación asíncrona que evita conflictos en los registros clave-valor.

- MongoDB: es un software que usa una serie de configuraciones que tienen como mínimo el uso de 3 servidores para realizar la fragmentación de datos. Cuenta con auto balanceo de carga y permite escalabilidad horizontal.
- CouchDB: es un proxy que recoge los datos y los envía a varias particiones o nodos y cada una de los nodos se compone de varias instancias de bases de datos, que contiene información replicada.

Cada una de las herramientas que maneja clústeres tiene sus propias características que se adaptan a las necesidades de la empresa, al manejar un *clúster* se puede tener más eficiencia en el manejo de los datos teniendo que manejar una tarea específica, los beneficios se muestran en la figura 9.

Figura 9. **Beneficios del uso de un *clúster***



Fuente: Elaboración propia. Basado en: <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/402826/5-ventajas-de-la-arquitectura-de-Hadoop>.

3.5. **Business Intelligence**

Conocido también como inteligencia de negocios (BI por sus siglas en inglés). Es una herramienta que brinda soporte para la toma de decisiones, ya que permite acceder a los datos para su análisis y manipulación.²³ Considerado como una habilidad y un grupo de metodologías que pueden transformar los datos en información con valor que se pueda usar para la toma de decisiones en una empresa, y permiten transformar la información de las bases de datos

²³<http://www.gestiopolis.com/business-intelligence-teoria-y-conceptos/>. Consulta: 25 de agosto de 2016.

relacionales e información no estructurada, en un conjunto de datos estructurados.²⁴

Tanto Big data como el uso del Business Intelligence, se usan para analizar datos, y por lo consiguiente, para auxiliar en el proceso de toma de decisiones, pero ambos procesos son diferentes ya que difieren en el modo en que analizan y en qué datos son los que analizan.

Los datos en el Business Intelligence se almacenan de una forma centralizada que se basa en el uso de cubos multidimensionales, y se extraen de un ambiente denominado *Data Warehouse*; en lugar de que los datos se tomen de una base No SQL, como es el caso del Big Data, se toman de una base de datos relacional. Aunque los dos sistemas sean tan diferentes también son complementarios; por un lado el Big Data muestra una visión global con un análisis profundo de los datos reconociendo tendencias y patrones, mientras que el Business Intelligence ofrece otros elementos como el uso de *dashboards*, reportes y métricas acerca del rendimiento de los datos. No obstante desde sus inicios el Big Data se consideró para el almacenamiento de datos estructurados por lo que con el tiempo se crearon aplicaciones de software con ese fin, como es el caso de Bigtable de Google.

El Business Intelligence tenderá a desaparecer debido al auge y flexibilidad que provee el Big Data y otro concepto tomará su lugar, como el Business Discovery, que en lugar de usar cubos multidimensionales para buscar datos concretos, se enfocará en las relaciones que se puede tener con cualquier

²⁴http://www.sinnexus.com/business_intelligence/. Consulta: 26 de agosto de 2016

tipo de datos y explora las mismas para poder tener una toma de decisiones más efectiva.²⁵

Tabla I. **Diferencias entre Big Data y Business Intelligence**

Big Data	Business Intelligence
Utiliza un sistema distribuido.	Utiliza un servidor central.
Maneja mayores cantidades de información más rápido.	Los datos son procesados para después aplicarles funciones para la toma de decisiones.
Permite variedad de formatos de datos para su análisis.	Solo soporta datos estructurados.
Las fuentes de los datos pueden ser históricas o bien en tiempo real.	Los datos deben estar almacenados previamente para poder analizarlos.
Permite procesamiento en paralelo en forma masiva.	

Fuente: elaboración propia, con base en: Principales diferencias entre Business Intelligence y Big Data. <http://artyco.com/principales-diferencias-entre-business-intelligence-y-big-data/>.

Consulta: mayo 2016.

3.6. **Aplicaciones de software para el uso del Big Data**

Estas herramientas permiten tomar decisiones que son útiles dependiendo del modelo del negocio de cada empresa que las utilice, pero principalmente se usa para las decisiones en donde la información es demasiada y los resultados son importantes para una empresa.

Existen varios tipos de licencias de código abierto, como también licencias pagadas, entre las herramientas más populares, están:

²⁵<http://www.baquia.com/emprendedores/2013-07-19-manuel-daza-bigdata-business-intelligence-bi-discovery-informacion-datos-empresas-olap-analisis-qlikview-ibm-smart-cities-redes-sociales>. Consulta: 30 de agosto de 2016

- Hadoop: utiliza modelos simples de procesamiento distribuido, por medio de clústeres, y fue diseñada para superar errores y fallos en la capa de aplicaciones, con lo que tiene alta precisión. Cuenta con una licencia de código abierto.
- Gridgain: proporciona métodos dinámicos de computación y de almacenamiento, balanceo de carga, tolerancia a fallos. Utiliza el algoritmo MapReduce. Orientada a soluciones web.
- HPCC: se le conoce como *clustering* computacional de alto rendimiento. Ofrece un rendimiento superior a Hadoop en entornos específicos, tiene dos tipos de versiones, la gratuita y la de pago.
- Apache Hive: es una opción para data warehouse que proporciona Hadoop, pero utilizan un almacenamiento distribuido, con lo cual proporciona resúmenes de datos, consultas y análisis.
- Apache Kafka: es un software que funciona como intermediario de mensajes en sistemas distribuidos. Puede manejar varios cientos de megabytes de información, con poca latencia para manipulación de datos en tiempo real.
- Cloud BigTable Beta: es una de las aplicaciones creadas por Google, en el cual presta el servicio de base de datos NoSQL. Se alimenta de los datos que recibe del navegador y de varias de las aplicaciones proporcionadas por la empresa. Diseñado para cargas masivas, baja latencia y un alto rendimiento. También posee la característica de que se integra de manera fácil como aplicaciones Hadoop y Apache Spark.

- Amazon Elasticsearch Service: es un servicio creado y gestionado por Amazon para operar y escalar servicios en la Nube. Cuenta con las opciones de monitoreo y análisis de datos en tiempo real, además ofrece servicios como alta disponibilidad, replicación, manejo de *clúster*, integración con otros servicios de Amazon, entre otros.

3.6.1. Análisis de datos en tiempo real utilizando Big Data

El uso del Big Data permite de una forma más fácil la localización geográfica del *target* que maneja cada empresa, por lo que se pueden obtener los puntos de venta, reducir los costos y realizar campañas de *marketing*, entre otras actividades. Se presentan los tipos de análisis en tiempo real:

- *Data Stream*: analiza los datos mientras están en movimiento, antes que se almacenen en un destino final.
- *Complex Event Processing*: se enfoca en el análisis basado en los eventos que puede desencadenar un proceso, por lo que se buscan eventos predefinidos en tiempo real.

Estas técnicas al ser ejecutadas pueden minimizar el tiempo y riesgos al obtener la información en el momento lo que produce que tengan los siguientes beneficios:

- Detección en tiempo real, se pueden hallar anomalías en los flujos comerciales utilizando algoritmos de aprendizaje automático.

- Categorización de operaciones, es decir búsqueda de los datos más importantes para su análisis.
- Uso de análisis estadístico y procesamiento del lenguaje natural de datos en redes sociales y otras fuentes.
- Realizar análisis de sentimientos, con lo cual se pueden dar recomendaciones de productos o servicios personalizados a los clientes.

