



FACULTAD DE TURISMO Y FINANZAS

GRADO EN TURISMO

Estudio del Big Data

Trabajo Fin de Grado presentado por Marta Laynez Quirós, siendo la tutora del mismo Ana María Domínguez Quintero.

Vº. Bº. de la Tutor/a:

Dra. Dña. Ana M^a Domínguez Quintero

Alumno/a :

Dña. Marta Laynez Quirós.

Sevilla. Mayo de 2017



**GRADO EN TURISMO
FACULTAD DE TURISMO Y FINANZAS**

**TRABAJO FIN DE GRADO
CURSO ACADÉMICO [2016-2017]**

TÍTULO:

EL ESTUDIO DEL BIG DATA

AUTOR:

MARTA LAYNEZ QUIRÓS

TUTOR:

DR. D. ANA MARÍA DOMINGUEZ QUINTERO

DEPARTAMENTO:

ECONOMIA APLICADA I

ÁREA DE CONOCIMIENTO:

MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y EMPRESA

RESUMEN:

En este trabajo se realiza un estudio sobre el Big Data y su posible aplicación en el sector turístico. En primer lugar, se define que es el Big Data y sus características, así como la historia cronológica de cómo se origina dicho fenómeno. A continuación, se explica la finalidad de esta técnica, detallando los objetivos que persigue, tanto los principales como los secundarios. Seguidamente se explican algunas herramientas necesarias para que las empresas puedan implementar el Big Data. Para finalizar, se desarrollan todas las ventajas competitivas que derivan de la aplicación de esta técnica para las empresas, posteriormente se observa la posible aplicación del Big Data en el sector turístico y se concluye analizando los resultados obtenidos por diferentes estudios sobre el Big Data en España.

PALABRAS CLAVE:

Big Data; Datos masivos; Base de datos; Aplicación informática.

ÍNDICE

CAPÍTULO 1: INTRODUCCIÓN, OBJETIVOS, JUSTIFICACIÓN Y METODOLOGÍA DEL TEMA A ESTUDIAR	3
1.1. INTRODUCCIÓN	3
1.2. OBJETIVOS	4
1.3. JUSTIFICACIÓN DE LA RELEVANCIA DEL TEMA	4
1.4. METODOLOGÍA	4
CAPÍTULO 2: QUE ES EL BIG DATA	5
2.1. DEFINICIÓN DEL BIG DATA	5
2.2. CARACTERÍSTICAS DEL BIG DATA. LAS ‘3V’	6
2.3. TIPOS DE DATOS UTILIZADOS EN EL BIG DATA	7
2.4. DESARROLLO DEL FENOMENO ‘BIG DATA’	8
2.4.1. Historia cronológica	9
CAPÍTULO 3: FINALIDAD DEL BIG DATA	14
3.1. OBJETIVO PRINCIPAL DEL BIG DATA	14
3.2. OBJETIVOS SECUNDARIOS DEL BIG DATA	14
CAPÍTULO 4: HERRAMIENTAS QUE UTILIZA EL BIG DATA	16
4.1. TÉCNICAS PRINCIPALES	16
4.1.1. Machine Learning	16
4.1.2. Lenguaje R	17
4.1.3. Python	18
4.1.4. MapReduce	18
4.2 PROGRAMAS INFORMÁTICOS UTILIZADOS EN EL BIG DATA	19
4.2.1. Hadoop	19
4.2.2. MongoDB	19
4.2.3. Elasticsearch.....	20
4.2.4. Apache Spark.....	20
4.2.5. Apache Storm	21
4.2.6. HPCC	21
4.2.7. Hana	22
CAPÍTULO 5: VENTAJAS QUE LE OFRECE EL BIG DATA A LAS EMPRESAS ...	23
5.1. VENTAJAS COMPETITIVAS EN CUALQUIER SECTOR EMPRESARIAL ..	23
5.2. APLICACIÓN DEL BIG DATA EN EL SECTOR TURISTICO	27
CAPÍTULO 6: EL BIG DATA EN ESPAÑA	30
6.1. ANÁLISIS DE DATOS SOBRE EL BIG DATA EN ESPAÑA.....	30
CAPÍTULO 7: CONCLUSIONES	35
7.1. CONCLUSIONES SOBRE EL ESTUDIO	35
BIBLIOGRAFÍA	36

CAPÍTULO 1

INTRODUCCIÓN, OBJETIVOS, JUSTIFICACIÓN Y METODOLOGÍA DEL TEMA A ESTUDIAR

1.5. INTRODUCCIÓN

El universo digital lleva formándose ya desde un largo tiempo atrás, pero recientemente los datos no paran de crecer a un nivel masivo y según IDC¹, crecerán un 33% más durante esta década. Los datos masivos que se generan, se recopilan y se almacenan lo cual nos permite acceder a una enorme cantidad de información que antes simplemente no estaba disponible y era totalmente desconocida para nosotros.

Es importante analizar los nuevos datos para desvelar el valor que encierran. Pero este proceso tomaría demasiado tiempo y sería muy costoso hacerlo con una base de datos de tipo relacional. Por ello, aparece el concepto del Big Data, para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.

Aparte del gran volumen de información, dentro de esta existe una gran variedad de datos representados de diferente manera como, por ejemplo, de dispositivos móviles, audio, video, sistemas GPS, sensores digitales en equipos industriales, automóviles, medidores eléctricos, etc. Por ello, las aplicaciones que realicen el procesamiento de estos datos necesitan tener una característica clave, que es que la velocidad de respuesta sea lo suficientemente rápida para lograr obtener la información correcta en el momento preciso. De ello se deduce que las características principales del Big Data sean volumen, variedad y velocidad.

El concepto de Big Data, pertenece a la lengua inglesa, por lo tanto, no es posible encontrar su definición en la Real Academia Española, aunque en el periódico *El Economista* el autor López. J. C (2014, p.1) lo definió como “la gestión y análisis de enormes volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software habitualmente utilizadas para la captura, gestión y procesamiento de datos” Este concepto engloba una serie de herramientas necesarias para llevar a cabo la técnica del Big Data. Por ello, las empresas que lleven a cabo dicha técnica tendrán que ponerse al día en las técnicas y aplicaciones informáticas que son necesarias para llevarlo a cabo.

La finalidad del Big Data es, que las empresas sean capaces de recolectar y aprovechar los datos relevantes, para usarlos en la toma de mejores decisiones de negocio. Por ello, las empresas que estén al día en las nuevas tecnologías y apliquen esta técnica conseguirán posicionarse sobre sus competidores consiguiendo un gran número de ventajas competitivas, como son la mejora de eficiencia y disminución de costes, la velocidad en la toma de decisiones, la vinculación de clientes y la retroalimentación entre otras.

¹ International Data Corporation (IDC) es la principal firma mundial de inteligencia de mercado, servicios de consultoría, y conferencias para los mercados de Tecnologías de la Información, Telecomunicaciones y Tecnología de Consumo.

El sector turístico, es uno de los sectores más importantes en nuestro país, se puede observar su importancia en la aportación que realiza al PIB. Según Exceltur, el último año, el sector turístico hizo una aportación del 10% aproximadamente al PIB español. Por ello, es muy importante que las empresas que forman este sector estén al día en las nuevas tecnologías y sobre todo en la técnica del Big Data para conseguir ser más eficiente y crecer cada vez más.

1.6. OBJETIVOS

A través de este trabajo, se pretende hacer un estudio de un concepto que está a la orden del día actualmente, como es el Big Data. Se pretende conocer cómo se puede realizar esta técnica dentro de una empresa y observar los beneficios que esta puede aportar a las diferentes organizaciones, y su posible aplicación en las empresas del sector turístico español. A continuación, se mostrarán los objetivos perseguidos en este estudio:

- Conocer el concepto Big Data.
- Estudiar las características que engloban dicho concepto y los datos que lo componen.
- Entender el nacimiento y la historia del Big Data.
- Conocer la finalidad y los objetivos del Big Data.
- Saber cuáles son las herramientas necesarias para que las empresas pongan en marcha esta nueva técnica.
- Conocer todas las ventajas competitivas que conseguirán las empresas que lleven a cabo el Big Data.
- Conocer la aplicación que se hace del Big Data en el sector turístico.

1.7. JUSTIFICACIÓN DE LA RELEVANCIA DEL TEMA

Este estudio ha sido llevado a cabo porque el término Big Data es nuevo en mi opinión, no hay demasiados estudios españoles llevados a cabo sobre el tema y aún menos realizados por estudiantes de grados como es el de turismo. En mi opinión, es un tema que debería ser más conocido por la cantidad de ventajas que aporta el uso de esta técnica y aun con más importancia el sector turístico, que es uno de los más importantes en nuestro país.

Con la realización de este estudio pretendo profundizar y favorecer que las empresas conozcan esta técnica y consigan crecer y mejorar gracias a ella y sobre todo que la técnica del Big Data se aplique en más empresas de nuestro país.

1.8. METODOLOGÍA

Este estudio se ha realizado gracias a la obtención de información a través de la consulta de libros, artículos de revistas, artículos de prensa y consultas en fuentes digitales.

Para el desarrollo del estudio se ha optado por ir de lo más general a lo más particular. Se ha ido conociendo cual es el tema de estudio, en qué consiste, cómo y por qué surge dicho concepto. Posteriormente la finalidad que persigue y como llevarlo a cabo. Por último, para finalizar se observan las ventajas competitivas que conseguirán las empresas a la hora de implementar el Big Data y conocer la aplicación que se hace de dicha técnica en el sector turístico.

CAPÍTULO 2

QUE ES EL BIG DATA

2.1. DEFINICIÓN DEL BIG DATA

La palabra Big Data, se refiere a datos masivos, datos a gran escala o macrodatos, aunque su definición también ha de englobar a las herramientas procesos y procedimientos que ayudan a obtener esos datos y procesarlos para conseguir información relevante.

Los principales conceptos que han definido al Big Data han sido las denominadas '3 V': volumen, variabilidad y velocidad.

Según Tascón, M. (2013), se pueden considerar la idea de las '3 V' como que los volúmenes, velocidad y variabilidad de los datos han crecido rápidamente. Lo que antes eran unos simples números que se procesaban con un simple PC y una hoja de cálculo, ahora han pasado a ser enormes cantidades de datos que están almacenadas en 'la nube' a lo largo de conjuntos de ordenadores que necesitan ser procesados con programas especiales que permitan la posibilidad de manejar estos datos masivos con rapidez. Esta nube (Cloud Computing), es un nuevo modelo de prestación de servicios de computación, información y aplicaciones a través de Internet donde la mayoría del software se ejecuta en la propia Red. Cada vez con más frecuencia, las aplicaciones, el software en la nube y el Big Data se unen para poder desarrollarse conjuntamente. De la hoja Excel hemos pasado a Hadoop, que es un software que permite trabajar con un gran tamaño de información.

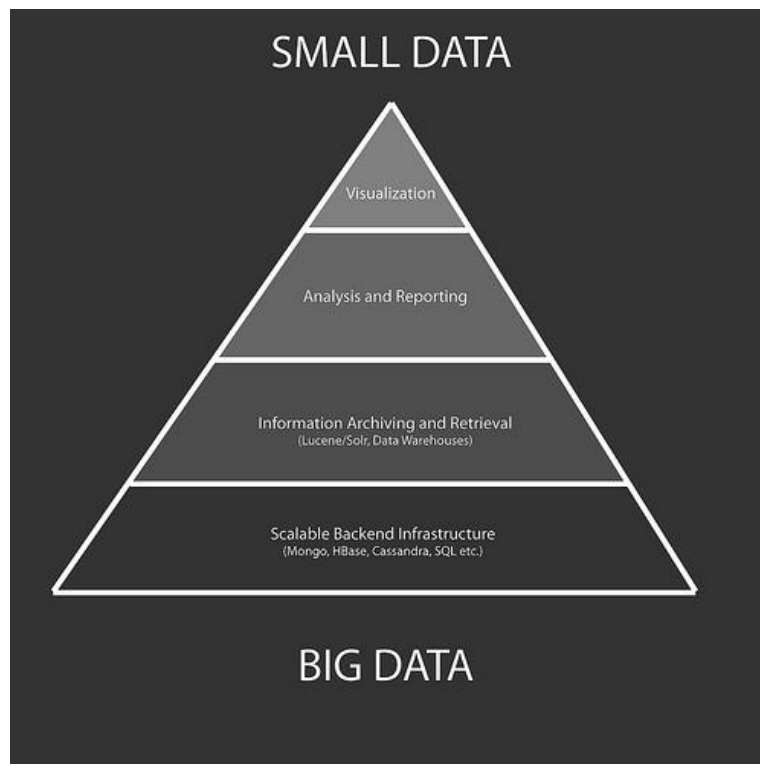


Figura 2.1. Big data

Fuente: <http://informationxdummies.blogspot.it/2014/03/recurso-web-para-practicantes-de-big.html>

2.2. CARACTERÍSTICAS DEL BIG DATA. LAS '3V'

Según el autor Tascón, M (2013, p.49), "Macrodatos es todo aquello que tiene que ver con grandes volúmenes de información que se mueven o analizan a alta velocidad y que pueden presentar una compleja variabilidad en cuanto a la estructura de su composición." Tascón, siempre ha considerado que debería añadirse una cuarta uve, la visualización, ya que muchas de las imágenes que nos traen a la memoria el trabajo con Big Data tienen que ver con estas nuevas formas de ver estos datos.

