

UNIVERSIDAD DE CANTABRIA
FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES



GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE
EMPRESAS

CURSO ACADÉMICO 2016-2017

TRABAJO FIN DE GRADO

BIG DATA 2.0

Autor

D^a. Laura de Cos González

Director

D. Pedro Solana González

Santander, 12 de Septiembre de 2017

UNIVERSITY OF CANTABRIA
FACULTY OF ECONOMIC AND BUSINESS SCIENCES



DEGREE IN BUSINESS ADMINISTRATION

ACADEMIC COURSE 2016-2017

DEGREE FINAL PROJECT

BIG DATA 2.0

Author

Mrs. Laura de Cos González

Director

Sr. Pedro Solana González

Santander, September 12, 2017

BIG DATA 2.0

Resumen

El auge de las nuevas tecnologías de la información en los últimos tiempos ha provocado que de una forma global y casi sin darnos cuenta, se generen cantidades de datos desorbitadas. Cualquier acción que se realiza a través de internet, ya sea la compra de un artículo, consulta de hoteles...deja una huella digital que almacenándose junto con el resto provoca una cantidad de datos masiva. Es por ello que es necesario no solo una herramienta sino una tecnología que permita diferenciar unos datos de otros, ya que estos se obtienen de diferentes fuentes.

El empleo de técnicas Big Data permite analizar los gustos de consumidores para ofrecer respuestas personalizadas de productos, analizar el éxito o fracaso de una campaña...con el objetivo de obtener una ventaja competitiva antes sus competidores.

Por ello, en este trabajo se pretende desarrollar de una manera sencilla, los puntos clave de este fenómeno que esta tan de moda y que seguirá evolucionando en los próximos años, convirtiéndose en el futuro empleo de muchas de las personas que a día de hoy están en edad de estudio.

Palabras clave:

Big Data, herramientas, tecnología, datos, conocimiento, información

BIG DATA 2.0

Abstract

The rise of new information technologies in recent times has caused that in a global way and almost without realizing it, generate exorbitant amounts of data. Any action that is done through the internet, whether the purchase of an article , hotel consultation ... leaves a fingerprint that stored together with the rest causes a massive amount of data. That is why it is necessary not only a tool but a technology that allows to differentiate some data from others, since these are obtained from different sources.

The use of Big Data techniques allows analyzing consumer tastes to offer customized product responses, analyze the success or failure of a campaign ... in order to gain a competitive advantage over competitors.

Therefore, this paper aims to develop in a simple way, the key points of this phenomenon that is so fashionable and that will continue to evolve in the coming years, becoming the future employment of many of the people that today are in age of study.

Keywords:

Big Data, tools, technology, data, knowledge, information

ÍNDICE

1. MARCO GENERAL DEL TRABAJO	7
1.1 INTRODUCCIÓN	7
1.2 JUSTIFICACIÓN	7
1.3 OBJETIVOS DEL TRABAJO.....	9
1.3.1 Objetivos específicos.....	9
2. ESTADO DEL ARTE Y REVISIÓN DE LA LITERATURA.....	9
3. MARCO TEÓRICO	15
4. METODOLOGÍA	20
5. LIMITACIONES	22
6. DESARROLLO EMPÍRICO	23
7. CONCLUSIONES.....	27
8. REFERENCIAS.....	29
9. GLOSARIO.....	31
10. ANEXOS	32

ÍNDICE DE FIGURAS

Figura 1. El embudo del conocimiento.	10
Figura 2. Definición de Big Data.....	11
Figura 3. Elementos de Big Data.....	12
Figura 4. Tecnologías de Big Data.	15
Figura 5. Características de Big Data.....	16

ÍNDICE DE TABLAS

Tabla 1. Tamaño y equivalencia de datos informáticos.	12
Tabla 2. Tipos de datos según su categoría.....	13
Tabla 3. Tipos de datos según su origen.	14
Tabla 4. Tipos de análisis Big Data	20

1. MARCO GENERAL DEL TRABAJO

1.1 INTRODUCCIÓN

Hoy en día las compañías y empresas trabajan con un gran volumen de datos e información la cual almacenan en diferentes bases de datos. Según West, 2013, el volumen de datos generados crece exponencialmente. En los dos últimos años la cifra mundial ha aumentado un 92%. Estos datos se generan cada vez a mayor velocidad, debido al uso generalizado y diario de las tecnologías, por lo que éstas bases de datos se quedan pequeñas y obsoletas, ocasionando grandes problemas no solo en la gestión de las mismas sino también en la toma de decisiones.

En el presente trabajo se intentará evidenciar la importancia del Big Data dentro del sector empresarial, ya que por su gran desconocimiento en el tema, las compañías en muchas ocasiones no saben sacar partido al gran volumen de datos e información del que disponen, mostrando así la velocidad y variedad con la que se generan esos datos.

Mediante la presente investigación se pretende acercar el mundo del Big Data a aquellas personas que desconocen tanto su concepto, sus características, sus alcances, tecnologías y cuál es su aprovechamiento, así como mostrar a las empresas cómo pueden gestionar los datos para convertirlos no solo en información, sino también en conocimiento útil con el objetivo de facilitarles sus tareas diarias.

1.2 JUSTIFICACIÓN

A medida que el tiempo pasa, se dispone de un volumen de datos mayor, pero los expertos señalan que una cosa es disponer de datos y otra la capacidad de transformar esos datos en conocimiento o información que nos permita la mejora de la toma de decisiones. En el conjunto de actividades que desarrollamos diariamente, como operaciones comerciales, operaciones de logística, capital humano...es necesario disponer de herramientas que nos permitan transformar los datos en información.

Según Beyer de Gartner y otros expertos: "*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*".

En una encuesta realizada por LogLogic (California) se encontró que el 49% de las organizaciones están algo o muy preocupados por la gestión de Big Data, pero que el 38% no entiende lo que es, y el 27% dice que tiene una comprensión parcial; además, la encuesta encontró que 59% de las organizaciones carecen de las herramientas necesarias para gestionar los datos de sus sistemas de TI. Khera, 2017, explicó que: "*Big Data se trata de muchos terabytes de datos no estructurados*", donde "*La información es poder, y Big Data, si se gestiona correctamente, puede proporcionar una cantidad de conocimiento...*". Según esta encuesta, el 62% ya había gestionado más de un terabyte de datos; eso indica que el volumen de datos en el mundo está creciendo a un ritmo casi incomprensible.

Pero, ¿por qué es importante Big Data? Cuando se habla de Big Data se hace referencia a la tendencia en el avance de la tecnología con respecto al entendimiento y la toma de decisiones que se utilizan para describir grandes cantidades de datos que

no pueden ser procesadas a través de las herramientas o bases de datos tradicionales.

Así pues, con la utilización de Big Data y tomando datos desde cualquier fuente y analizándolas se pueden obtener:

- Reducción de costes
- Reducción de tiempo
- Desarrollo de nuevos productos
- Toma de decisiones más eficientes e inteligente

Además, un estudio publicado por la firma alemana Digital Cologne en el año 2015 ya mostraba 10 tendencias que a medio plazo podían convertirse en nuevas oportunidades de negocio, algunas de ellas ya implantadas en la actualidad:

Big Data en la nube: Muchas organizaciones apuestan por el almacenamiento de la información en soportes que permitan tanto un acceso rápido como un acceso desde distintas plataformas de análisis.

Aplicaciones de Big Data: En el estudio se refiere a que las nuevas aplicaciones muchas de ellas ya instaladas en la vida de muchas personas, están orientadas a facilitarles la vida, ya sea organizándoles la agenda, controlando su economía, o aconsejándoles oportunidades de inversión.

Integración del Big Data en todos los departamentos de la empresa: Para que Big Data funcione de una forma eficiente, será necesario que este se implante en todos los departamentos, desde el departamento de producción hasta que el producto llega al consumidor final.

