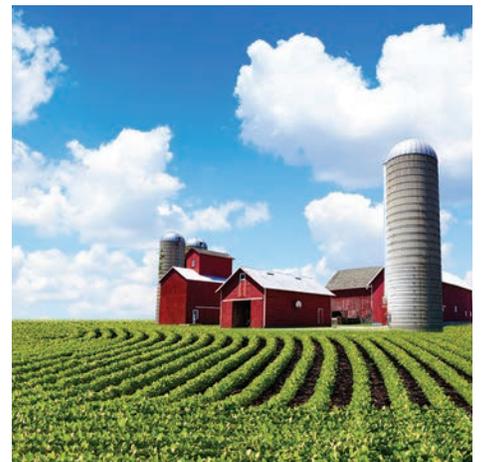
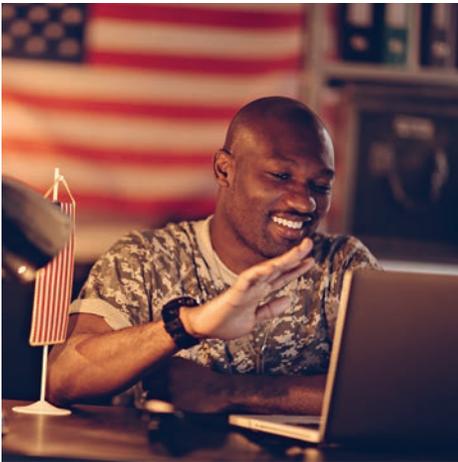


The  
Software  
Alliance

BSA

# ¿Por qué son tan importantes los datos?



# RESUMEN EJECUTIVO

La innovación del software continúa produciendo avances sin precedentes que transforman el mundo que nos rodea, les brindan autonomía a las personas y permiten el crecimiento de nuestras economías.

Sin embargo, esta transformación digital solo puede brindarnos su máximo potencial si explotamos el poder de los datos que estas innovaciones liberaron. En efecto, estamos atravesando una época de revolución de los datos. Esto es impulsado no solo por la abundancia de datos actual, sino por las tecnologías fundamentales que cambian la forma en que reunimos, almacenamos, analizamos y transformamos la información.

# Hoy, el 90 % de los líderes de negocios citan a los datos como uno de los recursos clave y un factor distintivo fundamental para los negocios, a la par de recursos básicos como las tierras, la mano de obra y el capital.

**No hace mucho, por ejemplo, la recopilación de datos requería observar patrones meteorológicos durante cientos de años para discernir los ciclos de precipitaciones. Era como sentarse junto a una calle para registrar la velocidad del tráfico a fin de planificar las redes de transporte. Implicaba reunir millas de carpetas de manila llenas de notas escritas a mano para estudiar de qué manera evolucionaban y se podían curar las enfermedades.**

Ahora, los datos son generados por sensores que se encuentran en millones de dispositivos, máquinas, vehículos e, incluso, las luces de la calle. Si bien mantener esta cantidad de datos fue alguna vez costoso y difícil, las capacidades de almacenamiento crecieron y los costos cayeron, y así los datos almacenados son ahora un recurso renovable. Con esta capacidad de reutilizar los datos y darles nuevos propósitos, podemos continuar con su análisis y transformarlos en nuevas formas de producir conocimientos valiosos que permitan ahorrar tiempo y dinero e, incluso, salvar vidas.

Algunos de estos datos capturados constituyen información personal y, por este motivo, se deben emplear los métodos de seguridad más vanguardistas así como modelos de administración responsables para garantizar que esta información esté segura y se utilice correctamente. Pero la gran mayoría de los datos proviene de muchos dispositivos y máquinas que se envían información entre sí y les envían información a quienes los operan. Desde la línea de ensamblaje de la planta de producción al avión de pasajeros en vuelo, se generan millones de bytes de datos que luego se analizan. Hacer esto mejora el rendimiento y aumenta la productividad en formas que en algún momento eran inimaginables.

Si bien los datos se encuentran en todas partes y su omnipresencia y utilidad están mejorando nuestras vidas de diversas maneras, muchas personas no comprenden qué son, de dónde provienen, cómo se pueden utilizar ni el potencial enorme que tienen por naturaleza.

El presente documento presenta a grandes rasgos algunos ejemplos concretos de cómo la innovación de los datos está produciendo un avance extraordinario respecto de algunos de los desafíos más complicados a escala mundial. Describe cómo los cambios fundamentales en la forma en que los datos se reúnen, almacenan, analizan y transforman nos ubican al borde de todo lo que es posible en nuestra economía digital del siglo XXI y más. También

aborda algunos de los mitos que se han instalado a medida que la gente continúa esforzándose por comprender plenamente la industria de la innovación de los datos en expansión. Para finalizar, el documento ofrece un glosario de términos que define el lenguaje de la innovación de los datos y sirve como una guía para aquellos que sean nuevos en cuanto a la comprensión de la economía de los datos.

La oportunidad que la innovación de los datos le ofrece al mundo es virtualmente incomparable. Las innovadoras herramientas de software ya están revolucionando nuestras vidas de maneras sorprendentes; ahora, estas herramientas están ayudando a las personas a descubrir las respuestas escondidas en una cantidad enorme y creciente de recursos de datos. Estas nuevas y transformadoras herramientas están convirtiendo los datos en nuevos productos, soluciones e innovaciones que existen para cambiarnos la vida. Desde una perspectiva económica, un mejor uso de los datos podría conducir a “dividendos de datos” de \$1,6 billones en los próximos cuatro años solamente. Los economistas calculan que las ganancias por la eficiencia que permiten los datos podrían sumarle casi \$15 billones al PIB global para el 2030.

Si realizamos elecciones inteligentes hoy, esta economía “centrada en los datos” que está surgiendo podría convertirse en una poderosa fuente generadora de empleos e industrias nuevas, avances nuevos y curas nuevas —y alimentará el crecimiento de la economía durante décadas.

## LA “INNOVACIÓN DE LOS DATOS” DEFINIDA

Se ha escrito bastante acerca de las “cuatro V” de la innovación de los datos: volumen, la cantidad de datos; velocidad, la rapidez con que se crean; variedad, los tipos de datos involucrados; y veracidad, su precisión. Sin embargo, se ha dedicado menos tiempo a analizar el poco valor que yace en los datos sin procesar —y la oportunidad para cambiar las reglas del juego que todos compartimos a fin de maximizar verdaderamente su uso.

Según lo examinado por el presente documento, los datos se deben reunir, almacenar, analizar y transformar para brindar beneficios que pueden ser prácticos y hasta salvar vidas. Estos procesos se encuentran en el centro de la innovación de los datos —la derivación de un valor inmenso a partir de cantidades enormes de información que es, de otro modo, improductiva.

# ÍNDICE

- 5** INTRODUCCIÓN
- 7** TRANSFORMAR LOS DATOS EN RESPUESTAS
- 14** UNA ECONOMÍA IMPULSADA POR LOS DATOS
- 17** SEPARAMOS LOS MITOS SOBRE LOS DATOS DE LOS HECHOS
- 28** DISCURSO DIGITAL: COMPRENSIÓN DEL LENGUAJE DE LOS DATOS
- 34** NOTAS FINALES
- 40** ACERCA DE BSA



6000 AC

Siglo 15



Década de 1850

Siglo 21



MILESTONES DE DATA

# INTRODUCCIÓN

A lo largo de la historia de los seres humanos, los hitos de la civilización estuvieron marcados por avances en nuestra capacidad para observar y reunir información. Nuestros ancestros desarrollaron herramientas para medir la distancia, el peso, el volumen, la temperatura, el tiempo y el lugar —cada una fue mejorando con el tiempo y cada una fue fundamental para la transición de cazadores y recolectores a agricultores y a residentes de ciudades.

Ya en el 6000 A.C., se utilizaron los datos del rendimiento de las cosechas y los ciclos de barbecho para incrementar la producción agrícola y alimentar a más gente. En el siglo XV, se utilizaron los datos del firmamento para navegar por el mundo y abrir los profundos mares al comercio global. En la década de 1850, se utilizaron los datos para relacionar los brotes de cólera con el mal clima y así salvar vidas.

A través de la historia moderna, incluso cantidades limitadas de datos nos han brindado perspectivas clave para encontrar soluciones inesperadas a algunos de nuestros más grandes desafíos. Independientemente de que estén grabados en una estela, un rollo de papiro, un volumen ilustrado o un libro impreso, los datos —y su creciente dominio e importancia— han sido factores clave de impulso del avance económico y humano.

En el siglo XXI, estamos experimentando un aceleramiento de este proceso. A medida que los datos empiezan a abundar más y su costo de almacenamiento baja, las nuevas tecnologías les están proporcionando a los científicos de los datos herramientas de vanguardia que dejan al descubierto valiosos conocimientos a partir de enormes cantidades de datos. A medida que esas tecnologías de procesamiento de datos adquieren características más transformadoras, sus efectos son más profundos y las oportunidades incluso más generalizadas.

Nos dirigimos a un mundo de información y posibilidades casi ilimitados. Se debe analizar cómo se están utilizando los datos para realizar predicciones que mejoren nuestras vidas cotidianas. Los datos predictivos nos ayudan a saber

con anticipación si traer un paraguas al trabajo o tomarnos el autobús. Los datos del tráfico se utilizan para sincronizar los semáforos, predecir los horarios de llegada de los trenes y ayudarnos a encontrar la vía más rápida para llegar al ensayo de nuestro hijo a tiempo. Los dispositivos que se pueden usar como accesorios nos ayudan a realizar un seguimiento de nuestro estado físico, de modo que podamos elegir opciones para tener vidas más largas y saludables, y los científicos están analizando terabytes de información genética para encontrar curas nuevas y desarrollar tratamientos más efectivos y personalizados.

## LOS DATOS MARCAN UNA DIFERENCIA

- + **Barcelona** está sacando provecho de los datos para construir una ciudad más inteligente, y darle la capacidad de examinar las pautas de tráfico de los turistas, ver dónde poner más estaciones de bicicletas públicas e identificar qué esquinas de la ciudad necesitan más cajeros electrónicos.
- + En los **Emiratos Árabes Unidos**, las nuevas herramientas de datos se están utilizando para diseñar el primer edificio de energía positiva del mundo que efectivamente produce más energía de la que consume.
- + En **Kenia**, se están utilizando los datos móviles para identificar los patrones de la infección de la malaria e identificar puntos de acumulación de casos que guíen al gobierno en sus esfuerzos de erradicación.
- + Los agricultores desde **Iowa** hasta la **India** están utilizando datos de semillas, satélites, sensores y tractores para tomar mejores decisiones acerca de qué cultivar, cuándo plantar, cómo hacer un seguimiento de la frescura de los alimentos de la granja a la mesa y cómo adaptarse a los cambios en el clima.

ANÁLISIS

TRANSFORMACIÓN

RECOLECCIÓN

ALMACENAMIENTO  
Y TRADUCCIÓN

Al comprar un automóvil, en lugar de brindarnos acceso solo al precio que figura en su etiqueta, los datos nos permiten ver en profundidad el consumo de combustible del vehículo, el mantenimiento, el seguro y los registros de seguridad para ayudarnos a tomar decisiones más informadas. Y el propio auto es ahora, en efecto, una supercomputadora sobre ruedas. Posee un procesador que interactúa con sensores que analizan el rendimiento, de modo que los conductores puedan estar informados acerca de cuándo deben cambiar el aceite, cuándo cambiar a un motor eléctrico o si hay un niño jugando en la entrada del auto mientras el vehículo retrocede.

Ya la creciente abundancia de datos deposita poder en nuestras manos al ponernos la información que tanto necesitamos en la punta de los dedos.

¿Pero qué son exactamente los “datos”? ¿Quién o qué los genera? ¿Qué posibilidades tienen de mejorarnos la vida? ¿Cómo hay que usarlos para obtener un máximo beneficio? ¿Y cómo podemos asegurarnos de usarlos de maneras que sean compatibles con nuestros valores e inquietudes?

Estas preguntas son importantes porque, dada su transición de un recurso que en algún momento fue escaso a un recurso cada vez más abundante, valioso y renovable, los datos se están convirtiendo en una fuente importante de beneficios económicos y sociales. Tradicionalmente, fue el acceso a recursos como las tierras, la mano de obra y el capital el que representaba una diferencia económica entre quienes tenían éxito y quienes fracasaban. Hoy, el 90 por ciento de los líderes de negocios citan a los datos como uno de los recursos clave y un factor distintivo fundamental para los negocios, a la par de recursos básicos como las tierras, la mano de obra y el capital.<sup>1</sup>

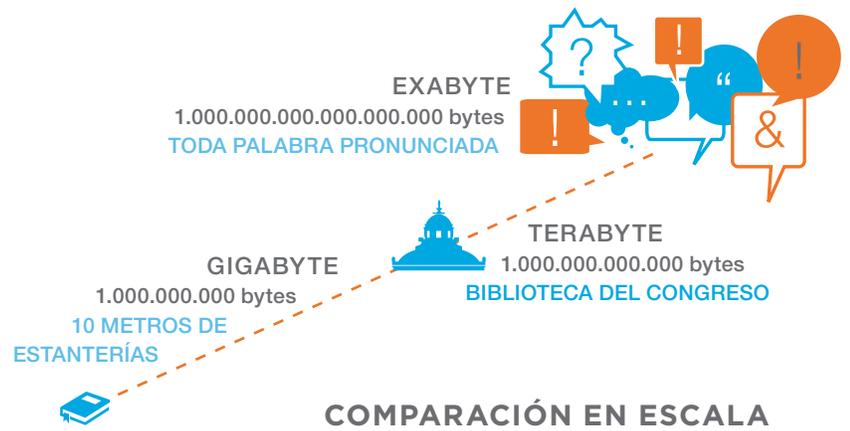
Un ejemplo: los economistas estiman con cautela que si el mejor aprovechamiento de los datos ha producido pequeñas ganancias que permitieron que las industrias fueran un 1 por ciento más eficientes, esto le sumaría cerca de \$15 billones al PIB global para el 2030.<sup>2</sup> La “próxima gran tendencia” puede provenir de los miles de millones de pequeños elementos conectados a Internet que producen datos cada vez mejores acerca del mundo que nos rodea y pueden generar soluciones aún más efectivas a partir de los datos.<sup>3</sup> Ya estamos encontrando respuestas a preguntas que ni siquiera sabíamos que teníamos.

Este enorme cambio se encuentra en marcha.

Prácticamente todo lo que hacemos genera datos, y cada día se crean caudales completamente nuevos de datos. De hecho, el 90 por ciento de los datos actuales del planeta se crearon solo en los dos últimos años, y ahora estamos duplicando la velocidad con que se producen los datos cada dos años. La mayoría de estos datos que se generan no son personales. Esta distinción es importante porque, si bien es crucial que protejamos la privacidad, la mayoría de las veces los datos que ayudan a mejorar nuestras vidas son generados por un sensor conectado a una máquina.

Nuestro desafío es sacar provecho de los datos y ponerlos a funcionar, valiéndonos de nuestra ingenio para entender los valiosos aprendizajes que guardan. Esta capacidad de procesar los datos y transformar las observaciones en conocimientos, y los conocimientos en respuestas, es la que nos permite obtener soluciones de verdadero significado a los considerables desafíos de la actualidad.