3.6.2. Software en análisis predictivo

En el mercado del software existen aplicaciones que analizan los datos en tiempo real para obtener un resultado efectivo en el momento que se le necesita. A continuación, se describen algunas herramientas de software de análisis que se relacionan con el uso del Big Data y son las siguientes:

- SAP Infinite Insight: es una herramienta que elabora predicciones basadas en la administración de un gran volumen de datos y con los resultados obtenidos se puede tener la capacidad de anticiparse a las necesidades de la empresa, además es una herramienta muy intuitiva y visual. También permite hacer simulaciones y no dejar datos en blanco, los cuales podría generar predicciones erróneas.
- Apache Spark: es un motor de procesamiento de código abierto que incluye lenguaje SQL, procesamiento gráfico y que corre sobre Hadoop. Usa un sistema de almacenamiento llamado Shark.
- Apache Storm: es un software libre y abierto para recuperar flujos de datos en tiempo real, el cual se puede utilizar con cualquier tipo de lenguaje de

programación. Permite particionar los datos de entrada y redistribuirlos por diferentes componentes.

Al usar las herramientas para el análisis predictivo sobre grandes volúmenes de datos se convierte en factor que marca la diferencia en el momento de comercializar un producto o servicio para que sea rentable y así obtener ganancias.

4. APLICACIÓN DE PRUEBAS A BASES DE DATOS NOSQL PARA MANEJO DE BIG DATA

Este capítulo contiene detalles de pruebas realizadas acerca de inserciones y consultas de datos masivos entre una base de datos SQL y una base de datos NoSQL; y comparar costos de implementación para elegir una opción entre las bases de datos presentadas para una toma de decisión orientada al marketing proactivo en tiempo real.

4.1. Definición de ambiente de pruebas

Las pruebas realizadas a las bases de datos se hicieron en las mismas condiciones en cuanto a hardware, software y disponibilidad de recursos, por lo cual las bases de datos cuentan con las mismas posibilidades sin ninguna desventaja proporcionada.

Se presentan las especificaciones técnicas con las que se realizaron las pruebas en un ambiente virtual, por lo que se describe el sistema operativo, los recursos utilizados y el software de virtualización utilizado:

- Sistema Operativo: Ubuntu 14.04 LTS (Trusty Tahr).
- Procesador: Intel® Core™ i7, utilizando 4 núcleos.
- Memoria RAM: 2048 MB.
- Capacidad de disco duro: 50 GB.
- Hipervisor: VirtualBox 5.12.

A continuación se describen las características en la tabla II de la base de datos Redis utilizada para el ambiente de prueba.

Tabla II. **Características técnicas base de datos Redis**

Característica	Descripción
Versión	2.8.4-2
Tipo	NoSQL clave-valor
ACID	No
Tamaño	ilimitado
Tipo de datos	Estructurados y no estructurados
Tamaño de la base de datos	Limitado
Tamaño máximo de RAM a usar	5 GB a 10 GB según el caso.
Sistemas operativos soportados	Linux, Mac OS, Unix

Fuente: Redis. <http://redis.io/documentation>. Consulta: mayo 2016.

Se describen los aspectos técnicos que se consideran importantes de la base de datos relacional, MySQL, que se utilizó en el ambiente de prueba.

Tabla III. **Características técnicas base de datos MySQL**

Característica	Descripción
Versión	5.6
Tipo	SQL Relacional
Tamaño	Limitado
ACID	Si
Tipo de datos	Estructurados
Tamaño de la base de datos	Limitado
Tamaño máximo de RAM a usar	Máximo de memoria RAM disponible
Sistemas operativos soportados	Linux, Unix, Windows y Mac OS

Fuente: MySQL. <http://dev.mysql.com/doc/refman/5.6/en/>. Consulta: mayo 2016.

Se define, en el siguiente apartado, el escenario de prueba utilizado para realizar la manipulación de los datos, de los cuales se toman estadísticas que se muestran más adelante, tomando en cuenta las especificaciones técnicas

descritas anteriormente y utilizando las bases de datos simultáneamente para el almacenamiento de datos.

4.1.1. Explicación técnica del ambiente de prueba

El escenario es una simulación de inserciones de datos en una red social simple que registra el envío de mensajes, posteriormente esos mensajes son consultados directamente sobre una base de datos NoSQL y una base de datos relacional.

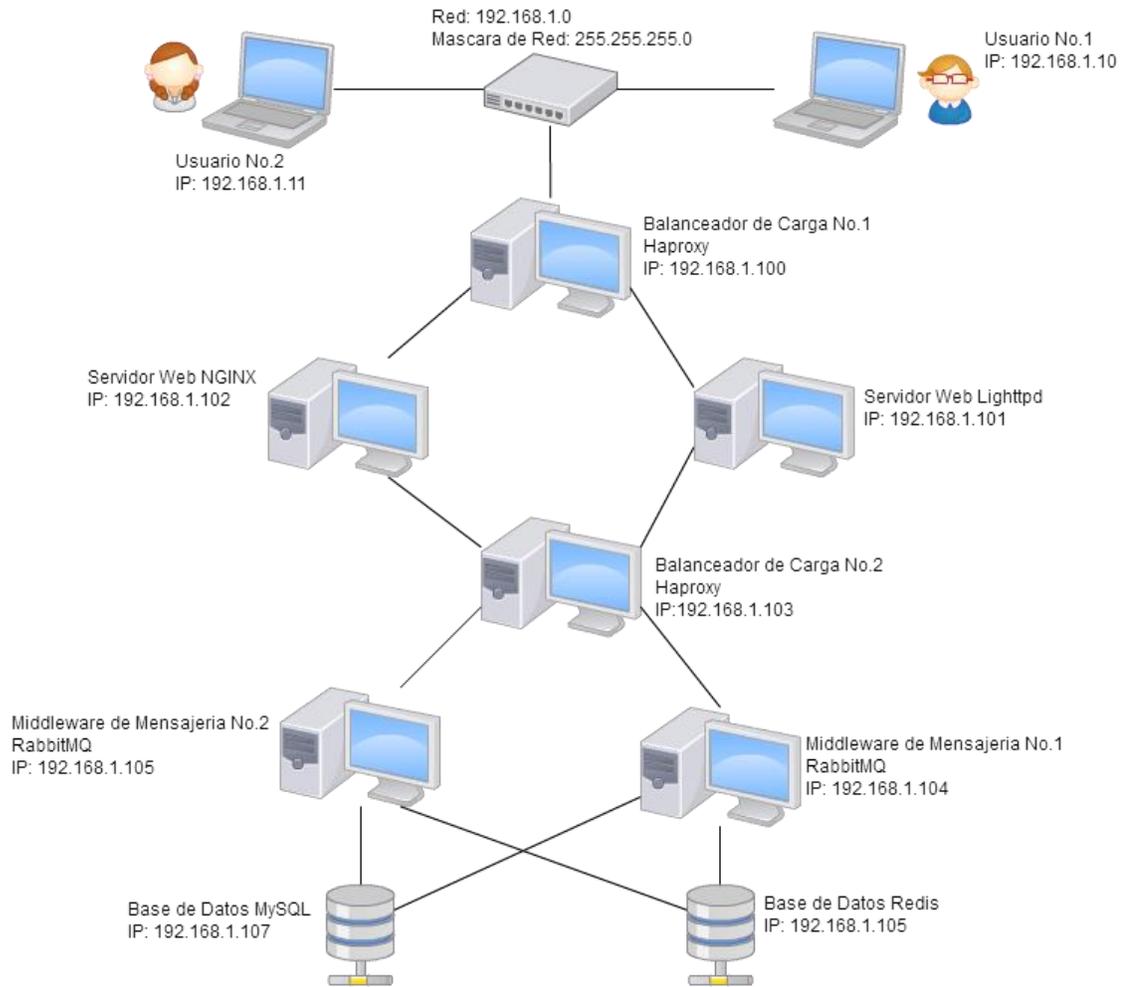
A continuación se detalla el funcionamiento de la infraestructura que se utilizó para realiza las pruebas.