La automatización o deep learning: El aprendizaje profundo hace que las máquinas y robots aprendan de manera inteligente y autónoma en tiempo real.

Análisis predictivo y análisis de tendencia: Permite anticipar las necesidades de los consumidores, así como sus expectativas aportando información de gran utilidad para la mejora de la estrategia de la empresa.

Consultoría de Big Data: El análisis de los datos a gran escala permite minimizar riesgos a la vez que mide el impacto de las decisiones tomadas en diferentes ámbitos.

Diagnósticos clínicos: Una de las aplicaciones mejor valoradas esta en la medicina ya que se beneficia para realizar diagnósticos.

Asistentes en la conducción: Asiste en la circulación dando información sobre el tráfico dependiendo de la situación de la persona, además de servir como guía.

Entrenadores personales: El cuidado personal, físico y de la salud está en auge. Cada vez más personas buscan a través de las aplicaciones, la manera de cuidarse. Según el estudio, en unos años hasta las camisetas podrán informar sobre el ritmo cardiaco del deportista.

Casa inteligente: Los electrodomésticos cada día mas potentes aportarán información tanto a los usuarios como a las empresas, pudiendo estas conocer mejor a sus cliente y por consiguiente prestar un mejor servicio.

1.3 OBJETIVOS DEL TRABAJO

El principal objetivo del presente trabajo es dar a conocer el concepto Big data 2.0 realizando un estudio teórico de sus características, tecnologías empleadas, ámbitos de aplicación, aportando los beneficios que aportan a las compañías que decidan implantarlo en su empres, así como limitaciones que dificultan su desarrollo.

Además se analizarán los puntos clave que hacen que ésta tecnología sea tan importante hoy en día en el mundo empresarial.

1.3.1 Objetivos específicos

Como objetivo específico, se pretende realizar un estudio de casos de empresas que hayan implantado el proyecto Big Data 2.0 comparando el proceso de implementación, las tecnologías empleadas, los objetivos de cada empresa y las limitaciones que se han encontrado en cada caso.

2. ESTADO DEL ARTE Y REVISIÓN DE LA LITERATURA

2.1. CONTEXTO DE LA GESTIÓN DEL CONOCIMIENTO, LAS TECNOLOGÍAS DE GESTIÓN DE INFORMACIÓN EN LAS EMPRESAS

La gestión del conocimiento es *“una nueva cultura empresarial, una manera de gestionar las organizaciones que sitúa los recursos humanos como el principal activo y sustenta su poder de competitividad en la capacidad de compartir la información y las experiencias y los conocimientos individuales y colectivos”*. (Josep M^a Rodríguez Rovira, 1999).

Cuando se habla de conocimiento hay que hacer diferencias entre los componentes que forman lo que se conoce como Pirámide del conocimiento. En la base de la pirámide se encuentran los datos, que por sí solos no tienen valor. En el siguiente paso está la información, formada por un conjunto de datos que hacen referencia a un determinado mensaje. Tras la información, el conocimiento, que es la capacidad para realizar y producir resultados, es decir, la inteligencia. Y por último en la parte que se encuentra más elevada está la sabiduría.

En el caso que se muestra se puede ver una pirámide invertida, en la que se puede apreciar la metáfora del embudo del conocimiento, es decir el filtrado de los datos y la información hasta llegar a la sabiduría.

Figura 1. El embudo del conocimiento



Fuente: La teoría del embudo del conocimiento.

<http://www.slideshare.net/escenaenelmar/gestion-del-conocimiento-presentation-591517>.

A partir de ahí se puede hacer referencia al concepto de Gestión del conocimiento como:

- "Arte de crear valor a partir de los activos intangibles, representados en clientes, proveedores y en el conocimiento de las personas que es tácito, compartido, dinámico y relevante para la empresa" (Sveiby, 1997).
- La gestión del conocimiento se relaciona con el "uso de la información estratégica para obtener los objetivos de negocio. La gestión del conocimiento es la actividad organizacional de creación del entorno social e infraestructura para que el conocimiento se pueda compartir y crear". (Logan & Stokes, 2004).
- Gestión del conocimiento es el "proceso de identificar, agrupar, ordenar y compartir continuamente conocimiento de todo tipo para satisfacer las necesidades presentes y futuras y para desarrollar nuevas oportunidades". (Saéz Vacas, 2003).

Las Tecnologías de la Información y las Comunicaciones se definen como "el conjunto de tecnologías desarrolladas para gestionar información y enviarla de un lugar a otro. Incluyen las tecnologías para almacenar información y recuperarla después, enviar y recibir información de un sitio a otro, o procesar información para poder calcular resultados y elaborar informes" (Servicios TIC, 2006).

Las TIC se consideran muy importantes para el mundo empresarial ya que añaden valor a las actividades operacionales y de gestión empresarial, permitiendo ganar competitividad frente a otras empresas, por lo que ayudan a mantenerse en el mercado.

A la hora de implementar un sistema de información es necesario realizar una planificación de lo que se quiere obtener con ello, además de tener en cuenta el factor humano ya que en algunas ocasiones puede ser reticente a adoptar nuevos procedimientos de trabajo.

Aún así, las TICs son esenciales para mejorar la productividad, la calidad, el control...de las empresas,; siempre y cuando se tenga gran conocimiento de los procesos de la empresa, una planificación detallada de las necesidades, así como realizar la implementación de una manera paulatina. De esta forma, han logrado transformar la manera de trabajar y gestionar recursos haciéndolos más productivos, agilizando las comunicaciones, gestionando existencias, realizando análisis financieros...por lo que permite a las empresas obtener mayor producción, más rápido y de mayor calidad, empleando menos tiempo, por lo que se mejora la competitividad.

2.2. DEFINICIÓN DE BIG DATA

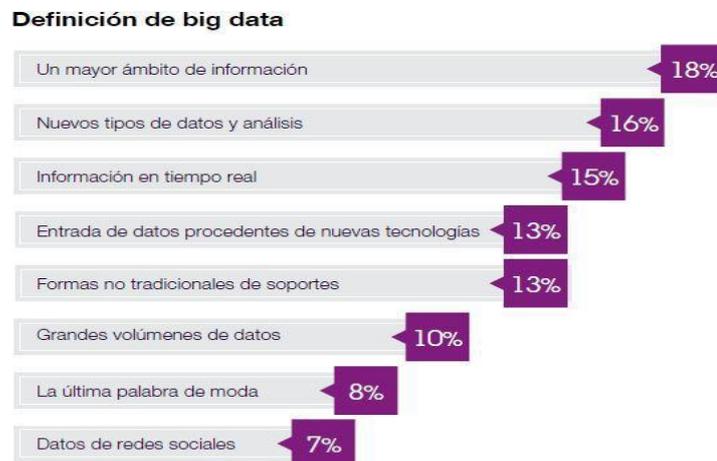
Big data (datos masivos o datos a gran escala) se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas del software de bases de datos típicas para capturar, almacenar, gestionar y analizar. (McKinsey Global Institute, 2011).

Por otra parte, según Gartner, Big Data se define como *“un gran volumen, velocidad o variedad de información que demanda formas costeables e innovadoras de procesamiento de información que permitan ideas extendidas, toma de decisiones y automatización del proceso”*.

