

BIG DATA, UN SISTEMA DE GESTIÓN DE DATOS

Autor: **Ángel Gómez Degrares**
PhD in Applied Statistics and Intelligent Organizations,
Tecana American University, of the USA

gomezdegraves@gmail.com

(2021)

ÍNDICE

ÍNDICE DE TABLAS.....	3
ÍNDICE DE FIGURAS.....	4
Resumen.....	5
INTRODUCCIÓN.....	6
Definición de big data.....	7
2. Tipos de datos big data.....	8
3. Dimensiones del big data.....	10
4. Herramientas para el tratamiento de big data.....	11
4.1. Tecnologías.....	11
4.2. Técnicas.....	13
Conclusiones.....	15
.Referencias bibliográficas.....	16

Índice de tablas

Tabla 1. Datos semiestructurados.....10

Tabla 2. Tipos de datos en big data.....10

Índice de figuras

Figura 1. La “V” de big data11

RESUMEN

La presente investigación tuvo por objeto analizar el auge del big data, una nueva revolución industrial en la transformación de las empresas como valor diferencial del análisis de los datos en los diversos sectores. Se realizó una investigación exploratoria documental, mediante un diseño bibliográfico, hacia la búsqueda, recuperación, análisis crítico y riguroso e interpretación de la información, a través de fuentes documentales impresas y electrónicas. La tecnología big data, cada día encuentra más usuarios y se está posicionando a nivel mundial en su aplicación, optimización y adaptación, encontrándose hoy resultados que presentar, en relación a la reducción de los tiempos de captación , organización, almacenaje, procesamiento y análisis de grandes volúmenes de datos.

Palabras clave: análisis de datos, procesamiento de datos, tecnologías, volumen de datos.

ABSTRACT

This Research aimed to analyze the rise of big data; a new industrial revolution in the transformation of companies as a differential value of data analysis in various sectors. An exploratory documentary research was carried out, through a bibliographic design, towards the search, recovery, critical and rigorous analysis and interpretation of the information, through printed and electronic documentary sources. Big data technology, every day find more users and is positioning itself worldwide, in its application, optimization and adaptation, finding today results to present, in relation to the reduction of the capture times, organization, storage, processing and analysis of large volumes of data.

Keywords: data analysis, data processing, technology, volumes of data.

,

Introducción

En la actualidad hay una realidad innegable, en la que se visualiza una aceleración exponencial y vertiginosa en las tecnologías de información y comunicación (TIC), en relación a la utilización de grandes volúmenes de datos, producto de nuevos conjuntos de ellos; estos datos proceden de las interacciones entre las empresas, instituciones, organizaciones, etc., sus clientes y la incorporación de nuevos instrumentos de comunicación e información, de registro de datos. La avalancha de datos es increíble, hasta el punto que la mayoría de las entidades no ofrecen respuestas importantes en el manejo de los mismos.

La necesidad sentida por parte de la gerencia de empresas y un uso intenso de datos por usuarios potenciales, mediante las famosas redes sociales, entre otras, han llevado a formular nuevas tecnologías que son capaces de realizar análisis que no se podrían efectuar por las metodologías tradicionales. Esta búsqueda incesante de gestionar grandes conjuntos de datos ha llevado al nacimiento de big data, considerado como un nuevo paradigma en la gestión de los datos, término que incluye diversas tecnologías ligadas a la gestión de grandes volúmenes de datos o macrodatos, provenientes de diferentes sectores.

Son innumerables las fuentes de donde procede el gran volumen de datos, sin embargo en la mayor parte de las referencias se incluyen: Twitter, Facebook, Instagram, páginas web, blogs, entre otros; comunicación entre equipos móviles y computadoras ; cámaras de seguridad; sensores de eventos a otros dispositivos,; el registro de las transacciones bancarias, compra, llamadas a empresas; mensajes de texto y voz, correos electrónicos; en ciencias de la salud, donde se realiza el escaneo de retina, los reconocimientos faciales, etc.