Las empresas y los gobiernos ahora deben trabajar activamente para hacer arrancar el motor de la innovación.



## LOS CUATRO PASOS CLAVE PARA TRANSFORMAR LOS DATOS EN RESPUESTAS

La revolución de los datos actual no es impulsada exclusivamente por la enorme y creciente cantidad de datos; es alimentada por tecnologías fundamentales que cambian la forma en que reunimos, almacenamos, analizamos y transformamos los datos. Juntos, estos factores de impulso nos permiten lograr percepciones claras y poderosas desde la profundidad de los datos y así extraer nuevos conocimientos, descubrir nuevas conexiones y realizar nuevas predicciones.

### 1

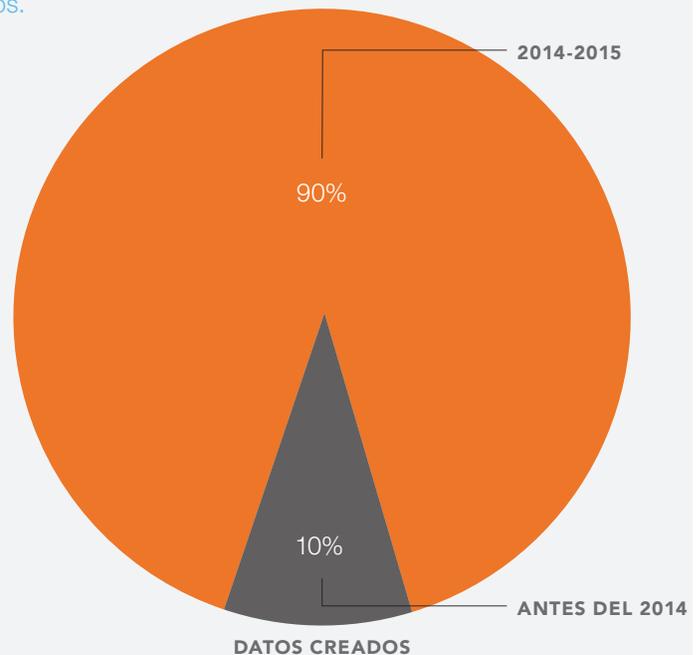
#### LA RECOPIACIÓN DE LOS DATOS

A lo largo de la historia, siempre hemos reunido datos y los hemos utilizado para colaborar con el avance de la sociedad. Pero muchas veces, los datos eran demasiado escasos. Hoy, tenemos la suerte de poder extraer más datos del mundo que nos rodea —datos cargados de mayor significado, recopilados en formas más útiles y que generan resultados más deliberados. Los datos dejaron de ser un recurso escaso para convertirse en un recurso fundamental, renovable y cada vez más abundante.

Esto fue posible no por nuestra capacidad de concentrar y extraer información personal; sino gracias a nuestra capacidad para conectar varios dispositivos y sensores a Internet, lo que hoy genera una gran cantidad de datos nuevos a velocidades exponenciales. Como resultado, los datos se producen en todos lados —mediante sensores en las calles para medir el flujo del tráfico; mediante la música

#### CRECIMIENTO ACTUAL DE LOS DATOS EN TODO EL MUNDO

El 90 % de los datos del mundo se crearon en los últimos dos años.



Fuente: IBM  
<http://www.ibm.com/software/data/bigdata/what-is-big-data.html>

digital y las películas que creamos; mediante los satélites que giran en el cielo; mediante los sensores y sistemas que controlan nuestras fábricas y mercados financieros; y mediante las herramientas que utilizamos para diseñar digitalmente la próxima novedad. Esta información está aumentando más rápido, llegando más lejos y creciendo en cuanto a su importancia.

Se calcula que ya se generan 2,5 cuatrillones de bytes de datos al día.<sup>4</sup> A nuestras mentes de la era analógica les cuesta incluso pensar en la enormidad de estos datos. Para tener un contexto, el año pasado el mundo creó suficientes datos digitales como para formar una pila de DVD que vaya desde la Tierra hasta la luna y vuelva.<sup>5</sup> Y también se está acelerando el ritmo con el que creamos

# La mayoría de los datos no cuenta con identificación personal. Pronto, los dispositivos que están conectados alrededor del planeta nos ayudarán a comprender claramente y mejorar nuestro entorno.

los datos. El volumen de los datos comerciales a escala mundial, en todas las empresas, se está duplicando ahora cada 1,2 años.<sup>6</sup> ¿De dónde vienen todos ellos? A continuación se presentan algunos ejemplos de fuentes, entre muchas:

- + Se espera que la información digital de los hospitales, mayormente de los exámenes clínicos de diagnóstico por imagen, trepe 665 terabytes por día para el 2015 —y ayude a encontrar curas y salvar vidas.<sup>7</sup>
- + Los aviones de las aerolíneas transcontinentales modernas están tan cargados de sensores en los motores, alerones y mecanismos de aterrizaje que pueden generar medio terabyte de datos por vuelo para mejorar el rendimiento del vuelo,<sup>8</sup> reducir las turbulencias, mejorar la seguridad e identificar posibles defectos en los motores, y hacerlo 2000 veces más rápido que antes.<sup>9</sup> Si multiplicamos eso por los más de 25.000 vuelos que se realizan por día logramos tener una idea de la inmensa cantidad de datos útiles que generan ahora los aviones comerciales solamente.
- + Los satélites meteorológicos, los observatorios meteorológicos, los radares y otros sensores capturan más de 2,25 mil millones de datos del clima 15 veces por hora —y recopilan 20 terabytes por día— lo que permite que se realicen pronósticos del tiempo más exactos en todo el planeta.<sup>10</sup>
- + Los mercados financieros generan entre cuatro y cinco terabytes de datos por día que se utilizan para realizar análisis en tiempo real y detectar problemas en las actividades comerciales, y simultáneamente ayudan a que los negocios crezcan y la economía sea más próspera.<sup>11</sup>
- + Los sensores telemáticos ubicados en decenas de miles de vehículos de reparto llevan un registro del rendimiento del motor, mejoran el recorrido y prevén los problemas de manera anticipada. Los datos del sensor del vehículo combinados con los análisis de asignación de datos les han permitido a las empresas ahorrar millones de galones de combustible y reducir las emisiones en una proporción equivalente a la de quitar de circulación miles de automóviles durante un año.<sup>12</sup>
- + El Gran Colisionador de Hadrones (Large Hadron Collider) de la CERN, la Organización Europea para la Investigación Nuclear, genera 40 terabytes de datos por segundo con cada experimento, y permite tener una percepción nueva y clara de los secretos más profundos relativos a la forma en que funciona el universo.<sup>13</sup> De manera similar, el Gran Telescopio para Rastreo Sinóptico de Chile genera 30 terabytes de datos sobre nuestro universo al observar el firmamento cada noche.<sup>14</sup>
- + El secuenciamiento de un solo genoma de ADN puede generar 200 gigabytes de datos. A medida que el costo del secuenciamiento del ADN baja, los científicos están construyendo bases de datos inmensas llenas de cientos de miles de estas secuencias con el objeto de encontrar las diferencias y similitudes que se correlacionen con los avances médicos y poder así salvar vidas.<sup>15</sup>

No solo la cantidad de datos está aumentando de manera exponencial, sino también las formas en que se producen. A medida que crece la cantidad de dispositivos que conectan Internet con el mundo que nos rodea, y dan lugar así a la “Internet de las cosas”, existe una infinidad de sensores que crean formas completamente nuevas de datos cada día. La próxima gran tendencia puede basarse en muchos elementos pequeños, puesto que se estima que 50 mil millones de dispositivos llenos de poderosos sensores estarán conectados a Internet para el 2020.<sup>16</sup>

Estos dispositivos crearán datos mediante tareas como la medición de la humedad del suelo, el rendimiento de los motores, la eficiencia del sistema energético y la localización de las crisis asmáticas. Los seres humanos contamos simplemente con cinco sentidos para comprender el mundo donde vivimos. Pronto, los dispositivos que están conectados alrededor del planeta detectarán un conjunto completo de características acerca del mundo físico para ayudarnos a comprender claramente y mejorar nuestro entorno —y, mientras tanto, producen exabytes de datos nuevos y beneficiosos.

Puesto que el costo de almacenamiento de los datos sigue en picada y la cantidad de datos continúa creciendo, los usos de los datos siguen expandiéndose.

## 2

### EL ALMACENAMIENTO DE LOS DATOS

La disminución de los costos de almacenamiento está permitiendo la innovación impulsada por los datos. En 1980, era difícil encontrar espacio de almacenamiento de un gigabyte de datos, esto costaba cientos de miles de dólares y requería de la administración de una persona a tiempo completo.<sup>17</sup> Hoy, un gigabyte de almacenamiento cuesta solo centavos, se administra fácilmente y permite el acceso en cualquier momento, desde cualquier lugar.<sup>18</sup> Desde la década del ochenta, el precio de almacenamiento cayó por más de un factor de 10 millones.<sup>19</sup> Para ponerlo en contexto, si el precio de la gasolina hubiera bajado de la misma manera, uno podría conducir un auto alrededor del mundo casi 10.000 veces con lo que se pagaba por un galón de gasolina en 1980.<sup>20</sup>

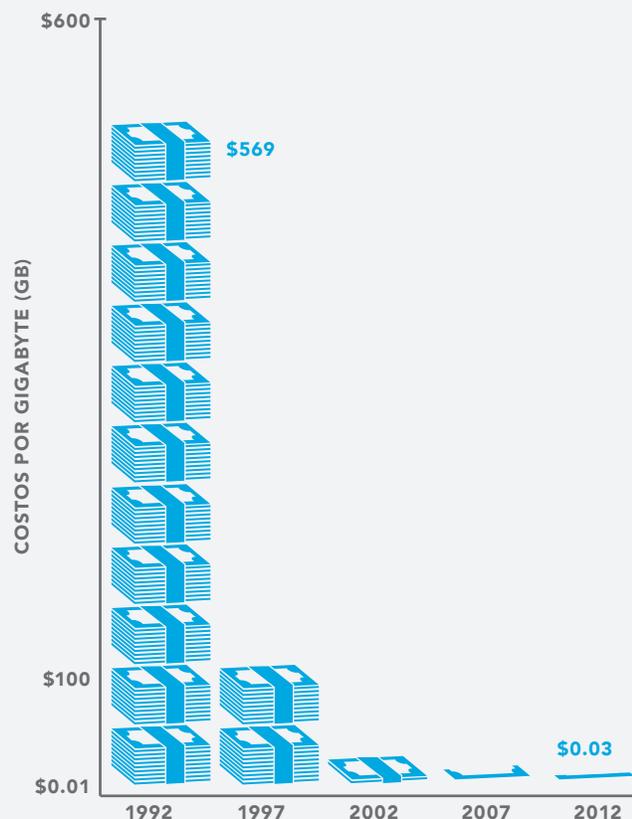
Dado que el costo de almacenamiento continúa bajando, podemos almacenar cantidades cada vez mayores de datos. En 1994, solo el 3 por ciento de los datos del mundo se almacenaban digitalmente.<sup>21</sup> Para 2007, el 94 por ciento se almacenaba digitalmente.<sup>22</sup>

Puesto que el costo de almacenamiento de los datos sigue en picada y la cantidad de datos continúa creciendo, los usos de los datos siguen expandiéndose. Las tecnologías en la nube —que permiten el almacenamiento remoto de los datos y el acceso mediante una serie de dispositivos— provocaron una reducción drástica en el precio de almacenamiento de los datos, de modo que ya no es necesario eliminar los datos después de su uso inicial para que haya más lugar para datos nuevos.<sup>23</sup>

Como resultado, y a diferencia de otros recursos, los datos no se agotan luego de ser utilizados una vez. Los datos constituyen un recurso renovable que se puede combinar con otros conjuntos de datos y utilizar muchas veces para generar respuestas a preguntas que no podían preverse al momento de la creación de los datos. Por ejemplo, los datos meteorológicos no se utilizan únicamente para predecir si necesitaremos llevar un paraguas; también puede ayudar a predecir el rendimiento de los cultivos.

Las tecnologías en la nube les ofrecen a los usuarios un acceso mejor, más confiable, más rentable y más flexible a sus datos —esto alivia la presión que las inmensas cantidades de datos pueden ejercer sobre la infraestructura interna de los departamentos de sistemas. Al haber transformado desde los cimientos la forma en que el almacenamiento de los datos se compra, vende y brinda —y al permitir que los datos estén disponibles prácticamente en cualquier lugar y en cualquier momento— la tecnología en la nube se está posicionando como una de las tecnologías más transformadoras de la década, y una de las grandes facilitadoras de soluciones generadas a partir de los datos.

#### DISMINUCIÓN DE LOS COSTOS DE ALMACENAMIENTO 38 % AL AÑO



Fuente: Hagel III, John et al. *From Exponential Technologies to Exponential Innovation*. Deloitte University Press, 2013. Publicación. 2013 Shift Index Series.



## DATOS EN TIEMPO REAL: REDUCCIÓN DEL TRÁFICO

# 3

## EL ANÁLISIS DE LOS DATOS

Los datos tienen valor solo cuando son comprensibles; de otra forma, no son más que un montón de observaciones aleatorias. Solo se puede lograr la comprensión de los conocimientos que contienen los datos si se combinan el ingenio humano con un software innovador.

Si bien el mundo es cada vez más autónomo, siguen siendo necesarias la curiosidad personal, las habilidades humanas y el trabajo intenso para liberar las respuestas del interior de los datos.

Primero, para que puedan ser útiles, es necesario limpiar los datos binarios. Según un cálculo aproximativo, los científicos de los datos pueden pasar entre el 50 y el 80 por ciento de su tiempo preparando rebeldes datos digitales antes de que puedan ser explorados para encontrar piezas útiles.<sup>24</sup>

Segundo, se necesita de la creatividad humana para formular las preguntas correctas y, luego, encontrar las repuestas mediante la revisión y el reconocimiento de los datos incorrectos, y la interpretación coherente de los resultados. El papel de los científicos de los datos se ha descrito como una función con tres partes: la de analista, la de artista y la de narrador de historias.<sup>25</sup> Cada dato individual es como un píxel en una pantalla. Solo, brinda únicamente una pequeña cantidad de información. Pero cuando se lo combina con una cantidad suficiente de píxeles en el orden correcto, el científico de los datos puede pintar un cuadro que vale mil palabras y derivar significados nuevos, y a veces inesperados, de los datos.

Al examinar cuidadosamente los datos, las herramientas analíticas pueden atravesar el conjunto de datos desordenados para ayudar a los usuarios a descubrir pautas y tendencias nuevas, encontrar ideas inesperadas a partir de datos aparentemente inconexos y descubrir automáticamente relaciones interesantes desde el punto de vista estadístico. Mediante bases de datos de creciente riqueza y algoritmos estadísticos cada vez más avanzados, las herramientas analíticas de software nos permiten

examinar cuidadosamente montañas de datos para encontrar pepitas de información de oro.

Afortunadamente, los recursos de procesamiento más poderosos de las computadoras actuales combinados con software inventivo les están brindando a los científicos de los datos herramientas de vanguardia para comprender cantidades inmensas de datos y descifrar los valiosos conocimientos que contienen.