1. Simulador de tráfico: aplicación creada utilizando el lenguaje de programación Python, compatible con cualquier sistema operativo con soporte para dicho lenguaje, que debe contar con las librerías wx-tools y urllib3. El simulador envía inserciones de datos por medio de una página web hacia la base de datos MySQL y Redis, durante 20 minutos, simulando 6 clientes.
2. Configuración de topología de red: se describen los pasos y el flujo que siguen los datos que son insertados, siendo el siguiente el proceso de ejecución.
 - Se coloca la URL de la página que envía las inserciones en el simulador de tráfico y el número de clientes que realiza las inserciones.

- Las solicitudes de inserción son enviadas a un balanceador de carga que distribuye el tráfico entre dos servidores web Nginx y Lighttpd.
- Los servidores reciben las solicitudes utilizando un *script* que emplea un lenguaje de programación php, el cual obtiene los parámetros que se insertaran de la URL completa y envía estos datos a un segundo balanceador de carga.
- El segundo balanceador de carga distribuye el tráfico entre dos *middleware* de mensajería RabbitMQ los cuales poseen dos colas de mensajes para Redis y MySQL.
- En ambas bases de datos se utiliza un script en lenguaje Python que recibe las peticiones y realiza las inserciones en las respectivas bases de datos.

En la figura 10 se muestra la descripción de la topología usada en el ambiente de prueba.

Figura 10. Topología de red



Fuente: elaboración propia.

4.1.2. Descripción del escenario de prueba

Las especificaciones que se evaluarán en el ambiente de prueba, se detallan a continuación:

Tabla IV. **Especificación del escenario de prueba**

Especificación	Descripción
Duración	20 minutos
Número de usuarios conectados	6
Total de datos a insertar	64000
Numero de repeticiones de prueba	10
Numero de computadores en red	8
Red utilizada	192.168.1.0

Fuente: elaboración propia.

Para almacenar los datos se utilizó una tabla en la base de datos relacional y una estructura de lista en la base de datos NoSQL, en la cual se simula el envío de mensajes entre usuarios, que tiene la siguiente estructura:

Tabla V. **Estructura de tabla/lista en las bases de datos**

Campo	Tipo	Descripción
id	Integer	Llave primaria/clave
idusr	Integer	Identificador de mensaje
username	Varchar	Nombre de usuario
msg	Varchar	Mensaje enviado

Fuente: elaboración propia.

4.1.3. Aspectos a evaluar

Se describen a continuación los aspectos que se evaluarán en el ambiente de prueba, y sobre los cuales se realizará el análisis respectivo para determinar la aplicación y el uso del tipo de bases de datos en el almacenamiento masivo de datos y su posterior consulta para tomar decisiones dependiendo del campo a inferir.

- Inserciones de datos efectivas a disco: es la cantidad de solicitudes enviadas por los usuarios y que se guardan en la base de datos y en disco duro.
- Porcentaje de consumo de CPU y memoria RAM: el consumo de recursos que se efectúen durante el uso de las bases de datos por medio de la inserción.
- Cantidad de Memoria RAM consumida: la memoria RAM física que se consume durante el uso de las bases de datos.
- Tiempo de respuesta: es el tiempo en el que los datos tardan en mostrarse al usuario para su consulta.
- Factores de costos de la implementación de bases de datos: es cuánto se debe invertir en el uso de una determinada base de datos para su uso y en que arquitectura se debe invertir.

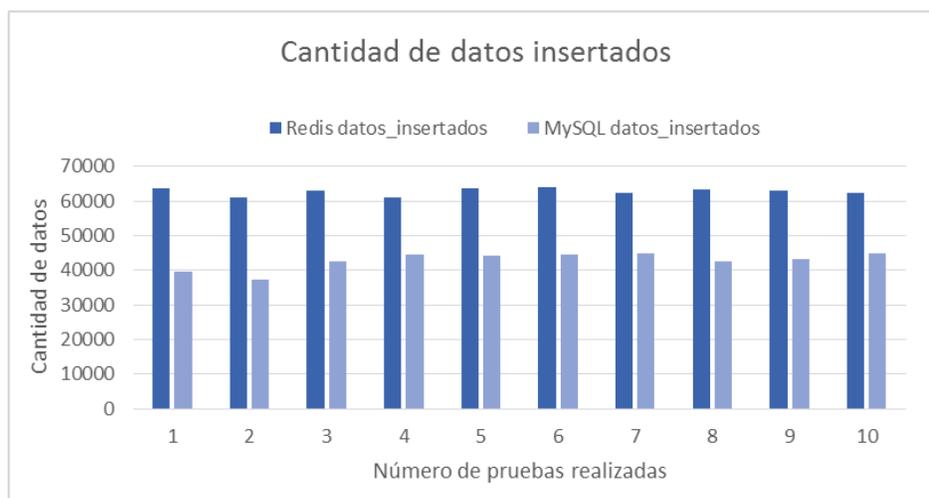
4.2. Análisis de los aspectos que se evalúan

Se describen los resultados obtenidos según el escenario descrito anteriormente, comparando las bases de datos SQL y NoSQL, en un mismo ambiente de prueba, los cuales se explican y se analizan realizando comparaciones de rendimiento.

4.2.1. Inserciones efectivas en disco

La prueba consiste en realizar diez intentos de inserción en las bases de datos, contando con un total de 64 000 datos a insertar en cada una, en un tiempo de 20 minutos; para que ningún dato se perdiera se utilizaron los servidores RabbitMQ para el manejo de colas de mensajes, por lo que en la figura 11 se muestra el número de datos que efectivamente fue almacenado en la base de datos. En la Tabla VI se puede observar la comparación entre los datos insertados entre las dos bases de datos MySQL y Redis

Figura 11. Cantidad de datos insertados en las bases de datos



Fuente: elaboración propia.

Tabla VI. **Diferencia en la inserción de datos**

Prueba	Redis		MySQL		Diferencia datos insertados Redis y MySQL
	Datos insertados	Datos no insertados	Datos insertados	Datos no insertados	
1	63649	351	39716	24284	23933
2	61136	2864	37180	26820	23956
3	63135	865	42442	21558	20693
4	61011	2989	44645	19355	16366
5	63664	336	44218	19782	19446
6	63993	7	44604	19396	19389
7	62448	1552	44908	19092	17540
8	63316	684	42448	21552	20868
9	62903	1097	43388	20612	19515
10	62302	1698	44761	19239	17541

Fuente: elaboración propia.