Todo este contexto ha llevado a los hombres de ciencia y tecnología a pensar en nuevas estrategias para el manejo de grandes volúmenes de datos, y en estos momentos, ha sido tan grande el desarrollo que ha aparecido la figura del científico de datos, de acuerdo a Aalst (2016 citado en Hernández-Leal, Duque-Méndez, y Moreno-Cadavid (2017), las personas capacitadas en este perfil deben saber del negocio, de las herramientas computacionales y de análisis e interpretación estadística y Data Mining. Una herramienta que ha venido a aceptar el reto planteado, es el big data, modelo, paradigma o como se le llame, que apoya el procesamiento y análisis de grandes volúmenes de datos estructurados, semiestructurados y no estructurados, en el que sus resultados se utilizan para conseguir valor del negocio o entidad, producto de la toma de decisiones correctas y justas en las labores diarias.

Por supuesto, el objetivo es procesar los grandes volúmenes de datos para obtener información que puede estar oculta, y obtener conocimiento, que al aplicarlo, va a producir mejoras en los beneficios de las entidades, empleados y clientes. Por ello se dice que el big data

es una herramienta para obtener conocimiento que produce beneficios. En este sentido, la información se convierte en elemento clave para que las empresas obtengan una ventaja competitiva. Hay que reconocer lo difícil que es tener muchos datos y no tener capacidad de procesamiento; y además, desconocer la forma de obtener información valiosa de ellos y seleccionarla para el logro de los objetivos de la empresa. En estos momentos se hacen grandes esfuerzos por parte de empresas para lograr una transformación que convierta su ambiente en un clima de big data, las llamadas empresas inteligentes como la Toyota, Google, Xerox, Amazon, IBM (International Business Machine), Facebook, Apple, PayPal, entre otras, llevan años usando big data, y han logrado avances importantes en ese aspecto.

El objetivo de esta revisión es aportar información relacionada con el auge del big data, una nueva revolución industrial en la transformación de las empresas como valor diferencial del análisis de los datos en los diversos sectores

1. Definición de big data

Un concepto que no está claro para muchas personas es el término big data, Schroeck, Shockley, Smart y Romero-Morales (2012) apuntan que significa muchas cosas para muchas personas, que las empresas lo utilizan para obtener resultados centrados en el cliente, aprovechar los datos internos y crear un mejor ecosistema de información. Mencionan que el big data se ha utilizado para trasladar al público todo tipo de conceptos, entre los que se incluyen: grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar los datos, datos en tiempo real y mucho más. Un grupo de investigadores de IBM, realizaron un estudio para el que encuestaron 1.144 negocios y profesionales de 95 países diferentes, y han entrevistado a más de dos docenas de académicos, expertos en la materia y directivos empresariales, llegando a concluir que al hablar de big data, las personas piensan en un mayor ámbito de información, nuevos tipos de datos y análisis, información en tiempo real, entrada de datos procedentes de nuevas tecnologías, grandes volúmenes de datos y datos de redes sociales, entre otros.

Otra definición de big data es un término que describe grandes volúmenes de datos de alta velocidad, complejos y variables que requieren técnicas y tecnologías avanzadas para obtener la captura de datos, almacenarlos, distribuirlos, procesarlos y obtener información valiosa (TechAmericaFoundation's Federal Big Data Commission, 2012). Esta definición es compartida por Zdnet (2010), quien asevera que Big Data trata de las herramientas, los procesos y procedimientos que permitan a una organización crear, manipular y gestionar conjuntos de datos muy grandes y las instalaciones de almacenamiento. Ambas definiciones comparten el criterio de que big data, prácticamente, es una herramienta de apoyo a la gestión del conocimiento.

En este mismo orden de ideas, el término aplica a los datos que no pueden ser procesados y analizados mediante tecnologías tradicionales aisladas. Para thinkupapp.com (2012) big data se refiere a grandes cantidades de datos acumulados y que para la empresa son difíciles de analizar utilizando las herramientas existentes de gestión de datos tradicionales y Dans (2011) apunta que Big Data se refiere no solo a los datos en sí, sino al tratamiento de ellos, donde los datos son cantidades desproporcionadamente grandes que resulta imposible tratarlos con herramientas de bases de datos y analíticas que sean convencionales.

2. Tipos de datos big data

El tipo de dato en el paradigma big data está clasificado como estructurados, semiestructurados y no estructurado; para estos variados datos, se ha venido utilizando tecnologías basadas en diversos tipos de software que permiten generar las soluciones big data, tecnologías conocidas como libres y propietarias. Los datos estructurados tienen definida su longitud, formato y tamaño, como las base de datos en hojas de cálculo o bases de datos relacionales, aparecen en formatos de programas computacionales usuales, se presentan como fechas, números cadenas de caracteres, generalmente aparecen en tablas, una base de datos estructurados muchas veces se le llama microdatos (Factor Humano Formación, 2014).

Los datos no estructurados se encuentran en el formato tal y como fueron recolectados y pueden tener cualquier estructura. Se puede encontrar en formatos como: texto, imagen, comentarios en redes sociales o logs de aplicaciones en redes sociales dispositivos móviles, audio, video, sistemas GPS (Sistemas de Posicionamiento Global), incontables sensores digitales en equipos industriales para medición eléctrica o de temperatura y sismos, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la velocidad de respuesta sea lo suficientemente rápida para lograr obtener la información correcta en el momento preciso (López, 2013).

Al respecto, también existen datos semiestructurados, los que no tienen un formato definido, se presentan como etiquetas que separan un dato de otro y presentan un nivel de complejidad para su análisis, en la tabla 1 se presenta un modelo muy sencillo de presentación para este tipo de dato, mientras que en la tabla 2, se presenta un ejemplo para los tres (3) tipos de datos.

Tabla 1. Datos semiestructurados

Nombre	Teléfono	Sexo	Correo
Pedro Pérez	2127409	M	pedroperez@gmail.com

Fuente: Camargo-Vega et al., 2015.

Tabla 2. Tipos de datos en big data

Datos Estructurados	Datos Semiestructurados	Datos No Estructurados
Fichas de clientes Fecha de nacimiento Nombre Dirección Transacciones en un mes Puntos de compra	Correos Electrónicos Parte estructurada: destinatario, receptores, tema Parte no estructurada: cuerpo del mensaje	Persona a Persona Comunicaciones a través de las redes sociales Persona a Máquina Dispositivos médicos Comercio electrónico Ordenadores Móviles Máquina a Máquina Sensores, dispositivos GPS, cámaras de seguridad

Fuente: María Gabriela De Freitas Padrón, Trabajo Especial de Grado, Universidad Central de Venezuela, 2016.

3. Dimensiones del big data

Una de las maneras que la mayoría de los autores se refieren a big data es mediante sus dimensiones o características, dos de ellos, Moreno y Calderón (2015) plantean las V de Big Data como: Volumen, Velocidad, Variedad, Valor y Veracidad, como se muestran en la figura 1. En cuanto al volumen, se refieren al tamaño de los datos, hay un crecimiento exponencial de los mismos. Los tamaños de big data se informan en varios terabytes y petabytes, que se generan a distintas velocidades, dependiendo de las fuentes, se trata de la velocidad con que los datos se generan y a la que los datos se deben analizar; la variedad viene dada por dos razones principales: la primera se debe a que los datos se generan desde diferentes fuentes geográficamente distribuidas; y la segunda, a la existencia de diferentes tipos de datos: estructurados, semiestructurados y no estructurados; la veracidad se refiere a la fiabilidad asociada a esos datos, donde se trata de eliminar la incertidumbre o mejorar la calidad de ellos; el valor de los datos tiene que ver con la generación de valor a partir de los datos iniciales, de manera que a partir del almacenamiento se logre realizar el procesamiento y análisis con el menor costo posible para la máxima información útil.