Wikipedia, por su parte define Big Data como *“un conjunto de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable.*

Sin embargo, según un estudio realizado por *IBM Institute for Business Value* en colaboración con *Saïd Business School* en 2012, en el cual se dió una serie de características sobre Big Data a más de 1144 negocios y profesionales de IT, para que escogieran cuál de ellas describía mejor el concepto de Big Data, se pudo comprobar lo siguiente:

Figura 2. Definición de Big Data

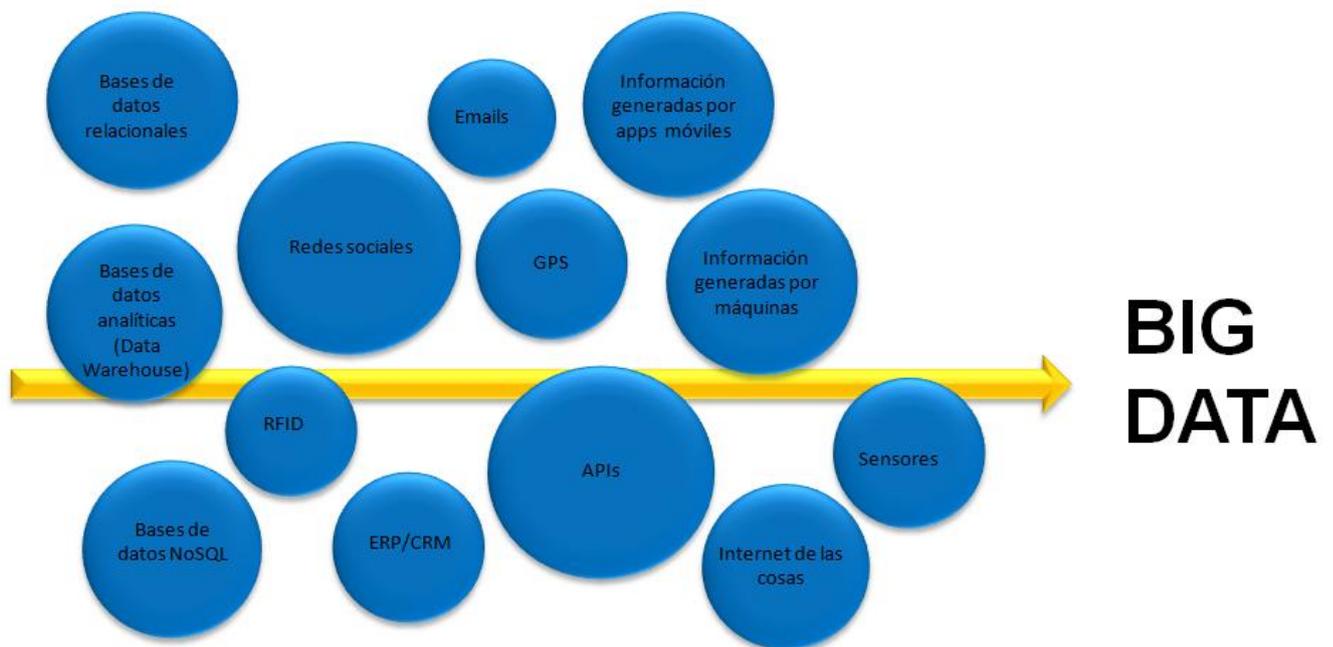


Fuente: “Definición de Big Data”. IBM Institute for Business Value

Según los encuestados, un mayor ámbito de información junto con nuevos tipos de datos y análisis es la definición que entienden por Big Data, dejando los datos de redes sociales como definición más alejada.

Por lo tanto, se puede decir que es una definición utilizada en tecnología para referirse a la información o grupo de datos que por su elevado volumen, diversidad y complejidad no pueden ser almacenados ni visualizados con herramientas tradicionales. Las dimensiones de estos datos obligan a las empresas a buscar soluciones técnicas para gestionarlos, pues un buen manejo del Big data puede representar nuevos métodos para la toma de decisiones y oportunidades de negocio.

Figura 3. Elementos de Big Data



Fuente: Introducción a Big Data. Consultec

Tipos de datos

Antes de nada, es importante conocer exactamente a que se refiere la cantidad de datos a la que se esta refiriendo:

Tabla 1. Tamaño y equivalencia de datos informáticos

MEDIDA	SÍMBOLO	ENQUIVALENCIA
Bytes	B	
Kilobytes	Kb	1 Kilobyte son 1024 Bytes
Gigabytes	Gb	1 Gigabyte son 1024 Kilobytes

Terabytes	Tb	1 Terabyte son 1024 Gigabytes
Petabytes	Pb	1 Petabyte son 1024 Terabytes
Exabytes	Eb	1 Exabyte son 1024 Petabytes
Zettabytes	Zb	1 Zettabyte son 1024 Exabytes
Yottabytes	Yb	1 Yottabyte son 1024 Zettabytes

Fuente: Tabla de equivalencias de informática. Elaboración propia basándome en es.tableworld.net

Existen dos clasificaciones dentro de Big Data, los relacionados con la estructura de los datos y las que dependen del origen de estos: (M. Pérez, 2015).

Tipos de datos por categorías:

Estructurados: Son datos con formato o esquema fijo y que tienen bien definido su longitud y su formato. Se almacenan en tablas:

- Creados: datos generados por nuestros sistemas de una manera predefinida (registros en tablas, hojas de cálculo...)
- Provocados: datos creados de manera indirecta a partir de una acción previa (valoraciones de restaurantes, películas, empresas Yelp, TripAdvisor...)
- Dirigido por transacciones: datos obtenidos al acabar una acción previa de manera correcta (facturas autogeneradas al realizar una compra, recibo de un cajero automático al realizar una retirada de efectivo...)
- Compilados: resúmenes de datos de empresa, servicios públicos de interés grupal. Entre ellos nos encontramos con el censo electoral, vehículos matriculados, viviendas públicas...)
- Experimentales: datos generados como parte de pruebas o simulaciones que permitirán validar si existe una oportunidad de negocio.

No estructurados: Son datos en el formato en el que fueron recogidos y que carecen de un formato específico. No se pueden almacenar dentro de una tabla.

- Capturados: datos creados a partir del comportamiento de un usuario (posición GPS, aplicaciones de seguimiento de actividades...)
- Generado por usuarios: datos que especifica un usuario (videos, publicaciones en redes sociales...)

Semiestructurados: Datos que no tienen formatos fijos, pero contienen etiquetas y otros marcadores para separar los diferentes elementos. Poseen sus propios metadatos semiestructurados. (Texto de etiquetas XML y HTML).

Tabla 2. Tipos de datos según su categoría

ESTRUCTURADOS	NO ESTRUCTURADOS	SEMIESTRUCTURADOS
Creados Provocados Dirigido por transacciones Compilados Experimentales	Capturados Generados por usuarios	Contienen etiquetas y otros marcadores

Fuente: Elaboración propia basándome en Big Data, Elaboración de los datos masivos, Mayer-Schönberger, V., & Cukier, K. (2013)

Tipos de datos dependientes de su origen:

Web and Social Media (Webs y redes sociales): Incluye información obtenida de las redes sociales.

Machine-to-Machine (M2M) (Comunicación entre máquinas): Son las tecnologías que nos permiten conectarnos a otros dispositivos a través de medidores.

Big Transaction Data (Transacciones): Pueden estar en formatos semiestructurados como no estructurados.

Biometrics (Biométricos): Los datos biométricos se consideran datos importantes en el área de seguridad de inteligencia para la investigación.

Human Generated (Generados por personas): Son datos generados por las personas en su rutina diaria.

Tabla 3. Tipos de datos según su origen

Web y Redes sociales	Entre máquinas	Transacciones	Biométricos	Personas
Clics en vínculos	RFID	Llamadas	Reconocimiento facial	Grabaciones de llamadas
Búsquedas en Google	GPS	Mensajería	ADN	E-mail
RRSS	Otros sensores	Facturación(online)		Registros médicos
Cont. web				

Fuente. Elaboración propia basándome en Big Data, Elaboración de los datos masivos, Mayer-Schönberger, V., & Cukier, K. (2013).

