



PROFESIONES



DIGITALES

# big data

hagamos  
hablar  
a los datos

*Telefonica*  
FUNDACIÓN

GARY KING

HARVARD UNIVERSITY

**El *big data*  
no va sobre  
números**



→ EL *BIG DATA* EN ACCIÓN

Barack Obama usó *big data* en la campaña para su reelección en 2012. Un centenar de personas trabajó en el departamento de analítica de la campaña y sus esfuerzos se centraron en tres temas: registro (recoger datos de los votantes convencidos), persuasión (dirigirse a los dudosos de una forma eficaz) y voto del electorado (asegurarse de que los partidarios fueran a ejercer el voto).

Ya es una tradición desde finales del siglo pasado que cada década traiga consigo una transformación social originada por la tecnología. En los ochenta apareció el ordenador personal, y en los noventa internet y la World Wide Web. La que abrió el siglo XXI conoció la génesis de los medios sociales, y esta en la que nos hayamos nos ha desvelado el poder de los datos.

En un mundo recubierto de redes digitales, los datos están por todas partes. Los objetos y las personas no paran de producir datos, y las cifras que refleja el tráfico son apabullantes. A modo de ejemplo, la International Data Corporation (IDC) aporta los siguientes números:

**En 2020 la información nueva generada cada segundo por cada ser humano será de alrededor de 1,7 megabytes.**

**Entre 2013 y 2020 el volumen de datos que creamos y copiamos anualmente aumentará de 4,4 zettabytes a unos 44 zettabytes o lo que es lo mismo casi 40 billones de gigabytes.**

Para cuantificar el volumen de datos producido ya, tenemos que recurrir a múltiplos como el exabyte, que es un 1 seguido de 18 ceros, es decir un millón de millón de millones de bytes, o al zettabyte (mil veces superior).

El reto del *big data* consiste en poder gestionar esa avalancha de información en bruto que nos rodea para darle una utilidad. El dato en sí, aislado, no tiene valor, pero al ser tratado, procesado y analizado se convierte en una fuente de conocimiento, de ventaja competitiva para las empresas y, en suma, de riqueza. La tecnología de que disponemos actualmente ha hecho posible el surgimiento de un modelo de negocio en torno a los datos que hace bien poco era inviable.



ARTHUR CONAN DOYLE

LAS AVENTURAS DE SHERLOCK HOLMES

Constituye un craso error el teorizar sin poseer datos. Uno empieza de manera inconsciente a retorcer los hechos para acomodarlos a sus hipótesis, en vez de acomodar las hipótesis a los hechos

GEOFFREY MOORE

CONSULTOR

Sin analítica de *big data* las empresas están ciegas y sordas, deambulando por la web como un ciervo por una autopista



# los datos importan

El crecimiento de la cantidad de datos que generan las empresas, los ciudadanos y las Administraciones es imparable a medida que la sociedad se hace cada vez más digital. Si a ello le sumamos el tráfico generado por objetos —el denominado internet de las cosas (IoT)—, la tendencia se convierte en exponencial.

¿Qué podemos hacer con tal cantidad de datos? Muchas más cosas de las que imaginamos en la actualidad. Por ejemplo, el *big data* ayuda a las empresas a conocer mejor a sus clientes. Los usuarios de internet van dejando trazos mientras navegan: información sobre sí mismos y su comportamiento, sobre sus preferencias y costumbres... Procesar y analizar todos esos datos sueltos nos ayuda a conocer a una persona y, en consecuencia, a desarrollar acciones precisas para atraerla como cliente o para fidelizarla, si ya lo es.

Igualmente, el *big data* nos puede sugerir en qué nuevos nichos de mercado operar y qué productos debemos desarrollar para satisfacer necesidades del consumidor identificadas que actualmente no están cubiertas. También es una herramienta útil para detectar ineficiencias en el funcionamiento interno de la empresa, cuya mejora se puede traducir en ahorro de costes.

Finalmente, la *data science* es una disciplina basada en métodos matemáticos, estadísticos y de programación informática, que permite responder a preguntas relacionadas con la estrategia de la empresa, describiendo, anticipando y prediciendo sucesos y recomendando acciones.

El *big data* es algo que la empresa debe tomarse en serio puesto que cada vez más es un factor clave de competitividad y creación de valor. De hecho, ya se habla de estrategias empresariales “dirigidas por los datos” (*data-driven strategies*).

# por qué precisamente hoy hablamos de *big data*





Tomar decisiones y actuar en base a la información disponible es algo tan antiguo como la humanidad. Desde mediados del siglo XX, la informática ha facilitado el almacenaje y procesado de información en formato digital, permitiendo manejar grandes volúmenes de datos. La gente de todo el mundo navega por internet desde la primera mitad de la década de los noventa, haciendo uso de servicios de red y dejando su huella digital. Entonces, ¿por qué no es hasta el momento actual en que hablamos de *big data*? ¿No era posible hacer lo mismo antes?

Dos son los factores que han impulsado el despegue del *big data* en esta década:

- Ya vivimos en un mundo digital: se ha digitalizado prácticamente todo, generando nuevos tipos de grandes datos en tiempo real procedentes de innumerables fuentes. Es decir, la cantidad de datos disponible se ha disparado y muchos de estos no están normalizados (por ejemplo, los procedentes de dispositivos del internet de las cosas) y no encajan en los sistemas de almacenamiento tradicionales.
- Los avances recientes que han tenido lugar en las tecnologías de análisis permiten extraer conocimiento de los datos a una velocidad sin precedentes y además con una precisión y sofisticación nunca vistas.

No obstante, la economía de los datos todavía tiene que hacer frente a varios retos desde el punto de vista tecnológico:

- El almacenaje de volúmenes enormes de datos.
- Recibir datos de múltiples fuentes: aparecen diferentes puntos de acceso a datos, con formas distintas de conexión, formatos, etc.
- Tasas de captura de información: en algunos casos, las fuentes no solamente generan un volumen muy grande, sino que lo hacen a velocidades desiguales a lo largo del tiempo, con picos enormes.
- Datos desestructurados: aparecen nuevas fuentes de datos que no aportan la información concreta semánticamente, sino que requieren un procesamiento previo para extraer su verdadero significado.

#### → EL *BIG DATA* EN ACCIÓN

La Liga Nacional de Fútbol Americano (NFL) dispone de una plataforma para ayudar con sus aplicaciones a los 32 equipos a tomar las mejores decisiones en los partidos en base a la analítica de datos: desde el estado de la superficie del campo a las condiciones meteorológicas, pasando por datos de la etapa universitaria de cada jugador, entre muchos otros factores.



El *big data* se define a menudo a través de sus características, que son resumidas como las tres uves: volumen, variedad y velocidad. Posteriormente, ha sido añadida una cuarta uve, veracidad.

# las cuatro uves

## Volumen

El *big data* supone aprovechar y explotar cantidades masivas de datos, si bien no hay una cantidad concreta que delimite lo que es *big data* de lo que no lo es. Además, los volúmenes de datos que somos capaces de procesar aumentan día a día a un ritmo sin precedentes.

## Variedad

A diferencia de los sistemas de almacenamiento de datos tradicionales, el uso de *big data* implica trabajar con datos de distintos tipos procedentes de fuentes diversas, que pueden ser estructurados, semiestructurados y no estructurados. Los datos pueden presentar innumerables formas, como texto, datos web, tuits, datos de sensores, audio, vídeo, secuencias de clic, archivos de registro, por mencionar unos pocos.

## Velocidad



La velocidad a la que se generan, procesan y analizan los datos continúa creciendo. Actualmente, los datos se crean continuamente a una velocidad a la que los sistemas tradicionales no pueden capturarlos, almacenarlos y analizarlos. Hay operaciones en tiempo real basadas en datos en las que el tiempo es un factor primordial, como la detección del fraude o el marketing instantáneo multicanal.

## Veracidad



La calidad de los datos se basa en su fiabilidad, y uno de los objetivos del *big data* es conseguir minimizar la imprevisibilidad que presentan ciertos tipos de datos y la incertidumbre asociada a ellos. Como ejemplos de datos que son inciertos se pueden mencionar los sentimientos y la sinceridad de las personas, las condiciones climáticas o la evolución futura de los sistemas económicos y financieros.



# cómo sacar valor de los datos

Tortura a  
los datos y  
confesarán  
cualquier cosa

El valor general que presenta el *big data* para una empresa u organización presenta varias facetas.

- Tiene la virtud de hacer transparente la información, y en consecuencia convertirla en una herramienta más accesible para ser utilizada con frecuencia.
- Al almacenar la empresa todo tipo de datos corporativos en formato digital, puede disponer de información más precisa y detallada sobre cualquier aspecto, desde los inventarios de los productos a las bajas de personal, y puede analizar los procesos y mejorar el rendimiento.
- El *big data* puede segmentar con mayor detalle a los clientes abriendo la posibilidad de ofrecer productos y servicios más personalizados.
- La analítica de datos más avanzada ayuda a mejorar el proceso de toma de decisiones.
- El *big data* puede utilizarse para mejorar el desarrollo de la siguiente generación de productos. Por ejemplo, hay empresas que introducen sensores en sus productos para poder ofrecer un servicio de mantenimiento proactivo, al detectar una avería o fallo antes de que se produzca o identifique.

## → EL *BIG DATA* EN ACCIÓN

Mabrian es una plataforma orientada al sector turístico que persigue comprender el comportamiento de los turistas y el mercado para poder segmentar adecuadamente la oferta y atraer a los clientes más interesantes.

## LA CADENA DE VALOR DEL DATO

### PROVISIÓN DEL DATO

Aprovisionamiento: identificar los conjuntos de datos necesarios.  
Captura y almacenamiento.

### TRANSFORMACIÓN

Perfilado: identificar posibles incoherencias en los datos.  
Limpieza: estandarización, normalización y eliminación de duplicados.  
Estructuración: dar forma a los conjuntos de datos.  
Enriquecido: mejorar el conjunto de datos con más información.

### DESCUBRIMIENTO Y MODELADO

Proceso de análisis por parte de los analistas de negocio  
y los científicos de datos.

### EXPOSICIÓN

Validación del proceso anterior e industrialización del mismo.

## → EL *BIG DATA* EN ACCIÓN

En 2002 el equipo de béisbol Oakland Athletics utilizó el análisis estadístico para contratar a su plantilla, cotejando las necesidades del equipo con los jugadores que mejor se adaptaban a ellas, aunque no fuesen muy destacados.

GARY KING

HARVARD UNIVERSITY

# aprendiendo a formular las preguntas adecuadas

El *big data*  
no va sobre  
números

No nos engañemos, disponer de millones de datos y de la tecnología adecuada para procesarlos no nos va a servir de nada si no sabemos qué queremos que nos cuenten. El *big data* consiste en que los datos ayuden a solucionar problemas de las empresas u organizaciones. Se trata de extraer valor de los datos, y la disciplina que se encarga de estos es la *data science* o ciencia de datos.

La *data science* se ocupa de plantear el problema que queremos que resuelvan los datos, es decir, de formular la pregunta adecuada. De esta forma, el proceso que se sigue comprende las siguientes fases:

- Plantear la pregunta que, de ser contestada, resuelve un problema de negocio.
- Estudiar las fuentes de datos relevantes para resolver el problema.
- Exploración y análisis de la información.
- Modelización de los datos: modelos matemáticos basados en millones de datos que establecen un patrón, por ejemplo, del comportamiento de los clientes.
- Comunicar las conclusiones a aquellos que toman las decisiones, traduciendo los resultados al lenguaje de negocio.

En 2012, un artículo de *Harvard Business Review* definió la de científico de datos como la “profesión más sexy del siglo XXI”. Se calcula que para 2020 un 40% de las empresas españolas tendrán incorporadas tecnologías *big data* en sus sistemas informáticos. Se trata de un campo que está cobrando una relevancia tal que la demanda de los perfiles profesionales asociados a él se ha disparado, creando una brecha entre las necesidades inmediatas de las empresas y una oferta de expertos insuficiente.

## las profesiones más sexys del siglo XXI

PUROMARKETING  
2017

Al 40% de las empresas les cuesta encontrar profesionales *big data*

EXPANSIÓN  
2017

Europa necesitará 8 millones de expertos en *big data* en 2020

CINCO DÍAS  
2018

Analista de datos, el puesto más demandado (y mejor pagado) por las empresas

LA VANGUARDIA  
2017

350.000 puestos sin cubrir en España por la falta de perfiles de *big data*

# quiénes trabajan los datos

Las actividades relacionadas con la economía de los datos están generando una demanda creciente de profesionales, y los perfiles más buscados son arquitecto *big data*, ingeniero *big data* y *data scientist* y visualizador datos.



## Arquitecto de *big data*

Se trata de la persona encargada de definir e implementar la arquitectura tecnológica que debe soportar un proceso de *big data*, desde la introducción de los datos, hasta la salida de información. A la hora de elegir una u otra solución tecnológica en cada caso, deberá tener en cuenta, por ejemplo, factores como si hay que capturar los datos en tiempo real, o el tipo de dato utilizado (no es lo mismo tratar un dato no estructurado, como un tuit, que uno totalmente estructurado, como los que contienen las bases de datos tradicionales). Debe, por tanto, conocer en profundidad las distintas soluciones tecnológicas disponibles.



## Ingeniero *big data*

También conocido como ingeniero de datos, a grandes rasgos es quien se ocupa de preparar los datos para que puedan ser analizados por el científico de datos. En concreto los identifica, limpia y clasifica para que puedan ser una fuente de información generadora de valor. Estos profesionales construyen el *software* necesario para integrar la información de los sistemas fuente en la plataforma *big data* de la empresa, de tal forma que quede preparada para su utilización. Se trata de un perfil que surge de la necesidad de liberar a los científicos de datos de la obligación de realizar tareas de preparación de los datos, de forma que solamente tengan que dedicar su tiempo a su estudio.

## Data scientist y visualizador de datos

El científico de datos es el encargado de convertir las ingentes cantidades de datos en información de utilidad para la empresa o la organización. Su cometido es analizar los datos, tanto cuantitativos como cualitativos, para sacar de ellos información relevante que apoye el proceso de toma de decisiones relacionadas con el negocio y los clientes.

Es importante que, aparte de los conocimientos técnicos que debe tener, conozca bien la empresa y la estrategia de negocio. Además, tiene que dominar el *storytelling* y las técnicas de presentación para poder contarle la información extraída de los datos a los responsables de tomar las decisiones.

Se trata de uno de los perfiles profesionales más difíciles de encontrar en el mercado. Ha sido definido como alguien mejor en estadística que cualquier ingeniero de *software*, y mejor ingeniero de *software* que cualquier estadístico.

### → EL BIG DATA EN ACCIÓN

En el campo de la seguridad, los sistemas predictivos Beware y PredPol pueden mapear los crímenes y altercados de un núcleo urbano antes de que se produzcan mediante el cruce de datos tales como la presencia de locales ruidosos en la zona, el tráfico, información del registro de propiedades o las matrículas de coches robados, entre muchos otros.

© Fundación Telefónica, 2018  
Gran Vía, 28. 28013 Madrid (España)  
<http://fundaciontelefonica.com/>

**Edita**

Fundación Telefónica

**Gerencia editorial**

Pablo Gonzalo Gómez

**Coordinador del proyecto y textos**

Pablo Rodríguez Canfranc

**Coordinación editorial**

Melisa Martínez Cíaurri

**Proyecto gráfico**

Lacasta Design



ISBN: 978-84-15282-38-9  
Depósito legal: M-35990-2018  
Impresión y encuadernación: CommerceGraf  
Primera edición: noviembre de 2018  
Impreso en España – Printed in Spain

Esta revista se ha impreso en papel reciclado Cyclus fabricado por Arjowiggins Graphic. Su uso ha reducido el impacto medioambiental en:

**29 kg de residuos**  
**713 litros de agua**  
**7 kg de CO2**  
**90 kWh de energía**  
**47 kg de madera**



¿Qué hace un científico de datos? ¿En qué se diferencia un desarrollador *backend* de uno *frontend*? ¿Quién es el *community manager* de la empresa? La revolución tecnológica ha traído consigo nuevas profesiones y ha cambiado por completo otras que ya existían.

**PROFESIONES DIGITALES** es una colección de monográficos de Fundación Telefónica que pretende dar a conocer aquellos perfiles profesionales más demandados por la economía digital.

Con un lenguaje sencillo y divulgativo, cada número acerca al lector una disciplina en la que desarrollan su actividad los trabajadores con los puestos más vanguardistas, ofreciendo de esta manera una breve orientación sobre nuevas áreas laborales.