Se hace referencia a las '3V' cuyas iniciales se refieren al volumen, velocidad y variabilidad de los datos que han crecido en exceso.

En referencia a los autores Camaro y Aguilar (2015), puede decirse que, volumen, es la característica que responde a la pregunta ¿Cuánta información? Ya que, los datos de las empresas van creciendo día a día y cada vez van un registrando un aumento más significativo de datos que son creados tanto por personas como por máquinas. Las redes sociales, también generan demasiados datos al día, por ejemplo, Twitter, por sí solo genera más de 7 terabytes (TB) diariamente, y Facebook, 10 TB de datos cada día. También, las empresas generan Terabytes de datos, por lo tanto, las empresas están desbordadas de datos y dichos datos con gran volumen conllevan costosos procesos de almacenamiento y análisis.

Variación, responde a la pregunta ¿Cuántos tipos de datos llegan? Variación está muy relacionado con el volumen, porque, gracias al volumen y el desarrollo de la tecnología, existen muchas formas de representar los datos; es el caso de datos estructurados y no estructurados. Un uso común del big data es tomar los datos no estructurados y extraer con un orden su significado. Es importante, que los datos no tradicionales no presenten ninguna estructura específica por su naturaleza, como puede ser imágenes, videos, etc. a diferencia de los tradicionales.

Velocidad, responde a la pregunta, ¿Cómo de rápida es? Esta característica, se refiere a la velocidad con que se crean los datos, que es la medida en que aumentan la creación y el desarrollo de nuevos softwares, como páginas web, redes sociales, etc.

Es importante recalcar que las tres características se complementan entre sí. Como podemos observar en la figura 2.2.

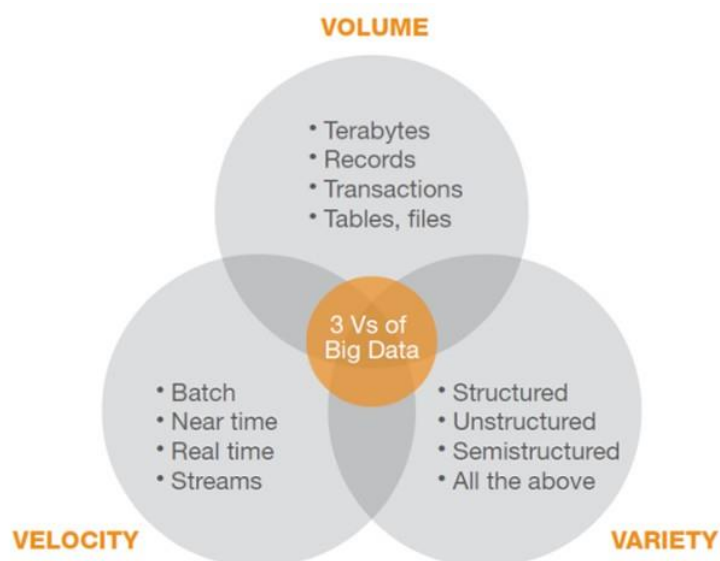


Figura 2.2. Las '3V'

Fuente: <http://informationxdummies.blogspot.it/2014/03/recurso-web-para-practicantes-de-big.html>

2.3. TIPOS DE DATOS UTILIZADOS EN EL BIG DATA

Anteriormente en la definición de variedad cuando estábamos analizando las características del Big Data, hemos hecho referencia a que existen diferentes tipos de datos, a continuación, vamos a analizar cada uno de ellos.

Los **datos estructurados** son aquellos que presentan una mayor facilidad para acceder a ellos, porque tienen una estructura bien especificada. Este tipo de datos se dividen dos subtipos que son los estáticos (array, cadena de caracteres y registros) un array es una colección limitada de elementos en formatos definidos del mismo tipo, es decir, son elementos homogéneos, y ordenados por un índice. Y los datos dinámicos como las listas, los árboles y archivos.

Los **datos semiestructurados** son aquellos que no tienen un formato definido a diferencia de los estructurados que tienen una estructura bien especificada. Los datos semiestructurados, lo que tienen son etiquetas que facilitan la separación entre un dato y otro. Un dato de este tipo se lee con un conjunto de reglas que engloban un cierto nivel de complejidad. A continuación, podemos observar un ejemplo de este tipo de datos.

Nombre	Apellido	Correo electrónico	Teléfono
Fco. Javier	Layne	jfl55@gmail.com	78954585265

Figura 2.3. Ejemplo de dato semiestructurado

Fuente: Elaboración propia

Los **datos no estructurados** según los autores Camaro y Aguilar (2015), son aquellos que no pueden ser normalizados, ni están organizados bajo ninguna pauta o patrón. Los datos no estructurados tampoco pueden ser almacenados racional o jerárquicamente, ya que, no son un tipo de datos predefinidos, es decir, no tienen un formato normalizado determinado. Sin embargo, este tipo de datos deben poder ser organizados, clasificados, almacenados, eliminados o buscados de alguna forma. Por ello, estos datos se pueden observar cada día en correos electrónicos, archivos de texto, en documentos de algún procesador de palabra, en hojas electrónicas, en una imagen, un objeto, archivos de audio, blogs, mensajes de correo de voz, mensajes instantáneos, contenidos Web y archivos de video, entre otros.

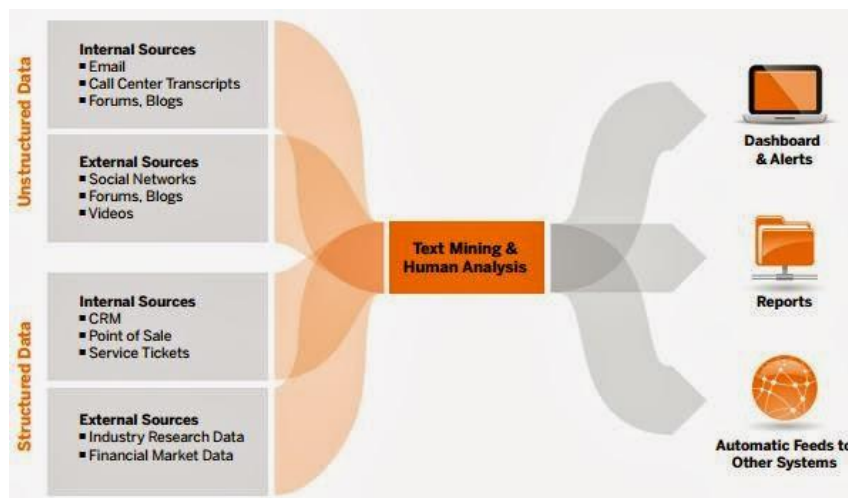


Figura 2.4. Tipos de datos

Fuente: <http://informationxdummies.blogspot.it/2014/03/recurso-web-para-practicantes-de-big.html>

2.4. DESARROLLO DEL FENOMENO 'BIG DATA'

Según el autor Luis Joyanes, en septiembre de 2016, IDC hizo público un informe sobre la situación actual y perspectivas del futuro del Big Data tanto en España como a nivel mundial. Según IDC, actualmente existen 2.8 zettabytes (ZB) y el volumen de datos que existen en el mundo digital, no paran de crecer. En 2020 habrá 40 ZB de información digital, para hacernos una idea existirán 5.247 GB por cada habitante del planeta.

Este crecimiento se verá empujado por la exponencial multiplicación de usuarios de internet, redes de sensores y objetos interconectados, dispositivos inteligentes que permiten nuevas formas de trabajo y de comunicación, así como las redes sociales, la cuales, dan lugar a nuevos modelos de negocio y nuevas formas de interaccionar con los clientes.

Según Fidelity Investment, empresa estadounidense especializada en gestión de activos y fondos de pensiones, más de un 90% de los datos digitales estarán desestructurados, envolviendo una gran cantidad de información de gran valor, pero muy difícil de analizar y comprender debido al rápido desarrollo del universo digital. En opinión del autor Aguilar, L. J, las empresas no pueden seguir confiando en los sistemas tradicionales para la toma de decisiones, porque ya no son capaces de ofrecer respuestas ágiles y precisas.

Cada vez más las empresas buscan la manera de conseguir el acceso unificado a la información, y el análisis para la mejor toma de decisiones de usuarios grupos y sistemas ganando así una mayor efectividad.

IDC recomienda que aquellas empresas que lleven a cabo iniciativas de Big Data, no solo conseguirán ser capaces de analizar grandes volúmenes de datos, sino que también aumentarán su capacidad para rediseñar los procesos de negocio, e incluso crear nuevos servicios basados en la Información, ya que, el Big Data es una herramienta muy útil para generar ventajas competitivas basadas en la información.

Como podemos observar en la figura 2.5. Los datos a medida que pasan los años, crecen con más rapidez, y el 80% de los datos actuales en el mundo están desestructurados, por tanto, las compañías que utilicen técnicas que ayuden a tener acceso a dichos datos y la oportunidad de analizarlos, podrán obtener ventajas competitivas sobre otras empresas que no practican dichas técnicas.

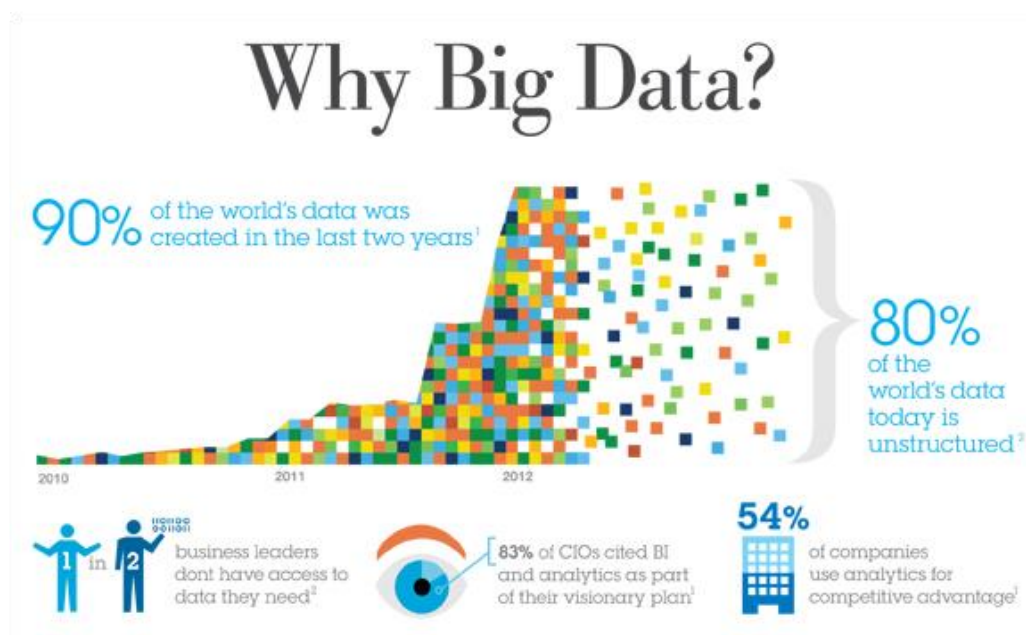


Figura 2.5. ¿Por qué el Big data?

Fuente: Pinterest

2.4.1. Historia cronológica

En el año 1880 el censo de los Estados Unidos tardó 8 años en tabularse y el de 1990 tardó 10 años, gracias a la introducción de avances, este proceso ha ido siendo cada vez más veloz. Tan solo un año después, el problema con los datos del censo dio lugar a la invención de la máquina tabuladora de Hollerith en 1881, con tarjetas perforadas. Este método fue capaz de abordar la masiva cantidad de información.

En 1932 la sobrecarga de información fue seguida del aumento de la población en EEUU, lo que conllevó a la emisión de más números de la seguridad social y avances en la investigación y el conocimiento. Pocos años después, las bibliotecas se vieron obligadas a adaptar sus métodos de almacenamiento para abordar a la creciente demanda de investigaciones y publicaciones. Posteriormente, surge el término “explosión de la información” en 1941 que significaba la expansión de la información.

Según Bell, D. (2000), el primer aviso sobre el crecimiento del conocimiento como problema de gran importancia a la hora de almacenar y recuperar los datos se da en 1944, cuando Fremont Rider, un bibliotecario de la Universidad Wesleyana, calculó que las bibliotecas de las universidades de EE. UU. duplicaban su tamaño cada dieciséis años.

Cuatro años después, Claude Shannon publicó la teoría matemática de la comunicación en 1948, donde se determinaban los requisitos de datos mínimos para transmitir la información a través de canales perjudicados por ruidos (imperfectos).

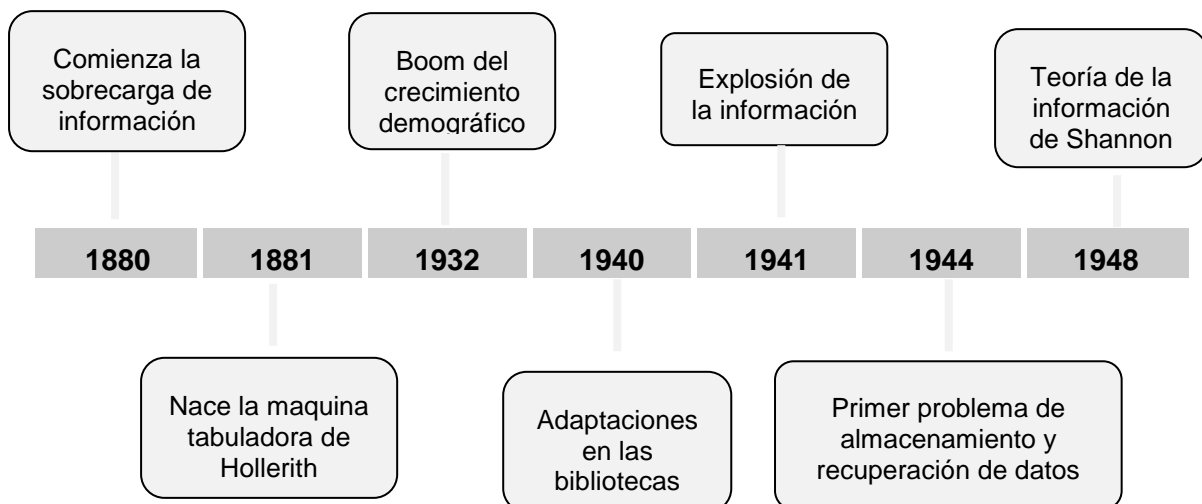


Figura 2.6. Cronología del Big Data

Fuente: Elaboración propia

Por otra parte, el físico Frits- Rudolf Güntsch desarrolla el concepto de memoria virtual en 1956 como una idea que abarcaba el almacenamiento tanto finito como infinito, y en 1963, debido a la enorme cantidad de documentos creados a finales de la década de 1800 se lleva a cabo la búsqueda de una solución organizativa para dicha cantidad de información, más tarde, las empresas desarrollaron sistemas informáticos que les permitían automatizar los sistemas de inventario y aparecen por primera vez los sistemas de computación centralizados, en el año 1966.

Según National Research Council (2004), en el año 1970, Edgar F. Codd, un matemático formado en Oxford que trabajaba en IBM Research Lab, publicó un artículo en el cual se explicaba la forma en la que podía accederse a la información

almacenada en bases de datos de gran tamaño sin saber cómo estaba estructurada la información o dónde se encontraba dentro de la base de datos.

Según el autor Duff A.S (2000), el Censo del Flujo de la Información, que fue realizado por el Ministerio de Correos y Telecomunicaciones de Japón, realizó un control del volumen de información que circulaba por el país en 1975. La unidad de medición que se utilizó fue el número de palabras que se utilizaron a través de todos los medios de comunicación, con el estudio se comprobó que el suministro de información superaba notablemente al volumen de información consumida, y que la demanda de comunicación unidireccional se había estancado. Ahora, la tendencia era el aumento de la demanda de comunicación bidireccional, más personalizada, que respondía a las necesidades de las personas.

A mediados de la década de 1970, se diseñaron los sistemas de Planificación de necesidades de material (MRP), era una herramienta que ayudaba a las empresas de fabricación a organizar y planificar su información. A partir de ahí, las empresas empezaron a usar datos para la toma de mejores decisiones.

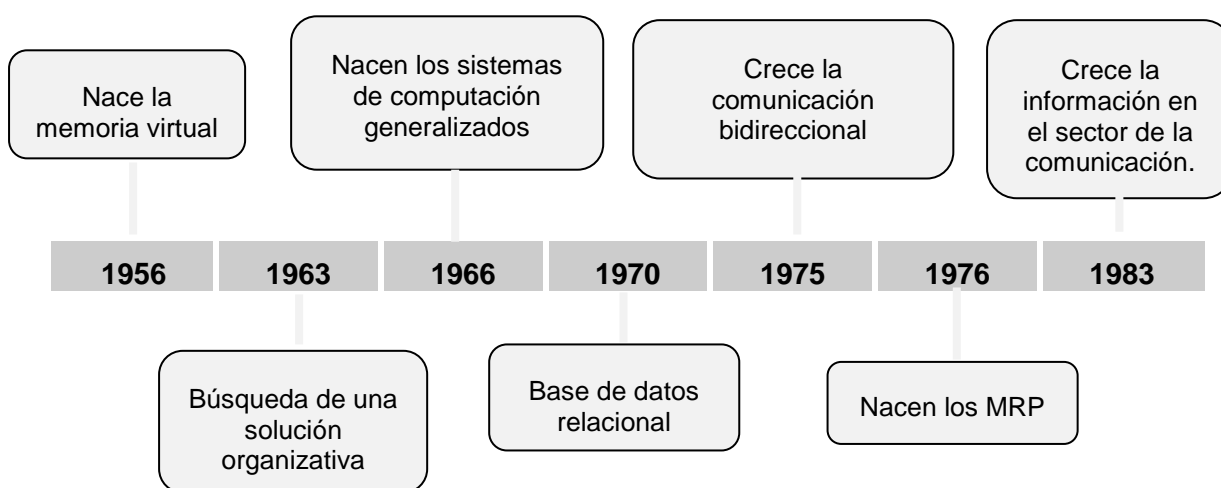


Figura 2.7. Cronología del Big Data

Fuente: Elaboración propia

Según Press, G (2003), tras el éxito de los sistemas de MRP, se introdujo la Planificación de recursos de fabricación (MRP II) en la década de 1980.

Conforme al autor Power, D.J. (2003), en 1989, Howard Dresner amplió el popular término «Business Intelligence (BI)» o Inteligencia empresarial, utilizado inicialmente por Hans Peter Luhn en el año 1958. Dresner lo definió como los “conceptos y métodos que mejoran la toma de decisiones de negocio mediante el uso de sistemas de apoyo basados en datos reales”.

Se crea el primer informe de base de datos por Crystal Reports con Windows en 1992. Sucesivamente, en 1995 se da la explosión de la World Wide Web, que es un sistema de distribución de documentos de hipertexto o hipermedios interconectados a los que se pueden acceder por Internet.

En referencia a los autores Cox y Ellsworth (1997), el término «Big Data» se empleó por primera vez en un artículo de los investigadores de la NASA Michael Cox y David Ellsworth. Ambos afirman en dicho artículo que el ritmo de crecimiento de los datos empezaba a ser un problema para los sistemas informáticos actuales. Esto se denominó como el «problema del Big Data».

En la mayoría de las empresas a finales de los 90, los trabajadores no eran capaces de acceder a los datos que necesitaban, ni de encontrar respuestas, por lo que, empezaron a creer que sus sistemas extracción de datos no funcionaban.

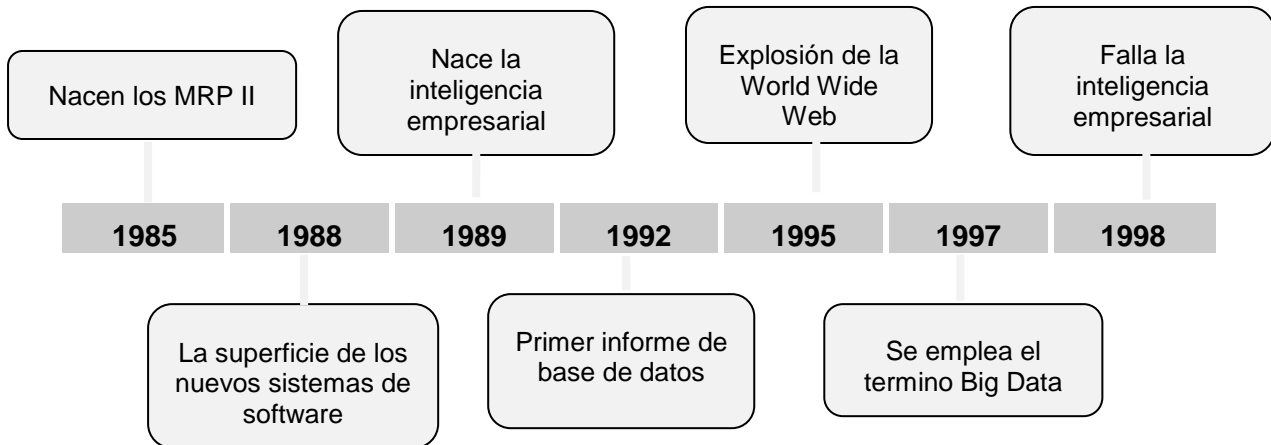


Figura 2.8. Cronología del Big Data

Fuente: Elaboración propia

Kevin Ashton, cofundador del Auto-ID Center del MIT, emplea por primera vez el término *IoT*, el cual significa, "Internet de las cosas". El autor Ashton, K. (1999) hace referencia a que este término menciona que, si tuviéramos equipos que supieran todo lo que hay que saber acerca de las cosas, a partir de datos que recopilan por ellos mismos, seríamos capaces de monitorizar y contar todo, y así poder reducir de esta manera costes, desperdicios y pérdidas.

En 2001, se emplea por primera vez el termino SaaS, que son las siglas de Software como servicio y se publica un artículo sobre las tres V, escrito por Doug Laney, llamado "3D Data Management: Controlling Data Volume, Velocity, and Variety"

Posteriormente, las empresas SaaS empiezan a cobrar importancia por su orientación al enfoque de la importancia que tiene el usuario final. En 2005 la gestión de datos comienza a cobrar una gran importancia. Tan solo un año después, se crea Hadoop debido a la necesidad que existía de crear nuevos sistemas de gestión para afrontar la gran cantidad de datos de la Web.

El SAP, que representan las siglas en alemán *Systeme Anwendungen und Produkte* que significa en español 'sistemas, aplicaciones y productos'. En 2008 desarrolla una estrategia de software destinada a grandes empresas.

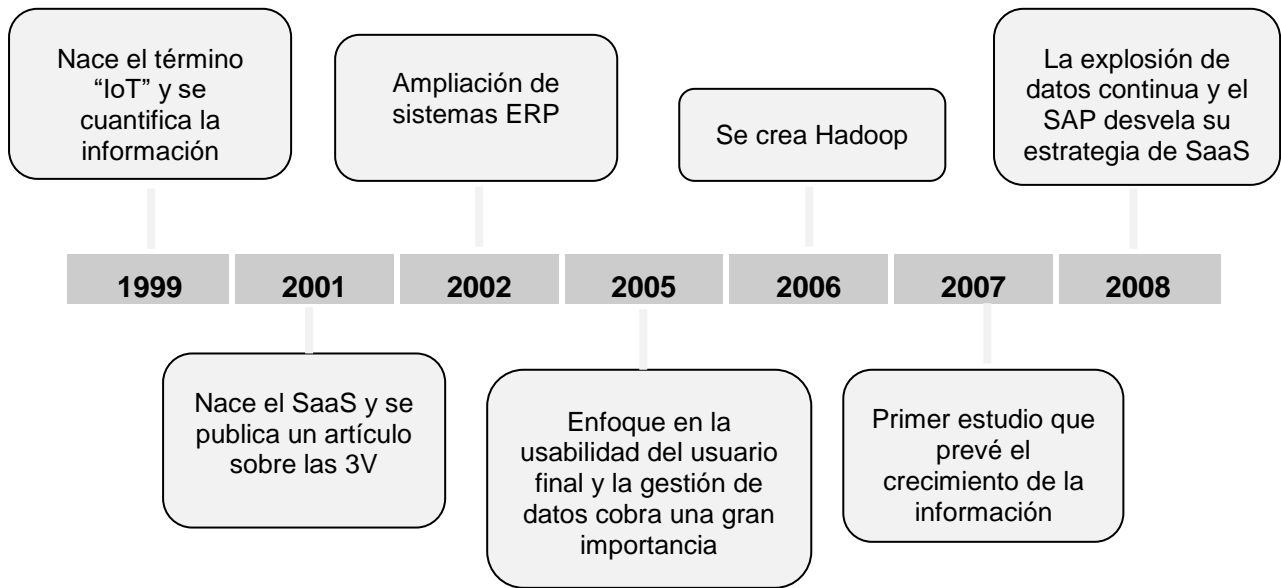


Figura 2.9. Cronología del Big Data

Fuente: Elaboración propia

Según los autores Heath, T y Bizer, C. (2009), la inteligencia empresarial, se convierte en una prioridad y en 2009, Tim Berners el inventor de la Word Wide Web usa el termino linked data por primera vez, Este término describe un método de publicación de datos estructurados, que está basado en protocolos, para que puedan ser interconectados, leídos automáticamente por ordenadores y enlazados.

En 2010 los sistemas ERP se comienzan a desarrollar en la nube, además estos sistemas ERP se comienzan a coordinar con los procesos de negocios. Tan solo un año después, se observa un gran despliegue de los datos, los datos comienzan a crecer velozmente y las empresas se ven obligadas a ampliar sus sistemas de almacenamiento. Y ya, en 2012, comienzan a surgir las dudas sobre el termino Big Data y su aplicación en el artículo "Critical Questions for Big Data", publicado en Information, Communications, and Society Journal.

Las empresas en 2013 ven la necesidad de implementar nuevas tecnologías en memoria como SAP HANA para optimizar y analizar cantidades de datos masivas.

Por último, Michael Lesk publicó "How much information is there in the world?". Conforme al autor Lesk, M. (1997), su conclusión fue que, en unos años, podremos guardarlo todo, no será necesario eliminar información, y que la mayoría de la información jamás será consultada por un ser humano.

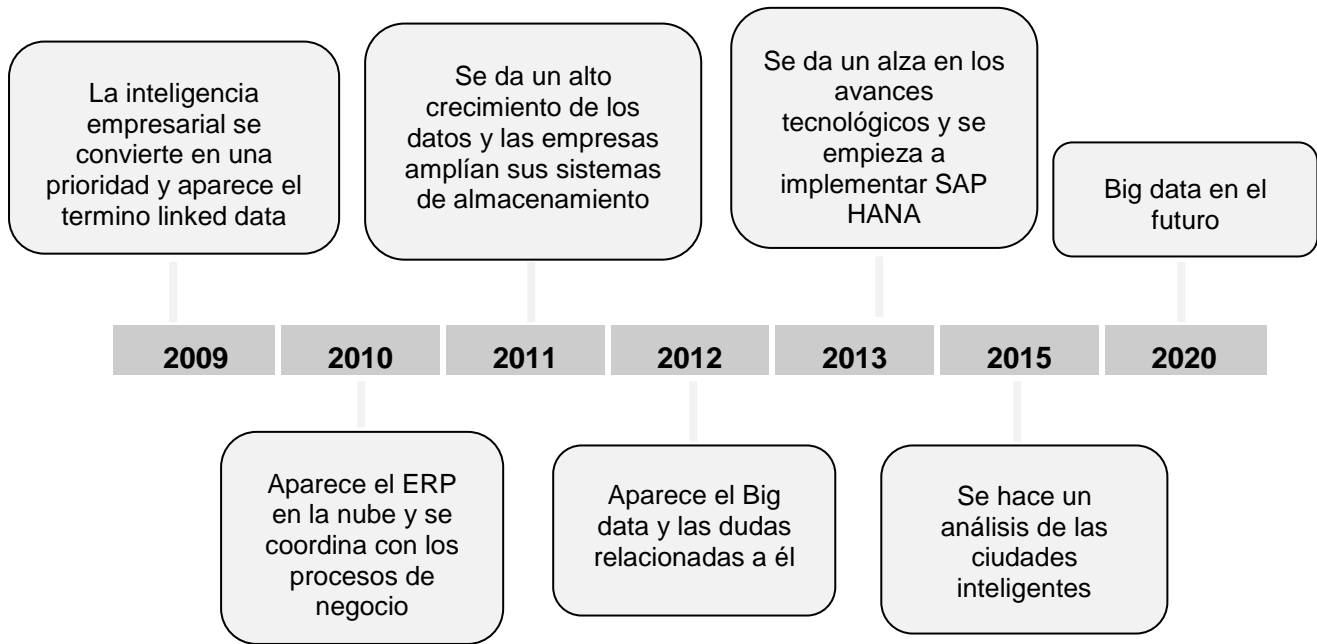


Figura 2.10. Cronología del Big Data
Fuente: Elaboración propia

CAPÍTULO 3

FINALIDAD DEL BIG DATA

3.1. OBJETIVO PRINCIPAL DEL BIG DATA

El objetivo principal del Big Data es, que las empresas sean capaces de recolectar y aprovechar los datos relevantes, para usarlos en la toma de mejores decisiones de negocio.

Big data, es un cambio de modelo en la idea que tenemos acerca de los conjuntos de datos y de cómo procesarlos. Su objetivo es recopilar la 'inteligencia', es decir lo necesario, lo que busca cada empresa a partir de los datos y traducir eso en una ventaja competitiva.

Por lo tanto, el Big Data, tiene la capacidad de transformar radicalmente a las empresas a la vez que realiza una amplia medición y conocimiento sobre el negocio. Posteriormente, traduce directamente ese conocimiento en una mejor toma de decisiones lo que lleva a las empresas a aumentar notablemente su rendimiento.

Según López, J.C. (2014), puede decirse que el objetivo de Big Data, al igual que los sistemas analíticos convencionales, es convertir el Dato en información que facilita la toma de decisiones, incluso en tiempo real. Es una oportunidad de negocio. Las empresas ya están utilizando *Big Data* para entender el perfil, las necesidades y la opinión de sus clientes respecto a los productos y/o servicios vendidos. Esto toma una especial importancia, ya que, les permite ajustar la forma de interactuar con sus clientes y mejorar la prestación del servicio.

3.2. OBJETIVOS SECUNDARIOS DEL BIG DATA

Aquellas empresas que apuesten por el Big Data, podrán crear nuevas oportunidades para diferenciarse de la competencia. Podrán llevar a cabo esta finalidad a partir de diversos objetivos, dependiendo de la finalidad que quieran llevar a cabo en su negocio. Los principales objetivos del Big Data son:

1. El uso de la información para crear nuevos productos.
2. La interacción con otros datos para complementarlos.
3. La segmentación.
4. La disminución del riesgo a la hora de hacer una inversión.

Uno de los objetivos del Big Data puede ser **el uso de la información para crear nuevos productos**, en la actualidad los nuevos productos son indispensables para el crecimiento de una empresa. Las empresas pueden optar por innovar nuevos productos o mejorar aquellos que ya existen. Para ello, es muy importante determinar las nuevas necesidades que les van surgiendo a los clientes y sus preferencias, ya que, nos encontramos en un momento en el que los rápidos cambios en hábitos, tecnología y competencia. Por ello, una compañía no puede confiar únicamente en los productos que ya tiene. Los clientes desean y esperan nuevos y mejores artículos. La competencia hace todo lo posible para producirlos y, por ello, muchas compañías se dan cuenta que necesitan desarrollar nuevos productos.

El Big Data, no solo puede servir para proporcionarnos nueva información, si no, para la intermediación de esta, es decir, a **la interacción con otros datos para complementarlos**, volviéndolos más útiles. La información es un recurso vital dentro de la empresa, ya que, sin ella no se podría llevar a cabo ningún tipo de trabajo. Con el nuevo uso de las tecnologías, las empresas utilizan su información para todas las actividades cotidianas, como por ejemplo, la búsqueda exacta de un estado financiero. La información, por tanto, es la parte fundamental de una empresa para tener un alto

nivel de desarrollo y competitividad, y hoy día, el Big Data hace que este proceso sea mucho más fácil y completo.

Big Data, permite la **construcción de redes para ofrecer productos en el momento y el lugar correcto**. El Networking, que es el arte de gestionar eficazmente una red de contactos, es esencial en el desarrollo de una empresa. Es un concepto muy reciente que se emplea como una herramienta que se basa en la relación de confianza y su aportación de valor para conseguir incrementar las ventas de las empresas que participan en dicho proceso.

Gracias al uso preciso que el Big Data hace de la información, este puede ser utilizado para la **segmentación**. La segmentación de mercado es un proceso que consiste en dividir el mercado total de un bien o servicio en varios grupos más pequeños y homogéneos. La clave para segmentar es conocer realmente a los consumidores, para ello el big data proporciona una información eficaz. La capacidad de segmentar puede llevar a una empresa al éxito ya que, las pymes pueden crecer más rápido si obtienen una posición firme en los segmentos especializados del mercado. Y la selección de canales de distribución y de comunicación se facilita mucho, además las empresas tienen menos competidores en un segmento específico.

Conforme a la Red de Centros SAT. (2014), también el Big Data **disminuye el riesgo a la hora de hacer una inversión**. Como hemos mencionado antes, el análisis de Big Data es el proceso de examinar a gran velocidad grandes volúmenes de datos de una amplia variedad para descubrir patrones ocultos, y otra información útil, de modo que los resultados del análisis puedan proporcionar ventajas competitivas a las empresas. El análisis del Big Data, trata esencialmente de hacer predicciones con los datos, no interesa tanto el “por qué” sucede una cosa, basta con saber “qué” va a suceder. Esto es lo que se conoce hoy día como analítica predictiva.

Por último, se finalizará este punto añadiendo algunos de los objetivos secundarios más frecuentes a la hora de llevar a cabo el desarrollo del Big Data. Estos pueden ser, identificar nuevos mercados, proporcionar un mejor servicio al cliente y retener clientes.

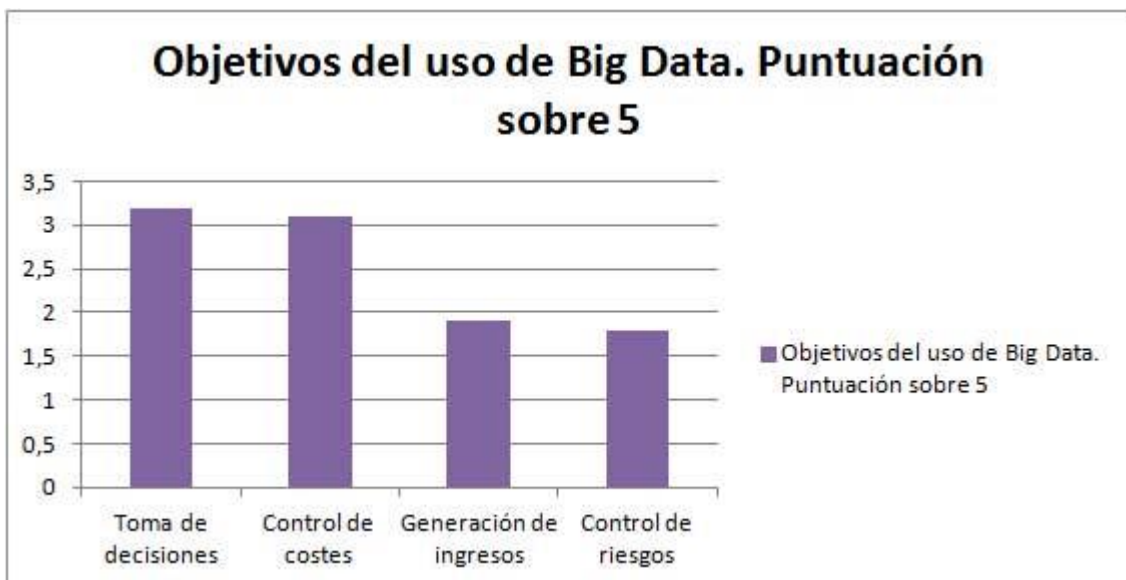


Figura 3.1. Objetivos del Big Data

Fuente: <http://www.sorayapaniagua.com/2012/06/27/tan-solo-un-48-de-las-empresas-espanolas-usa-big-data/>

CAPÍTULO 4

HERRAMIENTAS QUE UTILIZA EL BIG DATA

4.1. TÉCNICAS PRINCIPALES

Las técnicas más comunes llevadas a cabo para el correcto desarrollo del Big Data son Machine Learning, Lenguaje R, Python y MapReduce. A continuación, las estudiaremos detenidamente

4.1.1. Machine Learning

La tarea de almacenar datos es una tarea rutinaria y necesaria para las empresas. Dichas tareas se llevan a cabo por los sistemas de información. Estas tareas cobran aún más importancia en empresas de la nueva economía como la telefonía, el e-comercio, etc.

Los datos tienen un valor enorme para las empresas, ya que, es donde guardan las interacciones con los clientes, la contabilidad, etc. Lo que quiere decir que estos datos son la memoria de la empresa.

Según las palabras de Aluja Banet, T (2001, p.480), "Pero con tener memoria no es suficiente, hay que pasar a la acción inteligente sobre los datos para extraer la información que almacenan. Este es el objetivo de la Minería de Datos".

Para entender mejor que es la minería de datos, lo ideal sería centrarse en dos definiciones: *Data Mining* e *Iterative process*.

El autor Hand, (1998), lo definió así, *Data Mining*: «*the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners*»

Por otro lado, el autor Mena, (1999) lo definió así, *Iterative process*: «*the process of extracting hidden predictive patterns from large databases, using AI technologies as well as statistics techniques*»

Es decir, la primera definición define la minería de datos como el proceso de análisis secundario de grandes bases de datos con el objetivo de encontrar relaciones insospechadas que sean de interés o valor para los propietarios de bases de datos y la segunda definición define un proceso iterativo de extracción de patrones predictivos ocultos de grandes bases de datos, utilizando las tecnologías de AI ("artificial intelligent", inteligencia artificial), así como las técnicas estadísticas como análisis de clasificaciones múltiples y análisis multivariado de la varianza.