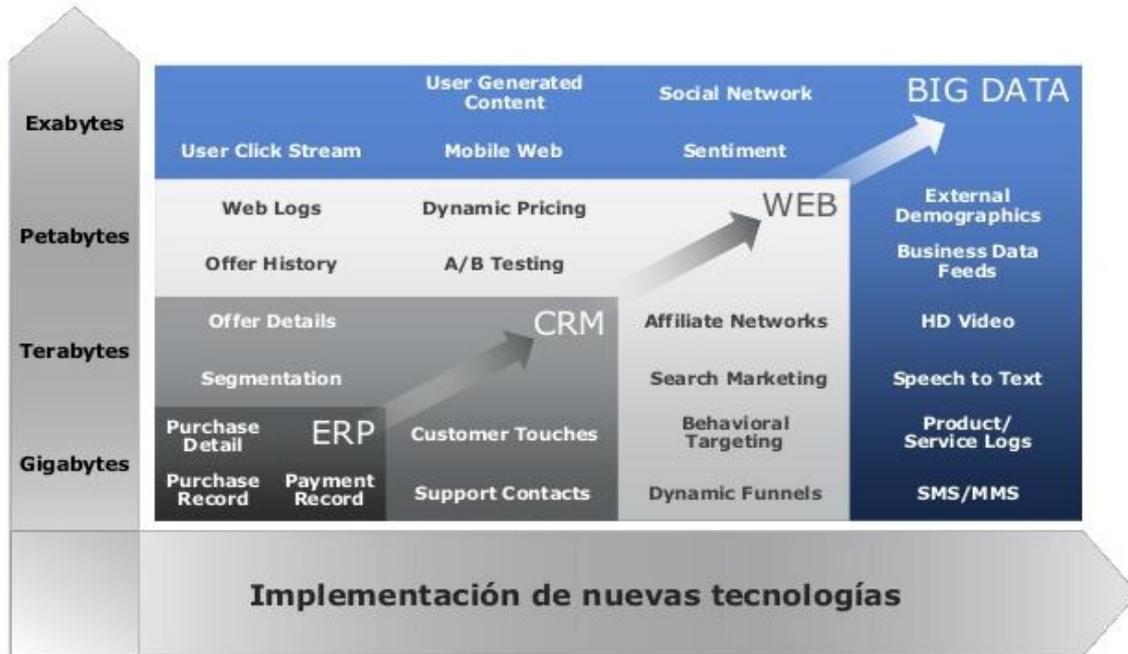
2.3. EVOLUCIÓN DE BIG DATA Y LAS TECNOLOGÍAS

El fenómeno Big Data nace de la unión entre compañías de Internet (Google, Facebook, Yahoo, LinkedIn, Twitter...). Estas compañías, debido al gran volumen de datos que se encontraron y carentes de las infraestructuras necesarias para su gestión, consiguieron crear tecnologías que pudieran soportar tanta información. Con el paso del tiempo, otras empresas más pequeñas empezaron a tener la misma necesidad para poder gestionar sus datos por lo que el fenómeno Big Data empezó a cobrar mayor importancia.

Si bien es cierto que Big Data es la solución a muchos problemas empresariales, también hay que tener en cuenta que no todas las empresas están dispuestas a este cambio ya que requiere de procesos, presupuestos, gestión de proyectos, preparación del personal, auditorías y la arquitectura de administración de la información en la mayoría de las empresas. (Beyer de Garner y McKinsey)

Actualmente, se están produciendo importantes innovaciones gracias a la actividad open source, y aunque se ha avanzado mucho en los últimos 3 años resolviendo problemas de raíz, todavía se están produciendo grandes innovaciones. Desde 2015, los grandes en Big Data como IBM han apostado por Spark dándole mayor credibilidad al proyecto.

Figura 4. Evolución de las tecnologías Big Data



Fuente: Big Data y Hadoop. MBA Luis Barragan, 2015. Disponible en: <https://es.slideshare.net/dataminingperu1/bigdatasummitperu>

3. MARCO TEÓRICO

3.1. CARACTERÍSTICAS DE BIG DATA 2.0

En un primer momento, las características de Big Data 2.0 se clasificaron en 4; las conocidas como las "4 V del Big Data". En ellas se podía distinguir entre Variedad, Volumen, Veracidad y Velocidad. Pero una de las características principales cuando se habla de Big Data es el Valor de los datos para su análisis, junto a ella se puede hablar también de Viabilidad y Visualización. Estas características conforman lo que actualmente se conoce como las "7 V del Big Data". (Russom, 2011)

- **Variedad de datos:** Existen diferentes tipos, formas y fuentes de registrar los datos. Estos se pueden clasificar en datos de fácil gestión denominados datos estructurados, entre los que se pueden encontrar bases de datos; y los no estructurados. Los datos no estructurados necesitan de una herramienta específica para el tratamiento de información. Los documentos de textos, correos electrónicos pertenecen a este segundo grupo, pero también los videos, audios, publicaciones en perfiles de redes sociales o cualquier tipo de datos que se pueda almacenar en un teléfono móvil, tablet u ordenador.
- **Volumen:** Hace referencia a la cantidad de datos que se generan en cada momento. Es la característica que más se asocia al Big Data (Datos masivos) ya que todos los datos se almacenan con la finalidad de transformarse en acciones.

- **Veracidad:** Tiene que ver con la incertidumbre de los datos, es decir, si la información recibida es fiable o no. Big data tiene como reto explorar esta incertidumbre existente y que puede influir en el comportamiento de los consumidores.
- **Velocidad:** El tiempo en que los datos son creados, almacenados y procesados es fundamental a la hora de la detección de fraude en una transacción bancaria por ejemplo, por ello esta característica es imprescindible a la hora de implantar un proyecto Big Data.
- **Valor:** El valor es la información que se obtiene de los datos recogidos. Esta información convertida en conocimiento será la responsable de la toma de decisiones o acciones tomadas.
- **Viabilidad:** La inteligencia empresarial es la capacidad de que disponen las empresas para que el uso del volumen de datos utilizados funcione de la manera más eficaz. El momento de comenzar el proyecto será una vez que se conozca la viabilidad de la organización.
- **Visualización:** Es la manera en que se presentan los datos para que se puedan ver de una forma legible. Para ello, además, existen herramientas que facilitan su comprensión.

El conjunto de estas 7 características son las que van a dar ventaja competitiva a las empresas frente a sus competidores. Por eso una buena metodología es importante a la hora de su implementación.

Figura 5. Características Big Data



Fuente: Elaboración propia.

3.2. TECNOLOGÍAS DE BIG DATA 2.0

Para manejar los datos en Big Data se necesitan dos componentes principales e imprescindibles como son: el software y el hardware.

Para el Hardware existen tecnologías como Procesamiento Paralelo Masivo (MMP), que ayudan de una manera rápida a su procesamiento. Pero es para el manejo de datos no estructurados o semiestructurados donde aparecen tecnologías como MapReduce o Hadoop, que están diseñadas para manejar tanto información estructurada, semiestructurada como no estructurada.

APACHE HADOOP

Para T. Olavsrud, 2012, Apache Hadoop es *“un marco de software de código abierto para aplicaciones intensivas de datos distribuidos originalmente creado por Doug Cutting para apoyar su trabajo en Nutch, una Web de código abierto motor de búsqueda. Hadoop es ahora una de las tecnologías más populares para el almacenamiento de los datos estructurados, semiestructurados y no estructurados que forman Big Data. Hadoop esta disponible bajo licencia Apache 2.0”*

1. Hadoop Distributed File System (HDFS: Sistema de archivos distribuido Hadoop). Sistema de archivos que permite que el sistema se recupere antes posibles fallos sin que afecte el proceso.

Cuando se creó HDFS se propusieron tres objetivos. (Cloudera.com, 2013)

- Permitir procesar archivos con tamaños de gigabytes (GB) hasta petabytes (PB).
- Poder leer datos a grandes velocidades.
- Capacidad para ser ejecutado en una máquina, sin solicitar hardware especial.