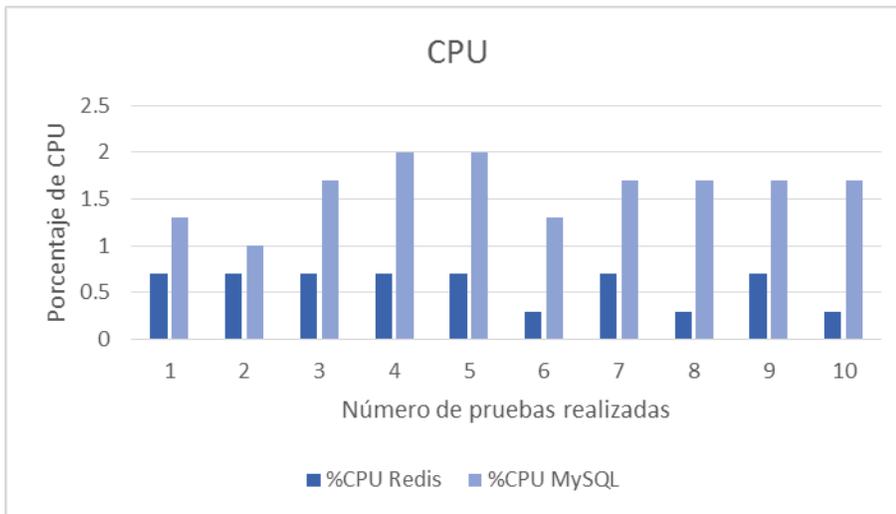
De acuerdo con los resultados se puede observar que las cantidades de datos insertados en ambas bases de datos difieren en cada intento que se efectuó. Por lo que se observa que las bases relacionales insertan menos datos que las bases de datos NoSQL. Se necesita consultar estos datos en tiempo real para realizar estrategias de *marketing*. La opción viable son las bases de datos NoSQL.

4.2.2. **Porcentaje de consumo de CPU y memoria RAM**

Durante el tiempo en que se realizaron las pruebas, se monitoreó el consumo del uso de memoria RAM y del CPU de los computadores en donde se alojaban las bases de datos para verificar el estado de cada uno y a cuanto se elevaba el consumo de estos recursos con el volumen de datos insertados. En las figuras 11 y 12 se muestra el porcentaje de consumo de recursos que se

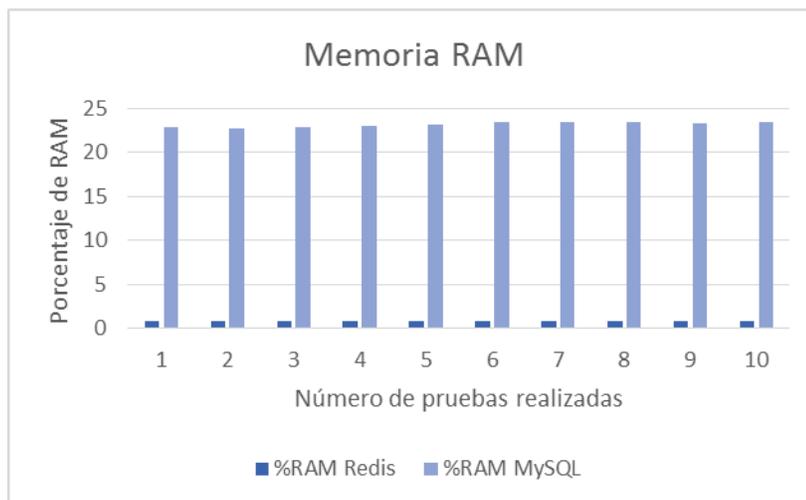
realizó en cada una de las pruebas por cada base de datos, siendo el máximo de CPU utilizado 2.5 % y 25 % de Memoria RAM.

Figura 12. **Porcentaje de consumo del CPU**



Fuente: elaboración propia.

Figura 13. **Porcentaje de consumo de memoria RAM**



Fuente: elaboración propia.

En la siguiente tabla se muestra el promedio de los porcentajes mostrados en la inserción de datos en las bases de datos.

Tabla VII. **Promedio de uso de CPU y memoria RAM**

Porcentaje uso CPU (%)		Porcentaje uso, memoria RAM (%)	
Redis	MySQL	Redis	MySQL
0.58	1.61	0.87	23.18

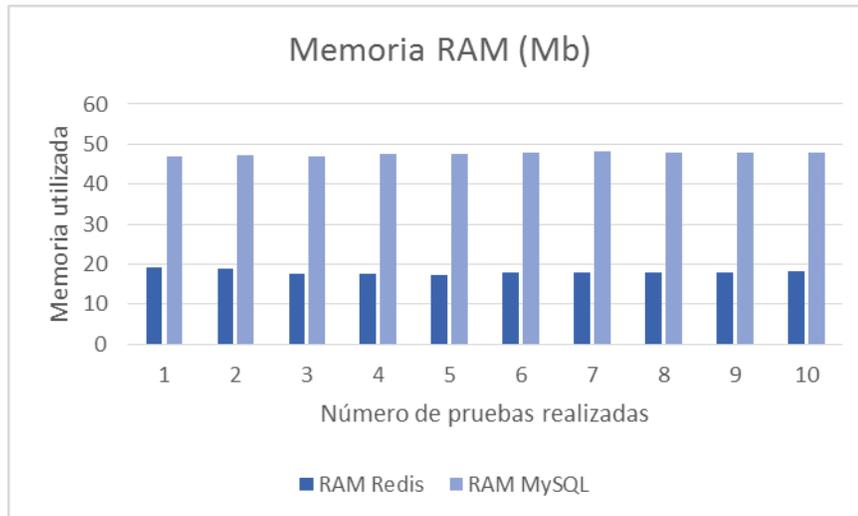
Fuente: elaboración propia.

Como se puede observar, según los resultados anteriores, el uso de una base de datos relacional como MySQL, puede afectar el rendimiento del equipo utilizado, ya que mientras el volumen de datos aumenta también el consumo de recursos. Se puede ver que el recurso más afectado es la memoria RAM. Mientras una base de datos NoSQL como Redis, se mantiene estable en el consumo de recursos de CPU y memoria RAM.

4.2.3. Memoria RAM física consumida

En el apartado anterior se observó que la memoria RAM es la más afectada en la inserción de datos masivos en una base de datos relacional, por lo que se muestra en la figura14 el comportamiento de esta en MB durante el tiempo de la prueba, siendo el máximo de memoria RAM 60 MB.

Figura 14. **Memoria RAM física consumida**



Fuente: elaboración propia.

En la tabla siguiente se muestra el promedio del consumo de memoria física que tienen los computadores que alojan las dos bases de datos durante la prueba.

Tabla VIII. **Promedio de uso de memoria RAM**

Memoria RAM (MB)	
Redis	MySQL
18.01	47.53

Fuente: elaboración propia.

Según los resultados obtenidos se observa que el uso de memoria es superado por la base de datos relacional por una diferencia de 29.52104 MB, en comparación con la base de datos NoSQL. Con esto se puede tener una mejor perspectiva en la momento de tomar la decisión de usar una u otra base de datos

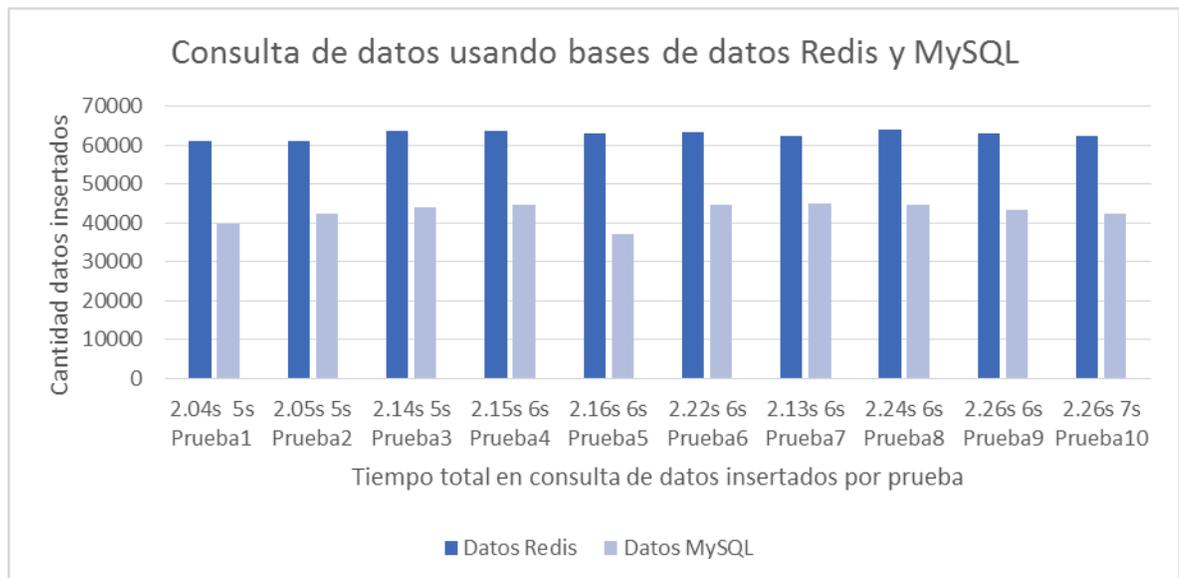
para el almacenamiento masivo, dado que mientras más datos sean almacenado o bien consultados, el consumo de la memoria RAM será mayor.