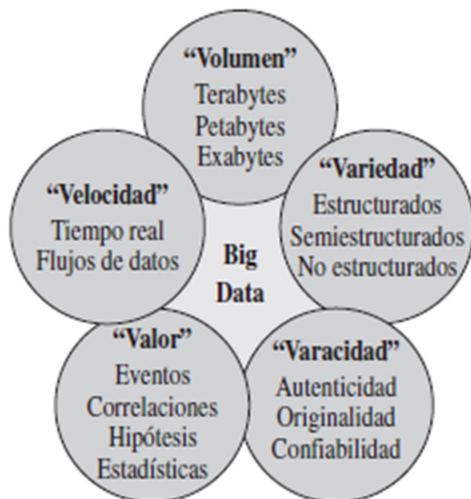


Figura 1. Las "V" de Big Data.

La veracidad asociada a los datos juega un rol muy importante, un aspecto que está siendo muy estudiado son los sentimientos de las personas o clientes en las redes sociales, ya que estos

datos; generalmente, son constructos de naturaleza incierta y compleja, a pesar de contener una información muy valiosa.

4. Herramientas para el tratamiento de los big data

4.1. Tecnologías

Para el procesamiento de los big data se han venido desarrollando diferentes soluciones de tipo computacional, las cuales se han diseñado teniendo en cuenta el volumen, variedad de fuentes y la velocidad con que se generan, siendo estas dimensiones, las mayormente nombradas por autores en esta revisión teórica (Hernández et al., 2017; Gandomi y Murtaza, 2015 y Camargo-Vega, Camargo-Ortega y Joyanes-Aguilar, 2015). En el mismo sentido, en searchstorage.techtarget.com (2012) señalan que: los grandes volúmenes de datos, o big data, requieren grandes cambios en el servidor, la infraestructura de almacenamiento y la arquitectura de administración de la información en la mayoría de las empresas. Es por ello que crear un ambiente big data en una empresa, es tarea ardua y costosa, y no cualquier empresa está en capacidad de asumirlo. Como tecnologías de big data se clasifican aquellas que dan soporte a la captura, transformación, procesamiento y análisis de los datos, ya sean estructurados, semiestructurados o no estructurados.

Ahora bien, la introducción de big data en una empresa debe contar con plataformas y herramientas computacionales y personal humano preparado para crear el ambiente big data, estas plataformas y herramientas apuntan hacia la gestión de altos volúmenes de datos y gente comprometida con la implementación, es decir, en la búsqueda de datos, el almacenaje, el procesamiento y análisis de los mismos, y su interpretación. La herramienta big data, en su aspecto de tecnologías big data, se clasifican como se mencionó anteriormente, en tecnología de captura, de transformación, procesamiento y análisis de cualquier tipo de dato, los llamados estructurales, no estructurales y semiestructurales. En la actualidad hay un sinnúmero de tecnologías, en este caso, se consideran las de mayor aplicación por parte de las empresas.

Uno de los documentos bien importante sobre este asunto, es el de Hurwitz, Nugent, y Halper (2013), quienes apuntan que para desarrollar soluciones big data se pueden utilizar varias herramientas de una arquitectura.

Entre las fuentes de big data del cual una arquitectura puede adquirir datos para su posterior trabajo y análisis son las siguientes: sistemas transaccionales, ERP (*Enterprise Resource Planning*), CRM (*Customer relationship management*), SCM (*Supply Chain Management*), redes sociales, correos electrónicos, señales de sensores, bolsas de empleo, puntos de venta, cajeros automáticos, páginas webs, blogs, telefonía móvil, logs, archivos XML, archivos JSON, comercios electrónicos, documentos de Word, Excel, datos en diferentes programas estadísticos, PDF,

imágenes y videos, entre otros. Relacionado con la captura de datos, igualmente se han creado interfaces de programación de aplicaciones API (Application Programming Interfaces) que permiten la comunicación entre diferentes programas computacionales para realizar la captación de datos (TAS Foundation., 2016 citado en Hernández-Leal et al., 2017).

