

Big data y los nuevos manejos de la información

Resumen

Con la llamada sociedad de la información y el uso de las nuevas tecnologías, los volúmenes de datos se han venido incrementando de manera exponencial en los últimos años, tanto así que los sistemas tradicionales de almacenamiento se han quedado limitados en sus capacidades frente al número de información generada. *El big data* busca gestionar de manera eficiente la información de tal manera que, a partir de su procesamiento y análisis las organizaciones públicas o privadas puedan tomar mejores decisiones. El presente artículo presenta una definición del *big data* como tendencia, apoyándose en diferentes fuentes documentales; de igual manera se ofrecen los escenarios presentes y futuros con el objetivo de entregar una visión de esta tendencia y sus posibles implicaciones.

Palabras clave

Big data, toma de decisiones, logística, análisis de datos.

Introducción

El *big data* es la expresión utilizada para referirnos a la gestión que se especializa en el manejo de datos masivos, llamada también ciencia de datos; sus métodos son una serie de herramientas informáticas y estadísticas para la agrupación y el análisis de datos recolectados, gracias a los volúmenes de información generados en la interacción de los individuos, muchos de ellos a través de plataformas virtuales y dispositivos electrónicos y móviles. La importancia del *big data* radica en que facilita la toma de decisiones para las organizaciones y empresas, a partir de la lectura y el análisis de los comportamientos y afinidades de sus diferentes grupos de interés, permitiéndoles generar estrategias y acciones centradas en los mismos.

La llamada sociedad de la información nombrada así por diferentes autores, no adquiere ese nombre en vano, en el sentido de que, las personas y las organizaciones que aprendan a agrupar y manejar la información, tendrán una ventaja significativa a la hora de tomar las decisiones importantes a nivel comercial, logístico o administrativo. Si en los grandes mercados del mundo han comprendido la importancia de utilizar el *big data*, es necesario entonces que Colombia en sus aspiraciones

comerciales desarrolle y fortalezca a nivel interno y en todos los niveles de la economía, el manejo de datos especializados.

1. Definición del concepto de *big data*

El *big data* se ha definido como un activo de información de alto volumen, alta velocidad y alta variedad, que exigen formas rentables e innovadoras de procesamiento de la información para una visión mejorada y la toma de decisiones (Chen, Preston, & Swink, 2015).

En el 2003 se da el “Boom” del *big data*, cuando Google revela los documentos que explican la manera en que gestionan sus datos, dando cabida a un desarrollo de la gestión de los mismos para crear valor a través de desarrollos de software de código abierto; debido a ello se ha mejorado la innovación tecnológica en función del desarrollo socioeconómico mundial, siendo así recibida de manera positiva por las empresas, que para 2018 serán la mitad más grandes del mundo las que dispongan del *big data* para apoyarse en su proceso de gestión empresarial.

2. Situación actual del *big data*

Diferentes mitos y versiones sobre lo que es el *big data* se han diseminado durante años en la internet, resultando de ello algunas definiciones como las recopiladas por Camargo, Camargo & Joyanes (2014) como son:

- Cantidades masivas de datos que se acumulan con el tiempo que son difíciles de analizar y manejar utilizando herramientas comunes de gestión de bases de datos.
- Tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales.
- Se refiere a las herramientas, procesos y procedimientos que permitan a una organización crear, manipular y administrar grandes conjuntos de datos e instalaciones de almacenamiento.
- “Técnicas y tecnologías que hacen que sea económico hacer frente a los datos a una escala extrema (...).
- Se refiere a las herramientas, los procesos y procedimientos que permitan a una organización crear, manipular y gestionar conjuntos de datos muy grandes y las instalaciones de almacenamiento.
- Gran volumen, velocidad o variedad de información que demanda formas costeables e innovadoras de procesamiento de información que permita ideas extendidas, toma de decisiones y automatización del proceso.
- Consiste en consolidar toda la información de una organización y ponerla al servicio del negocio.

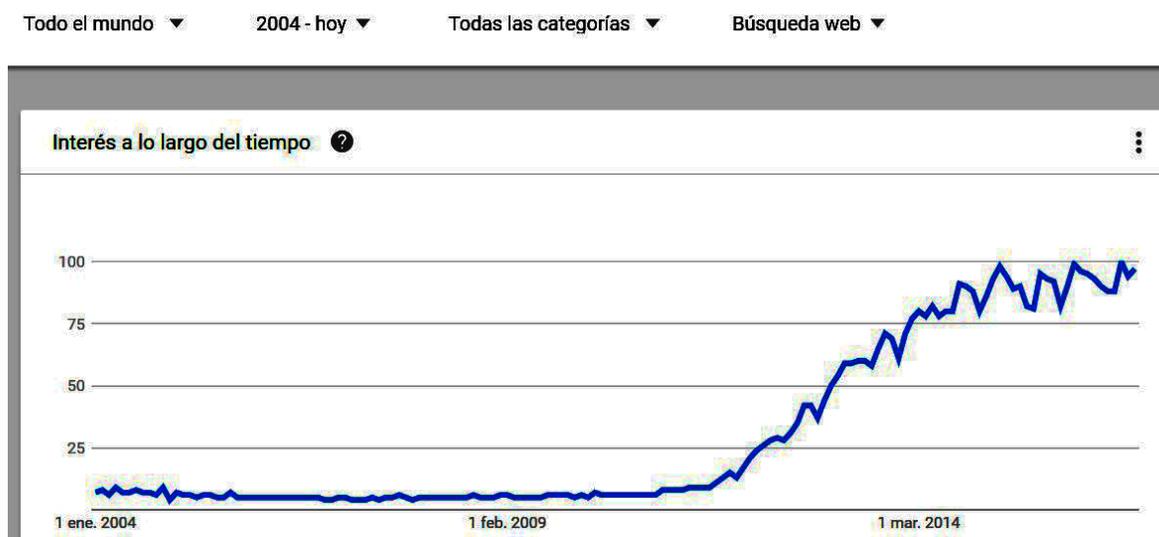
Todas estas definiciones, con algunas variaciones en cuanto a sus características, apuntan a una descripción de una gran cantidad de datos. De aquí, que se reconozca el *big data* por las tres uve (VVV), haciendo alusión al *volumen* de datos que en la actualidad es común sobrepasen los terabytes en las organizaciones; a la *variedad* de datos que significan imágenes, videos, texto estructurado y datos de redes sociales, entre otros; y a la *velocidad* que está relacionada con los requerimientos de procesamiento de algoritmos complejos que deben ser gestionados en cuestión de milisegundos (Paradigma Digital, 2016).

Adicionalmente, aparece una cuarta uve (V) que representa el verdadero sentido del *big data*, y que le da sentido a la última parte de la definición referente a la “toma de decisiones” y es el valor, del cual se deriva el *data science* o “ciencia de los datos” (Paradigma Digital, 2016).

El *data science* surge entonces como una herramienta que utiliza las tecnologías de Business Intelligence para proporcionar puntos de vista históricos, actuales y de predicción relacionados con la dinámica empresarial, para lo cual el *big data* posee funciones que van encaminadas a la presentación de informes, análisis de información, extracción de datos, gestión de rendimiento empresarial, análisis de predicción, etc. (Goyzueta, 2015).

El “Boom” del *big data* comienza en 2003, cuando google libera una serie de documentos que explican cómo gestionaban los datos en aquel entonces (Paradigma Digital, 2016). Uno de esos informes habla sobre el Google File System (GFS) o sistema de archivos de Google, que es el DFS o que en español significa sistema de archivos distribuidos; este da soporte a todas las aplicaciones de Google Inc (Riego, s.f.), y el otro referido al MapReduce, que es un modelo de programación que da soporte al sistema de procesamiento en paralelo que a su vez consiste en procesar diferentes paquetes de datos al mismo tiempo. Sin embargo, según la línea de tendencia marcada por búsqueda de “Big Data” a nivel mundial reportada por Google Trends, se nota un despegue importante a partir del 2011.

Grafico 1. Interés mundial en big data en Google



Fuente: Google Trends (2016)

Con relación a los hechos históricos que precedieron el *big data*, Raya (2015) ofrece un panorama de los diferentes proyectos que contribuyeron en la consolidación actual del *big data*. Para el año 2006 aparece el proyecto *Apache Hadoop*, desarrollando software de código abierto, adoptando y mejorando el MapReduce; donde gran cantidad de datos fragmentados en pequeños paquetes con el fin de solucionar un gran problema convirtiéndolo en pequeños problemas de más fácil solución y de manera simultánea, dando viabilidad a la adopción masiva del *big data*. En 2008 Facebook desarrolla una infraestructura de almacenamiento de datos sobre Hadoop para proporcionar agrupación, consulta y análisis de datos la cual se llamó Hive (Raya, 2015).

Poco después, el mismo año se crea HBase, una base de datos inspirada en Big Table de Google, que consiste en una serie de tablas que contienen filas y columnas en el que las primeras están consignados los registros, y en las segundas los atributos de un objeto permitiendo la agrupación de muchos atributos (familias de atributos). Este tipo de base de datos es recomendada para agrupar gran cantidad de datos en el que es más frecuente su lectura que su escritura y en el cual se requiere de muchas operaciones con los atributos como por ejemplo el manejo de datos estadísticos (Raya, 2015).

En la actualidad se busca que la información que se gestione a través de la web no discrimine la información en cuanto a raza, género, ingresos y/o ubicación barrial, con el fin de estar acorde con las políticas internacionales que apuntan a la inclusión, haciéndolo posible a través de un algoritmo que lo verifique. En cambio, sí se busca fortalecer la innovación tecnológica que tenga un impacto socioeconómico significativo y que vaya de la mano del tejido industrial tal y como lo hace el Centro Tecnológico de Cataluña, Eurocat; que dirige su actividad a siete ámbitos estratégicos relacionados con la alimentación, la energía y recursos, los sistemas industriales, las industrias basadas en el diseño, las industrias relacionadas con la movilidad sostenible, las industrias de la salud y las industrias culturales y basadas en la experiencia (Abajo, 2016).

En Colombia se ha venido fortaleciendo el tema del *big data*, Así lo muestra el proyecto CAOBA en que actores empresariales, académicos y sectores del Estado se articulan para promover y fortalecer el *big data* y la analítica de datos en el país, todo ello debido a las grandes afectaciones positivas que tiene esta tendencia para el desarrollo de las empresas, así lo afirma la viceministra del Ministerio de las Tecnologías de la Comunicación e Información, María Isabel Mejía, en entrevista para el periódico el Tiempo (2016, párr.4):

La estrategia consiste en la investigación aplicada, la formación de proyectos, la educación y el emprendimiento. Con la conformación de centros de excelencia y apropiación se dará acompañamiento a empresas interesadas en temas de *big data*. También se busca apoyar programas académicos que gradúen científicos de datos en el país. Esta profesión es considerada hoy en día 'la más sexí del siglo XXI' por el impacto que tiene en todo el mundo.

El principal objetivo es entonces, implementar una estrategia pedagógica para que las empresas y los directivos entiendan las potencialidades del *big data* y el impacto de las nuevas tecnologías en el desarrollo económico de las empresas. En concreto se promoverán becas para maestrías, doctorado y diplomados, de esta manera se hará una apropiación por parte de los profesionales nacionales de esta poderosa herramienta ya muy utilizada en otros países.

Las bondades del *big data* no son solo útiles para el sector empresarial, el Estado puede aprovechar la capacidad el *big data* para tomar mejores decisiones y diseñar mejor las políticas públicas tal y como afirmó en su momento el director del Departamento Nacional de Planeación, Simón Gaviria:

Sólo con datos de calidad se pueden hacer presupuestos informados por resultados y tomar mejores decisiones. La estrategia big data está transformando la construcción de las políticas públicas... existe la necesidad de optimizar los recursos del Estado y generar una cultura de datos en todo el país. Actualmente se toman decisiones con un porcentaje mínimo de los datos e información disponible. Big Data permite acceder y analizar la información a la cual antes no se tenía acceso e, incluso, conocer la opinión de la ciudadanía con precisión, sobre algún tema en particular, en cuestión de horas (DNP, 2016).

3. Situación futura del *big data*

El *big data* seguirá expandiendo su abanico de posibilidades; haciendo un ejercicio breve de retrospectiva es evidente el aumento progresivo de datos e información, y con el aumento de la información más posibilidades de hacer mejores análisis de datos (Esterá, 2016).



Es también lógico, que las empresas empiecen a utilizar más profesionales en *big data* en la medida que este se convierte en una herramienta eficiente para el análisis de datos masivos, y a partir de allí tomar mejores decisiones (Esterá, 2016).

Se abrirán áreas completamente nuevas: ahora mismo suena a ciencia ficción, pero se obtendrán datos de áreas en la que nunca se pensó que serían accesibles. ¿Podremos obtener datos de compra en la calle?, ¿El análisis de comportamiento humano a través de sus movimientos captados por cámaras de vídeo? Las posibilidades son infinitas (Esterá, 2016).

En el área de la logística, el *big data* encierra grandes posibilidades y aportes, a partir de la identificación de estos beneficios el *big data* será la ciencia que determine la mejor toma de decisiones, tal y como afirma Gómez.

Ofrece *control eficiente* de los activos con los que se cuenta en la empresa, por ejemplo, la flota de vehículos, la materia prima almacenada, los productos en stock, entre otras cosas. De tal forma, los procesos de distribución son optimizados, reduciendo considerablemente los gastos y tiempos de organización.

El *big data segmenta la demanda* de la empresa. Es decir, analizando la información de los clientes, se puede determinar cuál es la necesidad e interés de los consumidores, por ejemplo: los productos que utilizan o cuáles podrían interesarles. De igual forma, la información que se obtiene puede ayudar a rediseñar los productos actuales, brindándoles características personalizadas según los gustos e intereses. Al tomar todos estos aspectos en cuenta, se estará logrando mayor fidelización de los clientes con la marca. (Gomez, s.f.)

Una vez analizados todos los beneficios que el *big data* aporta a la logística podemos concluir que su aplicación significa una importante optimización del sector, además de obtener rentabilidad, incrementar el número de clientes y detectar futuros modelos de negocio (Juan, 2016).

En la sociedad actual, también llamada por diferentes académicos, la sociedad del conocimiento, es

precisamente el *big data* esa puesta que abre las posibilidades para obtener cualquier tipo de información, a partir del análisis de datos, algunos llaman al *big data* el nuevo petróleo, por las potencialidades en ahorro de procesos y toma de decisiones acertadas y eficaces.

Por lo tanto, en la medida en que las empresas se adapten a estos avances tecnológicos logran enfocar la toma de decisiones, por ejemplo, estudiar los clientes y sus prácticas de una manera detallada, a partir de allí diseñar la mejor estrategia comercial, para ello cada empresa deberá tomarse la tarea de cultivar la cultura de las TIC, investigar, cualificar y diseñar su estrategia *big data* según sus necesidades e intenciones.

Referencias

Abajo, C. G. (3 de Octubre de 2016). *El Big Data Congress aborda las oportunidades del tratamiento de datos*. Obtenido de tendencias21.net: El Big Data Congress aborda las oportunidades del tratamiento de datos

Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2014). *Conociendo Big Data*. Obtenido de Scielo: <http://www.scielo.org.co/pdf/rfing/v24n38/v24n38a06.pdf>

Chen, D. Q., Preston, D. S., & Swink, M. (2015). *How the use of big data analytics affects value creation in supply chain management*. Obtenido de Econbiz: <http://www.tandfonline.com/doi/full/10.1080/07421222.2015.1138364?scroll=top&needAccess=true>

Estera, À. (Diciembre de 2016). *10 predicciones sobre el futuro del Big Data*. Obtenido de Saimasolutions: <https://saimasolutions.com/10-predicciones-sobre-el-futuro-del-big-data/>

El Tiempo. (15 de Marzo de 2016). Así se apropiará el 'Big Data' en nuestro país. *EL TIEMPO*. Obtenido de <http://www.eltiempo.com/archivo/documento/CMS-16537461>

Gomez, E. (s.f.). *Cómo influye el Big Data en la logística de una empresa*. Obtenido de <https://neoattack.com/big-data-en-la-logistica/>

Goyzueta, S. I. (2015). *Big Data Marketing: una aproximación*. Obtenido de Scielo: http://www.scielo.org.bo/pdf/rp/n35/n36_a07.pdf

Grupo de Comunicaciones y Relaciones Públicas DNP. (22 de Marzo de 2016). *Big Data: Colombia entra en la revolución de los datos*. Obtenido de <https://www.dnp.gov.co/Paginas/Big-Data-Colombia-entra-en-la-revoluci%C3%B3n-de-los-datos-.aspx>

Juan, C. (2016). *Cómo el Big Data ha revolucionado la Logística*. Obtenido de <https://www.iebschool.com/blog/big-data-en-logistica-big-data/>

Paradigma Digital. (22 de Marzo de 2016). *Spark Wars*. Obtenido de YouTube: <https://www.youtube.com/watch?v=oE6jU3EPCW0>

Referencias

Raya, A. M. (1 de Diciembre de 2015). Big Intelligence. *Nuevas capacidades big data para los sistemas de vigilancia estratégica e inteligencia competitiva*. Obtenido de EOI: <https://www.eoi.es/es/savia/publicaciones/21151/big-intelligence-nuevas-capacidades-big-data-para-los-sistemas-de-vigilancia-estrategica-e-inteligencia-competitiva>

Riego, M. d. (s.f.). Google File System (GFS). *El sistema de archivos distribuido de Google !!* Obtenido de Fing: <https://www.fing.edu.uy/inco/cursos/sistoper/recursosEnlaces/presentacion-gfs.pdf>

Romero, A. C., Sanabria, J. S., & Cuervo, M. C. (2012). *revistas.uptc*. Obtenido de Utilidad y funcionamiento de las bases de datos NoSQL: <http://revistas.uptc.edu.co/index.php/ingenieria/article/view/2115/2078>