4.2.4. Tiempo de respuesta

Uno de los aspectos que se toman más en cuenta en el momento de inferir en una empresa, es que tan rápido se pueden consultar los datos y que tan actualizados están, por lo que se realiza una comparación para verificar que tan rápido se puede acceder a ellos mediante una base de datos relaciona y una base de datos NoSQL.

En la figura 15 se comparan los datos que se obtienen a través del tiempo, comparando MySQL con Redis.

Figura 15. Consulta de datos en la base de datos Redis y MySQL



Fuente: elaboración propia.

En la siguiente tabla se muestra el promedio de datos insertados y el promedio de tiempo que se tarda en consultarlos una base de datos.

Tabla IX. **Promedio de datos insertados versus tiempo**

Promedio de tiempo acceso (s)		Promedio de datos leídos	
Redis	MySQL	Redis	MySQL
2.15	5.8	62756	42831

Fuente: elaboración propia.

En la gráfica se puede observar que, aunque la base de datos MySQL tiene menos datos, la consulta se demora más. Y Redis siendo una base de datos NoSQL, puede acceder a una mayor cantidad de datos en menos tiempo. Por lo que se puede concluir que entre más datos se pueda almacenar en una base de datos relacional el consumo de memoria crecerá. Además, este tipo de bases de datos no soportan datos no estructurados.

4.2.5. Factores que influyen en costos de implementación

El costo de implementación es el precio que tiene una aplicación o herramienta de software en el momento en que se compra, se adapta y se implementa para su uso en un entorno de trabajo, ya sea comercial o no.

A continuación, se mencionan algunos de los factores más importantes que se consideran en el momento de comprar un servicio en la Nube orientado a bases de datos.

- Nodos o servidores.
- Manejo de zonas geográficas.

- Número de conexiones permitidas para consumir el servicio.
- Replicación de datos.
- Persistencia de los datos.
- Número de bases de datos que puede contener cada instancia del servicio en la Nube.
- Número de procesadores y la capacidad de CPU de la instancia en la Nube.
- Capacidad de almacenamiento memoria RAM.
- *Auto failover* del sistema de datos.
- Crear clúster de nodos o servidores.
- Cantidad de tiempo que el servicio estará activo y se podrá acceder a él.
- Licenciamiento del servicio a utilizar.
- Mano de obra encargada de administrar los servicios en la Nube.

Estos influyen en la elección de una base de datos NoSQL, pero hay otros costos ocultos que no se consideran hasta el momento en cuando se comienza a adaptar la base de datos a las necesidades del negocio.

4.3. Elementos que intervienen en el costo de implementación

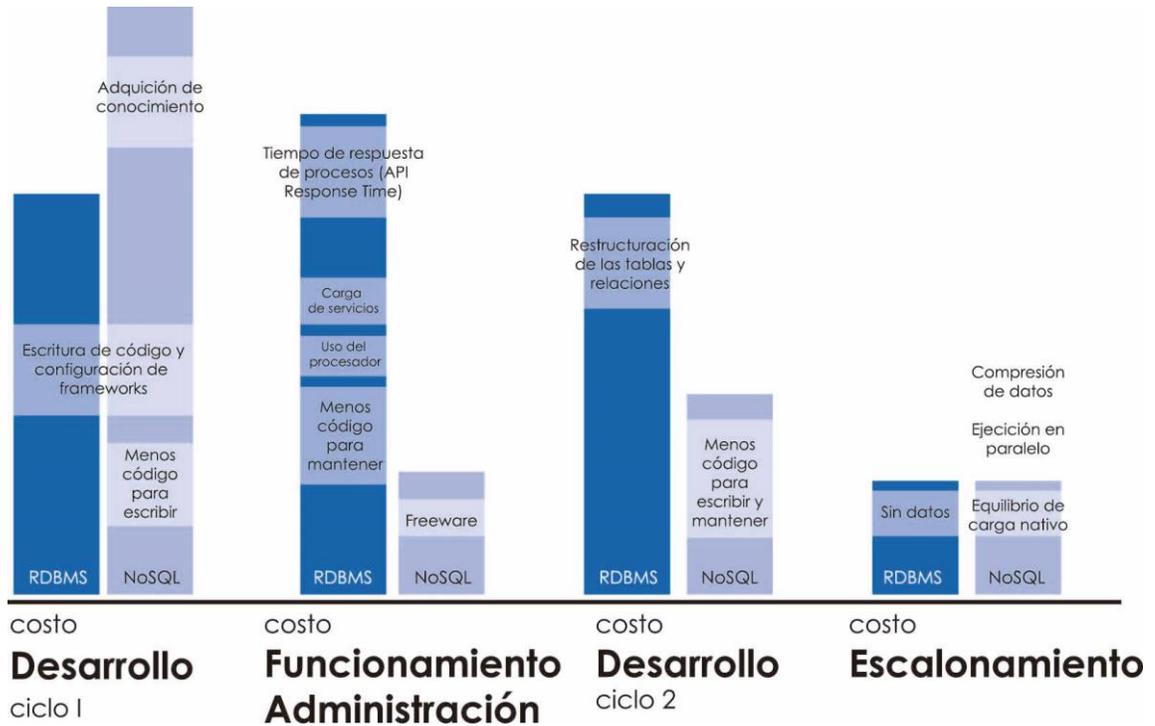
Existen cuatro elementos que intervienen en el costo de la implementación dentro del ciclo de vida que puede tener el uso de una base de datos NoSQL que se relaciona con el Big Data Marketing, los cuales se describen a continuación:

- Costo de desarrollo: incluye la escritura de código necesaria para adaptación de la base de datos y la configuración de los ambientes de trabajo. En esta etapa el costo es relativo, ya que depende de la empresa y del dominio que tenga para el manejo de un entorno nuevo.

- Costo de funcionamiento y administración: mide el consumo de hardware, la disponibilidad de recursos, el consumo de energía, los canales de comunicación. En síntesis la infraestructura necesaria para desarrollar el entorno.
- Costo de desarrollo: el entorno de la empresa cambia y la tecnología también. Se debe contar con una estructura que mantenga la flexibilidad de cambiar lo que sea necesario en el ambiente ya establecido y que su mantenimiento sea constante.
- Costo de escalonamiento de la base de datos: se escalona la infraestructura para poder tener los datos disponibles y que los mismos puedan crecer y tener un equilibrio en la carga.