Sin embargo, para Hurtwitz et al., 2013, en el procesamiento de datos no estructurados o semiestructurados hay que acudir a otras tecnologías; donde aparecen nuevas técnicas y tecnologías, En el procesamiento de los datos, la plataforma más utilizada es la llamada Apache Hadoop, esta es un marco de software de código abierto para aplicaciones intensivas de datos. Hadoop es ahora una de las tecnologías más populares para el almacenamiento de los datos estructurados, semi-estructurados y no estructurados que forman Big Data. Hadoop está disponible bajo la licencia Apache 2.0. Está compuesto por dos módulos: Hadoop Distributed File System (HDFS), y Hadoop MapReduce. Es una librería de Apache que se define como un framework que hace procesamiento de datos sobre volúmenes de datos de considerable tamaño sobre clúster. Está diseñado para brindar poder de escalamiento desde un par de servidores hasta cientos de máquinas o nodos, las cuales manejan almacenamiento y procesamiento local

Apache Hadoop es una solución de software libre diseñada para el tratamiento de hasta exabytes de datos distribuidos en múltiples nodos. Hadoop se ha convertido en un estándar sobre el que se desarrollan herramientas comerciales por compañías tradicionales. La solución de Hadoop se basa en un desarrollo de Google del año 2009 denominado MapReduce, Cabe destacar que Hadoop no es un programa en sí, es decir, no se puede descargar un programa denominado Hadoop directamente, ya que Hadoop es un ecosistema de productos bajo el paraguas de Apache, la cual es una organización no lucrativa creada para dar soporte a los proyectos de software bajo la denominación Apache, incluyendo el servidor HTTP. Apache. Hadoop se caracteriza por ser económico, escalable o adaptable, eficiente a la hora de realizar trabajos en forma paralela y confiable ya que mantiene automáticamente copias de los datos en nodos para la prevención de fallos.

En el mercado existen innumerables herramientas tecnológicas para buscar, manipular, organizar, almacenar y procesar gran cantidad de datos adquiridas, entre otras se tienen: Hadoop, MapReduce, HDFS, Apache Hive, Apache Flume, Pig, Sqoop, Cloudera, Hortonworks, Avro, Hbase, DataStax, Data Integrator, Cognos Decisionstream, MongoDB, Kettle, Oracle NoSQL, Apache Cassandra (Hernández et al., 2017; Hurwitz, et al., 2013; TAS Foundation., 2016, y Gandomi y Murtaza, 2015).

4.2. Técnicas

Se presentan varias técnicas o herramientas en el sistema big data, Mujeeb y Naidu. (2015) expresan que las principales técnicas que se utilizan en big data son: técnicas estadísticas, técnicas de minería de datos, aprendizaje Learning, técnicas de clasificación o clustering, reconocimiento de patrones, algoritmos genéticos, aprendizaje de reglas de asociación, entre otras.

Relacionado con la minería de datos, se presenta como la extracción de información valiosa en un big data, que se transforma en conocimiento, nace a partir de la base de datos o Datawarehouse, encontrándose múltiples soluciones en áreas de mercadeo, salud, comercio, banca, educación, transporte, entre otras. Técnicas como redes neuronales, arboles de decisión, cubos OLAP, clasificación en dos fases, series de tiempo con metodología Box-Jenkins, regresiones múltiples, regresiones logísticas, entre muchas otras.

Así mismo, otra técnica muy utilizada es Machine Learning, reconocida en ingeniería de computación, se relaciona con el descubrimiento de patrones, modelos e irregularidades en los big data. Esta herramienta presenta dos enfoques: los simbólicos, el cual trabaja con aprendizaje inductivo de descripciones simbólicas, y el segundo, el estadístico, que trabaja en métodos de reconocimiento de patrones y estadísticos. Es de mucha utilidad en predicciones financieras en la banca, mercadeo, detecciones de spam y fraudes, entre otras.