La segunda definición se centra en las raíces de la minería de datos, que es la inteligencia artificial más bien conocida como *Machine Learning*. Definida por el autor Mena, (1999) como, *Machine learning*: «*a branch of AI that deals with the design and application of learning algorithms*». Machine Learning, es una técnica para conseguir el aprendizaje automático basado en datos.

Hay dos tipos de aprendizaje, el supervisado y el no supervisado. Como ejemplo del supervisado, podría ser el análisis de sentimiento en redes sociales que se basa en clasificar las opiniones en favorables o desfavorables (puede haber más de dos tipos). Se parte de unas normas u órdenes que se han catalogado anteriormente a mano para que el programa pueda generalizar a partir de ellos. Esto requiere un tratamiento del lenguaje natural y la posibilidad de aplicar cálculos estadísticos como la obtención de frecuencias y la elaboración de gráficos.

En el caso del no supervisado, podemos pensar en la segmentación de clientes, agrupando los que tienen características comunes entre sí. De nuevo se

utiliza la estadística para obtener esa medición de cercanía o distancia entre ellos, pero sin utilizar a una clasificación previa realizada por una persona.

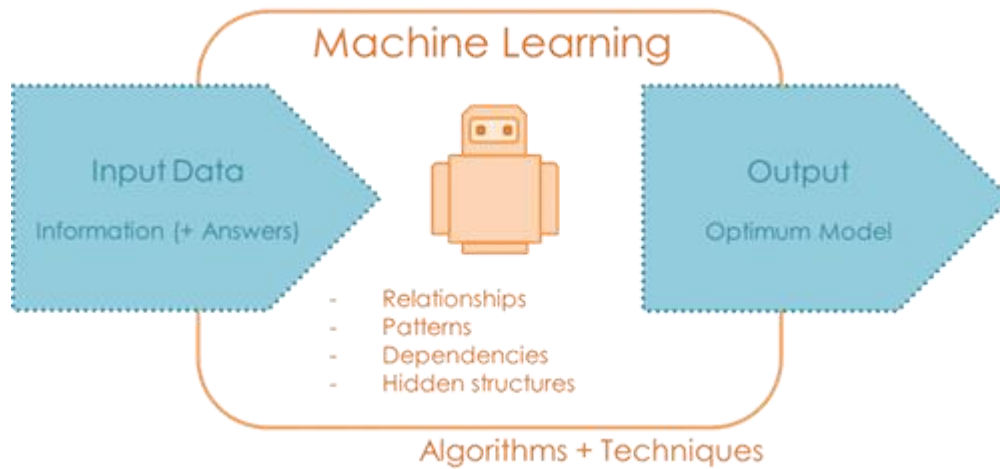


Figura 4.1. Machine Learning

Fuente: <http://quantdare.com/2016/03/machine-learning-a-brief-breakdown/>

4.1.2. Lenguaje R

El lenguaje R, es un lenguaje de programación. Este tipo de lenguaje es el más utilizado en la investigación estadística, cobra una gran importancia el campo de la minería de datos.

Este lenguaje proporciona una amplia gama de herramientas estadísticas y gráficas, además este lenguaje se integra muy fácilmente con distintas bases de datos, destinadas al procesamiento de datos masivos.

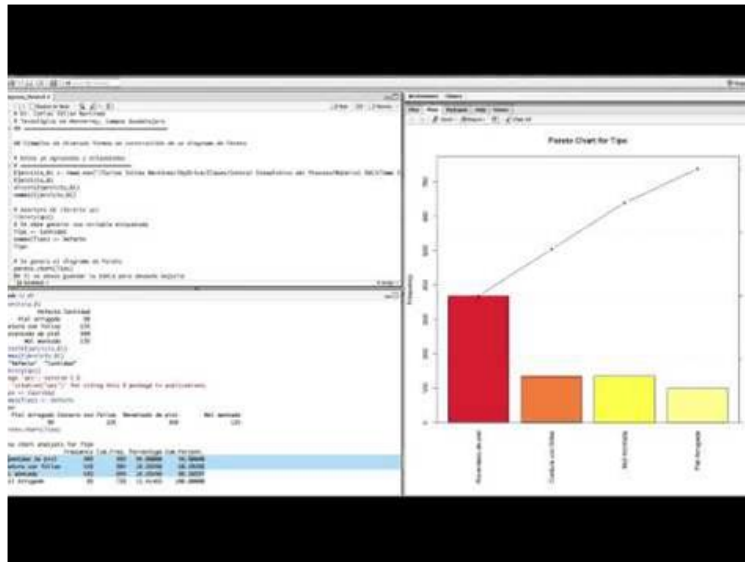


Figura 4.2. Ejemplo de programación en R

Fuente: <https://www.youtube.com/watch?v=8MDmYxuWxfs>

4.1.3. Python

Python, al igual que el lenguaje R, es un lenguaje de programación interpretado. El objetivo de Python es realizar una síntesis que ayude a obtener un código legible.

Es un lenguaje de programación multiprograma, ya que, permite varios estilos y no fuerza a las programaciones a adoptar un estilo particular de programación, es posible que tenga varios estilos como programación imperativa o la programación funcional. Python es un lenguaje interpretado que fue creado a finales de los ochenta por Guido Van Rossum, como un sucesor del lenguaje de programación ABC.

4.1.4. MapReduce

Debido al gran avance que se produjo en el campo de las tecnologías de la información se forzó a las organizaciones y empresas a llevar a cabo nuevos desafíos en el análisis de grandes cantidades de información, y de ahí surge el concepto Big Data, el cual está aplicado fundamentalmente a la información que no puede ser procesada o analizada mediante técnicas clásicas.

La propuesta más desarrollada para abordar este tipo de problemas de análisis de grandes cantidades de información es MapReduce. Propuesto por dos desarrolladores de Google, Dean y Ghemawat. El paradigma MapReduce, se centra en dos funciones fundamentales.

Por un lado, Función Map: Donde el nodo principal del clúster segmenta los datos de entrada en bloques de datos independientes y posteriormente los distribuye a los nodos de trabajo. A continuación, el nodo de trabajo procesa el problema más pequeño y lo devuelve al nodo maestro. Antes de pasar a la función reduce, se agrupan todos los valores intermedios con la misma clave para acelerar los cálculos de la función reduce. Posteriormente, en el apartado de HPPC, se detallará específicamente que es un clúster.

Por otro lado, Función Reduce: En la cual el nodo principal se encarga de recoger todos los mapas realizados por los nodos de trabajo y los combina para formar el producto final.

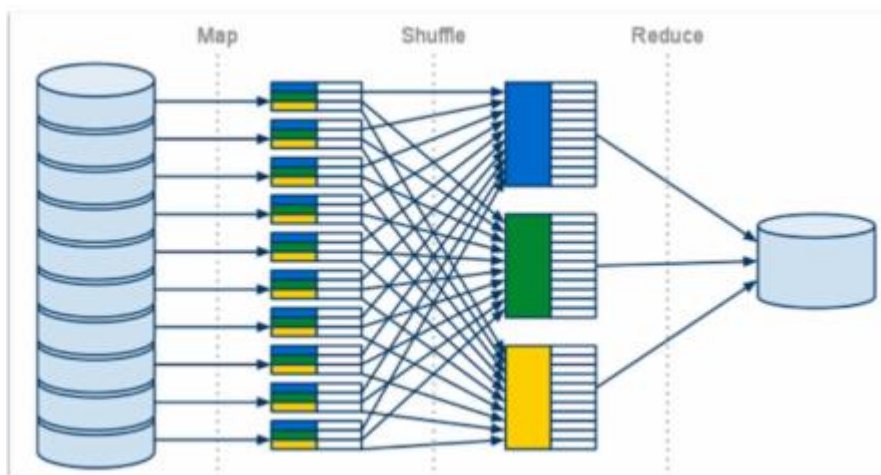


Figura 4.3. Representación del paradigma MapReduce

Fuente: Una primera aproximación al descubrimiento de subgrupos bajo el paradigma MapReduce

4.2. PROGRAMAS INFORMÁTICOS UTILIZADOS EN EL BIG DATA.

Para que las compañías puedan llevar a cabo el desarrollo de la técnica Big Data, necesitan instalar en sus empresas una serie de programas informáticos para llevar a cabo la recolección y procesamiento de datos masivos.

La elección de cada programa ha de adaptarse a las preferencias de cada empresa. A continuación, se desarrollan algunos programas posibles.

4.2.1. Hadoop

Conforme al autor Revilla, A (2013), Hadoop, es un sistema de código abierto que se utiliza para almacenar, procesar y analizar grandes volúmenes de datos; cientos de terabytes, petabytes o incluso más. Hadoop tiene la capacidad de almacenar toda clase de datos: estructurados, no estructurados, semiestructurados; archivos de registro, imágenes, video, audio, comunicación, etc. Por otra parte, Hadoop también se señala por tener una construcción con capacidad de asegurar una alta disponibilidad y recuperación de los datos que obtiene.

Hadoop nace gracias a Google. Porque, llega un momento en el que Google necesita una solución para seguir procesando datos a la vez que estos van creciendo tan rápidamente.

La solución que obtuvieron se basa en el lema 'divide y vencerás'. Esta solución fue denominada posteriormente "Hadoop". La cual, consistía en un gran número de ordenadores pequeños y cada uno de ellos estaba encargado de procesar una pequeña porción de información. Aunque ellos funcionaran de forma independiente y autónoma, todos actuaban en conjunto como si formaran un ordenador de una dimensión espectacular.



Figura 4.4 logotipo de hadoop

Fuente

https://www.google.it/search?q=hadoop+que+es&espv=2&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiZINOxutjSAhXEORoKHT7eBt0Q_AUIBigB&biw=985&bih=294#imgdii=k3GIPzt1pTqEAM:&imgsrc=Ea2kiIWNDZU8PM:

4.2.2. MongoDB

Es una de las bases de datos más famosas, esta base de datos, está orientada a los documentos, es decir, en vez de guardar los datos en registro, los guarda en documentos y los documentos se almacenan en BSON (*JavaScript Object Notation*) que es un formato de texto ligero para el intercambio de datos.

Lo que diferencia a MongoDB de otras bases de datos es que no hay que seguir un esquema, lo que es algo impensable en una base de datos de tipo relacional.

Es posible utilizar Mongo DB en la mayoría de los proyectos que se llevan a cabo en la actualidad. Se puede utilizar en cualquier aplicación que necesite almacenar datos semiestructurados, aunque no es posible utilizar esta aplicación para recoger datos de transacciones comerciales, ya que, en esta base de datos, no existen. Aunque se da la posibilidad de simular algunas, no tiene la capacidad de realizar transacciones.

4.2.3. Elasticsearch

Elasticsearch, es una herramienta que se dedica a indexar grandes volúmenes de datos para poder consultarlos después. Es decir, se dedica a realizar índices con los conjuntos de datos. Como los datos están indexados, los resultados se obtienen de forma muy rápida.

Al igual que MongoDB, esta aplicación no utiliza los propios esquemas de las bases de datos relacionales. Lo que en una base de datos de tipo relacional sería un esquema, en Elasticsearch es un índice. Esta herramienta utiliza conceptos propios para llevar a cabo el trabajo, dichos conceptos son muy similares a los de las bases de datos relacionales.

Elasticsearch trabaja recibiendo y enviando datos en formato BSON, los cuales comentamos anteriormente en el apartado de MongoDB.

Una de las diferencias que posee esta aplicación en relación a las bases de datos relacionales es que permite hacer búsqueda de textos completos a diferencia de las otras.

4.2.4. Apache Spark

Apache Spark, es un framework de código abierto, es decir, es una estructura desarrollada y distribuida libremente, que sirve para el procesar datos masivos.

Según Rous, M (s.f), en los sistemas informáticos, un *framework* es conocido como una estructura en forma de capas, que indica qué tipo de programas se pueden o deben ser construidos y cómo se interrelacionan entre sí.

Apache Spark, está diseñado en base a tres objetivos, velocidad, facilidad de uso, y capacidades avanzadas de analítica.

En opinión de Galván, P. (s.f), Spark se está volviendo muy popular, porque resuelve varias de las limitaciones que tienen de Hadoop y MapReduce. Según Galván, P. (s.f), puede decirse que, Spark puede utilizarse junto con Hadoop, pero no es necesario utilizarlos conjuntamente. Spark se encargará de extender el MapReduce, definido anteriormente, para mejorarlo y hacerlo más rápido.