Los costos son relativos dependiendo del modelo de negocio y del presupuesto de cada empresa que desee implementar Big Data Marketing utilizando bases de datos NoSQL, por lo que se presenta un estimado de la comparación entre bases de datos relacionales SQL y las NoSQL.

Figura 16. Costos de implementación de una base de datos NoSQL



Fuente: *Los costos ocultos de NoSQL*. <http://ehealthnosql.blogspot.com/2013/06/los-costos-ocultos-de-nosql.html>. Consulta: 27 mayo 2016.

En la gráfica anterior se puede analizar que los costos relativos de implementación en los que se involucran el hardware, la escritura de código en lenguaje de programación, así como la escalabilidad de los recursos y la comprensión de los mismos son menores en comparación con la implementación de una base de datos relacional en el ámbito del Big Data.

4.4. Evaluación de resultados

Según los resultados de las pruebas realizadas a las dos bases de datos se concluye que las bases de datos NoSQL son la mejor opción para su uso en el almacenamiento masivo de datos utilizados en el *marketing*.

En la siguiente tabla se resumen los aspectos evaluados en el ambiente de prueba, y se usan los números del 1 al 3 para indicar el mejor resultado, siendo 3 mejor, 2 regular y 1 deficiente.

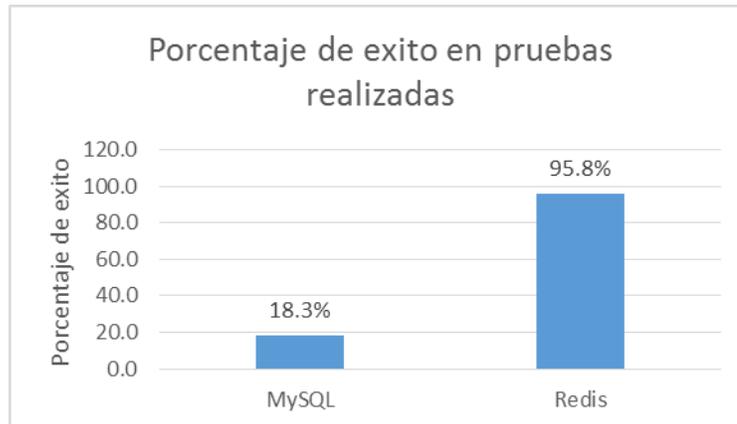
Tabla X. **Resumen de pruebas**

Aspecto evaluado	MySQL	Redis
Mayor número de inserciones efectivas a disco	1	3
Menor porcentaje de consumo de CPU	1	3
Menor porcentaje de consumo de memoria RAM	1	3
Menor cantidad de memoria RAM física consumida	1	3
Mejor tiempo de respuesta en consulta de datos	1	2
Menor costo de implementación	1	3

Fuente: elaboración propia.

El porcentaje de éxito en el ambiente de pruebas al implementarlas en una base de datos NoSQL, en este caso Redis, y una base de datos relacional MySQL se puede observar en la figura.

Figura 17. **Porcentaje de éxito en los aspectos evaluados**



Fuente: elaboración propia.

Como se muestra en la figura anterior la mejor opción para la implementación de bases de datos en el ámbito del marketing, es una base de datos NoSQL ya que tiene un porcentaje de un 95,8 % de resultados satisfactorios, el cual es mayor al porcentaje obtenido por la base de datos relacional con un 18,3 %.

Es por eso que grandes empresas como Facebook, Google y Twitter, entre otras, usan para administrar sus datos masivos las bases de datos NoSQL, ya que proveen una respuesta menor en el consumo de recursos y reduce los costos.

El uso de las bases de datos NoSQL depende de la naturaleza de los datos, es decir, de si se debe asegurar la atomicidad, consistencia e integridad referencial de los datos; por ejemplo, en las transacciones bancarias no es recomendable el uso de las bases NoSQL, ya que la mayoría de este tipo de bases de datos no considera estos aspectos. Por el contrario, si se desea tener gran escalabilidad de los datos, sin importar aspectos como la integridad,

atomicidad o la consistencia de los mismos, y se requiere la velocidad en la consulta de los datos para realizar análisis de los mismos, una buena opción son las bases de datos NoSQL.

CONCLUSIONES

1. Es posible usar las bases de datos NoSQL y sus aplicaciones de análisis como software con enfoque en el Big Data para tomar decisiones y estrategias de marketing en una empresa.
2. Las bases de datos NoSQL aportan la estructura necesaria para el almacenamiento masivo de datos, lo que produce que se posean más datos y que se consulten en menos tiempo con costos más bajos.
3. El análisis de datos debe realizarse correctamente para obtener información que aporte valor a una empresa en sus decisiones aplicadas a productos y servicios al cliente, por lo que existen varias tendencias actualmente que pretenden usar la inteligencia artificial como una ayuda en el Big Data.
4. El reconocimiento de patrones y tendencias por medio del uso de la minería de datos, beneficia el aprendizaje automático de las aplicaciones de software que realizan el análisis de los datos que aplican el Big Data Marketing.
5. Un profesional en informática puede aplicar conocimientos de programación para administrar o crear herramientas de software en el análisis de Big Data Marketing y seleccionar datos que den valor al negocio, como en el caso de la publicidad programática.

RECOMENDACIONES

1. Realizar un análisis del modelo del negocio de la empresa que pretende adoptar el Big Data, para que no tenga inconvenientes posteriores, así como también seguir las fases de la adopción del Big Data.
2. Proporcionar al ingeniero en informática un enfoque administrativo en el área de *marketing* para que conozca más campos de aplicación de su carrera.
3. Analizar las necesidades de la empresa y adaptarlas a un tipo de base de datos NoSQL conveniente, como lo puede ser documental, de grafos, columnar, clave valor, o bien orientada a objetos; y establecer el impacto que tiene al aplicarlo al Big Data, verificando el tipo de transacción para el cual se requiere utilizar la base de datos NoSQL.

BIBLIOGRAFÍA

1. ALONSO, Alberto. *Big Data derivado del marketing digital*. [en línea]. <<http://www.harvarddeustobt.com/big-data-marketing-digital/>> [Consulta: 21 de abril de 2016].
2. ALTIERI, Claudia. *Marketing Proactivo*. [en línea]. <http://www.deguate.com/artman/publish/gestion_merca/Marketing_Proactivo_2981.shtml#.V0Tw67h97Df> [Consulta: 18 marzo 2016].
3. AMAZON WEB SERVICE. *What Is Amazon Elasticsearch Service?* [en línea]. <<http://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/what-is-amazon-elasticsearch-service.html>> [Consulta: 5 de junio de 2016].
4. APACHE CASSANDRA. *What is Cassandra?* [en línea]. <<http://cassandra.apache.org>> [Consulta: 5 de junio de 2016].
5. APACHE HADOOP. *Hadoop Cluster Setup*. [en línea]. <<https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/ClusterSetup.html>> [Consulta: 15 de julio de 2016].
6. APACHE HIVE. *Apache Hive TM*. [en línea]. <<https://hive.apache.org/>> [Consulta: 11 de julio de 2016].

7. APACHE KAFKA. *Kafka™ is a distributed streaming platform. What exactly does that mean?* [en línea]. <<http://kafka.apache.org/>> [Consulta: 5 de mayo de 2016].
8. APACHE STORM. *Apache Storm*. [en línea]. <<http://storm.apache.org/>> [Consulta: 6 de mayo de 2016].
9. ARISER. *Historia del Big Data – Del comienzo del análisis de datos a nuestros días*. [en línea]. <<https://bigdataparacuriosos.wordpress.com/historia-big-data/>> [Consulta: 30 de abril de 2016].
10. BAQUIA, Manuel Daza. *¿Es el Big Data el fin de la Business Intelligence?* [en línea]. <<http://www.baquia.com/emprendedores/2013-07-19-manuel-daza-bigdata-business-intelligence-bi-discovery-informacion-datos-empresas-olap-analisis-qlikview-ibm-smart-cities-redes-sociales.>> [Consulta: 21 de agosto de 2016].
11. BARRANCO FRAGOSO, Ricardo. *¿Qué es Big Data?* [en línea]. <<https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>> [Consulta: 20 de febrero de 2016].
12. BBC. *La ciencia de encontrar perlas en el Big Data*. [en línea]. <http://www.bbc.com/mundo/noticias/2013/03/130310_tecnologia_big_data_datos_informacion_perla_dp.> [Consulta: 20 de abril de 2016].

13. BIGELOW, Stephen J. *¿Qué modelo de nube es mejor para su proyecto de big data?* [en línea]. <<http://searchdatacenter.techtarget.com/es/respuesta/Que-modelo-de-nube-es-mejor-para-su-proyecto-de-big-data>> [Consulta: 15 de mayo de 2016].
14. BORJA. *¿Qué es Redis?* [en línea]. <<http://www.antweb.es/servidores/redis-todo-lo-que-debes-saber>> [Consulta: 24 de abril de 2016].
15. BRAD HEDLUND. *Understanding Hadoop Clusters and the Network* [en línea]. <<http://bradhedlund.com/2011/09/10/understanding-hadoop-clusters-and-the-network/>> [Consulta: 3 mayo de 2016].
16. BREEUWER, Dick. *Big Data Marketing: Los 3 componentes esenciales para una estrategia exitosa.* [en línea]. <<http://www.inboundcycle.com/blog-de-inbound-marketing/big-data-marketing-los-3-componentes-esenciales-para-una-estrategia-exitosa>> [Consulta: 22 de febrero de 2016].
17. BROZINSKY, Murray. *The future of Big Data: Deep Learning* [en línea]. <<http://www.talix.com/blog/the-future-of-big-data-deep-learning/>> [Consulta: 10 de abril de 2016].
18. CANTABRIATIC. *NoSQL y el teorema de CAP* [en línea]. <<http://www.cantabriatic.com/nosql-y-el-teorema-de-cap/>> [Consulta: 17 mayo de 2016].

19. CIA, Juan. *Neo4j: qué es y para qué sirve una base de datos orientada a grafos* [en línea]. <<http://www.bbvaopen4u.com/es/actualidad/neo4j-que-es-y-para-que-sirve-una-base-de-datos-orientada-grafos>> [Consulta: 20 de abril de 2016].
20. CIF. *Cluster* [en línea]. <<http://suseos.weebly.com/uploads/1/8/7/0/18707470/23-cluster.pdf>> [Consulta: 14 mayo de 2016].
21. CLUB PLANETA. *De las 4P hacia las 4C de la mercadotecnia* [en línea]. <http://www.trabajo.com.mx/de_las_4p_hacia_las_4_cs_de_la_mercadotecnia.htm> [Consulta: 14 abril de 2016].
22. CORTES Díaz, Estuardo. *Dos de tres, el teorema CAP* [en línea]. <<http://www.lnds.net/blog/2012/05/dos-de-tres.html>> [Consulta: 16 mayo de 2016].
23. COUCH DB. *Data Where You Need It*. [en línea]. <<http://couchdb.apache.org>> [Consulta: 8 de abril de 2016].
24. CRECE NEGOCIOS. *Cadena de valor de Porter*. [en línea]. <<http://www.crecenegocios.com/la-cadena-de-valor-de-porter/>> [Consulta: 15 marzo de 2016].
25. DOMÍNGUEZ, Alexandra. *5 Grandes beneficios que el Big Data Aporta a las marcas*. [en línea]. <<http://www.merca20.com/5-grandes-beneficios-que-el-big-data-apota-a-las-marcas/>> [Consulta: 19 mayo de 2016].

26. GARCÍA BARBOSA, Julián. Cloud computing, *el mejor camino para acceder al big data*. [en línea]. <<http://aunclidelastic.blogthinkbig.com/cloud-computing-mejor-camino-para-acceder-al-big-data/>> [Consulta: 10 mayo 2016].
27. GARCÍA, Luis Miguel. *Arquitectura Lambda: Principios de Arquitectura para sistemas Big Data en tiempo real*. [en línea]. <<https://unpocodejava.wordpress.com/2013/09/07/arquitectura-lambda-principios-de-arquitectura-para-sistemas-big-data-en-tiempo-real/>> [Consulta: 10 de abril de 2016].
28. GENBETA DEV. *MongoDB: qué es, cómo funciona y cuándo podemos usarlo (o no)*. [en línea]. <<http://www.genbetadev.com/bases-de-datos/mongodb-que-es-como-funciona-y-cuando-podemos-usarlo-o-no>> [Consulta: 29 de abril de 2016].
29. GENBETA DEV. *NoSQL: Clasificación de las bases de datos según el teorema CAP* [en línea]. <<http://www.genbetadev.com/bases-de-datos/nosql-clasificacion-de-las-bases-de-datos-segun-el-teorema-cap>> [Consulta: 15 mayo 2016].
30. GESTIOPOLIS. *¿Qué es la cadena de valor?* [en línea]. <<http://www.gestiopolis.com/que-es-la-cadena-de-valor/>> [Consulta: 3 marzo 2016].
31. GIL, Gabriel. *Big Data y marketing digital: una relación imprescindible* [en línea]. <<http://www.educacionline.com/instituto-de-marketing-online/big-data-y-marketing-digital-una-relacion-imprescindible/>> [Consulta: 2 de febrero de 2016].

32. GÓMEZ Abajo, Carlos. *El big data y la nube unen sus fuerzas para revolucionar la sociedad* [en línea]. <http://www.tendencias21.net/El-big-data-y-la-nube-unen-sus-fuerzas-para-revolucionar-la-sociedad_a39130.html> [Consulta: 9 mayo de 2016].
33. GÓMEZ, Orlando. *Experiencia de Usuario y ROI: ¿Cuánto es suficiente?* [en línea]. <<http://orlandogomez.co/index.php/2016/01/23/experiencia-de-usuario/>> [Consulta: 12 de febrero de 2016].
34. GOOGLE CLOUD PLATAFORM. *Cloud Bigtable Documentation*. [en línea]. <<https://cloud.google.com/bigtable/docs/>> [Consulta: 12 de abril de 2016].
35. GTI. *5 diferencias entre Big Data y Bussiness intelligence* [en línea]. <<http://noticias.gti.es/productos/5-diferencias-entre-big-data-y-business-intelligence/>> [Consulta: 30 abril de 2016].
36. GUTIÉRREZ Amaya, Camilo. *Lo que representa Big Data para la seguridad de la información* [en línea]. <<http://www.welivesecurity.com/la-es/2014/01/29/que-representa-big-data-seguridad-informacion/>> [Consulta: 3 mayo de 2016].
37. HEREDIA, Ramón. *BI, Minería de Datos y Big Data: Oportunidades para la Banca Digital* [en línea]. <<http://reportedigital.com/m2m/bi-mineria-datos-big-data-oportunidades-banca-digital/>> [Consulta: 16 mayo de 2016].