En el mismo orden de ideas, el reconocimiento de patrones (Pattern Recognition) aplica en matemáticas e ingeniería de computación, extrae información de cúmulos de datos para establecer relaciones entre ellos. Se utilizan algoritmos de optimización con el fin de encontrar una de las mejores soluciones con un mínimo costo.

Por otra parte, se tiene una técnica de algoritmos genéticos (genetic algorithms), que se inicia en la concepción biológica de la Genética. Comprenden un enfoque de problemas matemáticos, que no pueden resolverse con las herramientas tradicionales, utilizan operaciones genéticas como la mutación. Cruce y recombinación, resuelven problemas estocásticos complejos. Los algoritmos genéticos subyacen en los algoritmos evolutivos, siendo uno de los componentes que se une a la programación genética. Como elementos esenciales de estos algoritmos se tiene:

- a) Estrategia de codificación que determina la forma en que se representará la solución en forma de cromosomas.
- b) Población de cromosomas o individuos.
- c) Mecanismo para la evaluación de cada cromosoma.
- d) Procedimiento de selección/reproducción.

- e) Operadores genéticos: cruce, mutación
- f) Probabilidades para los operadores genéticos
- g) Un criterio de finalización.

En correspondencia con las técnicas anteriores, otra de ellas es el aprendizaje de reglas de asociación, que se utilizan para encontrar relaciones entre variables en big data, se identifican reglas usando algunas mediciones de relación de intereses, en el caso de las redes sociales, se vuelve interesante detectar a las personas que quieren seguir a otras, dependiendo de sus amistades o seguidores, también en el caso que hay poblaciones de personas ocultas, ha dado buenos resultado. En mercadeo, en la revisión de productos que se compran con frecuencia al mismo tiempo.

En relación a las técnicas de clasificación o clustering, muy usadas por cierto, el análisis de conglomerados es un conjunto de técnicas que permiten clasificar una muestra de objetos, casos, variables (entidades) en grupos relativamente homogéneos llamados conglomerados (Clusters), donde los objetos de cada grupo tienden a ser similares entre sí y diferentes a los objetos de otros grupos. Al análisis de conglomerados se le conoce como taxonomía numérica o reconocimiento de patrones, los grupos no se conocen a priori, sino que se forman post-hoc. Se realiza la división de los objetos de tal manera, que los perfiles de ellos en un mismo grupo sean muy similares entre sí y muy diferentes al compararlos con los otros grupos (Malhotra, 1998; Gondar, 2000; Figueras, 2001, y Miquel et al., 1997).

Conclusiones

Después de la revisión de literatura, se llegó a las siguientes conclusiones:

- I. Big data se ha convertido en algo más que una tecnología, es una respuesta a la gestión de datos, obtención de información y conocimiento al alcance de todos los sectores interesados en que sus empresas logren el valor añadido de su negocio y la mejora en las relaciones con sus clientes, incrementar las oportunidades de negocio, incrementar su competitividad en el mercado, y potenciar su recurso humano.
- II. Relacionado con su definición, los autores consultados coinciden en que hay mucha diversidad en la definición e interpretación de las funciones de big data, sin embargo, esas definiciones comparten los aspectos de volumen, velocidad y veracidad como sus principales dimensiones o características en el procesamiento y análisis de grandes volúmenes de datos estructurados, semiestructurados o no estructurados, y que su importancia no solo radica en estas dimensiones, sino en el valor añadido de la información resultante.
- III. La aplicación de big data requiere de personal experto para su implementación, equipo de trabajo con conocimiento en computación, ingeniería de sistemas, Data Mining, Estadística, y conocimiento del negocio y sus clientes.
- IV. Son innumerables las técnicas y tecnologías que requiere big data, debido a sus dimensiones, sobre todo en la variedad de datos, este elemento ha venido a apoyar a todas las entidades donde se implemente, en virtud de la posibilidad no solo de manejar grandes volúmenes de datos con velocidad, sino en la altísima variedad de datos que se puede generar en esas entidades.
- V. Cada día se incorpora un mayor número de empresas al paradigma big data, por lo que se predice un futuro con mayor crecimiento de su mercado; en la actualidad, muchos gerentes y directivos de empresas se interesan por conocer en profundo sus bondades y obstáculos. Relacionado con este aspecto, es bien importante que esos actores estén convencidos que una implementación de big data en sus empresa no es algo fácil por los problemas de experticia, la escalabilidad de los datos, la arquitectura de big data y el alto costo que implica, sin embargo, toda empresa, cualquiera que sea su tamaño, debe acercarse a la aplicación de big data debido a los grandes beneficios que genera.

Referencias bibliográficas

- Camargo-Vega, J., Camargo-Ortega., y Joyanes-Aguilar, I. (2015). Conociendo Big Data. *Revista Facultad de Ingeniería*, 24 (38), 63-77.
- Dans, E. 2011. Disponible en: <http://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html>.
- De Freitas, M. (2016). SOLUCIÓN DE BIG DATA QUE APOYE A LA FASE DE RECLUTAMIENTO DE LA GESTIÓN DEL TALENTO HUMANO EN EL ÁREA DE TECNOLOGÍA DE LA INFORMACIÓN (Trabajo Especial de Grado, Universidad Central de Venezuela). Caracas.
- Factor Humano Formación. (2014). Factor Humano Formación Escuela Internacional de Postgrado. Obtenido de <http://factorhumanoformacion.com/big-data-ii/>
- Figueras, M. (2000). *Introducción al análisis multivariante*. [Página Web en línea]. Recuperado el 22 de Diciembre de 2010, de <http://www.5campus.com/leccion/anamul>
- Gandomi, A. y Murtaza, H. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35,137–144.
- Gondar, J. (2001). *Análisis de conglomerados*. Artículos Estadísticos [Página Web en línea]. Recuperado el 11 de Enero de 2010, de <http://www.estadistico.com>
- Hernández-Leal, E., Duque-Méndez, N., y Moreno-Cadavid, J. (2017). Big Data: una exploración e investigaciones, tecnologías y casos de aplicación. *Tecnológicas*, 22 (39). Medellín, Colombia.
- Hurwitz, J., Nugent, A., & Halper, D. (2013). Big Data for Dummies. IDC. (2012). Big Data: Un Mercado Emergente. Obtenido de http://www.portalidc.com/resources/ponencias/big_data12/Resumen_Ejecutivo.pdf
- Lopez, D. (2013). Analysis of the possibilities of use of Big Data in organizations (Trabajo Final De Grado. Master in Business and Information Technology. Universidad de Cantabria. España.

Malhotra, N. (1998). *Investigación de mercados: un enfoque práctico*. (Trad. V. de Parres). México D.F: Prentice Hall (Original en inglés).

Miquel, S.; Bigné E.; Levy, J. Y Miquel, A. (1997). *Investigación de mercados*. Madrid: Mc Graw-Hill/ Interamericana de España, S.A.

Moreno, L. y Calderón, C. (2017). Arquitectura referencial de Big Data para la gestión de las telecomunicaciones. *Ingeniare*, 25 (4), 566-577

Mujeeb, S. & Naidu, L. (2015). A Relative Study on Big Data Applications and Techniques. *Int. J. Eng. Innov. Technol.*, 4 (10), 133-138.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., y Tufano, p. (2012). Analytics: el uso de big data en el mundo. IBM Institute for Business Value. USA.

Searchstorage.techtarget.com. (2012). Examining HDFS and NameNode in Hadoop architecture. Disponible en: <http://searchstorage.techtarget.com/video/Examining-HDFS-and-NameNodein-Hadoop-architecture>.

T. A. S. Foundation, Welcome to ApacheTM. (2016).Hadoop®. Hadoop. Available: <http://hadoop.apache.org/>

TechAmerica Foundation's Federal Big Data Commission. (2012).Demystifying big data: A practical guide to transforming the business of Government. Retrieved from <http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf>VanBoskirk

Zdnet.com. (2010). Big Data. Disponible en: <http://www.zdnet.com/search?q=big+data>.