Si bien las redes actuales son impresionantes, mover cantidades enormes de datos por las redes hasta un lugar para poder procesarlos a todos al mismo tiempo es, a menudo, prohibitivo en términos económicos e imposible en términos de logística. Aun así, algunos de los motores analíticos más poderosos de hoy pueden existir y ser rentables gracias a grandes plataformas informáticas en la nube de distribución en paralelo. Estas plataformas permiten que los usuarios ejecuten herramientas de análisis de datos de primera categoría con datos almacenados en varios lugares al mismo tiempo.

¿Qué nos permite hacer este análisis de los datos? Predecir el futuro solía parecer inverosímil, pero ahora parece inevitable. Hoy, gracias a los análisis de datos modernos, realizamos predicciones confiables todo el tiempo. Los pronósticos del tiempo son más confiables incluso para los próximos 10 días. Los gerentes de flota pueden predecir qué motores necesitan reparación antes de que el auto sufra una avería. Cuando los datos del presente se pueden comparar con los del pasado, muchas veces se pueden usar para ayudar a predecir el futuro.

Los economistas están encontrando formas de realizar mejores previsiones de los mercados, el empleo y la inflación. Durante mucho tiempo, los datos económicos del gobierno obligaron a los encargados de tomar decisiones a mirar por el espejo retrovisor. Las estadísticas económicas del gobierno, como el crecimiento del PIB, siempre miraron los meses en retrospectiva para decirnos, después de una larga pausa, la forma en que se comportaron



IBM y la ciudad de Estocolmo se unieron para instalar 1.600 sistemas de GPS en taxis. Un software de transmisión de datos de IBM recolecta los datos de los dispositivos GPS y brinda información sobre el flujo del tráfico, los tiempos de viaje y las mejores rutas.

Se redujo la cantidad de emisiones en un

**10 %**

Se redujo el tráfico en la ciudad en un

**20 %**

Se redujo el promedio de tiempos de viajes en casi un

**50 %**

La proporción de vehículos ecológicos que no pagan impuestos aumentó un

**9 %**

Fuente: Bertolucci, Jeff. 'Dublin Points Big Data Tech At Traffic Jams'. InformationWeek 2013. Internet. Nusca, Andrew. 'Stockholm Uses Real-Time GPS Data To Manage Traffic Congestion ZDNet, 2010. Internet.

las economías en el pasado como el mejor punto de referencia para el futuro. Ahora, los economistas están combinando una variedad de datos en tiempo real, como nuevos puestos de empleo y órdenes de la industria, y los están comparando con los datos históricos para trazar un panorama más exacto de las dinámicas actuales, y formular mejores políticas para garantizar economías saludables.

El surgimiento de los análisis de datos en tiempo real está permitiendo también una toma de decisiones autónoma para ayudarnos a nosotros, o a las máquinas que utilizamos, a tomar decisiones mucho más rápido y con mayor precisión. Ya muchas empresas automotrices de Estados Unidos están diseñando vehículos cargados con cientos de sensores, telemática y conectividad en tiempo real para permitir los grandes avances logrados en el estacionamiento autónomo. Los fabricantes de autos están potenciando las herramientas de análisis en tiempo real que permiten evitar choques de manera autónoma y desarrollar autos que se conducen solos. Dichos avances podrán, algún día, salvar vidas con una reacción más rápida y confiable ante las situaciones en comparación con lo que podemos hacer los seres humanos.

Al tener a su disposición una cantidad de datos sobre nuestro entorno en tiempo real que crece de manera exponencial, aquellos que pueden comprender lo que averiguan a la misma velocidad que lo averiguan podrán maximizar el efecto de las herramientas de

**Los recursos de procesamiento más poderosos de las computadoras actuales combinados con software inventivo están brindando a los científicos de los datos herramientas de vanguardia para comprender cantidades inmensas de datos y descifrar los valiosos conocimientos que contienen.**

análisis de datos. El poder de las mejores herramientas de la actualidad yace en su capacidad para establecer correlaciones nuevas y encontrar respuestas inesperadas que están bien ocultas en los datos —incluso cuando la gente no sepa cuál es la pregunta correcta que debe hacer. En todo el planeta, las herramientas de análisis están encontrando correlaciones de gran impacto y produciendo resultados inesperados. Por ejemplo:

- + Mediante el seguimiento y la correlación de más de 1000 datos por segundo, los investigadores canadienses sorprendieron a los médicos al demostrar que los bebés prematuros con signos vitales extrañamente estables se correlacionaban con cuadros de fiebre serios al día siguiente —esto permitió que los médicos tomaran medidas preventivas.<sup>26</sup>
- + Se están utilizando dos décadas de noticias pasadas de periódicos para predecir dónde y cuándo se producirán brotes de cólera en lugares como Angola.<sup>27</sup>
- + Los departamentos de policía modificaron un algoritmo inicialmente diseñado para predecir terremotos, y ahora lo están utilizando para predecir, en un radio de 500 pies, dónde es probable que se produzcan delitos. En las zonas donde se utiliza este software, los robos disminuyeron un 33 por ciento y los delitos violentos se redujeron un 21 por ciento.<sup>28</sup>
- + Mediante análisis de datos y sensores marinos que monitorizan olas, corrientes y otros datos, los investigadores están utilizando los análisis de datos para predecir tsunamis y otros desastres naturales, así como sus efectos.<sup>29</sup>
- + Los datos provenientes de las visitas al médico y la información de las recetas revelaron cómo los pacientes con enfermedades autoinmunitarias tienen un mayor riesgo de epilepsia.<sup>30</sup>
- + Los datos de la capacidad crediticia se están utilizando para predecir qué pacientes necesitarán recordatorios “amistosos” para que tomen sus medicamentos recetados.<sup>31</sup>
- + Al usar una década de datos históricos de vuelos correlacionados con pautas meteorológicas, los pasajeros de avión pueden calcular qué vuelos tienen más probabilidades de llegar a horario.<sup>32</sup>

Así como los datos están a nuestro alrededor, también hay oportunidades. Cuando las personas que innovan actúan de manera responsable y creativa, la innovación de los datos puede producir respuestas tanto para los problemas cotidianos como para los desafíos más grandes del planeta.

# 4

## LA TRANSFORMACIÓN Y TRADUCCIÓN DE LOS DATOS

Nuevas y poderosas herramientas de software nos están brindando la posibilidad de utilizar los conjuntos de datos para tomar decisiones basadas en la información y no en una corazonada o intuición.

En particular, hay un conjunto nuevo de herramientas que les están dando un propósito a los datos al transformarlos de formas que nos ayudan a extrapolar, enfocar, visualizar, reflexionar, perfeccionar, modelar y predecir. Estas herramientas incluyen tecnologías de aprendizaje automático que entienden los datos para ayudarnos a responder mejor a ellos; tecnologías de modelado y simulación que pueden probar escenarios y transformar los datos en soluciones reales; y herramientas que reconocen y traducen sonidos, imágenes o videos a nuevas formas de mayor significado.

Este tipo de transformación de los datos conduce a mejores planes, diseños superiores y decisiones más inteligentes. Por ejemplo, los médicos que ejercen la medicina hoy son bombardeados con investigaciones nuevas y así es casi imposible mantenerse al corriente de los últimos desarrollos, mucho menos interpretar los datos de los pacientes en tiempo real.<sup>33</sup> Como resultado, los hospitales están recurriendo a sistemas de respaldo de decisiones clínicas. En esencia, se trata de sistemas de software que analizan datos de fuentes dispares para colaborar con diagnósticos más rápidos y confiables en un entorno de datos complejos —demostraron ser de beneficio en más del 70 por ciento de los casos.<sup>34</sup>

Otras herramientas informáticas están ayudando a traducir los datos a formas de mayor significado. El procesamiento en tiempo real de audios, imágenes y videos está conduciendo a avances que cambian la vida. A modo de ilustración: mientras se reúnen más datos acerca de la forma en que hablan las personas, la tecnología de reconocimiento de voz mejora continuamente. Esto ha permitido avances como la traducción bidireccional de idiomas en tiempo real durante las conversaciones en todos los continentes —y genera, de manera potencial,

nuevas oportunidades para el comercio global. De manera similar, con una cantidad estimada de 360 millones de personas que sufren de pérdida auditiva, los investigadores de China han recurrido al reconocimiento de patrones y al procesamiento en tiempo real de datos de un sensor Kinect 3D para desarrollar un sistema que comprenda los gestos del lenguaje de señas y los convierta, en tiempo real, en lenguaje oral y escrito —y viceversa.<sup>35</sup>

El procesamiento más rápido de las imágenes también está teniendo un efecto profundo en áreas como la detección del cáncer, la informática cognitiva, la neurobiología y la robótica. Por ejemplo, debido a su apariencia y forma impredecibles, es particularmente difícil identificar los tumores cerebrales en las imágenes médicas. Con la ayuda de la informática en la nube y los algoritmos de análisis avanzado de imágenes, los equipos de científicos ahora están compitiendo para encontrar los mejores algoritmos informáticos a fin de identificar con mayor precisión y velocidad los tumores cerebrales.<sup>36</sup>

El cambio de las imágenes 2D a las imágenes 3D para las mamografías está mejorando los índices de detección del cáncer de seno. Las mamografías tridimensionales utilizan un software para combinar muchas radiografías desde diferentes ángulos a fin de crear una imagen tridimensional que pueda aumentar los índices de detección del cáncer de seno y disminuir las estresantes falsas alarmas.<sup>37</sup>

En efecto, la capacidad de utilizar los datos para crear tanto visualizaciones como simulaciones permite que sea más fácil comprenderlos y usarlos. Ahora modelamos y simulamos sistemas complejos y evaluamos diseños con mayor precisión y velocidad, sin tener que construirlos realmente. Por ejemplo, en la década del ochenta, Boeing evaluó 77 de sus 767 prototipos utilizando túneles de viento físicos. Al 2005, Boeing realizó solo 11 pruebas físicas para sus 787 prototipos, evaluó los prototipos mediante túneles de viento virtuales y supercomputación a fin de ahorrar tiempo, ahorrar energía, ahorrar dinero y salvar vidas.<sup>38</sup>



Los túneles de viento virtuales son un ejemplo de herramientas que mastican cantidades inmensas de datos para que el uso de la mecánica de fluidos computacional 3D sea más sencillo y su implementación más rápida. Estas herramientas nos permiten modelar mejor el flujo de calor, el flujo de fluidos, el flujo de aire y el flujo de los procesos para un mejor rendimiento. Se están utilizando

para modelar los lugares donde podrían viajar sustancias contaminantes por el agua subterránea, cómo aumentar el rendimiento de las turbinas de viento y cómo diseñar mejores edificios que puedan soportar las peores inclemencias de la Madre Naturaleza.

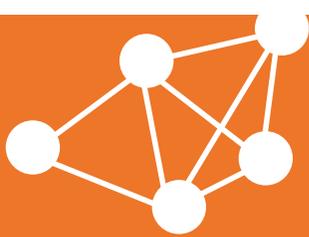
En pocas palabras, estas herramientas transforman los datos en soluciones.

En conjunto, las funciones de reunir, almacenar, analizar y transformar los datos se están uniendo para dejar al descubierto nuevas oportunidades de mejores soluciones.

En la práctica, sin embargo, cada una de estas cuatro fuerzas es ejercida, a menudo, por personas dispares, que manipulan conjuntos de datos diferentes, almacenados en lugares distribuidos. Sin embargo, esto es parte del poder de la revolución de los datos. Los conjuntos de datos diferentes, antes inconexos, se pueden combinar y analizar, incluso si se almacenan en lugares diferentes; se pueden acrecentar, incluso cuando los datos no estén estructurados; y se pueden manipular, incluso cuando las partes estén descubriendo respuestas fundamentales a preguntas que los creadores de los datos ni siquiera sabían cómo formular. Es posible que uno nunca conozca anticipadamente el poder exacto de la información que pueda extraerse posteriormente de un conjunto de datos en particular —en parte,

porque puede tornarse valiosa después para una combinación de conjuntos de datos aparentemente inconexos, y porque el algoritmo para explotar los conocimientos puede no haberse inventado aún.

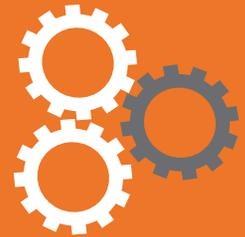
Así como los datos están a nuestro alrededor, también hay oportunidades. Cuando las personas que innovan actúan de manera responsable y creativa, la innovación de los datos puede producir respuestas tanto para los problemas cotidianos como para los desafíos más grandes del planeta. A fin de maximizar las oportunidades facilitadas por los datos de maneras que puedan acelerar nuevas olas de productividad, crecimiento económico y beneficios individuales, debemos hacer arrancar el motor de la innovación nuevamente y sentar las bases para las soluciones de punta y los efectos fortalecedores que está preparada para producir.



Las empresas de producción que saquen pleno provecho de sus datos podrían ahorrar

**\$371 mil millones**

en un período de cuatro años.



**DATOS COMERCIALES: AHORRO DE COSTOS**



# UNA ECONOMÍA IMPULSADA POR LOS DATOS

Los datos están surgiendo ahora como una de las nuevas fuerzas de beneficio económico más dinámicas. Los efectos económicos de los datos ya están extendiéndose a muchos sectores de la economía, tanto en las industrias de alta tecnología como en las de tecnología poco avanzada. Solo en los próximos cuatro años, el mejor uso de los datos podría dar lugar a \$1,6 billones de “dividendos de datos” en todo el mundo.<sup>39</sup>

La innovación de los datos ha demostrado que tiene la habilidad de aumentar la productividad. Las empresas que ya están realizando una toma de decisiones guiada por los datos informan un aumento de entre el 5 y 6 por ciento en la productividad.<sup>40</sup> Si, al utilizar los datos más efectivamente, podemos lograr ganancias aunque sean pequeñas en un amplio rango de industrias a fin de que puedan ser al menos un 1 por ciento más eficientes, los economistas estiman que pueden sumarse alrededor de \$15 billones al PIB global para el 2030. Esto equivale a incorporar otra economía de los EE. UU. Un aumento de la productividad del 1 por ciento puede parecer poco, pero tal como lo expresa el Director Ejecutivo de General Electric, Jeff Immelt: “dile a un hombre del sector del petróleo que puedes usar software para ahorrarle un uno por ciento en algo, y ese hombre será tu amigo para toda la vida”.<sup>41</sup>

## La innovación de los datos como creadora de empleo.

La innovación de los datos no implica solo aumentar el crecimiento económico. También se trata de alimentar un poderoso motor nuevo de creación de empleo y un multiplicador potente de la fuerza de trabajo. La innovación de los datos ya está creando miles de trabajos nuevos y bien pagos —desde analistas de datos y desarrolladores de software, hasta las personas que administran los depósitos

En EE. UU., cada puesto de trabajo **DRELACIONADO CON LOS DATOS** genera otros tres puestos de trabajo de manera indirecta.

**61 %** de ejecutivos Americanos reconocen que el poder de **ANÁLISIS DE DATA** es importante para el reclutamiento en sus **empresas.**

inmensos de datos que hacen posible la innovación de los datos. Además, el 61 por ciento de los ejecutivos senior de los EE. UU. y el 58 por ciento de los de Europa afirman que el análisis de los datos es importante para los planes de su empresa relativos a la contratación de nuevos empleados.<sup>42</sup> Por cada puesto de TI relacionado con los datos que se cree, se estima que se crearán otros tres puestos para personas ajenas a la TI —esto generará millones de puestos de trabajo adicionales en toda la economía.<sup>43</sup> El resultado final consistirá en beneficios económicos



Los edificios inteligentes solos pueden ahorrarles a las empresas

**\$25 mil millones**

al año en costos energéticos.



En un periodo de 15 años, una mejora de un 1 % de la productividad impulsada por los datos en la aviación podría dar lugar a un ahorro de

**\$30 mil millones**

en combustible a escala mundial.



nuevos y extendidos a partir de otra explosión del empleo producida por las tecnologías nuevas.

### Los datos permiten negocios más ágiles, competitivos y de mayor respuesta.

Tradicionalmente, los grandes aumentos de la productividad impulsados por la TI que hacían crecer nuestra economía y elevaban nuestros estándares de vida se limitaban solo a unos pocos sectores de la economía —como el propio sector de la tecnología. No obstante, a medida que más sectores generan más formas de datos, los datos están ahora preparados para producir ganancias nuevas en los sectores que solían quedar rezagados —de la agricultura a la salud, del transporte a la educación y de la energía a las finanzas. De hecho, en una encuesta reciente, el 79 por ciento de los ejecutivos senior de los Estados Unidos y el 80 por ciento de los ejecutivos de Europa afirmaron que el análisis de los datos es importante para los planes de sus empresas relativos a una mejor satisfacción de las necesidades de los clientes.<sup>44</sup> Asimismo, el 70 por ciento de los ejecutivos de los Estados Unidos y el 72 por ciento de los ejecutivos de Europa dicen que el análisis de los datos es importante para los planes de sus empresas relativos a la creación de nuevos productos o servicios.<sup>45</sup>

## ASISTENCIA MÉDICA

### La innovación de los datos está ayudando a las personas a vivir más tiempo y tener vidas más saludables mediante un mejor discernimiento.

La asistencia médica es uno de los entornos actuales con mayor riqueza de datos —genera cientos de terabytes de datos por hospital al día.<sup>46</sup> El uso maximizado de estos datos puede ser la receta para una mejor atención y curas más rápidas. Actualmente, los datos les están proporcionando a los médicos una percepción clara y sin precedentes de la salud de sus pacientes, y están permitiendo una mejor toma de decisiones. Con la ayuda de los análisis de datos, los médicos pueden hacer uso de los resultados de cada paciente previo a fin de formar opciones de tratamiento para cada paciente futuro. Y aun así, muchas veces no se hace un uso óptimo de los datos médicos. Si el sector de la asistencia médica utilizara los datos de manera más efectiva para generar eficiencia y calidad, se estima que el sector podría ahorrar más de \$300 mil millones de dólares por año —y reducir gastos en un significativo 8 por ciento.<sup>47</sup> Pero los mayores efectos no se miden solo a partir de los dólares ahorrados, sino de las vidas salvadas. Por citar solo un ejemplo, los investigadores

desarrollaron un algoritmo de aprendizaje automático que puede predecir los paros cardíacos con cuatro horas de anticipación y es exacto el 66 por ciento de las veces; lograron esto al combinar datos en tiempo real con los antecedentes médicos del paciente.<sup>48</sup>

## TRANSPORTE

### La innovación de los datos para ahorrar tiempo, ahorrar dinero, ahorrar combustible y salvar vidas.

Los datos nos ubicaron rumbo a un futuro con un transporte más inteligente. En todas nuestras aerovías, vías férreas y calles, los análisis de datos en tiempo real están conduciendo a sistemas de transporte más inteligentes que están mejorando nuestra capacidad de transportar bienes y personas de manera segura y eficiente. Actualmente, los automóviles nuevos están provistos de tantos sensores que pueden generar hasta 25 gigabytes de datos por hora, y contener más de 10 millones de líneas de códigos de software solo para procesar los datos.<sup>49</sup> Estos datos se están utilizando para impulsar sistemas nuevos de seguridad y prevención de accidentes que podrían tener un efecto de las mismas proporciones que tuvieron los cinturones de seguridad —al permitir una reducción de las lesiones y muertes nada menos que del 50 por ciento.<sup>50</sup> En una época donde la cantidad de automóviles en las calles crece constantemente, los datos se pueden utilizar también de maneras innovadoras para disminuir las congestiones y el tráfico en las rutas de forma más eficiente —y ahorrar millones de horas, miles de galones de combustible y toneladas de emisiones de gases de efecto invernadero.

Los datos de la aviación también están despegando. Hoy en día, una aerolínea moderna puede generar hasta medio terabyte de datos por vuelo a partir de los sensores ubicados en todo el avión; éstos se utilizan para mejorar el rendimiento del vuelo, reducir las turbulencias, mejorar la seguridad e identificar defectos en el motor 2000 veces más rápido que antes.<sup>51</sup> Los datos de la aviación también están mejorando la planificación de las rutas de vuelo, y le informan a la tripulación qué pieza debe ser reemplazada antes de que tenga un desperfecto. Estos beneficios suman. Una mejora de la productividad en la aviación impulsada por los datos de tan solo un 1 por ciento podría dar lugar a un ahorro de \$30 mil millones en combustible a escala mundial.

Si el sector de la asistencia médica utilizara los datos de manera más efectiva para generar eficiencia y calidad, se estima que el sector podría ahorrar más de **\$300 mil millones** de dólares por año y reducir gastos en un significativo **8 por ciento**.

## ENERGÍA/MEDIOAMBIENTE

**La innovación de los datos está reduciendo el consumo de energía y mejorando nuestro medioambiente.**

Los datos también están facilitando ahorros enormes de energía. Un mejor uso de los datos puede producir ahorros de miles de millones de dólares en varias industrias, y mejorar al mismo tiempo el medioambiente. Los conocimientos impulsados por los datos permiten un uso más inteligente de la energía en las redes eléctricas, los edificios, los hogares, las fábricas, las granjas y todo el medioambiente. Mediante las herramientas de diseño de datos, por ejemplo, ahora los edificios pueden tener diseños que consuman menos energía, y pueden estar provistos de sistemas inteligentes que combinen abanicos de datos de sensores con herramientas de análisis y activadores para lograr operaciones más eficientes. En conjunto, se calcula que los edificios inteligentes solos pueden ahorrarles a las empresas \$25 mil millones al año en costos energéticos.<sup>52</sup>

## PRODUCCIÓN

**La innovación de los datos está mejorando la forma en que los productos se diseñan, elaboran y distribuyen.**

El sector de la producción almacena más datos que ningún otro sector.<sup>53</sup> Como resultado, los productores tienen mucho que ganar a partir de un mejor uso de los datos para aumentar la eficiencia, producir calidad y mejorar la forma en que los productos se diseñan, elaboran y distribuyen. Según un cálculo aproximativo, el mejor uso de los datos de producción puede dar lugar a una reducción de hasta un 50 por ciento en el tiempo de desarrollo del producto y los costos de ensamblaje.<sup>54</sup> En efecto, IDC (International Data Corporation) calcula que las empresas de producción que saquen pleno provecho de sus datos están preparadas para alcanzar dividendos de datos por \$371 mil millones en un período de cuatro años.<sup>55</sup> Al usar datos en tiempo real, las empresas pueden realizar un mejor seguimiento y una mejor gestión de las cadenas de suministro globales, y reducir los defectos de los productos.

Los datos también están ayudando a mejorar los diseños. Los diseños digitales facilitados por los datos están mejorando el proceso de diseño repetitivo —permiten que los diseñadores prueben incontables ideas y las modifiquen en el entorno virtual hasta alcanzar finalmente un diseño

de producto óptimo. Toyota, Fiat y Nissan recortaron el tiempo de desarrollo de modelos nuevos entre un 30 y un 50 por ciento mediante el uso coordinado de los datos y las técnicas de modelado.<sup>56</sup>

## SECTOR FINANCIERO

**La innovación de los datos está aumentando la eficiencia, mejorando el cumplimiento normativo y reduciendo los fraudes.**

En el sector de los servicios financieros, la explotación de las crecientes cantidades de datos puede mejorar la eficiencia operativa, mejorar el cumplimiento normativo e identificar los fraudes. En una encuesta, el 71 por ciento de las firmas de los mercados bancario y financiero informaron que el uso de la información y las herramientas de análisis está creando una ventaja competitiva para sus organizaciones.<sup>57</sup> Las inversiones en datos pueden pagar enormes dividendos. Los análisis de datos ayudaron a una empresa de tarjeta de crédito a identificar \$2 mil millones en fraudes de tarjeta de crédito evitables al año.<sup>58</sup>

## AGRICULTURA

**La innovación de los datos está produciendo mejores alimentos usando menos recursos.**

Los datos están brotando por todos lados en las granjas —ayudan a producir alimentos más deliciosos y nutritivos para una cantidad cada vez mayor de personas. Gracias a un concepto denominado agricultura de precisión, los agricultores de todo el planeta pueden usar los datos de las semillas, los satélites, los sensores y los tractores para tomar mejores decisiones que aumentan el rendimiento, reducen los costos y alimentan a más gente. Al tener un cuarto de la población mundial involucrada en la agricultura y la producción de alimentos, se pueden lograr grandes beneficios a partir de un mayor uso de los datos. Por ejemplo, al aprovechar las herramientas de análisis de datos, los agricultores pueden reducir los gastos de producción, el uso de pesticidas y el uso de sustancias químicas, y tener una mejora de cinco o 10 bushels por acre en las cosechas.<sup>59</sup> Para los productores lecheros, un emprendimiento croata fue el primero en utilizar una plataforma de análisis informático basada en la nube que brinda datos en tiempo real sobre aspectos como el efecto de la calidad del alimento de los animales en la producción y los índices de concepción de cada animal; esto, a la vez, aumentó nada menos que un 50 por ciento el rendimiento y la eficiencia del productor.<sup>60</sup>

## Mediante el uso de más de 1.000 datos

por segundo, los científicos canadienses descubrieron que los bebés prematuros con signos vitales atípicamente estables tenían un mayor riesgo de presentar fiebre; esto permitió que los médicos tomaran medidas con antelación y salvaran vidas.

CANADÁ



# SEPARAMOS LOS MITOS SOBRE LOS DATOS DE LOS HECHOS

Hay muchos mitos alrededor de las recientes innovaciones de los datos y la economía de los datos.

Esto incluye mitos acerca de:

- + La información personal y la protección de los datos,
- + El efecto económico de la economía de los datos,
- + La confiabilidad de los datos,
- + Las innovaciones de los datos del siglo XXI,
- + Los beneficios globales de la innovación de los datos, y
- + La función del gobierno en la regulación de los datos.

## LA INFORMACIÓN PERSONAL Y LA PROTECCIÓN DE LOS DATOS

### MITO

**La innovación de los datos beneficia únicamente a las empresas de TI —no a los individuos.**

### REALIDAD

La innovación de los datos les da poder a los consumidores para tomar mejores decisiones y permite que los comerciantes personalicen sus bienes y servicios para brindarnos una mejor atención. Puede conducir a importantes avances mediante el aprendizaje personalizado y la medicina personalizada. Puede conducir a una mejor experiencia del consumidor mediante el entretenimiento personalizado. Puede ayudarnos a evolucionar de una economía impulsada por la producción en masa a una facilitada por la personalización en masa. En efecto, existe un conjunto de empresas en el centro del desarrollo y de la producción de las nuevas tecnologías de innovación sobre las que se apoya la revolución de los datos. Sin embargo, si hacemos elecciones inteligentes hoy, esta economía centrada en los datos que está surgiendo puede crear nuevos trabajos e industrias, y poner al mundo en las manos de los consumidores.

EN LA INDIA, LOS CENTROS DE CONEXIÓN  
A INTERNET PERMITEN QUE MÁS DE

# 4 millones

de granjeros tengan acceso a los precios de cultivo,  
datos del clima y otros tipos de información en los

## idiomas locales.

INDIA



### MITO

**Todos los datos son datos personales.**

### REALIDAD

Algunos datos pueden consistir en información personal (p. ej., los datos que generamos con nuestros dispositivos móviles o que creamos al usar las redes sociales). La mayoría de los datos, sin embargo, no son personales.

La inmensa cantidad de datos que se crea cada día incluye información como la monitorización del clima por satélite, el rendimiento de los motores de los aviones de pasajeros, las transacciones de la bolsa generadas por computadora y los sensores no relacionados con los individuos. Aun cuando los datos sí se relacionan con un individuo, a menudo no accede a ellos otro ser humano y es probable que estén desidentificados —en esencia, datos almacenados y utilizados sin información que revele la identidad de la persona involucrada.

### MITO

**A las empresas no les preocupa la protección de los datos personales.**

### REALIDAD

Cuando se generan datos personales, deben ser protegidos de manera adecuada. A fin de expandir las oportunidades que generan los datos, la confianza pública en ellos debe ser alta. Las empresas y las organizaciones que utilizan datos deben ejercer una buena administración de los datos. Estas prácticas podrían estar estandarizadas mediante esfuerzos guiados por la industria para crear pautas voluntarias respecto del uso responsable de los datos. Muchos líderes en el campo ya están dando un paso adelante para dejarles claro a los consumidores cómo se recopilan sus datos y si se comparten. Muchas empresas siguen buenas prácticas que les exigen anonimizar la información personal cada vez que resulte práctico.

### MITO

**La innovación de los datos provocará que pierda toda mi privacidad.**

### REALIDAD

El éxito de la economía de los datos depende de la confianza de los consumidores. Las personas deben sentir que su información personal está segura. Los principales desarrolladores de software ya integran protecciones de privacidad en sus sistemas desde el inicio, esto se denomina “privacidad desde el diseño”. Asimismo, los desarrolladores a menudo emplean la anonimización, la desidentificación y herramientas de cifrado para poder minimizar aún más el efecto de cualquier posible violación de los datos. Cuando los datos se agrupan para que los usuarios particulares no sean identificados de manera específica, en general, los datos pueden seguir analizándose para detectar pautas de comportamiento sin violar la confianza ni privacidad del usuario. Asimismo, las políticas de privacidad aplicables pueden tener en cuenta el contexto y los riesgos relativos que implican cualquier exposición o uso incorrecto de los datos, de modo que los datos más sensibles (como datos financieros o de asistencia médica) obtengan el nivel más alto de protección. Esto significa que datos como los meteorológicos o los análisis comerciales que no incluyen información personal no requieren el mismo nivel de protección que los datos de asistencia médica de pacientes específicos.

### MITO

**Los datos nunca se pueden desidentificar por completo. La desidentificación de los datos no es efectiva.**

### REALIDAD

La desidentificación de los datos es un proceso que se utiliza para evitar la conexión entre la identidad de una persona y la información. Una vez que los datos son desidentificados, se los puede analizar sin conexión alguna con una persona. Los expertos han desarrollado técnicas que permiten desidentificar los datos en formas que pueden maximizar tanto la privacidad como la calidad de los datos.<sup>61</sup> Según los expertos, si la desidentificación de los datos se realiza correctamente, el riesgo de reidentificar a las personas a partir de los datos anonimizados es inferior al 1 por ciento en la mayoría de los casos.<sup>62</sup>

EN EE. UU., LAS EMPRESAS AUTOMOTRICES  
MÁS IMPORTANTES están diseñando vehículos  
nuevos cargados con

# cientos de sensores

y dispositivos de análisis para permitir grandes avances, como el estacionamiento autónomo y la prevención de accidentes.

EE. UU.



## MITO

**No se puede confiar en las empresas que utilizan los datos.**

## REALIDAD

La industria está escuchando las inquietudes sobre la privacidad y les está prestando atención. En este momento, hay signos de una animada competencia entre las empresas líderes que se encuentran en una carrera hacia las mejores soluciones de protección de la privacidad. Por ejemplo, las dos empresas responsables de los sistemas operativos que hacen funcionar, de manera combinada, al 96,4 por ciento de los smartphones a escala mundial anunciaron recientemente mejoras en sus configuraciones de privacidad. Les están brindando controles adicionales a los usuarios y se están cambiando al cifrado de datos para proteger la privacidad personal.<sup>63</sup> Las empresas líderes de software apelan ahora directamente a los consumidores al publicitar su compromiso con la protección de la privacidad.<sup>64</sup> Las empresas están cambiando incluso sus servicios y políticas para que faciliten más la privacidad, por ejemplo, al no examinar las comunicaciones de los clientes para dirigir la publicidad.<sup>65</sup>

## MITO

**Las personas no tienen control alguno sobre sus datos.**

## REALIDAD

Si bien a veces puede parecer que no tenemos control sobre los datos, hay muchas herramientas que los consumidores tal vez desconozcan aún y que los ayudan a controlar mejor sus datos. Por ejemplo, algunas empresas líderes de software han elegido habilitar funciones con la orden de “No rastrear” de manera predeterminada en sus navegadores web, de modo que los sitios web visitados y las empresas de publicidad de terceros utilizadas por esos sitios reciben automáticamente una solicitud de “No rastrear”. Les informa a los sitios web que no deseamos ser rastreados y esto puede ayudar a protegernos contra formas de rastreo en la web. Asimismo, algunos agentes de datos han creado páginas web donde los consumidores pueden ver los tipos de información que se recopilaron sobre ellos, elegir que ya no se recopilen sus datos y corregir información errónea.<sup>66</sup> En conjunto, estas herramientas permiten que los consumidores controlen mejor la forma en que se recopila y utiliza su información, u optar por no ser parte de determinados usos de marketing.

## EL EFECTO ECONÓMICO DE LA ECONOMÍA DE LOS DATOS

### MITO

**La innovación de los datos no creará empleos nuevos e incluso puede eliminar puestos de trabajo.**

### REALIDAD

La innovación de los datos puede ser un poderoso factor de generación de crecimiento económico. De hecho, el 61 por ciento de los ejecutivos senior de los Estados Unidos y el 58 por ciento de los ejecutivos de Europa encuestados recientemente afirman que el análisis de los datos es importante para los planes de su empresa relativos a la contratación de nuevos empleados.<sup>67</sup> Aun así, la innovación de los datos no implica la creación de empleos dentro del sector de la TI exclusivamente. Cada función relacionada con los datos creará empleo para una cantidad estimada de tres personas ajenas a la TI, esto generará incluso más puestos en toda la economía.

Si bien habrá algunos cambios en el empleo a medida que la innovación de los datos ayude a encontrar nuevas formas de llevar a cabo las tareas tradicionales, es probable que este cambio se parezca a la proliferación de Internet, cuando se estima que se crearon 2,6 puestos de trabajo por cada puesto suspendido y perdido.<sup>68</sup> Una de las principales áreas de posible crecimiento laboral es la del análisis de datos. Con frecuencia, la comprensión de los conocimientos presentes en los datos se puede lograr únicamente mediante el ingenio de los seres humanos. Actualmente, hay una escasez mundial de analistas y administradores de datos especializados que puedan ayudar a comprender los datos. Según McKinsey, solo en Estados Unidos existe una falta de entre 140.000 y 190.000 personas con la capacidad de análisis profundo necesaria para descubrir el poder que se esconde en los datos, y 1,5 millones de administradores y analistas con la capacidad de comprender y tomar decisiones sobre la base del análisis de la economía de los datos.<sup>69</sup>

EN BRASIL, uno de los mayores productores

# de soja

está invirtiendo en software y en el poder de las herramientas de análisis de datos para incrementar la eficiencia de las técnicas de control de daño, reducir costos e impulsar las técnicas de cosecha.

BRASIL



El salario básico medio de 2014 para un científico de datos de plantilla es de \$120.000 y de \$160.000 para un administrador.<sup>70</sup> Muchos han sugerido que para capturar plenamente el poder de la innovación impulsada por los datos, las empresas y los legisladores de todo el mundo deben tomar medidas para superar la gran escasez de talentos.

## MITO

**El análisis de datos implica la eliminación del criterio humano del proceso.**

## REALIDAD

Si bien algunas preguntas se pueden responder con los datos (por ejemplo, ¿la población de mi ciudad está creciendo o disminuyendo?), muchas de las repuestas más penetrantes no son tan claras. Es posible que uno no sepa siempre cómo se relacionan entre sí los diferentes datos individuales. Y como tal vez uno no sepa anticipadamente cuál es la pregunta correcta, los análisis de datos son, a menudo, un proceso reiterativo de formulación de preguntas sucesivas para encontrar finalmente la respuesta. Por estos motivos, nunca podremos suprimir el criterio y el aporte humanos para conciliar las diferencias y revisar las posibles incongruencias.

Los datos solos no son una panacea, y no pueden hacer milagros. De hecho, con frecuencia los datos tienen poco valor por sí solos. A menudo están desordenados, son desorganizados por naturaleza y no están prolijamente estructurados. El esfuerzo empieza al tratar de comprenderlos y descubrir su importancia. El hecho de que los datos puedan o no resolver problemas depende de la ejecución efectiva de una estrategia de datos inteligente que pueda conducir a soluciones más rápidas y mejores. También depende de que se formulen con precisión las preguntas correctas. Pero si aprovechamos los datos en las formas adecuadas, podemos ayudar a descubrir las respuestas a algunos de los desafíos más apremiantes de la sociedad, ayudar a atizar la fogata de la innovación y alimentar una poderosa ronda nueva de empleos impulsados por la TI y el crecimiento económico.

## MITO

**La innovación de los datos es solo para las grandes empresas y no para los pequeños negocios.**

## REALIDAD

Gracias a que los datos están cada vez más extendidos, los costos de almacenamiento están cayendo y las herramientas de análisis se están tornando más poderosas y rentables, en este momento incluso las empresas más pequeñas pueden sacar provecho de los análisis avanzados de datos —herramientas que alguna vez estuvieron a disposición de las empresas más grandes únicamente. Por ejemplo, la función Trends incluida en QuickBooks Online de Intuit permite que los negocios pequeños se beneficien a partir de la sabiduría colectiva de otros usuarios de Intuit —esto permite que los negocios pequeños comparen sus ingresos y gastos para destacar oportunidades. Les permite tomar decisiones más inteligentes acerca de la forma en que desarrollan sus actividades. Si bien el uso de las soluciones de inteligencia empresarial y de análisis no se ha generalizado entre los emprendimientos de pequeño y mediano tamaño, se prevé que su adopción crezca rápidamente.<sup>71</sup> Un estudio reciente reveló que los análisis de datos son importantes para el 60 por ciento de las empresas pequeñas.<sup>72</sup> Esto incluye al 57 por ciento de las empresas de los EE. UU. con 50 empleados o menos y al 62 por ciento de las empresas europeas del mismo tamaño, según los encargados senior de tomar decisiones. En las empresas de tamaño mediano (las que tienen entre 51 y 500 empleados), el 87 por ciento de los ejecutivos de los EE. UU. y el 79 por ciento de los ejecutivos europeos afirman que los análisis de datos son importantes.

CIENTÍFICOS QUE ESTUDIAN los patrones

# de la malaria

utilizaron recientemente datos de teléfonos celulares en Kenia para precisar los lugares de concentración de casos donde se estaban transmitiendo enfermedades y, así, guiar al gobierno en sus

**esfuerzos de erradicación.**

KENIA



## MITO

**Los datos benefician únicamente al sector de la TI y no a otros sectores de la economía.**

## REALIDAD

Los datos se encuentran ahora en el centro de una importante transición tecnológica que promete transformar y mejorar casi todos los sectores de la economía. De hecho, muchos creen que la innovación de los datos tiene la capacidad de mejorar la productividad en todos los sectores de la economía. Si bien las empresas que emplean la toma de decisiones guiada por los datos informan un aumento de entre un 5 y 6 por ciento en la productividad, si la innovación de los datos pudiera lograr una mejora de la eficiencia de tan solo el 1 por ciento, habría efectos considerables en todos los sectores de la economía —generaría ahorros de energía y combustible, produciría mejores resultados médicos con menores costos y aumentaría el rendimiento y la vida útil de los activos físicos.<sup>73</sup> En la industria de la aviación comercial, por ejemplo, GE prevé que al poder capturar datos en tiempo real para mejorar la eficiencia de los motores y transportar a los pasajeros por rutas más efectivas, con un aumento de solo el 1 por ciento en el ahorro de combustible se lograría ahorrar \$30 mil millones en un período de 15 años.<sup>74</sup>

## MITO

**La correlación implica siempre causalidad.**

## REALIDAD

Esta búsqueda de las “incógnitas desconocidas” es una de las mayores penetraciones que se puede derivar de los datos. Sin embargo, a veces las correlaciones simplemente no tienen sentido, y no siempre implican causalidad. Por ejemplo, si bien se observó que el índice de delitos violentos y asesinatos sube cuando lo hacen las ventas de helado, es altamente improbable que comprar helados convierta a las personas en asesinas.<sup>75</sup> Aun así, encontrar relaciones causales entre los datos es uno de los conocimientos más valiosos que podemos descubrir

en ellos. Y cada día se descubren muchas correlaciones y causalidades interesantes a velocidades cada vez mayores, y con una creciente importancia. La capacidad de distinguir la diferencia entre una mera correlación y la causalidad es uno de los motivos por los que el buen aprovechamiento de los datos requiere de científicos de datos especializados que sepan cómo separar una mera correlación de la causalidad y eliminar resultados que no pasen la prueba de improbabilidad.

## MITO

**La innovación de los datos requiere de un gran presupuesto.**

## REALIDAD

Sacar provecho de los datos no necesariamente requiere de grandes presupuestos. A menudo solo se necesitan inversiones menores, y esas inversiones pueden conducir a considerables ganancias. Por ejemplo, se proyecta que las organizaciones con un enfoque más holístico para sus datos producirán algunas de las mayores ganancias —y lograrán enormes “dividendos de datos” por \$1,6 billones aproximadamente en ingresos adicionales, costos menores y una mejor productividad solo en los próximos cuatro años.<sup>76</sup> De hecho, el uso de conjuntos reducidos de datos sumados a los datos de los que ya se dispone puede ser un buen punto de partida. El análisis de datos de transacciones financieras de una semana para detectar fraudes puede arrojar luz sobre tendencias más fácilmente que el análisis de datos históricos de cinco años. Según un análisis, las empresas que utilizaron análisis de economía de datos efectivamente fueron un 26 por ciento más lucrativas que sus competidores del sector, generaron un 9 por ciento más de ingresos a través de sus empleados y activos físicos, y gozaron de coeficientes de valoración de mercado un 12 por ciento más altos.<sup>77</sup>

EN LOS EMIRATOS ÁRABES UNIDOS, las nuevas herramientas de datos se están utilizando para diseñar el

# primer edificio de

energía positiva del mundo que produce más energía de la que consume.

## EMIRATOS ÁRABES UNIDOS



## CONFIABILIDAD DE LOS DATOS

### MITO

**Los conocimientos a partir de los datos son siempre precisos por sí solos.**

### REALIDAD

Los conocimientos son precisos solo si los datos subyacentes son precisos, si las herramientas son inteligentes y si un científico de los datos desarrolla el modelo para evitar resultados sesgados. Por ejemplo, la innovadora aplicación de Boston, Street Bump, utiliza smartphones que incorporan acelerómetros y localización por GPS para ubicar baches a partir de una gran cantidad de usuarios que informan cada vez que alguien encuentra un bache en la calle. Sin embargo, si los usuarios de smartphones tienden a ser más jóvenes y pudientes, los datos de la aplicación podrían no capturar con precisión los datos de los baches de todas las zonas demográficas de la ciudad —y la respuesta de la ciudad frente a los baches correría el riesgo de estar sesgada. Es por eso que muchas veces se necesita de la participación de científicos de los datos para garantizar la relevancia estadística, y evitar un sesgo en los resultados de los datos.

### MITO

**Las decisiones tomadas por instinto normalmente son las decisiones correctas.**

### REALIDAD

Desde el momento en que nacemos, aprendemos a confiar en nuestra intuición para tomar decisiones. De hecho, se estima que el 19 por ciento de los gerentes mundiales se describen como “referentes intuitivos” que se apoyan casi exclusivamente en el instinto y la intuición.<sup>78</sup> Aun así, en la actualidad los datos pueden ayudarnos a tomar decisiones más informadas, con mayor precisión, velocidad e impacto. Según una encuesta realizada a responsables de TI de una serie de industrias, el 59 por ciento manifestó que mejorar la calidad de la toma de decisiones es el objetivo principal que impulsa las inversiones en las tecnologías de los datos.<sup>79</sup> No solo beneficia al lugar de trabajo; nosotros también tomamos decisiones intuitivas en nuestras vidas

cotidianas. En algunos casos, podemos beneficiarnos mucho con las herramientas que nos ayudan a tomar decisiones más rápidas y mejores. Por ejemplo, se estima que el error humano es responsable del 93 por ciento de los accidentes automovilísticos, pero los sensores de los vehículos nuevos que recopilan cantidades gigantes de datos sobre sus entornos combinados con análisis en tiempo real y una toma de decisiones automatizada (incluso sin ser completamente autónoma) podrían reducir las lesiones y muertes relacionadas con los automóviles nada menos que en un 50 por ciento.<sup>80</sup>

### MITO

**Una mayor cantidad de datos equivale siempre a mejores datos.**

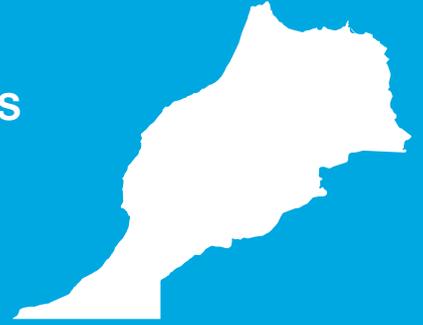
### REALIDAD

Los conjuntos de datos no siempre necesitan ser más grandes para ser mejores. De hecho, el tamaño puede ser uno de los factores menos importantes. En el caso de muchas preguntas y respuestas, incluso cantidades pequeñas de datos, si se las analiza con las herramientas correctas, pueden conducir a observaciones penetrantes. El desarrollo de estas observaciones depende a menudo de la calidad de los datos subyacentes y de la calidad de las herramientas utilizadas para manipularlos. Aun así, a veces se cree que una mayor cantidad de datos conlleva un mayor grado de verdad, y que cuanto más grande sea el conjunto de datos, más objetivo será. En ocasiones, es el acceso a datos simples el que puede tener el efecto más inmediato; por ejemplo, saber a qué hora cierra una tienda local, cuántas millas recorrió un automóvil desde el último cambio de aceite o cuánto costará la factura de teléfono este mes. De acuerdo con la pregunta formulada, los datos presentes en páginas web, en hojas de cálculo de Excel o en bases de datos de Gestión de relaciones con los clientes (CRM, por sus siglas en inglés) pueden ser reducidos, pero tan poderosos para brindar respuestas como los macrodatos. Lo más importante es crear datos sólidos, almacenarlos de manera segura, tener acceso a ellos y poder procesarlos —independientemente del tamaño—, de modo que se puedan utilizar cuando y donde se necesiten para resolver problemas.

# Los datos meteorológicos

no se utilizan solo para predecir si necesitamos llevar un paraguas; también se pueden usar para predecir el rendimiento de las cosechas, predecir la probabilidad de un brote de gripe y estimar cuánta capacidad de desalinización se necesita construir en Marruecos.

## MARRUECOS



### MITO

**Los datos no estructurados son inutilizables.**

### REALIDAD

Los datos no estructurados pueden ser útiles. De hecho, algunos de los datos más poderosos no siempre encajan bien en las tablas estructuradas de columnas y filas. Pueden estar estructurados como documentos de texto o radiografías. La clave es transformar los datos a sus formas más útiles. Si la empresa promedio de Fortune 1000 pudiera aumentar la capacidad de uso de sus datos solo un 10 por ciento, la empresa podría esperar un aumento de más de \$2 mil millones en ingresos.<sup>81</sup>

### MITO

**Los datos solo deben utilizarse con el propósito original por el que se recopilaron.**

### REALIDAD

Las tecnologías en la nube provocaron una reducción drástica en el precio de almacenamiento de los datos, de modo que ya no es necesario eliminar los datos después de su uso inicial para que haya más lugar para datos nuevos. Como resultado, y a diferencia de otros recursos, los datos no se agotan luego de ser utilizados por primera vez. Los datos constituyen un recurso renovable que, a menudo, se puede reutilizar combinado con otros conjuntos de datos y usar muchas veces para generar respuestas a preguntas que no podían preverse al momento de la creación de los datos. Por ejemplo, los datos meteorológicos no se utilizan solo para predecir si necesitamos llevar un paraguas; también se pueden usar para predecir el rendimiento de las cosechas, predecir la probabilidad de un brote de gripe y estimar cuánta capacidad de desalinización se necesita construir en Marruecos.

## INNOVACIONES DE LOS DATOS DEL SIGLO XXI

### MITO

**Los beneficios más profundos de los datos están alejados en el futuro.**

### REALIDAD

Si bien la innovación de los datos continuará generando beneficios durante décadas futuras, sus fuertes resultados ya se pueden ver casi en todo nuestro entorno. Los terabytes de datos diarios ya se están empleando para brindarnos pronósticos del tiempo más precisos; los datos de millones de señales de GPS se usan para predecir los tiempos de viaje en transporte público y hacernos llegar más rápido a nuestro destino; millones de datos médicos se están utilizando para identificar las posibles causas de las enfermedades; y se están empleando terabytes de datos financieros para ayudar a prevenir los fraudes con las tarjetas de crédito.

De hecho, cuando la Intelligence Unit de The Economist les pidió a los encuestados que describieran el efecto que tuvieron los datos en sus organizaciones en los últimos cinco años, casi el 10 por ciento afirmó que había cambiado por completo la forma en que llevan adelante los negocios.<sup>82</sup> El 46 por ciento de los encuestados dijo que se había convertido en un factor importante que impulsa las decisiones comerciales. Las nuevas investigaciones sugieren que solo en los próximos cuatro años, el mejor uso de los datos podría conducir a “dividendos de datos” equivalentes a \$1,6 billones en todo el mundo gracias a que los datos ayudan a descubrir nuevas oportunidades y soluciones.<sup>83</sup> En una encuesta diferente, al enfocarse en este año, el 33 por ciento de los ejecutivos senior de los Estados Unidos y el 24 por ciento de los ejecutivos de Europa preveían que un 10 por ciento o más del crecimiento de sus empresas estaría relacionado con los análisis de datos.<sup>84</sup> Al mirar los próximos cinco años, el 58 por ciento de los ejecutivos de los EE. UU. y el 43 por ciento de los de Europa hicieron la misma predicción.

Luego del desastre del tsunami de 2004 en el sur de Asia, a los pescadores indonesios se les proporcionaron teléfonos celulares. Sus ingresos

**subieron un 30 %,**

ya que por primera vez en su vida contaban con datos sobre el verdadero precio de mercado del pescado.

## INDONESIA



### MITO

**Se exagera demasiado la importancia de los datos.**

### REALIDAD

El uso de los datos para resolver los problemas de las personas difícilmente sea una novedad. Hemos usado los datos para tomar mejores decisiones desde el inicio de la civilización moderna —lo que condujo al uso de ciclos de barbecho para el cultivo a fin de alimentar más gente, técnicas de navegación que facilitaron el comercio global y descubrimientos médicos que evitaron millones de muertes por el cólera. Sin embargo, en el pasado, los datos eran un recurso escaso que resultaba oneroso y de difícil manipulación. La diferencia es que hoy los datos abundan más, los costos de almacenamiento bajaron considerablemente y las herramientas para manipularlos son más poderosas. Como resultado, al enfrentarnos a la aparición de un conjunto nuevo de desafíos, las poderosas tecnologías de análisis de datos pueden ayudarnos a revisar volúmenes crecientes de datos a fin de descubrir conocimientos sólidos y soluciones inesperadas a algunos de nuestros desafíos más apremiantes.

### MITO

**La era del crecimiento económico impulsado por la TI terminó, la innovación de los datos no puede aumentar la productividad.**

### REALIDAD

La innovación de la TI y su capacidad para hacer crecer las economías, crear empleos y elevar los niveles de vida en todo el planeta se apoya en su demostrada capacidad para aumentar la productividad —por ejemplo, incrementó la productividad entre un 1 y 2 por ciento en la década del noventa.<sup>85</sup> Un destacado economista de Northwestern University afirma que los mayores beneficios de la innovación de la TI han quedado a nuestras espaldas.<sup>86</sup> Sin embargo, la era de la productividad está sana y salva. De hecho, las empresas que emplean la toma de decisiones guiada por los datos informan un aumento de entre un 5 y 6 por ciento en la productividad.<sup>87</sup> Incluso si esta oportunidad en desarrollo a partir de los datos aumentara la productividad solo en los EE. UU., por ejemplo, en un 1,5 por ciento, en un período de 20 años podría permitir ahorrar dinero suficiente para elevar los ingresos nacionales promedio en nada menos que un 30 por ciento.<sup>88</sup> Si, al utilizar los datos de manera más efectiva, podemos lograr pequeñas ganancias en un amplio rango de industrias a fin de que sean solo un 1 por ciento más eficientes, los economistas estiman que pueden sumarse alrededor de \$15 billones al PIB global para el 2030, esto equivale a incorporar otra economía de los EE. UU.<sup>89</sup>

Barcelona está utilizando los datos para construir una ciudad

# más inteligente,

mejorar los servicios gubernamentales y brindar soluciones de transporte

# más sostenibles.

ESPAÑA



## LOS BENEFICIOS GLOBALES DE LA INNOVACIÓN DE LOS DATOS

### MITO

**En la actualidad, solo los Estados Unidos se benefician a partir de los datos.**

### REALIDAD

En todo el mundo, los datos ya se están empleando para resolver desafíos importantes. Por ejemplo:

- + Mediante el seguimiento de más de 1000 datos por segundo, los científicos canadienses descubrieron que los bebés prematuros con signos vitales atípicamente estables tenían un mayor riesgo de presentar fiebre — esto permitió que los médicos tomaran medidas con antelación y salvaran vidas.
- + En Brasil, uno de los mayores productores de soja está invirtiendo en software y en el poder de las herramientas de análisis de datos para incrementar la eficiencia de las técnicas de control de daño, reducir costos e incrementar la productividad de la cosecha.
- + Barcelona está utilizando los datos para construir una ciudad más inteligente, mejorar los servicios gubernamentales y brindar soluciones de transporte más sostenibles.<sup>90</sup>
- + En los Emiratos Árabes Unidos, las nuevas herramientas de datos se están utilizando para diseñar el primer edificio de energía positiva del mundo que produce más energía de la que consume.<sup>91</sup>
- + Los agricultores de la India están utilizando datos de semillas, satélites, sensores y tractores para tomar mejores decisiones acerca de qué cultivar, cómo cultivarlo, cómo hacer un seguimiento de la frescura de los alimentos de la granja a la mesa y cómo adaptarse a los cambios en el clima.<sup>92</sup>

### MITO

**Los países en desarrollo no están listos aún para sacar provecho de los análisis de datos.**

### REALIDAD

La revolución de los datos y los beneficios que crea constituyen un fenómeno global. Algunos de los beneficios más importantes y algunas de las mayores oportunidades de los datos se encuentran en los países en desarrollo, donde la tecnología muchas veces quedó retrasada. Según IDC, en el universo digital, los mercados emergentes superarán a los mercados establecidos para el 2017 —un crecimiento del 36 al 62 por ciento del universo digital en expansión entre 2012 y 2020.<sup>93</sup> De manera similar, una encuesta de las ONG en los países en desarrollo reveló que más del 90 por ciento cree que los análisis de datos son la herramienta más importante para lograr percepciones más profundas a fin de ayudar a sus beneficiarios finales.<sup>94</sup>

Abundan los ejemplos de innovación de los datos en los países en desarrollo:

- + Los científicos que estudian los patrones de la infección de la malaria utilizaron recientemente datos de teléfonos celulares en Kenia para precisar los lugares de concentración de casos donde se estaba transmitiendo la enfermedad, y así guiar al gobierno en sus esfuerzos de erradicación.<sup>95</sup>
- + Luego del desastre del tsunami de 2004 en el sur de Asia, a los pescadores indonesios se les proporcionaron teléfonos celulares. Sus ingresos subieron un 30 por ciento, ya que por primera vez en su vida contaban con datos sobre el precio de mercado del pescado.<sup>96</sup>
- + En Perú, los sitios históricos se encuentran bajo amenaza por el desarrollo. Mediante tecnologías aéreas y un poderoso software que une las imágenes, Perú creó nubes de datos tridimensionales detallados para cartografiar, supervisar y salvaguardar sus tesoros en peligro.<sup>97</sup>
- + Vietnam está utilizando escáners 3D a fin de preservar su herencia cultural, y se fijó el objetivo de digitalizar 40.000 utensilios históricos en los próximos cinco años.<sup>98</sup>
- + En la India, los centros de conexión a Internet permiten que más de 4 millones de granjeros tengan acceso a los precios de cultivo, datos del clima y otros tipos de información en los idiomas locales. Utilizan análisis de datos avanzados y tecnologías móviles para seguir datos de granjas particulares y, después del análisis, les pueden ofrecer suministros a los granjeros de acuerdo con sus necesidades, como fertilizantes y semillas.<sup>99</sup>

PARA PRESERVAR SU HERENCIA CULTURAL,  
Vietnam está utilizando escáneres  
3D a fin de digitalizar

**40.000**  
**objetos históricos**  
en los próximos cinco años.

VIETNAM



## LA FUNCIÓN DEL GOBIERNO EN LA REGULACIÓN DE LOS DATOS

### MITO

**Los datos se utilizarán como una herramienta de exclusión a partir de la capacidad de reforzar los obstáculos que enfrentan las comunidades desfavorecidas y de bajos ingresos.**

### REALIDAD

Un desafío que debería ser una de las principales prioridades tanto para la industria como para el gobierno es calmar la preocupación de que se usen los datos de manera injusta respecto de algunas personas o clases de personas. Al mismo tiempo, hay oportunidades que no deben pasarse por alto a la hora de usar los datos para combatir la discriminación injusta y darles poder a los grupos. De hecho, si se usan de manera responsable, los datos pueden ser una poderosa herramienta nueva para dejar al descubierto los actos de discriminación actuales que afectan el acceso a los empleos, las finanzas, la educación y las oportunidades.<sup>100</sup>

### MITO

**Los gobiernos no tienen ningún papel que cumplir.**

### REALIDAD

Con el enorme crecimiento del poder de la informática remota, del almacenamiento, de los análisis y de los servicios de software, surgió una serie de cuestiones normativas nuevas. La mayoría de los tratados, leyes y regulaciones no preveían estas posibilidades cuando se redactaron. Esta ambigüedad puede ser problemática para los gobiernos y el crecimiento de la economía de los datos. Las cuestiones normativas tradicionales, como el equilibrio entre la privacidad y la seguridad, el libre flujo del comercio, las investigaciones básicas y el desarrollo de la fuerza laboral deben considerarse como si fueran nuevas, ya que el crecimiento de estas tecnologías ha tenido una explosión.

Hay varias medidas concretas que los legisladores pueden adoptar para ayudar a capturar todo el poder de la revolución de los datos y acelerar las nuevas olas de productividad, el crecimiento económico y los beneficios para el consumidor que se encuentran apenas sobre el horizonte. Pueden invertir en el avance de las soluciones de datos que aborden algunos de los desafíos más grandes de la sociedad —como mejorar la forma en que aprendemos, producimos alimentos, vivimos nuestras vidas, viajamos de un lugar a otro y hacemos crecer las economías. Pueden establecer reglas claras para la colaboración internacional en cuanto a la aplicación de las leyes. Pueden promover las libertades de mercado que afectan la capacidad propia de las industrias de la tecnología para invertir, innovar, realizar transacciones comerciales y crecer. Pueden ayudar a promover el libre flujo de los datos sin límites físicos. Pueden ayudar a garantizar una fuerza de trabajo talentosa y superar la gran falta de científicos de datos especializados. Y pueden ayudar a atizar la fogata de la innovación para acelerar los beneficios de los datos tanto para los negocios como para los consumidores. Con políticas pragmáticas que exploten el talento y la tenacidad de las personas; aprovechen la innovación y la inversión; y expandan la capacidad y los recursos; los legisladores pueden ayudar a maximizar los beneficios que esta economía centrada en los datos puede producir.

**EN PERÚ, LOS SITIOS HISTÓRICOS  
SE ENCUENTRAN BAJO AMENAZA  
POR EL DESARROLLO.**

Mediante tecnologías aéreas y un poderoso software que une las imágenes, Perú creó detalladas nubes de datos

# tridimensionales

para cartografiar, supervisar y salvaguardar sus tesoros en peligro.

PERÚ



## MITO

**La localización de los datos ayuda a proteger la privacidad y mejora la seguridad.**

## REALIDAD

Algunas personas creen que la localización de los datos —que requiere que los datos se almacenen dentro de los límites de las fronteras de un país determinado— puede mejorar la privacidad y la seguridad. Sin embargo, los beneficios tecnológicos actuales son facilitados por la fuerza global que es Internet, y alimentados por los datos que se entrecruzan en todo el planeta entre centros de datos dispares. El tráfico de Internet que cruza las fronteras aumentó más del 50 por ciento desde 2005.<sup>101</sup> Permitir que los datos fluyan libremente por las fronteras está dando lugar a que incluso las empresas y los emprendimientos más pequeños sean la tienda de la esquina para el planeta entero a medida que empiezan a atravesar las fronteras para vender y adquirir productos, servicios e ideas. No obstante, los gobiernos de todo el mundo a menudo analizan políticas que restringen el libre flujo de datos o exigen que los servidores de datos se encuentren dentro de sus jurisdicciones como condición para prestar servicios al mercado.<sup>102</sup> Estas restricciones socavan la enorme eficiencia de escala y los beneficios económicos que pueden surgir de la innovación de los datos, y la capacidad de combinar diferentes conjuntos de datos de diversos lugares para descubrir conocimientos beneficiosos a partir de la creciente abundancia de datos. Puede socavar también la seguridad al evitar que se hagan copias de datos valiosos en varios lugares para protegerlos en caso de un desastre natural o un desperfecto técnico. Para lograr los beneficios que los datos pueden brindar, no es necesario que las leyes de cada país sean idénticas, pero sí compatibles. Permitir que los datos crucen las fronteras libremente es uno de los principios fundamentales para habilitar los beneficios impulsados por los datos.

## MITO

**La única forma de que los datos puedan estar protegidos es que los gobiernos intervengan para exigir su protección.**

## REALIDAD

Las normas gubernamentales actuales en materia de privacidad se pueden combinar con avances rigurosos e innovadores en la privacidad y buenas prácticas voluntarias de la industria para garantizar que los datos estén seguros y que se proteja la información personal. Por el contrario, las órdenes gubernamentales que intentan abordar la privacidad y la seguridad mediante la exigencia de que los datos se almacenen localmente podrían inhibir la innovación y limitar los tipos de beneficios sociales que la innovación de los datos puede producir.



## DISCURSO DIGITAL

# COMPRENSIÓN DEL LENGUAJE DE LOS DATOS

### ABUNDANCIA DE DATOS

Alguna vez fueron escasos, pero en la actualidad es posible que abunden los datos gracias a la creciente capacidad de recopilar formas coherentes de datos digitales de maneras completamente nuevas, en combinación con la caída en los costos de almacenamiento de los datos y las nuevas formas de crear valor a partir de ellos.

### ACUMULACIÓN DE DATOS

La acumulación de datos es el acto de recopilar datos de varias fuentes con el objetivo de proporcionar un análisis de nivel superior.

### ALGORITMO

Un algoritmo es un procedimiento paso por paso o una serie de instrucciones informáticas que utiliza la matemática para analizar datos con el objeto de resolver problemas. Los algoritmos se utilizan prácticamente en todos los programas de software.

### ANÁLISIS

Los análisis son el uso simultáneo de estadísticas y algoritmos basados en software para descubrir ideas, patrones y conexiones coherentes dentro de los datos.

### ANÁLISIS DE DATOS

El análisis de datos es la aplicación de software como una forma de transformar los datos y darles forma a fin de derivar información útil, conocimientos o significado de ellos. Se utiliza con frecuencia para descubrir patrones escondidos o correlaciones desconocidas, y como una ayuda para la toma de decisiones.

### ANÁLISIS DE LA CAUSA PRINCIPAL

El análisis de la causa principal es un método de resolución de problemas que enfoca la mirada en la relación entre la causa y el efecto para identificar la causa principal de una falla o problema. La causa es una causa principal si, una vez retirada de una secuencia de eventos, evita que un evento no deseado se repita.

### ANÁLISIS DE REGRESIÓN

El análisis de regresión es un proceso estadístico para utilizar datos a fin de calcular la relación entre dos o más variables.

### ANÁLISIS DE RIESGOS

Un análisis de riesgos es el uso de herramientas informáticas de análisis de datos para identificar los riesgos probables de un proyecto, una acción o una decisión. Las nuevas herramientas de datos pueden ayudar a identificar posibles riesgos con anticipación, moldear mejor una serie de escenarios para reducir los riesgos que enfrentan las organizaciones y controlar los sistemas para identificar problemas si algo empieza a desviarse de su curso normal.

### ANÁLISIS POR AGRUPACIÓN EN CLÚSTERES

El análisis por agrupación en clústeres es el proceso de identificación de datos que son similares entre sí y su agrupación en clústeres para poder comprender mejor las diferencias y las similitudes entre los datos.

### ANÁLISIS PREDICTIVOS

Los análisis predictivos implican el uso de algoritmos de software en uno o más conjuntos de datos para predecir tendencias o eventos futuros. Cuando los datos del presente se pueden comparar con los del pasado, con frecuencia, es posible emplearlos para predecir el futuro.



## **ANALISTA DE DATOS**

Un analista de datos es alguien responsable de preparar, limpiar y procesar los datos.

## **ANÁLISIS DE TEXTO**

El análisis de texto es el uso de técnicas estadísticas, lingüísticas y de aprendizaje automático con datos basados en texto para derivar significados, extraer conceptos o descubrir conocimientos. En general, los análisis de texto se realizan con textos en lenguaje natural incluidos en documentos, transcripciones, publicaciones web, comentarios o formularios. Pueden ser de utilidad para el resumen, el descubrimiento o la clasificación de contenido.

## **ANONIMIZACIÓN**

La anonimización de los datos implica eliminar toda la información de identificación personal que pudiera conducir a la identidad de una persona.

## **APRENDIZAJE AUTOMÁTICO**

El aprendizaje automático es el uso de algoritmos para permitir que una computadora analice datos con el objetivo de “aprender” a partir de la experiencia las acciones que debe tomar cuando se produce un determinado patrón o evento.

## **ARQUITECTURA Y DISEÑO DE DATOS**

La arquitectura de los datos se realiza, en general, durante la etapa de planificación de un nuevo sistema para diseñar y estructurar la forma en que los datos se procesarán, almacenarán y utilizarán, y cómo se accederá a ellos. Al definir desde el inicio cómo se relacionarán entre sí y pondrán en movimiento datos específicos, es posible diseñar la manera en que los datos fluirán, así como controlar el flujo de datos para garantizar su protección en todo el sistema.

## **BASE DE DATOS**

Una base de datos es un conjunto grande y estructurado de datos digitales organizados de manera tal que se permita una búsqueda, acceso y actualización rápidos de los datos que contiene.

## **BRONTOBYTE**

Un brontobyte es un término de medición no oficial que corresponde a una cantidad extraordinariamente enorme de datos. Generalmente, se considera que un brontobyte es una unidad equivalente a 1000 yottabytes y se representa con un 1 seguido de 27 ceros.

## **CALIDAD DE LOS DATOS**

La calidad de los datos es una métrica que se utiliza para definir el valor de los datos para el usuario. Hace referencia a la fiabilidad, eficiencia y mérito de los datos en la toma de decisiones, la planificación y las operaciones.

## **CAMBIOS DISRUPTIVOS**

Los cambios disruptivos son los cambios grandes y fundamentales en la sociedad y los negocios, a menudo facilitados por tecnologías nuevas y transformadoras que establecen todo un contexto nuevo para la forma en que trabajamos, vivimos, jugamos y creamos valor. La innovación de los datos se describe a menudo como una tecnología que posibilita cambios disruptivos.

## **CENTRO DE DATOS**

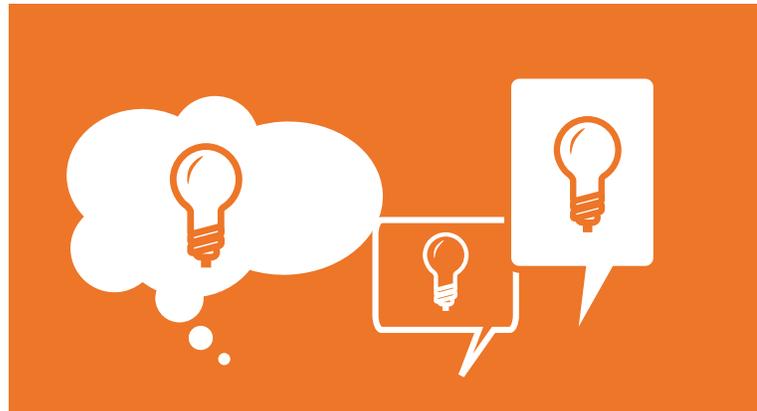
Un centro de datos es un lugar físico que alberga una gran cantidad de servidores en red y repositorios de almacenamiento de datos; normalmente, se utiliza para el almacenamiento y procesamiento remotos de grandes cantidades de datos a los que se puede acceder de manera remota. Se estima que hay medio millón de centros de datos en todo el mundo, muchos de ellos conforman la nube.

## **CIENCIA DE LOS DATOS**

La ciencia de los datos es una disciplina que incorpora estadísticas, visualización de los datos, programación informática, minería de datos, aprendizaje automático e ingeniería de bases de datos para extraer conocimientos coherentes que puedan resolver problemas complejos.

## **CIENTÍFICO DE LOS DATOS**

Un científico de los datos es alguien que puede combinar el conocimiento humano, el know-how matemático y las herramientas tecnológicas para extraer significado de los datos; por ejemplo, al desarrollar y utilizar algoritmos informáticos.



## CONJUNTO DE DATOS

Un conjunto de datos es un grupo de conjuntos relacionados de información, normalmente elementos independientes, en forma tabular que se puede manipular como una unidad.

## DATOS

Los datos son información sin procesar ni organizar que se pueden manipular digitalmente para representar condiciones, objetos o ideas. Los tipos comunes de datos incluyen cifras de venta, resultados de investigaciones de marketing, lectura de sensores meteorológicos, o una lista de ciudades y sus poblaciones. Actualmente, generamos una cantidad estimada de 2,5 cuatrillones de bytes de datos por día.

## DATOS EN TIEMPO REAL

Los datos en tiempo real son datos que se manipulan a medida que se crean. Normalmente se crean, procesan, almacenan y analizan en milisegundos. Los datos en tiempo real pueden incluir de todo, desde los precios de la bolsa hasta la velocidad de una rueda cuando se utiliza en un sistema de frenos antibloqueo.

## DATOS ESTRUCTURADOS

Los datos estructurados están sumamente organizados y, en general, dispuestos en filas y columnas que facilitan su búsqueda y manipulación.

## DATOS GENERADOS POR COMPUTADORA

Los datos generados por computadora hacen referencia a datos que una computadora produce de manera automática sin la intervención de los seres humanos — como un archivo de registro de una computadora, datos de telemetría de un satélite o datos de una máquina industrial.

## DATOS INCORRECTOS

Son datos incorrectos o faltantes. Puede tratarse de algo tan simple como una dirección de calle incorrecta, pero los datos incorrectos les cuentan miles de millones de dólares a las empresas de Fortune 1000 cada año.

## DATOS NO ESTRUCTURADOS

Los datos no estructurados no tienen una estructura predefinida —por ejemplo, las notas de una reunión. Según algunos cálculos, la información no estructurada podría representar más del 70 al 80 por ciento de todos los datos de una organización.

## DATOS OSCUROS

Los datos oscuros consisten en datos no estructurados y sin explotar que se almacenan sin que se los analice ni procese, y se cree que, de alguna manera, se los descuida o no se los utiliza bien.

## DATOS REDUCIDOS

Los datos reducidos implican aprovechar incluso cantidades reducidas de datos, como los que se encuentran en una encuesta al cliente, para lograr resultados utilizables. En general, se refieren a tamaños de datos tan reducidos que un ser humano puede comprenderlos y analizarlos.

## DATOS SEMIESTRUCTURADOS

Los datos semiestructurados no están estructurados por un modelo de datos formal, como los que se utilizan en las bases de datos, pero proporcionan otros medios para describir los datos y las jerarquías. A menudo, los datos semiestructurados utilizan etiquetas u otros marcadores de datos en lo que a veces se conoce como una estructura autodescriptiva.

## DATOS TRANSACCIONALES

Los datos transaccionales son datos derivados de actividades específicas, como compras financieras, facturas, pagos y datos de envío. Generalmente incluyen una marca de tiempo y brindan soporte para las operaciones diarias de una organización.

## DESIDENTIFICACIÓN

La desidentificación de los datos es el proceso de eliminación de la información que vincula a una persona con un dato particular.



## DETECCIÓN DE ANOMALÍAS

La detección de anomalías es la identificación de datos individuales dentro de un conjunto de datos que no coinciden con un patrón esperado. Las anomalías se denominan también valores atípicos, excepciones o contaminantes en los datos y, con frecuencia, brindan información crítica y útil.

## DETECCIÓN DE DATOS ATÍPICOS

Un dato atípico es un dato que se desvía de manera significativa del promedio general dentro de un conjunto de datos más amplio. Es numéricamente lejano al resto de los datos; por lo tanto, el valor atípico indica que algo está sucediendo y, por lo general, requiere un análisis adicional. (Ver también la Detección de anomalías.)

## EXABYTE

Un exabyte es una unidad de almacenamiento de datos enorme —un 1 seguido de 18 ceros. Para ponerla en contexto, hoy en día creamos un exabyte de información nueva a diario.

## FUENTE DE DATOS

Una fuente de datos es el lugar principal de donde provienen los datos, por ejemplo, de una base de datos, una hoja de cálculo o un flujo de datos.

## HADOOP

Hadoop es un marco de software de origen abierto que se creó para permitir el procesamiento y almacenamiento de cantidades gigantes de datos entre sistemas de archivos distribuidos.

## HERRAMIENTAS DE ACUMULACIÓN DE DATOS

Las herramientas de acumulación de datos transforman los datos esparcidos de múltiples fuentes en un conjunto nuevo y único de datos.

## INFORMÁTICA COGNITIVA

La informática cognitiva es el proceso mediante el que se combinan grandes cantidades de información con técnicas de aprendizaje automático, tecnologías de reconocimiento de patrones y, a veces, el procesamiento del lenguaje natural para imitar la forma en que funciona el cerebro. Con frecuencia, estos sistemas pueden aprender e interactuar con las personas al combinar fuentes de información con el contexto y el conocimiento.

## INTELIGENCIA ADAPTATIVA

La inteligencia adaptativa es la inteligencia computacional que no solo implica el procesamiento estadístico de los datos, sino que lo combina con datos que contienen inteligencia de dominios específicos. Al combinar modelos de comportamiento inteligente con conocimientos expertos, los sistemas pueden aprender mejor a partir de los ejemplos y adaptarse a situaciones nuevas.

## INTELIGENCIA AMBIENTAL

La inteligencia ambiental representa la visión de que la tecnología será cada vez más invisible y pasará más inadvertida a medida que la creciente cantidad de informática extendida y de bajo costo se arraigue en el mundo que nos rodea. La inteligencia ambiental estaría a nuestra disposición cada vez que la necesitemos, ya que, literalmente, todo estará conectado, será inteligente y tendrá un buen nivel de respuesta.

## INTELIGENCIA EMPRESARIAL (BUSINESS INTELLIGENCE, BI)

La inteligencia empresarial hace referencia al conjunto de tecnologías y aplicaciones que transforman los datos sin procesar en conocimientos operativos que pueden mejorar el desempeño comercial y la toma de decisiones.



## INTERNET DE LAS COSAS

La Internet de las cosas describe un mundo donde los dispositivos ordinarios se tornan mucho más inteligentes, y se conectan a Internet para extender la revolución inteligente desde la palma de la mano hacia el mundo que nos rodea. Como todas las cosas que se puedan conectar, se conectan, algunos la describen más acertadamente como la Internet de todas las cosas. Según un cálculo aproximativo, solo hemos conectado cerca de un 1 por ciento de las cosas que se pueden conectar a escala mundial. Para el 2020, se estima que 50 mil millones de dispositivos estarán conectados a Internet.

## LIMPIEZA DE DATOS

La limpieza de datos es el proceso de examinación y revisión de los datos sin procesar para encontrar y eliminar datos duplicados, corregir errores, incorporar datos faltantes, eliminar datos dañados y brindar mayor coherencia.

## LIMPIEZA DE LOS DATOS

La limpieza de los datos es el proceso de detección y corrección o eliminación de datos erróneos, incompletos o duplicados de una base de datos.

## MACRODATOS

Se trata de un término abarcador que, con frecuencia, se refiere al proceso de aplicación de análisis informáticos a cantidades inmensas de datos a fin de descubrir conocimientos nuevos y mejorar la toma de decisiones. A menudo, describe conjuntos de datos con un volumen tan grande, de una naturaleza tan variada y que se mueven a una velocidad tan rápida que es difícil procesarlos mediante las herramientas de procesamiento de datos tradicionales.

## METADATOS

Los metadatos son datos sobre datos. Pueden incluir información de resumen básica acerca de los datos, como su autor, la fecha de creación, el tamaño del archivo y la última fecha de modificación.

## MINERÍA DE DATOS

La minería de datos es el proceso mediante el cual se utilizan algoritmos informáticos poderosos para encontrar patrones o conocimientos en conjuntos grandes de datos.

## MODELADO DE PREDICCIÓN

El modelado de predicción es el proceso de desarrollo de un modelo que, muy probablemente, predirá una tendencia, comportamiento futuro o resultado —a menudo, mediante la comparación de eventos actuales con eventos del pasado.

## MOTOR DE RECOMENDACIONES

Un motor de recomendaciones es un algoritmo informático que realiza recomendaciones y sugerencias o puede personalizar algo para el usuario sobre la base de una variedad de patrones de datos que frecuentemente se derivan mediante técnicas de aprendizaje automático.

## NUBE

La nube es un término amplio que se refiere a cualquier aplicación, servicio o dato hospedados de manera remota. En general, su existencia es posible gracias a grupos grandes de servidores remotos en red que permiten un acceso extendido y a petición a los recursos informáticos o de almacenamiento.

## PETABYTE

Un petabyte es una medida enorme de capacidad de almacenamiento que se representa con un 1 seguido de 15 ceros, o un millón de gigabytes. Un petabyte es, aproximadamente, cuatro veces la cantidad de datos que contiene la Biblioteca del Congreso.

## RECONOCIMIENTO DE PATRONES

El reconocimiento de patrones es el proceso de búsqueda e identificación de patrones en los datos. Puede ser una tarea simple, como identificar un conjunto de secuencias que se repite dentro de una secuencia de ADN; puede tratarse de encontrar un patrón en la forma en que dos conjuntos de datos interactúan para descubrir si hay un patrón que conecte un evento con otro; o, con la ayuda del aprendizaje automático, puede tratarse de buscar patrones más complejos, como encontrar caracteres numéricos en una foto.



## SEGURIDAD DE LOS DATOS

La seguridad de los datos es la práctica de proteger los datos de la destrucción, el uso incorrecto o el acceso no autorizado. Las medidas adecuadas de seguridad de los datos pueden ayudar a prevenir violaciones de los datos, garantizar la integridad de los datos y proteger la privacidad. Con frecuencia, implica un enfoque combinado en la gente, los procesos y la tecnología.

## SISTEMA HEREDADO

Un sistema heredado es cualquier computadora, aplicación o tecnología que pasó de moda o es obsoleta, pero se sigue usando porque realiza una función necesaria de manera adecuada.

## TERABYTE

Un terabyte es una medida de datos representada por un 1 seguido de 12 ceros. Ahora es común encontrar discos duros de terabytes en equipos domésticos y laborales, o acceder a ellos mediante la nube. Para ponerlo en contexto, un terabyte puede almacenar alrededor de 300 horas de video de alta definición.

## TOMA DE DECISIONES GUIADA POR DATOS

Las empresas que utilizan la toma de decisiones guiada por los datos recopilan, procesan y analizan los datos para respaldar decisiones críticas. Una investigación de Eric Brynjolfsson, un economista de la Sloan School of Management, Massachusetts Institute of Technology, muestra que las empresas que utilizan la toma de decisiones guiada por datos tienen un incremento de entre el 5 y 6 por ciento en la productividad.

## VARIEDAD

La variedad, una de las cuatro “V” que definen la innovación de los datos, representa los diversos tipos de datos, a menudo de diferentes fuentes, que se combinan y analizan para producir conocimientos. La variedad de tipos de datos que se procesan actualmente en las aplicaciones puede incluir bases de datos textuales, datos de transacciones, datos de transmisión por secuencia, imágenes, audio y video.

## VELOCIDAD

La velocidad, una de las cuatro “V” que definen la innovación de los datos, es la rapidez con que se crean, almacenan, analizan y visualizan los datos. Por ejemplo, los grandes depósitos de datos pueden recibir miles de millones de filas de información nueva por día. Los datos que dependen del tiempo deben utilizarse a medida que se transmiten para que se maximice su valor.

## VERACIDAD

La veracidad, una de las cuatro “V” que definen la innovación de los datos, se refiere a la exactitud, certeza y precisión de los datos.

## VIRTUALIZACIÓN DE LOS DATOS

La virtualización de los datos es el proceso para recuperar y manipular diferentes fuentes de datos sin tener que conocer detalles técnicos sobre el lugar donde se encuentran ni cómo están formateados.

## VISUALIZACIÓN DE LOS DATOS

La visualización de los datos implica crear una representación visual de los datos a fin de derivar significado o comunicar información de manera más efectiva.

## VOLUMEN

El volumen, una de las cuatro “V” que definen la innovación de los datos, se refiere a la cantidad de datos procesados —que va de los megabytes a los brontobytes.

## YOTTABYTES

Un yottabyte es una medida muy grande de almacenamiento de datos representada por un 1 seguido de 24 ceros. Para ponerlo en contexto, un yottabyte representa la cantidad de datos almacenada en 250 billones de DVD.

## ZETTABYTES

Un zettabyte es una medida de almacenamiento representada por un 1 seguido de 21 ceros. Hasta el 2013, se estimaba que la World Wide Web había alcanzado los 4 zettabytes. Para el 2016, se proyecta que más de un zettabyte de datos atravesará nuestras redes globalmente de manera diaria.

# NOTAS FINALES

- <sup>1</sup> Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. <https://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making>
- <sup>2</sup> Un aumento de la productividad del 1 por ciento puede parecer poco, pero, tal como lo expresa Jeff Immelt, Director Ejecutivo de GE: “tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life.” Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. Web. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- <sup>3</sup> Para el 2020, se estima que 50 mil millones de dispositivos estarán conectados a Internet, y esto posibilitará la Internet de las cosas. Evans, Dave. *The Internet Of Things How The Next Evolution Of The Internet Is Changing Everything*. Cisco Internet Business Solutions Group (IBSG), 2011. Web. White Paper. [http://www.cisco.com/web/about/ac79/docs/innov/IoT\\_IBSG\\_0411FINAL.pdf](http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf)
- <sup>4</sup> “Bringing big data to the enterprise.” IBM. <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- <sup>5</sup> Golub, Ben. “Enough Data to Fill a Stack of DVDs to the Moon (and Back).” *ComputerWorld* 2011. Web. [http://blogs.computerworld.com/18351/a\\_stack\\_of\\_dvds\\_to\\_the\\_moon\\_and\\_back](http://blogs.computerworld.com/18351/a_stack_of_dvds_to_the_moon_and_back)
- <sup>6</sup> Corry, Will. “BIG Data / The Volume Of Business Data Worldwide, Across All Companies, Doubles Every 1.2 Years, According To Estimates.” *The Marketing Blog* 2012. Web. <http://www.themarketingblog.co.uk/2012/10/big-data-the-volume-of-business-data-worldwide-across-all-companies-doubles-every-1-2-years-according-to-estimates/>
- <sup>7</sup> “Digital Imaging in the Cloud.” *There Magazine* 2012: 16. Web. [http://www.agfahealthcare.com/he/global/en/binaries/THERE\\_12\\_tcm541-95647.pdf](http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf)
- <sup>8</sup> El 787 utiliza sensores de datos para reducir el uso de combustible y controlar los sistemas e, incluso, emplea acelerómetros en la parte delantera del avión para contrarrestar las turbulencias. Si los sensores registran una caída abrupta, le indican de inmediato a los alerones que se ajusten (en nanosegundos) y, al hacer eso, lo que solía ser una caída de 9 pies en un avión más antiguo se puede reducir a tan solo 3 pies en el 787, para lograr un vuelo mucho más cómodo. Humphries, Matthew. “The Boeing 787 produces over 500GB of Data during Every Flight.” *Geek.com* 2013. Web. Gosling, Kevin. “E-Enabled Capabilities of the 787 Dreamliner.” *Aero Quarterly* 2009: 22-24. [http://www.boeing.com/commercial/aeromagazine/articles/qtr\\_01\\_09/pdfs/AERO\\_Q109\\_article05.pdf](http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf)
- <sup>9</sup> El fabricante de motores de aviones para pasajeros, GE, afirma que los datos del motor le permiten detectar aspectos como posibles desperfectos 2000 veces más rápido que antes. Hardy, Quentin. “What Cars Did for Today’s World, Data May Do for Tomorrow’s.” *New York Times* 2014. B7. Print. [http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?\\_php=true&\\_type=blogs&\\_php=true&\\_type=blogs&module=BlogPostTitle&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&\\_r=1&](http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPostTitle&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&_r=1&)
- <sup>10</sup> Henschen, Doug. “Bad Winter Weather Meets Big Data Prediction.” *InformationWeek* 2014. Web. <http://www.informationweek.com/big-data/software-platforms/bad-winter-weather-meets-big-data-prediction/d/d-id/1114192>
- <sup>11</sup> Groenfeldt, Tom. “At NYSE, The Data Deluge Overwhelms Traditional Databases.” *Forbes* 2013. Web. <http://www.forbes.com/sites/tomgroenfeldt/2013/02/14/at-nyse-the-data-deluge-overwhelms-traditional-databases>  
“NYSE Euronext Adapting To Market Changes with Near-Real-Time Insight into Information.” IBM Corporation, 2013. Print. NYSE Euronext. <http://www.ibmbigdatahub.com/sites/default/files/document/NYSE-Euronext-IMC14787USEN.PDF>
- <sup>12</sup> Mayyasi, Alex. “Why UPS Trucks Don’t Turn Left.” *Priceonomics.com* 2014. Web. <http://priceonomics.com/why-ups-trucks-dont-turn-left/>
- <sup>13</sup> *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper*. Cisco Visual Networking Index, 2015. Print. [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white\\_paper\\_c11-520862.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html)
- <sup>14</sup> “LSST General Public FAQs.” LSST.org 2015. <http://www.lsst.org/lsst/faq>
- <sup>15</sup> Quazi, Ed. “Next Generation Sequencing and the Big Data Bottleneck.” *Arkivum* 2014. Web. <http://arkivum.com/next-generation-sequencing-and-the-big-data-bottleneck/>
- <sup>16</sup> Valerio, Pablo. “Internet Of Things: 50 Billion Is Only The Beginning.” *EE Times* 2014. Web. [http://www.eetimes.com/document.asp?doc\\_id=1321229](http://www.eetimes.com/document.asp?doc_id=1321229)

- <sup>17</sup> En 1980, la regla general era que se necesitaba un administrador de datos para 1 GB de almacenamiento. En esa época, un GB de disco costaba alrededor de un millón de dólares, de modo que era comprensible contar con alguien que optimizara y supervisara el uso del espacio de disco. Gray, Jim, and Prashant Shenoy. "Rules Of Thumb in Data Engineering." Redmond, WA: Microsoft Research Advanced Technology Division, 2009. Print. Technical Report. [http://research.microsoft.com/pubs/68636/ms\\_tr\\_99\\_100\\_rules\\_of\\_thumb\\_in\\_data\\_engineering.pdf](http://research.microsoft.com/pubs/68636/ms_tr_99_100_rules_of_thumb_in_data_engineering.pdf)
- <sup>18</sup> Wohlsen, Marcus. "Dropbox Slashes Its Price As The Cost Of A Gigabyte Nears Zero." *Wired* 2014. Web. <http://www.wired.com/2014/08/dropbox-plan-to-stay-relevant/>
- <sup>19</sup> De más de \$200.000 por gigabyte en 1980 (incluso hasta millones de dólares) a \$0,02 por gigabyte en 2013. Meecker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. Presentation. <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- <sup>20</sup> En función del nivel de eficiencia promedio en el consumo de combustible de los autos de turismo en 1980 (24,3 mpg), esto permite que una persona compre 10 millones de veces más cantidad por el mismo precio; serían 10 millones de galones de combustible, que es la cantidad que se necesitaría para 243 millones de millas de viaje. Si la circunferencia de la Tierra es de 24.901 millas, una persona podría circular alrededor de la tierra 9758 veces, o casi 10.000 veces. Departamento de Transporte de los EE. UU. Table 4-23: Average Fuel Efficiency Of U.S. Light Duty Vehicles. Washington, DC: Bureau of Transportation Statistics, 2013. Print. [http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national\\_transportation\\_statistics/html/table\\_04\\_23.html](http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national_transportation_statistics/html/table_04_23.html)
- <sup>21</sup> Savitz, Eric. "Big Data: The Hidden Opportunity." *Forbes* 2012. Web. <http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/>
- <sup>22</sup> La capacidad tecnológica mundial per cápita para almacenar información se duplicó aproximadamente cada 40 meses desde la década del ochenta según una investigación realizada por Martin Hilbert y Priscila López. Hilbert, M., and P. Lopez. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332.6025 (2011): 60-65. Web. <http://www.sciencemag.org/content/332/6025/60>
- <sup>23</sup> Los costos de almacenamiento han caído de \$569 por gigabyte de almacenamiento en 1992 a \$0,02 por gigabyte en 2013 —a un índice de alrededor del 38 por ciento anual. Meecker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. Presentation. <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- <sup>24</sup> Lohr, Steve. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle To Insights." *New York Times*. 2014: B4. Print. [http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?\\_r=0](http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0)
- <sup>25</sup> "Data, data everywhere," *The Economist*, 25 de febrero del 2010. <http://www.economist.com/node/15557443>
- <sup>26</sup> Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal*. 2013. Print. <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- <sup>27</sup> Los informes de sequías en Angola de 2006 desencadenaron una advertencia acerca de posibles brotes de cólera en el país, debido a que eventos previos le habían enseñado al sistema que los brotes de cólera eran más probables durante los años siguientes a las sequías. Las advertencias del sistema fueron correctas entre el 70 y 90 por ciento de los casos. Simonite, Tom. "Software Predicts Tomorrow's News by Analyzing Today's And Yesterday's." *MIT Technology Review* 2013. Print. <http://www.technologyreview.com/news/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/>
- <sup>28</sup> "Ten Big Data Case Studies in a Nutshell." TechTarget, 2013. Print. Essential Guide. <http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>
- <sup>29</sup> "Big Data to Predict Offshore Accidents, Tsunamis and Other Natural Disasters." *Predictive Analytics Today*. 2013 Web. <http://www.predictiveanalyticstoday.com/big-data-predict-shore-accidents-tsunamis-natural-disasters/>
- <sup>30</sup> "New Developments in Big Data Visualization." *USTelecom Media* 2014. Web. <http://www.ustelecom.org/blog/new-developments-big-data-visualization#sthash.HefD5H52.dpuf>
- <sup>31</sup> