

## ¿Qué es eso del 'big data'?

**Lo mencionan en conferencias, charlas y facultades. Aclaremos el concepto de moda. O lo que es lo mismo: lo que todas las empresas quieren saber de usted**

[Eva van den Berg 31 MAR 2015 - 12:28 CEST](#)



En los años noventa, el informático teórico estadounidense John Mashey publicó un artículo titulado [Big Data and the Next Wave of Infrastrress](#) (*Big Data y la próxima ola de Infrastrress*), popularizando el término que hoy nos ocupa. En él, Mashey hacía referencia al estrés que iban a sufrir las infraestructuras físicas y humanas de la informática debido al imparable tsunami de datos que ya se oteaba en el horizonte, inmanejable con los instrumentos de gestión al uso. Desde luego no se equivocaba. Hoy, recién iniciado el siglo XXI, se generan, según la [Unión Europea](#), 1.700 nuevos billones de *bytes* por minuto. Equivale, dicen, a unos 360.000 DVD, lo que de media vienen a ser seis *megabytes* por persona y día (más o menos la cantidad de datos que generaba en toda su vida una persona del siglo XVI). Pero esto no ha hecho más que empezar: en los próximos cinco años duplicaremos ese chorreo desahogado de dígitos binarios. Algunas cifras: cada día se realizan, por ejemplo, más de un billón de [consultas en Google](#), más de 250 millones de *tuits* en Twitter, 800 millones de actualizaciones en Facebook, 60 horas de vídeos subidos [por minuto en YouTube](#), 10.000 transacciones mediante tarjeta de crédito por segundo...

### **Cada clic es una confesión**

¿El objetivo? Predecir el tiempo, analizar parámetros de salud, mejorar la eficiencia energética y, sobre todo, vender más y mejor

La ingente cantidad de información que producimos procede de un sinfín de dispositivos que forman parte de nuestra vida cotidiana. Con ellos emitimos una retahíla

de datos que van conformando nuestros avatares digitales. Cada vez que *clicamos* en Amazon, ese gesto queda consignado. Cuando pagamos con la tarjeta la compra del supermercado, dejamos rastro de qué hemos comprado y a qué precio. Cuando realizamos las lecturas de los contadores de electricidad o del gas constatamos digitalmente el consumo que hacemos. Cuando opinamos en la Red, colgamos imágenes, realizamos compras online o utilizamos una *app*, el *big data* sabe un poco más de nuestros gustos. Lo mismo sucede cuando subimos a un avión, mandamos un paquete por mensajería, [accionamos el GPS del coche](#) o el sistema domótico de casa. O cuando llamamos a una empresa y nos dicen que, por seguridad, la llamada quedará grabada. O cuando nos recetan medicinas, nos ingresan la nómina o pagamos la hipoteca. ¿Se imaginan los ajustadísimos retratos que se podrían extraer de cada uno de nosotros optimizando el análisis de todos esos datos? De eso trata el *big data*. Es el término que afronta la labor de almacenar, clasificar, analizar y compartir ese cúmulo masivo de información. De lidiar con las denominadas “tres uves” del *big data*: gestionar un volumen de datos descomunal a la mayor velocidad posible considerando su extraordinaria variedad.

## ¿Los datos nunca mienten?

## El lado bueno del 'big data'

Pese a los riesgos, del *big data* saldrán infinitos beneficios para el común de los mortales. La lista de aplicaciones en funcionamiento es inmensa. Muchas se explican en el libro *The Human Face of Big Data*, de Rick Smolan y Jennifer Erwitte, o en su *app* derivada. Una gran cantidad de ellas funcionan gracias a los datos emitidos por las tarjetas SIM de los celulares. Hoy, la mayor parte de las personas, vivan en Suecia o en Sierra Leona, disponen de teléfono móvil, aunque sea de gama baja. Las tarjetas SIM permiten localizar a los usuarios, lo que, en combinación con imágenes de satélite, resulta de gran utilidad para seguir en tiempo real, por ejemplo, campañas de vacunación o descubrir pueblos enteros que no aparecían en el mapa y a los que las vacunas no podían llegar. O para localizar grupos de personas tras una catástrofe: en 2010, tras el terremoto de Haití, muchas personas huyeron de la capital y gracias a la interpretación de los datos pudieron ser ubicados y atendidos por los equipos de emergencia. Los dispositivos inteligentes también incrementan la predicción y prevención de las pandemias globales. ¿Y las ciudades inteligentes? Cada vez tendremos *cities* más *smart* como resultado de los artilugios conectados a la Red. Optimizaremos los problemas de tráfico, la red de transportes públicos, la gestión del agua y la energía, la seguridad ciudadana y muchas cosas más.

Según Viktor Mayer-Schönberger y Kenneth Cukier, autores del libro [Big Data: a revolution that will transform how we live, work and think](#), los peligros que se deducen de la gestión de esos datos son, lógicamente, los relacionados con la privacidad de las personas, en pleno proceso de regresión. Pero hay otros, como el robo de datos para causar daños a terceros o la tendencia a penalizar a personas que, sin haber cometido ningún delito, son, según la correlación de datos disponibles, propensos a cometerlos. ¿Recuerdan [Minority Report](#), la película en la que se planteaba una sociedad cuya policía era capaz de detener a los criminales justo antes de delinquir sobre la base de unas predicciones llamadas el “precrimen”? Todavía no hemos llegado tan lejos, pero sí que existe el riesgo de que pensemos que los datos nunca mienten y que sustituyamos las decisiones humanas por predicciones deterministas. Por ejemplo, ¿debería permitirse

que un dron decida por sí mismo, sin supervisión humana, solo en función de una correlación de datos y algoritmos, si debe o no “eliminar a un objetivo”?

Mayer-Schönberger y Cukier definen el *big data* como “la capacidad de la sociedad para asimilar la información mediante vías novedosas con el objetivo de producir conocimientos, bienes y servicios de valor significativo”. Pero mientras para unos un “valor significativo” puede ser frenar la propagación de una epidemia, para otros, lo es optimizar la operatividad de un cártel de narcotraficantes o asegurar el éxito de un ataque terrorista a gran escala. [Marc Goodman](#) hablaba de ello en su aterradora conferencia en TEDGlobal 2012. Goodman, experto en cómo las tecnologías amenazan la seguridad mundial, es consultor de entidades como la Interpol, la OTAN o la ONU, y dirige el Instituto de Crímenes Futuros, dedicado a sopesar la vulnerabilidad derivada de nuestra extrema conectividad.

Se recaban datos en tierra, en las profundidades del océano, en la atmósfera, en el espacio y dentro del cuerpo humano. Existe una minipastilla que es algo así como la versión actualizada del submarino miniaturizado que aparecía en la película *Viaje alucinante* (1966). Hoy, ese improbable sumergible tripulado se ha sustituido por un sensor ingerible (Ingestible Event Marker, IEM) diseñado por la empresa [Proteus Digital Health](#). Tras ser engullida, esta píldora del tamaño de un grano de arena envía señales eléctricas de alta frecuencia a través del cuerpo que son recogidas por un parche que el paciente lleva de forma externa o implantado bajo la piel. Los variados parámetros que controla son recogidos por su *smartphone* o desde el ordenador de su médico personal.

La combinación de datos, además, puede revelar hechos insospechados. Eso es lo que ocurrió cuando [la cadena Walmart descubrió](#), a partir del cruce de sus informes de ventas con los datos meteorológicos, que las galletas *Pop-Tarts* se vendían muchísimo tras la alerta de un huracán, ya que la gente hacía acopio de provisiones. Esa información inesperada provocó que la multinacional posicionara esas galletas de forma destacada cada vez que el hombre del tiempo alertaba de un fuerte temporal, lo que ha redundado en un importante aumento de las ventas. Otro ejemplo: hace un par de años la cadena Target envió a una adolescente de Minneapolis cupones de descuento para productos premamá y de bebé. El padre acudió al centro hecho una furia. Pero a los pocos días la chica confesó que, efectivamente, estaba embarazada. Había comprado complementos vitamínicos, toallitas sin perfume y otro tipo de productos que llevaron al programa informático de la tienda a considerarla una madre en potencia.

## **Minería de datos**

Los datos son el petróleo del futuro, el oro del mañana, afirman los expertos. Sirven para predecir tormentas, huracanes, terremotos, erupciones volcánicas o tsunamis. Para monitorear la evolución del cambio climático. Para conocer el estado de nuestros mares, discernir el comportamiento de los animales o realizar el seguimiento de especies amenazadas... Y por supuesto sirven para vender más y más de la forma más óptima y personalizada. Por ello ya han surgido nuevas disciplinas y profesiones destinadas a encontrar, entre esa ingente amalgama de datos, patrones ocultos y correlaciones inesperadas. Como la denominada “minería de datos”, que recluta especialistas capaces de bucear y descubrir esas conexiones, o a expertos que crean nuevos software para almacenar y gestionar esa avalancha digital. En eso anda Albert Bifet. Este investigador

formado en la [Universitat Politècnica de Catalunya](#) trabaja actualmente en el laboratorio asiático de referencia en lo que al *big data* se refiere: el [Noah's Ark Lab](#) de la empresa Huawei, en Hong Kong.

“Me apasiona trabajar con equipos que desarrollan nuevas herramientas de software libre para analizar el *big data* en tiempo real, sobre la base de los datos procedentes de las redes sociales y de la telefonía fija y móvil”, explica. Este ingeniero participó primero en el proyecto MOA en Nueva Zelanda, luego en el SAMOA, originario de los [Yahoo Labs](#) de Barcelona, y ahora implementa este nuevo proyecto para Huawei. “El *big data* permitirá, entre otras cosas, hacer recomendaciones cada vez más precisas, basadas en el análisis de nuestro rastro y en la comparación con el de otros usuarios”, explica. Las empresas podrán ofrecernos los productos que mejor nos encajan: libros, películas, ropa, música, viajes... cualquier cosa que nuestro terminal *smart* nos propondrá acertadamente. Somos fábricas de datos andantes. “La infinidad de *wearables*, esos dispositivos electrónicos llevables que se incorporan al cuerpo o en la ropa para que interactúen con el usuario y con otros dispositivos, son una fuente inagotable de datos y por tanto, una fuente de innumerables funcionalidades, como también lo son todos los objetos que tenemos en casa conectados a la Red”, añade Bifet.

## Objetos conectados

En ese uso de Internet a nivel doméstico ha encontrado su nicho profesional la empresa española [Green Momit](#), que diseña y fabrica dispositivos que se adaptan a los hábitos de consumo energético del consumidor. Con los datos recabados, pueden automatizar el comportamiento de los sistemas de control de energía para mejorar su eficiencia y el confort de los hogares, a la vez que comparan sus hábitos con los de otros usuarios. Además, ofrecen a las empresas la posibilidad de consultar esos datos para que puedan optimizar sus sistemas de producción, predecir comportamientos en su red y prever demandas de consumo de energía. “Nuestro primer producto es un termostato Wi-fi inteligente que puede ser controlado desde cualquier dispositivo móvil y permite ahorrar hasta un 30% en la factura energética”, dice Eduardo Rodríguez, cofundador junto a Miguel Sánchez de Green Momit, la iniciativa empresarial con la que ganaron la edición española del [SmartCamp de IBM](#) el pasado mes de noviembre.

Se trata de la revolución de los objetos conectados. La dependencia de la tecnología. La fusión de los humanos con los dispositivos. ¿Podremos, sabremos, querremos vivir sin ellos? Todo apunta a que no, no y no. Dependemos de los omnipresentes *devices* a todos los niveles. ¿Incluso a nivel emocional? Puede que sí. Quizá descubramos, finalmente, que nuestra media naranja es, precisamente, uno de estos dispositivos tan *supersmart*. Como en la película *Her*, de Spike Jonze, donde el protagonista (Joaquin Phoenix) adquiere un nuevo sistema operativo llamado Samantha al que se acaba engancho. Ella, que no tiene entidad física, es solo una voz (la de Scarlett Johansson) que se erige como su asistente personal, su consejera, su amiga, su compañera, su hiperculta interlocutora. ¿Su amor? Quién sabe.

Crucemos los dedos para que lo que empezó como una fantasía utópica no acabe por ser una distopía total. El tiempo y los datos lo dirán... Mientras tanto, cuidado. En EE UU, uno de cada tres recién nacidos tiene presencia online desde antes de nacer (a través de su ecografía) y el 92% de los niños de dos años tiene ya su pequeño avatar digital, esculpiéndose *non stop* a base de las informaciones que los eufóricos papás cuelgan de

forma desahogada. Nuestros avatares crecen con el tiempo, haciéndose cada vez más completos, más robustos, más precisos. Y han llegado para quedarse. Puede que para siempre.