38. HO, Ricky. *CouchDB Cluster*. [en línea]. <<http://horicky.blogspot.com/2008/10/couchdb-cluster.html>> [Consulta: 30 de julio de 2016].
39. HPCC SYSTEMS. *Why HPCC Systems?* [en línea]. <<https://hpccsystems.com/why-hpcc-systems>> [Consulta: 20 de junio de 2016].
40. IBM. *What is big data?* [En línea]. <<http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>> [Consulta: 21 de febrero 2016].
41. IBM. *What is cloud computing?* [en línea]. <<https://www.ibm.com/cloud-computing/learn-more/what-is-cloud-computing/>> [Consulta: 8 de agosto de 2016].
42. IEEEEXPLORE. *Big Data Deep Learning: Challenges and Perspectives* [en línea]. <<http://ieeexplore.ieee.org/ielx7/6287639/6705689/06817512.pdf?arnumber=6817512>> [Consulta: 17 mayo de 2016].
43. INTERACTIVE Advertising Bureau. *Libro Blanco de la compra programática*. [en línea]. <<http://www.iabspain.net/wp-content/uploads/downloads/2014/09/Libro-blanco-de-Compra-Program%C3%A1tica-y-RTB.pdf>> [Consulta: 22 de abril de 2016].
44. LANTARES. *Big Data: ventajas de la revolución de los datos masivos*. [en línea]. <<http://www.lantares.com/blog/big-data-ventajas-de-la-revolucion-de-los-datos-masivos>> [Consulta: 17 marzo de 2016].

45. LO, Frank. *Big Data Technology. What is Hadoop? What is MapReduce? What is NoSQL?* [en línea]. <<https://datajobs.com/what-is-hadoop-and-nosql>> [Consulta: 15 de febrero de 2016].
46. LOBOS, Edgardo. *Las 4 V's de Big Data* [en línea]. <<https://medium.com/@elobosbable/las-4-vs-de-big-data-f7cd441ac31a#.q9gvfheuj>> [Consulta: 1 de mayo de 2016].
47. MARQUINA, Julián. *La vida en la nube: Big data y cloud computing* [en línea]. <<http://www.julianmarquina.es/la-vida-en-la-nube-big-data-y-cloud-computing/>> [Consulta: 23 de abril de 2016].
48. MARTÍNEZ, Desiree. *Big Data, Customer Intelligence, Recopilatorio 2015*. [en línea]. <<http://artyco.com/principales-diferencias-entre-business-intelligence-y-big-data/>> [Consulta: 8 de julio de 2016].
49. McCLEAN, Roy. *The C's and P's of marketing. What's the difference?* [en línea]. <<https://www.customfitonline.com/news/2012/10/19/4-cs-versus-the-4-ps-of-marketing/>> [Consulta: 2 mayo de 2016].
50. MONGODB. *Real-time Analytics* [en línea]. <<https://www.mongodb.com/use-cases/real-time-analytics>> [Consulta: 27 abril de 2016].
51. MORALES, Aurelio. *Sobre MongoDB, bases de datos NoSQL y GIS* [en línea]. <<http://mappinggis.com/2014/07/mongodb-y-gis/>> [Consulta: 23 de abril de 2016].

52. O'NEILL, Colin. *Calculando el ROI para la mejora de proceso* [en línea]. <<https://www.ibm.com/developerworks/ssa/rational/library/edge/09/mar09/>> [Consulta: 22 de febrero de 2016].
53. PIXEL CREATIVO. *Las 4C del Marketing* [en línea]. <<http://pixel-creativo.blogspot.com/2011/10/las-4-c-del-marketing.html>> [Consulta: 28 de abril de 2016].
54. POWER DATA. *5 ventajas de la arquitectura de Hadoop* [en línea]. <<http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/402826/5-ventajas-de-la-arquitectura-de-Hadoop>> [Consulta: 14 de marzo de 2016].
55. PULSO Negocios. *Big Data y segmentación: ¿cómo podemos aprovechar esta tendencia para conocer mejor al consumidor?* [en línea]. <<http://negocios.pulzo.com/big-data-y-segmentacion-como-podemos-aprovechar-esta-tendencia-para-conocer-mejor-al-consumidor/>> [Consulta: 15 de mayo de 2016].
56. PUROMARKETING. *Big Data: ¿de dónde salen todos esos datos?* [en línea]. <<http://www.puromarketing.com/12/22562/big-data-donde-salen-todos-esos-datos.html>> [Consulta: 1 de abril de 2016].
57. QUANTIC SOLUTIONS. *Las 5 V's del Big Data* [en línea]. <<http://www.quanticsolutions.es/quantic/las-5-vs-del-big-data/>> [Consulta: 20 febrero de 2016].

58. QUESADA Palacios, José Antonio. *La 'Nube', Big Data y otros retos para las empresas* [en línea]. <<http://expansion.mx/opinion/2016/01/05/hacia-una-nueva-economia-un-enfoque-disruptivo-en-negocios>> [Consulta: 4 de mayo de 2016].
59. REDIS. *Redis Cluster Specification*. [en línea]. <<https://redis.io/topics/cluster-spec>> [Consulta: 29 de julio de 2016].
60. REVISTA Cloud Computing. *MongoDB, ¿son las bases no relacionales el futuro?* [en línea]. <<http://www.revistacloudcomputing.com/2014/06/mongodb-son-las-bases-de-datos-no-relacionales-el-futuro/>> [Consulta: 15 de abril de 2016].
61. SÁNCHEZ, Jorge. *Neo4j una base de datos NoSQL orientada a grafos* [en línea]. <<http://xurxodeveloper.blogspot.com/2014/03/neo4j-una-base-de-datos-nosql-orientada.html>> [Consulta: 20 de marzo de 2016].
62. SCALE Unlimited. *What is Web Mining?* [en línea]. <<http://www.scaleunlimited.com/about/web-mining/>> [Consulta: 13 de mayo de 2016].
63. SINNEXUS. *¿Qué es Bussiness Intelligence?* [en línea]. <http://www.sinnexus.com/business_intelligence/> [Consulta: 4 mayo 2016].

64. SKYTREE. *Why do Machine Learning on Big Data?* [en línea]. <<http://www.skytree.net/machine-learning/why-do-machine-learning-big-data/>> [Consulta: 3 de mayo de 2016].
65. STUNTEBECK, Vincent. *Segmentación de clientes B2B* [en línea]. <<http://www.ibm.com/developerworks/ssa/library/ba-b2b-custseg-spss/>> [Consulta: 12 de abril de 2016].
66. TIRADOS, Maribel. *Análisis de datos en tiempo real: Spark y Storm* [en línea]. <<http://www.bigdatahispano.org/noticias/spark-y-storm/>> [Consulta: 6 de mayo de 2016].
67. TIRADOS, Maribel. *Big Data y seguridad en la nube* [en línea]. <<http://www.bigdatahispano.org/noticias/big-data-y-seguridad-en-la-nube/>> [Consulta: 3 de mayo de 2016].
68. VALENCIA, Edison. *Los Costos ocultos de NoSQL* [en línea]. <<http://ehealthnosql.blogspot.com/2013/06/los-costos-ocultos-de-nosql.html>> [Consulta: 27 de mayo de 2016].
69. VERNIA, Sonia. *Qué es el Big Data y Cómo usarlo en el marketing.* [en línea]. <<http://comunidad.iebschool.com/iebs/marketing-digital/como-aplicar-el-big-data-en-el-marketing/>> [Consulta: 30 de marzo de 2016].
70. WAXTER, Cindy. *Big Data: Los peligros de la minería de datos* [en línea]. <<http://cioperu.pe/articulo/14644/big-data-los-peligros-de-la-mineria-de-datos/>> [Consulta: 2 de mayo de 2016].