La arquitectura de HDFS se compone de un nodo principal, el servidor maestro, que es el que gestiona y controla el acceso a los archivos usuarios y los nodos esclavos que son los que gestionan tanto la lectura como la escritura de los archivos usuarios.

2. Hadoop MapReduce

En la página web hadoop.apache.org lo definen como "Un sistema basado en hilados para el procesamiento paralelo de grandes conjuntos de datos".

Es un modelo de programación que da soporte a la computación paralela sobre grandes volúmenes de datos. Una condición para que pueda ser usado MapReduce es que sea usado en soluciones donde se pueda procesar de forma paralela y, además, con grandes cantidades de información, de lo contrario no sería una solución adecuada. Otra condición es que se puede usar MapReduce en procesos que se puedan disgregar en operaciones map () y reduce (), las cuales se definen en función de datos estructurados. MapReduce fue diseñado como un modelo de programación, para que se pudiera realizar procesamiento de datos de gran tamaño, y, de la misma forma, para que resolviera el problema existente de escalabilidad. MapReduce "es un modelo de programación para el procesamiento de datos". (P. Zikopoulos, C. Eaton, D. DeRoos, T. Deutsch, & G. Lapis, 2012).

3.3. BENEFICIOS DE BIG DATA 2.0

Una vez conocido el significado de Big Data gracias a la mejora que supuso en los modelos relacionales y por qué es importante conocerlo en el entorno empresarial, se van a citar los beneficios más habituales del Big Data, siendo esto no aplicables a todas las organizaciones, ya que cada organización cuenta con unas características diferentes (Vauzza, 2013):

Mejora de la estrategia:

- A través de la búsqueda de nuevas oportunidades de negocio a través de segmentación mejorada.
- Predecir con mayor exactitud qué productos son los más apropiados para cada cliente a través del análisis de consumo de los servicios y productos.
- Ofrecer combinaciones adecuadas de productos y servicios para mejorar la relación con los clientes, mejorando así la lealtad.

Mejora de la segmentación

- Basándose en la probabilidad de que el cliente contrate servicios o productos complementarios o de mayor valor.

Análisis de navegación web y hábitos de consumo online

- Análisis de redes sociales: Se crea una visión de los clientes y se identifica el papel que desempeñan en sus entornos y su grado de influencia.
- Marketing viral: Mejora el conocimiento de clientes y del mercado en redes sociales.
- Análisis de datos de navegación: Analiza la navegación Web y los hábitos de consumo online.
- Información en tiempo real.

Anticipación a los problemas

- A través de un predictivo análisis y cruce de datos permite anticiparse a problemas que puedan surgir.

Mejoras de procesos

- Reduce coste ya que permite la simplificar procesos actuales y controla el negocio, y permite detectar patrones de fraude en tiempo real analizando los datos históricos, el patrón de uso de información de geolocalización, analizando transacciones...

Reduce tiempos:

- Proporcionando un feedback mucho más rápido

Desarrolla nuevos productos:

- Abarcan desde la optimización de procesos hasta la anticipación de tendencias, pasando por la reducción de costes.

Toma de decisiones más inteligentes

Estos beneficios permitirán que las empresas a través del Big Data mejoren su estrategia, permaneciendo en el mercado posicionadas, haciendo uso de nuevos conocimientos que en un primer momento no se manejaron de la forma debida por falta de herramientas que permitieran procesarla. Por lo tanto, podrán ofrecer mejores productos, con excelentes relaciones con los clientes, llegando a ser más competitivas.

3.4. ÁMBITO DE APLICACIÓN DE BIG DATA 2.0

La capacidad que tiene Big Data de gestionar y analizar eficientemente enormes volúmenes de datos de diferentes fuentes, pone en valor la ventaja competitiva y diferencial de Big Data haciendo que se pueda aplicar a campos muy diversos: (Tascón, 2013)

Logística: En este campo el Big Data permite la posibilidad de utilizar servicios específicos como movilidad, tracking, seguridad...

Seguridad: Analiza enormes volúmenes de datos de una manera ágil y óptima. (Por ejemplo, existen modelos de análisis del comportamiento para prevenir ataques terroristas a través del análisis de cámaras y sensores de seguridad).

Internet: La cantidad de datos que genera internet no podrían gestionarse sino fuera por el Big Data. (Redes sociales)

Investigación Médica: Mediante la estructuración de datos se pueden establecer diagnósticos de forma prematura (monitorización, historiales, tratamientos...)

Gobierno y toma de decisiones: Ofrece una mejor optimización en procesos de gobernanza de empresas y gobiernos, permitiendo el soporte a la toma de decisiones.

CRM: La integración de Big Data y CRM permite a la empresa gestionar la información del cliente, permitiendo el análisis de grandes volúmenes de datos, filtrándolos y segmentándolos con el fin de que se muestren de forma que se puedan entender. A través de esta segmentación se pueden analizar los gustos y preferencias del cliente sobre productos y servicios, pudiendo así mejorar tanto la calidad, como el desarrollo del producto o la atención al cliente o marketing. Estas campañas de marketing están enfocadas a la fidelización del cliente, que a fin de cuentas es uno de los principales objetivos de las empresas.

4. METODOLOGÍA

El presente trabajo presenta un estudio exploratorio en el que se desgranar las principales características de Big Data y en él se realiza un análisis de casos de estudio de 2 empresas que utilizan las herramientas tecnológicas propias de un proyecto Big Data.

A la hora de crear un proyecto Big Data que detecte, gestione, organice y presente los datos de una manera óptima, se debe responder a una serie de preguntas:

- ¿De dónde obtengo los datos?
- ¿Qué datos aportan más información a mi negocio?
- ¿Qué datos hay disponibles fuera de mi organización que me puedan ayudar?
- ¿Qué volumen de datos tenemos que manejar?
- ¿Qué formato tienen?

Pero la pregunta más importante de todas antes de comenzar con el proyecto es:
¿Qué problema se quiere resolver?

En la metodología Big Data se pueden diferenciar 4 fases:

1. Identificar los datos disponibles o recogidos y detectar las necesidades:

Se trata de analizar cuáles son los datos de que dispone la empresa y cuáles son los que necesita realmente. Con este análisis se pretende cuantificar el volumen de datos a tratar y el universo digital que se debe gestionar (Internet, redes sociales...). Los datos pueden recopilarse de operaciones comerciales, financieras, públicos o los que provienen de redes sociales que deben ser analizados. En esta primera fase es importante diferenciar los datos que realmente proporcionan valor identificando las metodologías y técnicas de almacenamiento y análisis de los datos recogidos.

Para realizar el análisis de datos se pueden distinguir 4 tipos de análisis:

Tabla 4. Tipos de análisis Big Data



Fuente: Tipos de análisis Big Data. Disponible en: <https://es.slideshare.net/BigDataColombia/business-analytics-la-culpa-es-del-big-data>. Big Data Colombia, 2015

En esta fase se deben detectar las necesidades de la empresa y las carencias que tiene en cuanto a tecnología. Consiste en pensar qué es lo que se quiere mejorar. Para ello se puede investigar cuáles son las estrategias que utilizan los competidores y compararlas para crear una idea de lo que se quiere conseguir y cómo se puede llegar a ello.

Es importante también justificar la inversión, ya que se pretenden mejorar cuestiones técnicas pero también crear un entorno de alto rendimiento que reduzca costes.

2. Planificación y diseño:

En esta fase se va a determinar el presupuesto con el que se va a contar durante el proceso y cuáles son los recursos que van a intervenir (gestores, diseñadores y arquitectos de datos, implementadores, operadores de datos...)

El diseño se comenzará en función de las necesidades encontradas, optimizándose progresivamente, teniendo en cuenta el coste, la escalabilidad y las opciones del mercado.

3. Implementación:

En este punto se debe llevar a cabo la instalación de servidores y puesta en marcha de la infraestructura, así como su configuración para su correcto funcionamiento.

A la hora de implementar el sistema existen unos puntos principales a los que atender:

- Se deben proponer las soluciones de Big Data más adecuadas teniendo en cuenta los análisis efectuados.
- El diseño y el establecimiento de sistemas de seguridad para prevenir riesgos.
- La dimensión del sistema en relación a estas necesidades.
- Prevención de posibles incidencias en la implementación del sistema Big data y la resolución de las que se puedan derivar de su integración definitiva.
- EL seguimiento de las tareas que se asocian al proyecto.

Una vez implantado el proyecto, es necesario realizar el seguimiento de los sistemas y apoyar la implantación, pudiendo:

- Realizar comprobaciones para el buen funcionamiento del sistema, pruebas piloto, calidad...
- Invertir en la formación del personal de los sistemas utilizados en Big Data.
- Participar en el desarrollo estratégico de la empresa, proporcionando nuevas soluciones.
- Elaborar documentación del proyecto Big Data, incluyendo diseño de la estructura, justificación de su planteamiento y descripción de las funcionalidades, es decir, crear una documentación técnica. (López-Cantos, F., 2015).

4. Fase de recopilación, análisis de datos y generación de valor:

Durante el proceso, los datos, que son los componentes básicos del Big Data, se transformarán en información aportando conocimiento, por lo que ya se habrán obtenido datos inteligentes.

La investigación y el posterior análisis de los datos permitirán la elaboración de conclusiones y la identificación de nuevas vías de desarrollo del negocio.

5. LIMITACIONES

A la hora de llevar a cabo un proyecto Big Data se presentan algunas limitaciones y desafíos que se deben de tener en cuenta:

Déficit metodológico:

Los enormes volúmenes de datos mostrarán todos los patrones ocultos del mundo externo, por lo que presentará debilidad en el proceso de toma de decisiones correctas.

No sirve para pronosticar el futuro:

Big Data entiende que el futuro está contenido en datos pasados, ya sea por la cantidad de datos perabytes como exabytes; pero las Ciencias Sociales apuntan a que el futuro nunca es una prolongación de datos pasados.

A mayor cantidad de datos, mayor cantidad de errores:

Influirá la interpretación de los datos que en muchos casos será deficiente. Un ejemplo es la crisis de 2007, había suficientes datos para poder darse cuenta de ello pero esos datos no fueron bien interpretados; como también algunos atentados que no ha sido posible pararlos a tiempo si de seguridad se habla.

Lo imposible sucede:

Aún con las mejores técnicas de Big Data, resulta muy difícil diferenciar entre lo imposible y lo improbable. Por ejemplo, la central nuclear Fukushima fue diseñada para soportar terremotos de algo más de 8 en la escala Richter porque era imposible de más escala, pero finalmente sucedió

6. DESARROLLO EMPÍRICO

A continuación se presentan dos modelos Big Data implantados en dos organizaciones distintas, en las que se puede comprobar cuáles qué objetivos se buscaban y cuáles han sido los beneficios obtenidos:

BBVA	
Título de estudio de caso o proyecto	La cultura de la innovación abierta
Descripción	<p>Implantar modelos de innovación abierta con el objetivo de acelerar los procesos de transformación y mantenerse cerca del talento e ideas que surgen en los ecosistemas de innovación y emprendimiento.</p> <p>Para ello está creando una red de conocimiento en torno a los Centros de Innovación, espacios en los que se están desarrollando los proyectos más disruptivos del Grupo BBVA. Los Centros de Innovación tienen un papel muy importante en la transformación digital que se está llevando a cabo en BBVA.</p> <p>El Grupo se está apoyando en los ecosistemas de innovación y talento de todo el mundo, propiciando una cultura de colaboración con emprendedores, startups y desarrolladores basándose en las nuevas ideas que les ayudan a crecer.</p>
Objetivos	El objetivo principal es capturar valor de los datos que tiene el banco, conocer mejor al cliente y extraerle información de manera que ese conocimiento generado permita evaluarles mejor de cara a productos para darle mejor servicio y evitar fraudes.
Tecnologías aplicadas	<ul style="list-style-type: none"> • Software de gestión de terceros <p>Hay terceros que venden software para utilizar en la gestión Hadoop, y sus productos son normalmente agnósticos en cuanto a las distribuciones a las que dan soporte.</p>
Desarrollo del caso, dificultades encontradas y resultados	<p><u>Internamente</u>, uso de los datos dentro del banco:</p> <ul style="list-style-type: none"> • Analiza riesgos, por ejemplo, en el caso de las PYMES, donde la idea es tener una visión más amplia de los clientes, que permita no limitarse a su contabilidad e historia financiera. • Ubicación de infraestructuras de servicios como cajeros y oficinas bancarias, de forma que sean lo más convenientes posible para los clientes. • Adecuación de la oferta de productos en función de las necesidades reales de los clientes, o “customer centricity”. <p><u>Externamente</u>, en la que otras organizaciones son las que se benefician del valor de los datos:</p> <ul style="list-style-type: none"> • Ayudar a “retailers” (comercio al por menor) a mejorar su rendimiento, sus clientes y su contexto geográfico y

	<p>temporal.</p> <ul style="list-style-type: none"> • Ayuda a los gestores urbanos a tomar mejores decisiones gracias a un mejor conocimiento de la ciudad. • Mide el impacto real de eventos o de decisiones. (www.mwcimpact.com). • Permite que terceros puedan crear nuevos servicios de valor sobre datos anónimos proporcionados por BBVA, en combinación o no con otras fuentes de datos.
<p>Limitaciones encontradas</p>	<p>Costes excesivos: La gestión tradicional de datos bancarios y de las prácticas de gestión de datos son incapaces de soportar las lo que requiere Big Data, lo que puede desembocar en proyectos de análisis de datos costosos y con retrasos.</p> <p>Límites presupuestarios: Implementar un modelo de coste-beneficio puede ser difícil cuando con herramientas como Hadoop los altos costes de desarrollo son habituales. Las nuevas plataformas analítica y basadas en la nube para Big Data hacen que poner en marcha una plataforma y lograr su rentabilidad sea más viable que nunca.</p> <p>El “momento Eureka”: Los vendedores y proveedores de servicios deben de aportar liderazgo intelectual, modelos de datos concretos y plantillas específicas el “momento Eureka” en las organizaciones, y aportar un modelo y una visualización mejor sobre cómo la tecnología puede resolver un problema de negocio de un modo más significativo.</p> <p>Alineación con el negocio: Los principales accionistas tienen unos objetivos empresariales muy claros. A menudo, estos objetivos no están alineados con las ideas relativas al Big Data, lo que genera un gran obstáculo para las organizaciones de servicios financieros.</p> <p>Lagunas de conocimiento: Las estrategias tecnológicas y los procesos de negocio para Big Data son muy distintos. Las lagunas en el almacenamiento de datos y las estrategias de procesamiento, así como la falta de conocimiento o dirección por parte de los gerentes de sistemas pueden generar dudas.. Además, los profesionales de la tecnología bancaria pueden desconocer las herramientas de gestión de Big Data. La formación técnica y de usuario final también puede impedir a los bancos apostar por Big Data.</p>

MEDIAMATH

Descripción	MediaMath es una compañía de tecnología que ofrece una plataforma abierta y extensible a las marcas para que puedan disfrutar del marketing de una manera transparente y centrada en conseguir sus objetivos.
Objetivos	<ul style="list-style-type: none"> • Aplicaciones especiales como el análisis de atribución entre canales que permite a anunciantes para recopilar y eliminar todos los datos de usuario a través de • Búsqueda, de una multitud de fuentes y plataformas técnicas • Informes internos / financieros • Alcance y frecuencia del intervalo dinámico • Análisis de embudo de compra • Análisis y clasificación de sitios profundos • Detección de fraude a nivel IP • Informes y atribuciones en tiempo casi real (por ejemplo, 15 min)
Tecnologías aplicadas	<ul style="list-style-type: none"> • Plataforma DSP Omnicanal: Para acceder a campañas omnicanal a gran escala para garantizar la pertinencia del contexto y del mensaje. • Plataforma de publicidad: Para acceder de forma privilegiada a medios de gran calidad en un ecosistema global de plataformas de publicidad

RETOS	SOLUCIONES	BENEFICIOS
<p>Optimización / promediación con datos de clientes. Un cliente de servicios financieros busca diferenciar aún más la oferta de productos basada en el valor de vida del consumidor (LTV)</p>	<ul style="list-style-type: none"> • Las puntuaciones e indicadores clave de LTV pasaron a MediaMath en tiempo real utilizando MediaMath Servicio flexible de integración de datos de clientes para uso en optimización • Establecimiento de un vínculo de integración de datos fuera de línea del cliente • Modelos robustos de predicción y apariencia Desarrollado para ampliar la audiencia de prospección 	<ul style="list-style-type: none"> • El rápido periodo de marca de agua permite al cliente Desactivar una conversión genérica y Mejor oferta para cada consumidor • Mejor perfiles de clientes habilitados a través de intercambios. • En una campaña que ya estaba en la cima del plan, la rápida integración de los datos del cliente resultó un 50% más de incremento en el mercado.
<p>Tasa de avance del embudo / análisis de costos En una gran iniciativa de marca, un cliente debe</p>	<ul style="list-style-type: none"> • Análisis de nivel de usuario y medida de velocidad de avance a los clientes en cada etapa del ciclo de vida. 	<ul style="list-style-type: none"> • Análisis de ROI real de cada etapa en el embudo (Conciencia, familiaridad, consideración y acción) • Seguimiento de usuario y

<p>obtener visibilidad de la tasa de avance y coste.</p>	<ul style="list-style-type: none"> • Proceso estándar establecido para proveer a todos los clientes con la conversión más básica del proceso. 	<p>análisis de exposición únicos que informan una inversión más efectiva</p> <ul style="list-style-type: none"> • Transparencia total en el sitio y la conducta clave <p>Segmentos de avance a través de las etapas.</p>
<p>Exposición al área comercial y Análisis del ROI. El gran cliente de la CPU necesita comprender el impacto de una amplia campaña al nivel del mercado. Mientras soporta nuevos productos en los mercados de prueba.</p>	<ul style="list-style-type: none"> • Estrategia de campaña de mercado oscuro Aislar y comparar como mercados pre y post campaña. • Cientos de millones de impresiones de campaña. Frecuencia y punto de calificación bruto (GRP) en los niveles de cierre y zona comercial. • Datos de investigación de mercado estándar de la industria integrado para proporcionar métricas de nivel SKU pre y post campaña. 	<p>El anunciante recibe un alcance significativo y medidas de audiencia</p> <ul style="list-style-type: none"> • El cliente puede controlar y monitorear el alcance y frecuencia diaria para asegurar los objetivos.

Fuente: www.ibm.com/software/data/netezza

7. CONCLUSIONES

Debido al aumento del uso de las tecnologías de la información a través de Internet, y por medio de cámaras, escáneres médicos, sensores, micrófonos... los datos que se generan han aumentado de una manera desmesurada. Las herramientas y tecnologías empleadas por las compañías hasta hace algunos años no eran capaces de transformar esos datos en información y por consiguiente en conocimiento, lo que se entiende por información útil, por eso este nuevo fenómeno llamado Big Data ha revolucionado el mundo empresarial permitiendo a las empresas conseguir información de los clientes de una manera rápida, y así poder conocer las necesidades de estos, fidelizarlos y por consiguiente, aumentar sus beneficios.

Pero, ¿de dónde salen los datos? Si bien es cierto que la tecnología dentro del mundo empresarial lleva muchos años utilizándose, el uso de las redes sociales por parte de los consumidores ha hecho que mediante un solo clic, las compañías pueden conocer los gustos de una persona. Facebook, twitter, instagram... son aplicaciones conocidas y usadas por casi todas las personas, pero si además de eso se añaden las compras online, las descargas a través de internet... el volumen de datos que se pueden esperar se multiplica.

A través de este trabajo se ha podido comprobar cómo esta tecnología que está emergiendo poco a poco, cada vez tiene más cabida en las empresas, ya que al generar no solo información sino conocimiento ofrece grandes ventajas competitivas en el mercado frente a otros competidores.

La utilización de Big Data ha favorecido al descubrimiento de cosas que en su ausencia se habría tardado años en conocer. Las herramientas que proporciona Big Data a través de la velocidad de análisis, hacen que el analista de datos pueda cambiar sus ideas basándose en el resultado obtenido y modificar el procedimiento hasta conseguir llegar al valor que persigue.

Por tanto se puede decir que:

- Los datos son cada vez más importantes en el desarrollo de las empresas y organizaciones ya que el uso de estos han hecho evolucionar sus negocios.
- El correcto uso de Big Data hace que se cree valor, aportando mayor transparencia y mejor utilización de los datos.
- A través de los datos las empresas se pueden anticipar a las necesidades y los gustos de los consumidores, mejorando la segmentación y posibilitando la creación de estrategias más concretas.
- Automatiza el proceso de toma de decisiones haciéndolo más simple y proporciona nuevos procesos de innovación.

Además, aunque se haya podido comprobar que aún tiene algunas limitaciones, ya que su implantación no es nada fácil, pues requiere de una gran inversión, seguridad... Big Data es una tecnología que progresa y que estará muy presente en el futuro ya que muchos de los trabajos que se generarán en un futuro están ligados a las tecnologías de la información, adquiriendo así profesionales en este ámbito, lo que propiciará la creación de conocimiento, seleccionando aquello que sea de utilidad y dejando de lado aquella información que provoca "ruido".

Sería de gran interés para el mundo Big Data, la formación de nuevos "Científicos de datos" que aportaran sus capacidades para continuar con la innovación de estos

procesos dentro de las organizaciones, ya que el mundo actual requiere tanto de innovación como de creatividad.

8. REFERENCIAS

BIBLIOGRAFÍA

Eaton, C., Deroos, D., Deutsch, T., Lapis, G., & Zikopoulos, P. (2012). Understanding big data. *Analytics for Enterprise class Hadoop and Streaming Data*.

Khera, R., & Krumholz, H. M. (2017). With Great Power Comes Great Responsibility: Big Data Research From the National Inpatient Sample. *Circulation: Cardiovascular Quality and Outcomes*, 10(7), e003846.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), 21.

Logan, R. K., & Stokes, L. W. (2004). *Collaborate to compete: Driving profitability in the knowledge economy*. New York: Wiley.

López García, D. (2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones* (Master's thesis).

López-Cantos, F. (2015). La investigación en comunicación con metodología BigData. *Revista Latina de Comunicación Social*, (70).

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. McKinsey Global Institute.(2011). *Big data: The next frontier for innovation, competition, and productivity*, 156.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: la revolución de los datos masivos*. Turner.

P. Zikopoulos, C. Eaton, D. DeRoos, T. Deutsch, & G. Lapis, *Understanding Big Data*, USA: McGraw-Hill Books, 2012.

Pérez, M. (2015). *BIG DATA-Técnicas, herramientas y aplicaciones*. Alfaomega Grupo Editor.

Rodríguez Rovira, J. M. (1999). La gestión del conocimiento: una gran oportunidad. *Profesional de la Información, El: Information World en Español*, 8(3), 4-7.
Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, 19, 40.

S. Montoro, Server and Cloud Platform. Disponible en: <http://lapastillaroja.net/2012/02/NoSQL-for-non-programmers/>, 2012.

Sáez Vacas, F., García, O., Palao, J., & Rojo, P. (2003). Temas básicos de innovación tecnológica en las empresas. *Documentación de la asignatura Innovación Tecnológica*, www.dit.upm.es/~fsaez/innovaciontecnologica.html.

searchstorage.techtarget.com, *Examining HDFS and NameNode in Hadoop architecture*. Disponible en: <http://searchstorage.techtarget.com/video/Examining-HDFS-and-NameNode-in-Hadoop-architecture>, 2012.

Servicios, T. I. C. (2006). Definición de TIC. *Recuperado de <http://www.serviciostic.com/lastic/definicion-de-tic.html>*.

Sveiby, Karl-Erik (1997). *The new organizational wealth: Managing and measuring knowledge-based assets*. San Francisco: Berrett-Koehler.

T. Olavsrud, *Big Data Causes Concern and Big Confusion*. Disponible en: http://www.cio.com/article/700804/Big_Data_Causes_Concern_and_Big_Confusion?page=2&taxonomyId=3002, 2012

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

West, G. (2013). Big data needs a big theory to go with it. *Scientific American*, May, 15.

WEBGRAFÍA

<https://www.bbva.es/>

<https://digitalcologne.de/>

Cloudera.com, Cloudera, Inc. Disponible en: <http://www.cloudera.com/content/cloudera/en/why-cloudera/hadoop-and-big-data.html>, 2013.

http://www.decideo.com/Las-5-etapas-a-realizar-para-que-todo-proyecto-Big-Data-sea-exitoso_a1325.html

<https://www.centrodeinnovacionbbva.com/proyectos/big-data>

<https://www.veinteractive.com/es/blog/que-es-big-data-fases-elementos/>

<https://www.wikipedia.com>

<https://www.ibm.com/es-es/>

<http://vauzza.es/es/>. Vauzza: Expertos en estrategia e implementación de proyectos tecnológicos. <http://vauzza.es/es/>

searchstorage.techtarget.com, *Examining HDFS and NameNode in Hadoop architecture*. Disponible en: <http://searchstorage.techtarget.com/video/Examining-HDFS-and-NameNode-in-Hadoop-architecture>, 2012.

9. GLOSARIO

Ada: es un lenguaje de programación orientado a objetos y fuertemente tipado de forma estática que fue diseñado por Jean Ichbiah. Ada se usa principalmente en entornos en los que se necesita una gran seguridad y fiabilidad como la defensa, la aeronáutica (Boeing o Airbus), la gestión del tráfico aéreo (como Indra en España) y la industria aeroespacial entre otros.

Apache Hadoop: es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop se inspiró en los documentos Google para MapReduce y Google File System (GFS).

Customer centricity: es una estrategia cuyo objetivo primordial es alinear la conceptualización, desarrollo y comercialización de los productos y servicios de una marca, con las necesidades y deseos de sus clientes más valiosos.

MMP: es una forma de cómputo en la que muchas instrucciones se ejecutan simultáneamente,¹ operando sobre el principio de que problemas grandes, a menudo se pueden dividir en unos más pequeños, que luego son resueltos simultáneamente (en paralelo).

Open Source (código abierto): es el software desarrollado y distribuido libremente. Se focaliza más en los beneficios prácticos (acceso al código fuente) que en cuestiones éticas o de libertad que tanto se destacan en el software libre.

Pirámide DIKW: Conjunto de modelos para representar las relaciones aparentemente estructurales entre Datos, Información, Conocimiento, y en algunos casos sabiduría.

RFID o identificación por radiofrecuencia (del inglés Radio Frequency Identification) es un sistema de almacenamiento y recuperación de datos remoto que usa dispositivos denominados etiquetas, tarjetas o transpondedores RFID. El propósito fundamental de la tecnología RFID es transmitir la identidad de un objeto (similar a un número de serie único) mediante ondas de radio. Las tecnologías RFID se agrupan dentro de las denominadas Auto ID (*automatic identification*, o identificación automática).

ROI (Retorno sobre la inversión): es una razón financiera que compara el beneficio o la utilidad obtenida en relación a la inversión realizada, es decir, representa una herramienta para analizar el rendimiento que la empresa tiene desde el punto de vista financiero.

Spark: es un lenguaje de programación especialmente diseñado para sistemas de alta integridad. Es un subconjunto anotado de Ada desarrollado por la empresa británica Praxis High Integrity Systems, Inc que elimina ciertas características del lenguaje consideradas peligrosas en este tipo de sistemas (como las excepciones o la sobrecarga de operadores), y que añade anotaciones formales para realizar automáticamente análisis de flujo de datos y de información. *un conjunto de herramientas destinadas al análisis de flujo de datos y de información. De hecho, el nombre «SPARK» deriva de «SPADE Ada Kernel».*

BIG DATA

www.wipro.com

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

The "three V's", i.e. the Volume, Variety and Velocity of the data coming in is what creates the challenge.

VOLUME

- > 3,500 NORTH AMERICA
- > 2,000 EUROPE
- > 250 CHINA
- > 200 MIDDLE EAST
- > 50 LATIN AMERICA
- > 400 JAPAN

VELOCITY

- 2.9 MILLION EMAILS SENT EVERY SECOND
- 20 HOURS OF VIDEO UPLOADED EVERY MIN
- 50 MILLION TWEETS PER DAY

VARIETY

- PEOPLE TO PEOPLE:** NETZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...
- PEOPLE TO MACHINE:** ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...
- MACHINE TO MACHINE:** SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...

CASE STUDY - Healthcare

\$900 billion is the potential annual value to Healthcare

- \$165B CLINICAL:** TRANSPARENCY IN CLINICAL DATA AND CLINICAL DECISION SUPPORT; AGGREGATION OF PATIENT RECORDS, ONLINE PLATFORMS AND COMMUNITIES
- \$108B R&D:** RESEARCH AND DEVELOPMENT, PERSONALIZED MEDICINE, CLINICAL TRIAL DESIGN
- \$9B PUBLIC HEALTH:** PUBLIC HEALTH AND RESPONSE SYSTEMS; ADVANCED FRAUD DETECTION, PERFORMANCE BASED DRUG PRICING
- \$47B ACCOUNTS:** RESEARCH AND DEVELOPMENT, PERSONALIZED MEDICINE, CLINICAL TRIAL DESIGN
- \$5B BUSINESS MODEL:** RESEARCH AND DEVELOPMENT, PERSONALIZED MEDICINE, CLINICAL TRIAL DESIGN

- 57.6% OF ORGANIZATIONS SURVEYED SAY THAT BIG DATA IS A CHALLENGE
- 72.7% CONSIDER DRIVING OPERATIONAL EFFICIENCIES TO BE THE BIGGEST BENEFIT OF A BIG DATA STRATEGY
- 50% SAY THAT BIG DATA HELPS IN BETTER MEETING CONSUMER DEMAND AND FACILITATING GROWTH

VALUE

Industry	Productivity Increase	Sales Increase
Retail	49%	\$9.68B
Consulting	38%	\$5.06B
Air Transportation	21%	\$4.38B
Construction	20%	\$4.2B
Food Products	20%	\$3.4B
Steel	20%	\$3.4B
Automobile	19%	\$2B
Industrial Instruments	18%	\$1.2B
Publishing	18%	\$0.8B
Telecommunications	17%	\$0.4B

40% PROJECTED GROWTH IN GLOBAL DATA CREATED PER YEAR

5% PROJECTED GROWTH IN GLOBAL IT SPENDING PER YEAR

The estimated size of the digital universe in 2011 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

DO BUSINESS BETTER

WIPRO
Apply the Power

Sources: • PricewaterhouseCoopers (PwC) Big Data: The Next Frontier for Innovation, Competition and Productivity. McKinsey Global Institute Report • International Data Corporation (IDC) Smart Data: The Next Frontier for Innovation, Competition and Productivity. • IBM Business Value Report • IBM Business Value Report • IBM Business Value Report