

**BIG DATA. CIENCIA DE LOS DATOS.  
EL CIENTÍFICO DE DATOS SE  
CONVIERTE EN UNA PROFESIÓN  
CON FUTURO**

**BIG DATA. DATA SCIENCE.  
THE DATA SCIENTIST BECOMES A  
PROFESSION WITH FUTURE**

Claudio Chavez  
Universidad Pedagógica de El Salvador  
“Dr. Luis Alonso Aparicio”

claudio.chavez@uped.edu.sv  
pp. 26 - 35

Recibido: 17-03-2023 Aceptado: 20-11-2023

## RESUMEN

Este documento trata de mostrar la notoriedad que está teniendo la tecnología actualmente. Cualquier persona con o sin conocimientos tecnológicos se pregunta cómo se almacena la información que se genera en el mundo: en Facebook, Twitter, Instagram o cómo Google es capaz de manejar todas las consultas que se hacen a diario. Todo esto nos lleva al concepto de *Big Data*.

PALABRAS CLAVE: *Big data*, tecnología, información, redes sociales, datos.

## ABSTRACT

This document tries to show the notoriety that technology is currently having. Anyone with or without technological knowledge wonders how the information generated in the world is stored: on Facebook, Twitter, Instagram or how Google is able to handle all the queries that are made daily. All this brings us to the concept of Big Data.

KEYWORDS: Big data, technology, information, social networks, data.

## Introducción

En estos últimos años, los avances tecnológicos producen una avalancha de datos que es necesario ordenar y procesar en todos los ámbitos: empresarial, académico, investigador y de la administración, con un nuevo concepto que ha dado en denominarse *Big Data*.

Por la simple denominación, se entiende que se trata de grandes volúmenes de información que no es sencillo tratar con las herramientas y procedimientos tradicionales. Encierra esta idea el tratamiento de información que hace evolucionar los métodos y recursos habituales para hacerse cargo de grandes volúmenes de datos (de terabytes pasamos a zetabytes). Estos se generan a gran velocidad (pasamos de datos en lotes/archivos a datos en “streaming”) y además se añade un posible componente de complejidad y variabilidad en el formato de esos datos (pasamos de datos estructurados a datos semiestructurados o no estructurados). Todo ello requiere de técnicas y tecnologías específicas para su captura, almacenamiento, distribución, gestión y análisis de la información.

También recientemente se añade una nueva “v” de valor: los datos por sí mismos, aun siendo muchos, no proporcionan valor a una empresa u organización. Es su tratamiento, a través de un proceso de planteamiento de hipótesis, creación de modelos estadísticos y semánticos, y definición de algoritmos de corta o larga duración, lo que permite descubrir el significado oculto en esos grandes volúmenes de datos.

Un ejemplo de esto es conocer, a ciencia cierta, cuántos clientes llegan a un supermercado, qué compraron, si pagaron en efectivo o con tarjeta de crédito, cuáles son las horas de mayor tráfico, entre otros; son los datos que un negocio de este rubro necesita tener en sus manos para elaborar sus planes mercadológicos, financieros, de inversiones y más.

Pero, ¿qué pueden hacer las empresas con el resultado del procesamiento de todos esos datos? El uso de la información es diverso, desde la inteligencia del mercado, la planificación estratégica, el “merchandising”, hasta alianzas con socios estratégicos, como por ejemplo proveedores o bancos, con quienes pueden definir promociones dirigidas al consumidor, que al final es un cliente común.

La información también permite a las organizaciones adaptar sus productos y servicios a las necesidades de los clientes y crear nuevas oportunidades de negocios.

La tarea de clasificar, procesar, interpretar toda esa masa de datos y hasta predecir comportamientos, ha propiciado, a nivel mundial, el apareamiento de una nueva profesión: el científico de datos, que es una persona formada en las ciencias matemáticas y las estadísticas, con dominios de programación y sus diferentes lenguajes: las ciencias de la computación y la analítica.