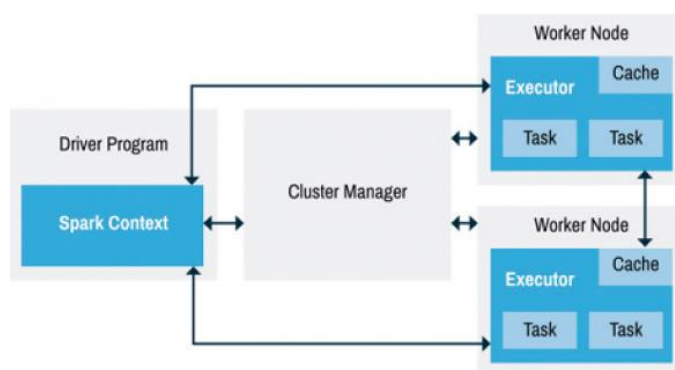


Figura 4.5. Modelo de funcionamiento de Apache Spark

Fuente: <https://sg.com.mx/revista/50/un-vistazo-apache-spark-streaming#.WMrKUDs1-00>

4.2.5. Apache Storm

Apache Storm, es un sistema de computación distribuido en tiempo real. Es tolerante a fallos, es decir, el sistema seguirá funcionando correctamente en caso de fallo de uno o varios de sus componentes.

Este programa se encarga de facilitar la tarea de procesar streams de datos, es decir, datos que requieren una gran capacidad para procesarlos. En este sistema, se pueden utilizar cualquier lenguaje de programación y en general es bastante simple.

Storm trabaja utilizando tuplas como modelo de datos. Una tupla es simplemente una lista de valores con nombre.

Las principales ventajas de Apache Storm son las siguientes. Soporta una variedad usos posibles. Garantiza el no perder ningún dato. Se puede configurar de forma sencilla. Es tolerante a fallos, como mencionamos anteriormente, si existiera algún fallo el sistema reasigna las tareas. Por último, los componentes de Storm pueden ser definidos en cualquier lenguaje de programación.

4.2.6. HPCC

HPCC son las siglas de '*High Performance Computing Cluster*', es decir, Cluster de computación de alto rendimiento. Cluster es un término que se utiliza para definir un conjunto de ordenadores interconectados entre sí por una red de alta velocidad, estos ordenadores que trabajan juntos se comportan como solo un ordenador de grandes dimensiones.

El objetivo principal de un Cluster es conseguir el mayor rendimiento en la velocidad de proceso de datos.

Este tipo de tecnología lo que permite es, que un conjunto de ordenadores trabaje a la vez, dividiendo el trabajo en varias tareas más pequeñas, dichas tareas se pueden desarrollar de forma paralela. Como mencionamos anteriormente en el apartado de Hadoop, se basa en el lema 'divide y vencerás'.

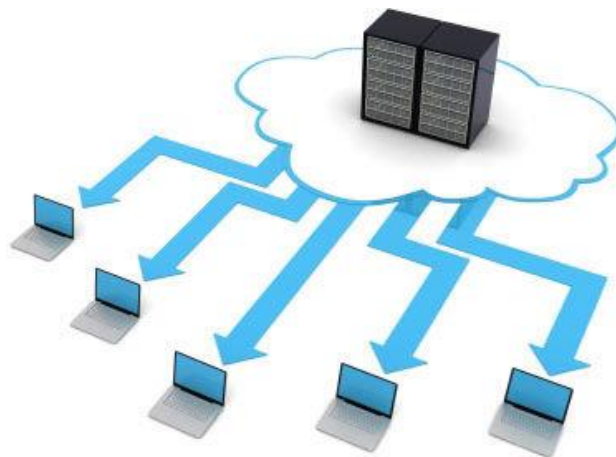


Figura 4.6. Clúster

Fuente: <https://www.raffaelechiatto.com/category/microsoft/microsoft-cluster/>

4.2.7. Hana

Hana es una plataforma de datos que se puede llevar a cabo tanto en un dispositivo local como en la nube.

Es una plataforma innovadora, que se adapta perfectamente a la tarea de realizar análisis y llevar a cabo operaciones a tiempo real.

Las empresas que trabajen con esta base de datos tienen la posibilidad de realizarle preguntas complejas e interactivas y Hana podrá responder efectivamente ya que, esta base de datos está trabajando con bases de datos enormes.

CAPÍTULO 5

VENTAJAS QUE LE OFRECE EL BIG DATA A LAS EMPRESAS

5.1. VENTAJAS COMPETITIVAS EN CUALQUIER SECTOR EMPRESARIAL

La técnica de la aplicación del Big Data es llevada a cabo por las empresas para tomar mejores decisiones y conseguir ventajas competitivas sobre sus competidores.

Las principales ventajas competitivas son las siguientes:

- La gran distribución.
- La mejora en la eficiencia y los costes.
- La mejora de la gestión empresarial.
- Permite el almacenamiento en la nube.
- Permite describir necesidades y puntos de mejora para las compañías.
- La inmersión de nuevas variables en la toma de decisiones.
- La innovación en la creación de productos y servicios.
- La segmentación.
- Una dinámica de trabajo más eficaz y rápida.
- La velocidad en la toma de decisiones.
- Permite desarrollar planes estratégicos inteligentes de Marketing.
- La vinculación de clientes.
- La agregación de datos del entorno que afectan al proceso principal.
- El Feedback y la retroalimentación.

A continuación, desarrollaremos las diferentes ventajas competitivas que surgen gracias a la aplicación del Big Data en las diferentes empresas.

Para comenzar, una de las ventajas competitivas que se dan es **la gran distribución**, ya que, esta técnica permite optimizar, actualizar y ajustar inventarios en tiempo real en función de la demanda. Gracias a la extracción de datos o *data mining* se puede analizar el comportamiento de los clientes y gracias a ello, fijar precios y ofrecerle los incentivos que corresponda para atraer a un mayor número de clientes. Es posible llevarlo a cabo, mediante la digitalización de elementos como cajas registradoras o almacenes.

El análisis del Big Data, **mejora la eficiencia y los costes**, ya que, dicho análisis puede dar lugar al aumento de la velocidad con la que se desarrolla un producto. Además, permite realizar simulaciones de producto y compartir datos muy rápidamente.

Otra ventaja competitiva que nos aporta la adopción del Big Data en una empresa es **la mejora de la gestión empresarial**. A parte de optimizar el inventario y la cadena de suministro, el Big data reduce el ciclo de conversión de efectivo, ayuda a la toma de decisiones empresariales fundamentadas en datos recogidos a tiempo real, que son clave para el futuro de la empresa y controla factores de riesgo.

Uno de los grandes problemas en la gestión de altos volúmenes de datos es el alto coste que soporta la infraestructura de almacenamiento. Por ello, la elección de la técnica Big Data aportara una nueva ventaja competitiva a las empresas, ya que, **permite el almacenamiento en la nube**. Muchos proveedores de almacenamiento de datos masivos, alquilan servidores potentes a través de los cuales ponen a disposición

del cliente la posibilidad de almacenarlos en una especie de nube a la que se puede acceder en línea. El resultado de todo este procedimiento, es, que se puede acceder a ella mediante aplicaciones diseñadas para el manejo de grandes volúmenes de datos y como consecuencia se pueden obtener soluciones a tiempo real y de forma sencilla.

Si una empresa decide llevar a cabo la aplicación del Big data en su organización, implementará mejoras tecnológicas que le permitan la adquisición de datos y con ellos podrá **describir necesidades y puntos de mejora para su compañía.**

Del mismo modo, otra ventaja competitiva es **la inmersión de nuevas variables en la toma de decisiones** a través de algoritmos automatizados. Las tomas de decisiones no son siempre las mismas, algunas organizaciones optimizan sus decisiones mediante el análisis de datos de empleados, clientes o incluso sensores que incorporan a los productos. De este modo, el análisis de los datos puede mejorar la toma de decisiones dentro de una compañía reduciendo los riesgos.

Es muy importante tener en cuenta que el Big Data permite a las empresas evaluar a fondo sus productos mediante el análisis de datos. Gracias a este análisis, las compañías obtienen una valiosa información que les permite crear nuevos productos o rediseñar y mejorar los productos ya existentes. Por tanto, otra ventaja competitiva es **la innovación en la creación de productos y servicios.**

La aplicación del Big Data permite a las empresas llevar a cabo **la segmentación de los clientes** para personalizar acciones. De esta manera las compañías tienen mayor facilidad para orientar los servicios y satisfacer las necesidades de sus clientes de una forma más específica.

Así mismo, la aplicación del Big Data, mejora la accesibilidad y la fluidez de la información dentro de la propia empresa. Las empresas que habiliten herramientas para facilitar la búsqueda de información y que digitalicen los datos, crearan una **dinámica de trabajo más eficaz y rápida.**

Hay que mencionar, además que, otra ventaja comparativa es **la velocidad en la toma de decisiones.** El Big Data nos proporciona un manejo dinámico de la información, lo cual es fundamental como base para la correcta toma de decisiones. Gracias a esta técnica, podremos asumir decisiones veloces e inteligentes que favorezcan a nuestro negocio, puesto que estas, se han basado en un fundamento firme y eficaz. Además, es posible realizar un análisis de oportunidad antes de poner cualquier producto o servicio en el mercado.

A través de los datos que nos proporciona el Big Data, se llevan a cabo **planes estratégicos inteligentes de Marketing.** En la actualidad, es posible conocer que piensan los clientes sobre una marca o producto, analizar y predecir el comportamiento que un usuario tendrá en la red y cuáles son sus necesidades reales a la hora de adquirir bienes o servicios. Es posible analizar parámetros relativos al perfil específico de cada usuario, sus tendencias, su vinculación a la marca o sus preferencias, de forma que sea posible para las compañías elaborar campañas de marketing con un alto nivel de personalización.

Cada día es más complicado captar nuevos clientes, y a su vez, esta adquisición es cada vez más costosa que fidelizar a los que ya tenemos en la empresa. Una de las ventajas competitivas que aporta el Big Data es **la vinculación de clientes.** Por lo tanto, es esencial utilizar correctamente los datos de los que disponemos para ofrecerle a nuestros clientes personalmente lo que desean y para ello es esencial el Big Data. De ahí que, conocer a través de los datos las necesidades de nuestros clientes, el nivel de satisfacción, etc. es una de las claves fundamentales para tener a los clientes fidelizados.

Una característica de bastante importancia que debemos de tener en cuenta en la utilización del Big Data es **la agregación de datos del entorno que afectan al proceso principal**. Estas variables responden al ecosistema que puede afectar a un cliente en el momento de la decisión de compra. Por lo tanto, dichas variables no están dirigidas precisamente al análisis de un cliente, ni a su comportamiento directo. Esta característica se basa en introducir en la fórmula de análisis provenientes de fuentes externas como pueden ser la zona donde vive el usuario, la meteorología o las estaciones del año. Estas variables externas hacen que el cálculo aplicado sea más eficiente, dirigiendo mejor nuestros productos y servicios no solo a comunidades específicas, si no a temporadas concretas.

Una de las aplicaciones de mayor importancia en la utilización del Big Data es el **Feedback y la retroalimentación** que genera el uso de esta técnica. Gracias a esta aplicación, es posible analizar los resultados de las acciones realizadas en tiempo real, lo que nos posibilita corregir rápidamente las desviaciones en las estrategias diseñadas u observar los aspectos que más éxito han tenido. El continuo análisis del feedback de las acciones llevadas a cabo con una tecnología veloz y dinámica permite no tener costes extras que no son necesarios para continuar con la estrategia de comercialización.

Las ventajas del Big Data

Comparativa

Hace 10 años

Un gran almacenador de datos tenía muy poca capacidad.

Hoy

Utilizan capacidades cercanas a los **10 Zettabytes**.

80%

De dichos datos se han generado en los últimos años.

¿Dónde van a parar todos estos datos?

Tradicionalmente: Volúmenes elevados de datos que podían manejarse de manera fluida utilizando:

Fuentes del tipo ERP.

Sistemas de gestión de logística, producción y facturación.

CRM o sistemas de gestión de clientes.

Uso masivo de la web

Multiplicado exponencialmente el volumen de datos (contenidos, tiendas online y redes sociales) con respecto al que gestionaban nuestras tecnologías de proceso e inteligencia.

se han quedado obsoletas

Capacidad de manejar Big Data

Ventajas competitivas a las empresas:



Velocidad en la toma de decisiones
Gracias a un análisis de oportunidad previo



Planes estratégicos Inteligentes de Marketing
Analizar y predecir el comportamiento de usuarios.



Mejora en la eficiencia y en costes
Los plazos y los costes de desarrollo de producto se acortan



Vinculación de clientes
Datos para adquirir y fidelizar a los usuarios.



Alcance de variables del entorno
Que pueden afectar en la decisión de compra.



Feedback y retroalimentación
Tiempo real, con una tecnología dinámica y veloz.

Estamos en la era del Big Data

Uno de los elementos que más valor añadido aporta a las empresas en el momento de definir sus estrategias.

OFF

Permanecer desconectado supone:

Un coste demasiado elevado para las empresas en su objetivo de crecimiento

Una pérdida de posicionamiento estratégico.

Principales aliados para las compañías:

Almacenamiento + procesamiento + correcto aprovechamiento de los datos

Figura 5.1. Las ventajas del Big Data

Fuente: <http://impulsandoexperiencias.com/ventajas-competitivas-del-big-data/>

Para finalizar, en mi opinión una de las preferencias del Big Data es mejorar la experiencia del cliente, este es un factor por el cual muchas empresas deciden implementar esta técnica, al igual que la eficiencia de procesos. Es importante mencionar que nos encontramos en la era del Big Data. Ya que este, es uno de los elementos más competitivos y que más valor añadido aporta a las empresas a la hora de definir sus estrategias. Por ello, esta técnica se ha convertido en la mejor opción de las organizaciones, para posicionarse mejor que sus competidores. Permite el procesamiento y el correcto aprovechamiento de los datos. Las compañías que se

sitúen ajenas a esta realidad, soportaran un coste elevado en su objetivo de crecimiento y sobre todo una pérdida de posicionamiento estratégico.

5.2. APLICACIÓN DEL BIG DATA EN EL SECTOR TURISTICO

Con la adopción de la nueva tecnología Big Data, el sector turístico tiene la oportunidad de transformar la forma de ofrecer servicios a sus clientes. Gracias al Big data, ofrecerán una experiencia de viaje más personalizada y maximizarán sus beneficios.

Una de las ventajas competitivas que ofrece la adopción del Big Data es mejorar la relación con los clientes. Con la posibilidad de seguir las **huellas digitales** de los clientes, las empresas conseguirán un trato más cercano y de confianza. Actualmente, la mayor parte de los consumidores de servicios del sector turístico, ya sean, compañías aéreas, agencias de viajes, compañías hoteleras, etc. Hacen sus compras y reservas por Internet. Cuando un cliente accede a Internet para realizar búsquedas específicas, reservar en algún alojamiento, comprar billetes de transporte, pagar por cualquier servicio etc. los sistemas detectan las elecciones que han tomado los clientes. Pero, es posible añadirle a esta ventaja de determinar sus gustos y preferencias la opción de conocer las tendencias de las noticias que forman parte del contexto, la ubicación geográfica, el clima, su correo electrónico, su nombre, etc.

Las aplicaciones móviles aportan un gran valor a la compañía, ya que a través de ellas es posible conocer la experiencia del cliente a tiempo real gracias a las valoraciones, puntuaciones y opiniones de estos. Por lo tanto, gracias a la agregación de todas estas huellas digitales que deja el usuario de servicios turísticos es posible crear un producto personalizado a la medida de los clientes.

Como nombramos anterior mente el Big Data hace posible que las empresas que lo implementen lleven a cabo el desarrollo de nuevos productos y servicios. Gracias a la tecnología big data la toma de decisiones es mucho más fácil para el empresario. Sobre todo, lo ayuda a mejorar el enfoque de la estrategia comercial, orientar las acciones de marketing o determinar las políticas de calidad gracias a la opinión de los clientes sobre los servicios ofrecidos.

En este caso, como mencionamos anteriormente, las aplicaciones móviles tienen un papel crucial, ya que, desde los dispositivos inteligentes los clientes tienen la posibilidad tanto de obtener como de generar información en cualquier momento del día, alquilar servicios, consultar u descubrir actividades de ocio, contratar viajes, realizar y anular reservas, recibir ofertas, consultar guías de viaje virtuales y una infinidad de acciones posibles en el sector turístico.

La autonomía que posee hoy en día el cliente de servicios turísticos para estar continuamente conectado a tiempo real desde cualquier parte del mundo, fuerza a las compañías a redefinir la planificación, ejecución y análisis de las estrategias con los datos que se han generado, debidamente cruzados, ordenados y gestionados por procesos informáticos del Big Data.

Si aplicamos el Big Data al Sector turístico, conoceremos con una mayor precisión en que invierte su tiempo cada viajero, el volumen real de afluencia de clientes en un determinado destino, sus necesidades, sus gustos, los idiomas más hablados, etc. De donde resulta que es más fácil calcular el nivel real de demanda con un sistema reglado de reservas. Por lo tanto, **el Big Data es el mejor soporte para la toma de decisiones**.

De ahí que, la recogida de datos permita optimizar los resultados de la gestión turística de un destino concreto, ya que, le ofrece al turista lo que más se acerca a sus deseos tras conocer sus necesidades y comportamientos.

Algunos ejemplos de nuevas prácticas que actualmente se están llevando a cabo son, unificar los datos a través de múltiples sistemas para aplicar las medidas necesarias y ofrecer los productos que interesan a cada segmento (todavía en vías de desarrollo). Usar la tecnología Big Data como base del sistema de gestión de ingresos a nivel corporativo. Predecir el precio de los vuelos en un periodo de tiempo entre 7 y 10 días para ofrecer la tarifa más conveniente a los clientes y mejorar el negocio a partir de la experiencia del cliente y estrechar más la relación con el mismo.

El turismo es un sector de gran peso en España, como mencionamos anteriormente, representa el 10% del PIB, por lo tanto, es bastante importante que este sector tome una mayor fuerza y crezca. La mejor opción para dicho crecimiento, es que España realice el esfuerzo de estar informado de las nuevas tendencias tecnológicas, como es el caso del Big Data.

En este sector tradicionalmente se realizaban estudios basados en encuestas para conocer mejor los gustos y preferencias del viajero y con ello adaptarse mejor a la demanda del cliente. Pero los dos grandes inconvenientes de esta práctica eran, el tamaño de la muestra y los sesgos de deseabilidad social.

Actualmente, cada vez va creciendo el número de empresas que confían en el uso de análisis Big Data. Ya que, es posible trabajar con una muestra de datos exponencialmente mayor y además les aportan una mayor fiabilidad en los resultados, puesto que están basados en la huella digital que dejan los clientes en la web, las cuales describen lo que el viajero o futuro viajero hace efectivamente, no lo que cree o desearía hacer.

Para finalizar este punto realizaremos una recopilación de los beneficios o ventajas comparativas que una empresa del sector turístico podría conseguir si implementara esta nueva tecnología.

- **Anticiparse a las necesidades del cliente.** Saber lo que va a pasar gracias al estudio de los datos ayuda a las empresas a anticiparse a los comportamientos tanto del cliente como del mercado. Como hemos comentado anteriormente con las huellas digitales, el análisis predictivo puede ser utilizado para conocer mejor al viajero
- **Mejorar el impacto de las acciones de marketing.** El departamento de marketing de las empresas que integran el sector turístico es uno de los que más se beneficia del análisis de datos, ya que, permite hacer un seguimiento de las acciones de marketing, y mantenerlas o modificarlas en función de los resultados obtenidos.
- **Atraer y fidelizar al viajero.** Gracias a una correcta segmentación del público, es posible medir el momento y manera perfectos para lanzar un producto o servicio según las preferencias y acciones observadas en los usuarios, prediciendo sus gustos y su propensión de compra.
- **Mejorar la toma de decisiones.** Como el Big Data les facilita a las empresas el conocimiento de qué está ocurriendo a tiempo real, es decir, en cada momento. Permite que las empresas reaccionen de una forma más eficiente e inmediata, mejorando así la toma de decisiones.
- **Encontrar nuevas oportunidades en el sector turístico.** El sector turístico está experimentando una evolución gracias al mundo digital que nos permite conocer información privilegiada sobre los usuarios, la cual nos permite encontrar nuevas oportunidades de negocio en este sector.
- **Reducir el riesgo de las inversiones.** Como llevamos diciendo desde el principio de este estudio sobre el big data, este nos proporciona información valiosa. Esta información se puede utilizar para abrirnos a nuevos proyectos, ya que, dicha información nos da un mayor conocimiento del sector, por consiguiente, reduce la sensación de inseguridad que les

surge a los emprendedores al darles la oportunidad de anticiparse aproximadamente a lo que podrá pasar una vez que realicemos la inversión.

CAPÍTULO 6

EL BIG DATA EN ESPAÑA

6.1. ANÁLISIS DE DATOS SOBRE EL BIG DATA EN ESPAÑA.

Según un estudio realizado por el instituto nacional de estadística (INE), INE, (2016 np 978), en el primer cuatrimestre de 2016, en España el 8,5% de empresas realizaron análisis de Big Data en el año 2015 y este porcentaje alcanzó el 21,5% en el caso de las empresas con 250 o más empleados. La mitad de las empresas que realizaron el análisis del Big Data, lo decidieron utilizar como fuente de Datos por geolocalización a partir de dispositivos portátiles. El 82,6% de empresas, decidieron realizar el análisis con su propio personal, incluido el de la empresa matriz o filiales.

En la figura 6.1. podemos observar el porcentaje sobre el total de empresas que realizaron el análisis del Big data.

Ese mismo estudio realizado por el INE en el primer cuatrimestre de 2016, refleja que en España el 19.3% de las empresas compuestas por más de 10 empleados compran soluciones de *Cloud Computing*, es decir, computación en la nube. Las empresas que lleva a cabo esta solución lo hacen pagando servicios que se encuentran en servidores compartidos. Además, como podemos observar en la figura 6.2., en la cual se muestra el porcentaje de empresas que llevan a cabo el uso de las tecnologías de la información y la comunicación (TIC), por comunidades autónomas y ciudades autónomas en que se ubica la sede social de la empresa. Madrid es la comunidad autónoma que más utiliza el *Cloud Computing* con un 27,4%, seguido de Cataluña (25.6 %) y Canarias con un 20.4%. En esta figura, se refleja que el tipo de TIC que menos es utilizada por las empresas españolas es el *Cloud Computing* con un 19.3%, en comparación con la Conexión a Internet (98.4%). Aunque hay datos que afirman que este panorama cambiara ya que en el año 2016 el 26,3% de las empresas emplearon a especialistas en TIC y el 13,2% contrataron a nuevos especialistas en el año 2015.

	Número de empleados			
	TOTAL	10 a 49	50 a 249	250 o más
Año 2015				
Datos por geocalización a partir de dispositivos portátiles	50,4	53,1	43,1	46,4
Datos generados por medios sociales	48,6	50,9	43,3	41,8
Datos de la propia empresa con sensores o dispositivos inteligentes	34,5	27,6	48,4	61,0
Otras fuentes de Big Data	20,7	18,8	23,2	32,1

Figura 6.1. Análisis del Big Data
Fuente: INE

Primer trimestre de 2016	Conexión a Internet		Internet y Página Web ⁽¹⁾		Banda Ancha móvil ⁽¹⁾		Uso de Firma Digital ⁽¹⁾		Uso de Medios Sociales ⁽¹⁾		Cloud Computing ⁽¹⁾	
	Internet		Página Web ⁽¹⁾		móvil ⁽¹⁾		Digital ⁽¹⁾		Sociales ⁽¹⁾		Computing ⁽¹⁾	
TOTAL	98,4		77,5		82,9		75,1		42,9		19,3	
Andalucía	99,1		71,9		80,8		78,6		42,4		16,2	
Aragón	99,8		84,1		87,2		69,3		39,4		15,3	
Asturias, Principado de	99,4		81,8		80,5		75,1		41,1		12,4	
Baleares, Illes	99,6		73,2		74,6		65,9		45,2		16,1	
Canarias	96,7		68,1		79,4		74,2		47,5		20,4	
Cantabria	96,9		73,9		73,9		67,0		36,6		14,6	
Castilla y León	98,6		74,7		79,0		74,4		33,3		11,2	
Castilla-La Mancha	94,7		68,4		79,4		77,4		41,3		9,2	
Cataluña	98,6		82,3		84,0		76,1		46,8		25,6	
Comunitat Valenciana	98,2		75,9		84,5		72,4		45,5		16,5	
Extremadura	99,3		74,2		86,1		77,2		38,3		8,5	
Galicia	97,9		76,3		79,4		71,9		35,5		11,5	
Madrid, Comunidad de	98,7		80,8		85,4		78,1		45,3		27,4	
Murcia, Región de	99,2		71,1		80,2		71,8		42,7		13,8	
Navarra, Comunidad Foral de	99,2		80,9		81,6		81,9		36,2		11,6	
País Vasco	96,6		81,7		88,0		70,4		36,9		18,1	
Rioja, La	97,5		75,3		81,2		80,0		43,2		11,7	
Ceuta	96,4		74,6		78,4		77,8		25,4		10,3	
Melilla	100,0		56,4		72,9		48,8		35,8		9,3	

(1) Porcentaje sobre el total de empresas con conexión a Internet

Figura 6.2. Uso de las TIC

Fuente: INE

Se ha ido notando un gran crecimiento en las empresas que están enfocadas en el sector del Big Data en nuestro país. Según un estudio realizado en 2015 por LogicFin, empresa que se dedica a proveer información y análisis para la toma de decisiones empresariales. En el año 2015 en España había más de 300 empresas Big Data, la mayor parte son pymes recién creadas, es decir, empresas nuevas que tienen alrededor de 50 trabajadores. Como hemos podido observar en el estudio anterior realizado por el INE esta tendencia continúa en aumento, gracias al interés de las empresas en formarse con especialistas.

Como podemos observar en la figura 6.3. en el año 2013 se observa un notable crecimiento en el número de empresas Big Data. Se pasa de tan solo 20 empresas creadas en 2012 a 44 empresas en 2013 y más de 50 en 2014.

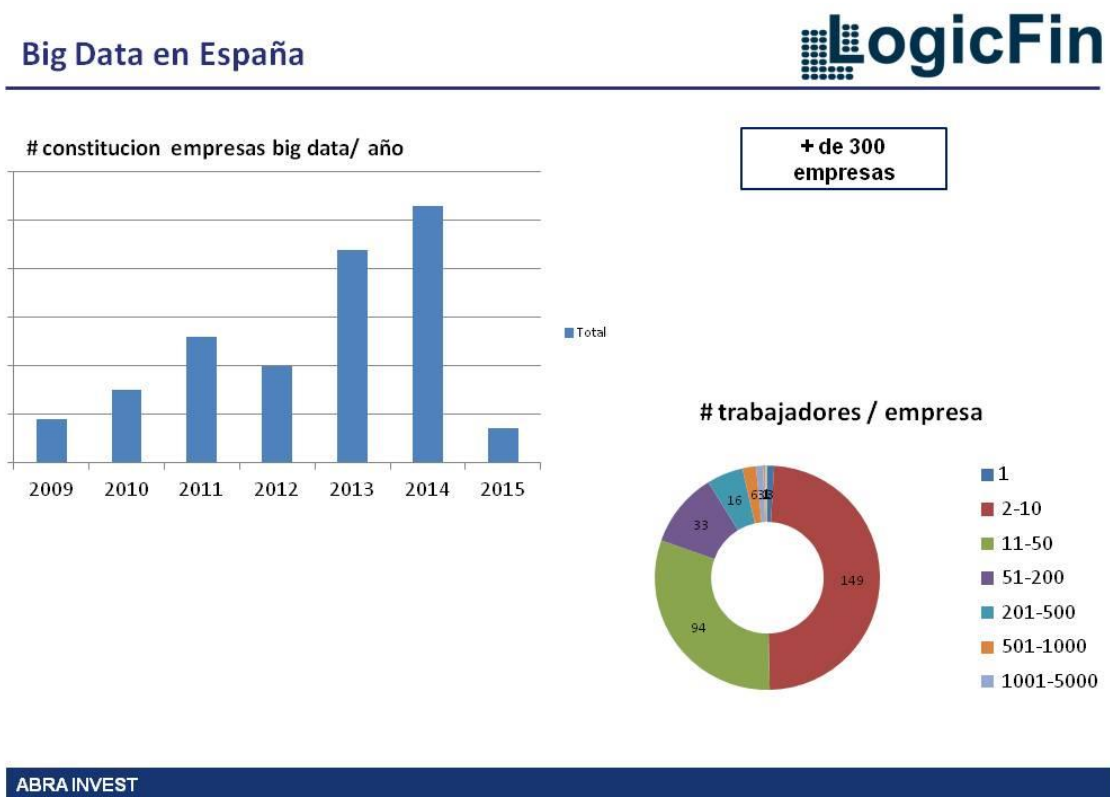


Figura 6.3. Creación de empresas Big Data

Fuente: LogicFin

Según otro estudio realizado por la universidad de Alcalá en el año 2016, aproximadamente un 70% de las empresas españolas llevan a cabo soluciones Big Data y un porcentaje aún mayor de empresarios afirman que esta técnica es esencial para realizar la transformación digital de una empresa. A partir del año 2013, fue cuando comenzó a crecer este sector como hemos podido observar en graficas anteriores. Y actualmente ya existen un gran número de empresas punteras que están liderado esta forma de hacer negocio en nuestro país. Algunas de las empresas Big Data con gran importancia en España son: Stratio, Capside, Iwatrix, Pragsis, Power

Data, Synergic Partners, GMV, Treelogic, Core Networks, Daedalus y Geoblink entre otras.

Por ejemplo, Iwatrix se encarga de ofrecerle a sus clientes la posibilidad de tener un mayor conocimiento de sus usuarios y trabajadores a través de su software el cual analiza el comportamiento que tienen las personas cuando utilizan las diferentes tecnologías. Geoblink trabaja optimizando las redes de tiendas de empresas de Retail, es decir empresas que llevan a cabo la comercialización masiva de productos. Geoblink, lleva a cabo este trabajo con la utilización de técnicas avanzadas de análisis de datos small y big data.

En lo que se refiere al Big Data en el sector turístico, Antonio López de Ávila, el presidente de Seguittur, una sociedad estatal española que se dedica a la gestión de la innovación y las tecnologías turísticas, afirmó en el debate sobre 'Fuentes de información: Claves en el sistema de Inteligencia Turística', que España es un país pionero en el uso del Big Data en el sector turístico y en su desarrollo e integración dentro de un plan de ámbito estatal, en referencia al Plan Nacional de Turismo 2012-2015 y que considera el desarrollo de los destinos turísticos inteligentes.

Según el autor Ávila, L. A. (2016), los turistas hoy día están conectados a Internet, lo que conlleva que, este consumo genere muchos datos que puedan analizarse y convertirse en conocimiento que ayude a reinventar los destinos y empresas y poder así, garantizar tanto el desarrollo sostenible de un determinado territorio turístico como una oferta personalizada a cada tipo de visitante. Por ello, Seguittur ha desarrollado SIT, que significa Sistema de Inteligencia Turístico. Lo que pretende es que a través de la recolección y transformación de los datos se mejore la capacidad, eficiencia y la toma de decisiones de las pymes. Y España, es la primera en aplicar este sistema en la parte turística, ya que, es el sistema que está impulsando el crecimiento económico en el país.

CAPÍTULO 7

CONCLUSIONES

7.1. CONCLUSIONES SOBRE EL ESTUDIO.

Tras el estudio realizado sobre el Big Data y su aplicación en las empresas españolas, es posible llegar a diferentes conclusiones:

- El Big Data es capaz de soportar la recolección y análisis de datos de cualquier tipo.
- Actualmente existe una amplia gama de programas que permiten que las empresas lleven a cabo esta técnica.
- El Big Data es una de las técnicas que más ventaja comparativa aporta a las empresas hoy día.
- Un tema que en mi opinión es uno de los más importantes, es, que España haya conseguido ser pionera en la utilización del Big Data en el sector turístico. Este logro aportara grandes beneficios al país, ya que el sector turístico es uno de los más importantes en España.
- Las empresas cada vez apuestan más por las nuevas tecnologías, y desde 2013, las empresas centradas en ellas no han parado de crecer.
- El Big Data es de gran utilidad en cualquier sector empresarial, pero en España específicamente, es de gran importancia su aplicación en el sector turístico.
- Las empresas están invirtiendo en estudios y análisis del Big Data para implementarlo en sus empresas.
- Las empresas de nuestro país realizan el esfuerzo de ponerse en manos de expertos para conseguir los medios y la información necesaria para digitalizar sus empresas.
- No solo las empresas están implementando esas técnicas, si no, que se están creando cada vez más empresas nuevas que proveen los medios necesarios para que las empresas lleven a cabo el Big Data.
- El Big Data es una técnica que se puede implementar en cualquier empresa, ya que, se adapta a las necesidades de cada una de ellas.
- En nuestro país, la principal comunidad autónoma que apuesta por esta técnica es Madrid.
- Es una tendencia que está en desarrollo y crecimiento hoy día, y en mi opinión, solo está en proceso de despegue aún tiene que desarrollarse mucho más.

Bibliografía

- Aguilar, L. J. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.
- Aluja Banet, T. (2001). *La minería de datos, entre la estadística y la inteligencia artificial*.
- Application-Controlled Demand Paging for Out-of-Core Visualization Michael Cox and David Ellsworth1 Report NAS-97-010, July 1997 NASA Ames Research Center MS T27A-2 Moffett Field, CA 94035-1000
- Ashton, K. "That Internet of Things Thing". RFID Journal (2009). URL: <http://www.rfidjournal.com/articles/view,4986>.
- Baos Analytics Everywhere: <http://www.baoss.es/10-herramientas-para-manejar-big-data-analytics/>
- Bell, D. (2000). The coming of post-industrial society. *The Urban Lawyer*, 6, 738.
- Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2015). "Knowing the Big Data". *Facultad de Ingeniería*, 24(38), 63-77.
- CIFF. Universidad de Alcalá: <http://www.ciff.net/blog/cuales-son-las-principales-empresas-big-data-en-espana.html>
- Danah Boyd & Kate Crawford (2012). "CRITICAL QUESTIONS FOR BIG DATA", *Information, Communication & Society*, 15:5, 662-679, DOI: 10.1080/1369118X.2012.678878
- Duff A. S (2000). *Information society studies*. Rotledge.
- España, "pionera" en el uso de Big Data para el sector turístico, (10 de octubre de 2016). *Europapress* . <http://www.europapress.es/turismo/nacional/noticia-espana-pionera-uso-big-data-sector-turistico-20161019130547.html>
- Exceltur: <https://www.exceltur.org/>
- Fundación general de la Universidad de Salamanca: <https://fundacion.usal.es/es/empresas-amigas/205-contenidos/actualidad/1163-el-networking-empresarial>
- Galván, P. (s. f). "Un vistazo a Apache Spark Streaming". *SG Buzz*.
- Granados, O. (4 de Junio de 2016). "Big Data: la nueva materia prima". *El País*. http://economia.elpais.com/economia/2016/06/03/actualidad/1464954943_672966.html
- GRAY, J. (1995). Database systems: A textbook case of research paying off. *Committee on the Fundamentals of Computer Science: Challenges and Opportunities, editors, Computer Science: Reflections on the Field, Reflections from the Field*, 80-88.
- Heath, T, Hepp, M., and Bizer, C. (eds.). "Special Issue on Linked Data, International Journal on Semantic Web and Information Systems". *Freie Universität Berlin, Germany*
- IANUX, IT and security solution: [http://www.ianux.com/Solucion%20HPCC%20\(Cluster%20de%20Alto%20Rendimiento\)](http://www.ianux.com/Solucion%20HPCC%20(Cluster%20de%20Alto%20Rendimiento))
- IBM. Que es el Big Data. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- IDC: <http://mx.idclatin.com/about/>
- INE:
- Infocif. Ventajas del Big Data: <http://noticias.infocif.es/noticia/5-ventajas-que-nos-puede-aportar-el-big-data>
- Instituto de Economía Digital: <http://www.icemd.com/digital-knowledge/articulos/las-ventajas-del-big-data/>
- Instituto de Ingeniería del Conocimiento: <http://www.iic.uam.es/digital/5-beneficios-big-data-sector-turismo/>
- Instituto de Ingeniería del conocimiento: <http://www.iic.uam.es/innovacion/herramientas-big-data-para-empresa/>

José Julio López. Ideas de un Project Manager:

<https://josejulilopezsantos.wordpress.com/2014/07/11/las-principales-tecnicas-big-data-y-sus-aplicaciones/>

José. J.L. "Ideas para Project Management, PMP, PMO, ITIL, Business Intelligence, Big Data"
<https://josejulilopezsantos.wordpress.com/>

Leo-Revilla, A. 2013. "¿Qué es Hadoop?". *Momentotic*

Logicalis, Busines and technologic working as one. Big Data: ventajas de "la revolución de los datos masivos" <https://blog.es.logicalis.com/analytics/big-data-ventajas-de-la-revolucion-de-los-datos-masivos>

LogicFin: <http://logic-fin.com/>

López, J.C. (2014). "La moda del Big Data: ¿En qué consiste en realidad?". *El economista*

Muy Pymes. Cinco ventajas competitivas para que se desarrolle una estrategia adecuada de Big Data: <http://www.muypymes.com/2013/11/16/estrategia-big-data>

Power data. El valor de la Gestión de datos.: <http://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/328884/Qu-es-el-Big-Data-y-por-qu-es-tan-importante>

Power, D.J., "A Brief History of Decision Support Systems, DSSResources.COM, Word Wide Web, URL. DSSResources.COM/history/dsshistory2.8.html, version 2.8, May 31, 2003".

Press, G (2003). "A very short history of big data". Forbes

Pulgar-Rubio, F., Carmona, C. J., Rivera-Rivas, A. J., & González, P. "Una primera aproximación al descubrimiento de subgrupos bajo el paradigma MapReduce". *Departamento de Informática, Universidad de Jaén*

Red de Centros SAT. (2014). *¿Qué es el Big Data?* CTIC

Red de centros Sat: <http://www.fundacionctic.org/sat/articulo-que-es-el-big-data>

Rick Sherman. Techtarget: <http://searchdatacenter.techtarget.com/es/cronica/Programas-de-analitica-de-big-data-exigen-conocimiento-del-negocio-y-experiencia-en-TI>

Rous, M. (s. f.) "Framework". *Techtarget*

Tascón, M. (2013). "Pasado, presente y futuro". *Big Data*, 95, 47.

Ticbeat. Los beneficios económicos y las ventajas competitivas del 'Big Data' llegan a la pyme: <http://www.ticbeat.com/bigdata/los-beneficios-economicos-las-ventajas-competitivas-del-big-data-llegan-la-pyme/>