Los científicos de datos son profesionales que logran desarrollar sus habilidades analíticas para obtener información relevante a partir de gran cantidad de datos que producen las organizaciones, ya sean públicas o privadas.

El aporte de este tipo de profesionales es útil en diversos sectores de la economía, la educación, la salud, la política y hasta en el combate del crimen.

La repentina demanda por parte de las empresas e instituciones públicas de este perfil profesional que hasta hace poco no existía, está sustentando en la gran cantidad de información que generan, segundo a segundo, los avances tecnológicos, incluyendo las redes sociales.

De hecho, los datos son el recurso más abundante del planeta, ya que se calcula que el *Big Data*, la masa de información disponible en el mundo, genera 2.5 trillones de bytes de información nueva cada día.

Las personas, los animales, las cosas, los hábitos, los sentimientos, todos son datos que se pueden medir; hoy en día hay tanta cantidad de datos generada por los avances tecnológicos y las redes sociales que las empresas no saben cómo almacenarla, y los métodos tradicionales ya no son suficiente.

El científico de datos resulta una combinación de conocimientos estadísticos que un programador no suele tener, y conocimientos informáticos que un estadístico tampoco posee. Pero son ellos los que en forma individual han realizado esta tarea.

Según Burtch Works (s.f.), una firma que se especializa en el reclutamiento de ejecutivos en esta área y en el análisis de *Big Data*, indica que el 32 % de los científicos de datos activos vienen del mundo de las matemáticas y la estadística, el 19 %, de la ingeniería informática, y el 16 % de otras ingenierías.

Un indicador que refleja la demanda que ha cobrado este tipo de profesionales en el mundo son las encuestas especializadas de empleo, así como también el número de instituciones académicas que están ofreciendo cursos especializados en esta área.

En Estados Unidos, España y México, por citar algunos países, hay instituciones que han incluido en su oferta académica maestrías y postgrados en la ciencia de los datos, como el *CS109 Data Science*, de Harvard; el curso de *Machine Learning*, de Stanford; el *The Analytics Edge*, de MIT, o el *Learning for data*, del Instituto de Tecnología de California (CalTech).

## Áreas de trabajo para el científico de datos

1. Mercadeo  
Monitorear tendencias, analizar el rendimiento de campañas y productos nuevos o existentes; realizar segmentación de mercados.
2. Ventas  
Definir perfiles de clientes según el consumo, gustos y preferencias de los clientes para el diseño de estrategias.
3. Finanzas  
Evaluar riesgos crediticios, detección de fraudes, realizar proyecciones financieras, correr simulaciones de eventos de mercado, construcción de modelos predictivos.
4. Recursos Humanos  
Evaluaciones de productividad, desempeño laboral y clima organizacional.
5. Producción  
Algoritmos para la optimización de procesos, técnicas de control de calidad.
6. Sector Público  
Análisis de grandes volúmenes de datos para diseño de políticas públicas.

## Definición de *Big Data*

El *Big Data*, Macrodatos o Datos masivos, es un concepto que hace referencia al almacenamiento de grandes cantidades de datos y a los procedimientos usados para encontrar patrones repetitivos dentro de esos datos. El fenómeno del *Big Data* también es llamado datos a gran escala.

La disciplina dedicada a los datos masivos se enmarca en el sector de las tecnologías de la información y la comunicación. Esta disciplina se ocupa de todas las actividades relacionadas con los sistemas que manipulan grandes conjuntos de datos. Las dificultades más habituales vinculadas a la gestión de estas cantidades de datos se centran en la

recolección y el almacenamiento, búsqueda, compartición, análisis, y visualización. La tendencia a manipular enormes cantidades de datos se debe a la necesidad, en muchos casos, de incluir dicha información para la creación de informes estadísticos y modelos predictivos utilizados en diversas materias, como los análisis de negocio, la publicidad, los datos de enfermedades infecciosas, el espionaje y seguimiento a la población o la lucha contra el crimen organizado.

La importancia de *Big Data* radica en que éste impacta tanto en la industria como en el negocio e, incluso, en nuestra sociedad y además ofrece una ventaja competitiva considerable.

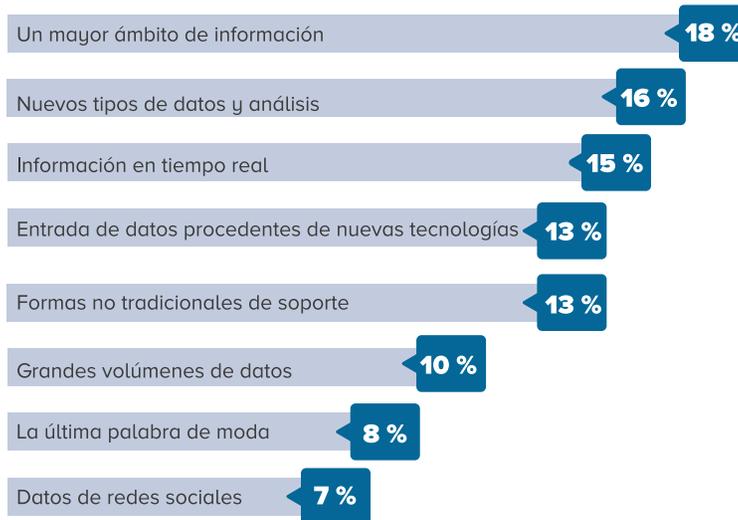
En efecto, es precisamente en ese tipo de datos donde las empresas han detectado que se encierra mayor valor. Hoy en día, para muchas empresas puede llegar a ser más importante detectar al cliente que más influye al resto de posibles compradores, que al que mayor volumen de compra realiza.

En la actualidad, la cantidad de datos que se generan es abismal y de una casuística extremadamente compleja para su análisis. Como hemos comentado, las empresas cada vez exigen que el análisis sea lo más cercano posible al tiempo real. Y en el *Big Data* está la clave, al traducirse el mismo en las variables de velocidad, variedad y volumen que requiere el mercado actualmente.

El informe de TicBeat (2012) define *Big Data* como: “la enorme cantidad de datos que desde hace unos años se genera constantemente a partir de cualquier actividad”; más adelante dicho informe recalca que: “el *Big Data* bien entendido en la búsqueda del mejor camino para aprovechar dicha avalancha de datos” TicBeat (2012).

Sin embargo, un estudio realizado por IBM Institute for Business Valué, junto con la colaboración de Saïd Business School (2012) el cual consistió en dar a los encuestados (más de 1,144 negocios y profesionales de TI de 95 países y docenas de expertos en la materia) una serie de características sobre *Big Data* para que escogieran las dos que mejor describiera el concepto, como se muestra en la Figura 1.

Como se puede observar en la Figura 1, según los encuestados, la definición de *Big Data* tiene sentido con un mayor ámbito de actuación de información y nuevos tipos de datos de análisis, mientras que no tiene sentido con datos en redes sociales.

**Figura 1***Definición de Big Data*

*Nota.* Información obtenida del IBM Institute for Business Value

Otras definiciones hacen referencia a la tendencia en el avance de la tecnología que han abierto las puertas hacia un nuevo enfoque del entendimiento y toma de decisiones

## Importancia del *Big Data*

Con el término *Big Data* se hace referencia a la tendencia del avance de las tecnologías que han abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que llevaría demasiado tiempo cargarlos en una base de datos relacional para su posterior análisis. Por lo tanto, el *Big Data* se aplicará para toda aquella información que no pueda ser procesada por los métodos tradicionales.

## Beneficios del *Big Data*

Una vez que se sabe la importancia de *Big Data*, sobre todo gracias a la mejora que supuso respecto a los modelos relacionales, se citarán los beneficios más habituales del *Big Data*; no obstante, estos beneficios no se tienen porque aplican a todas las organizaciones, ya que cada organización tiene y actúa en diferentes condiciones.

---

**Big Data. Ciencia de los Datos.**

**El científico de datos se convierte en una profesión con futuro.**

A continuación, se citan los beneficios e inconvenientes más relevantes que han sido extraídos de un artículo de Vauzza (2013):

### **Gestión del cambio**

Búsqueda de nuevas oportunidades de negocio a través de segmentación mejorada y venta cruzada de productos (mejora de la estrategia).

Mediante la aplicación de análisis y modelado predictivo a los datos de cuentas de clientes e historial de transacción, la solución permite a los agentes llevar a cabo una segmentación basada en la probabilidad de que el cliente contrate servicios o productos complementarios, o contratar servicios de mayor valor (mejora de segmentación).

Mediante el análisis de consumo de los servicios y productos de los clientes, la empresa puede optimizar las estrategias de venta cruzada, afinar mensajes de marketing y proporcionar ofertas específicas. Se puede predecir con mayor exactitud qué productos son los más apropiados para cada cliente (mejora de la estrategia).

Ofrecer la combinación adecuada de servicios y productos mejora la eficacia y la eficiencia de la fuerza de ventas de la compañía, mientras que el toque más personalizado ayuda a los agentes a forjar lazos más estrechos con clientes, lo cual mejora la lealtad (mejora de la estrategia).

### **Mejoras operativas**

Mayor capacidad de visibilidad del negocio a través de informes más detallados.

## **Inconvenientes *Big Data***

No obstante, no hay que olvidarse de los inconvenientes del *Big Data*. Siendo el principal de ellos el proceso de adopción de *Big Data*: software y hardware necesario y su coste. Pero además existen otros muchos, de menor peso, como, por ejemplo:

- Rechazo por parte del personal.
- Gasto de formación.
- Colaboración necesaria por parte de todos los departamentos.

- La denominada “Toma de decisiones pasivas”; esto hace referencia antes de la instalación de *Big Data*, a que las empresas primero esperan a que lo instalen sus competidores para ver que errores cometes con la creencia de que ellos lo podrán adoptar mucho más rápido.
- Coste.
- Problemas de privacidad.
- Problemas de información desactualizada.
- Filtrado (no todos los datos son información).

## Conclusiones: una reflexión

Es posible concluir que la disciplina del *Big Data*, en el manejo grandes cantidades de datos, enmarca a un sector de la tecnología y la información de la comunicación que hace referencia a los procedimientos usados para encontrar datos a grandes escalas.

Por otro lado, del gran volumen de información. Esta existe en una gran variedad de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, video, sistemas GPS, incontables sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la velocidad de respuesta sea lo demasiado rápida para lograr obtener la información correcta en el momento preciso. Estas son las características principales de una oportunidad para *Big Data*.

*Big Data* ya no es una promesa ni una tendencia, ya es una realidad y provoca cambios profundos en la industria. En la actualidad se necesitan personas que puedan entrar en interacción con grandes cantidades de datos, en distintas formas, que los avances tecnológicos generan.

Es debido a ello que se puede concluir que el termino *Big Data* se emplea, hoy en día, para describir el conjunto de procesos, tecnologías y modelos de negocio que están basados en datos y capturar el valor que los propios datos encierran. Esto se puede lograr tanto en la mejora en la eficiencia, gracias al análisis de los datos, como la aparición de nuevos modelos de negocios que supongan un motor de crecimiento. Se habla mucho de aspectos tecnológicos, pero hay que tener en cuenta la forma de valorar los datos para crear nuevos modelos de negocios o ayudar a los existentes.

## Referencias

O'Brien, J. (2008). *Sistemas de Información Gerencial* (7ª ed). Mc Graw Hill.

Laudon, K. C. (2008). *Sistemas de Información Gerencial* (10 ed). Mc Graw Hill.

MCleod, R. (2000). *Sistemas de Información Gerencial*. Prentice Hall

Burtch Works (s.f.). Industry Insights. <https://burtchworks.com/industry-insights>

Vauzza, M. (2013). "Todo lo que necesitas saber sobre Big. Eureka-Startups. <http://www.eureka-startups.com/blog/2013/05/28/todolo-que-necesitas-saber-sobre-bigdata/>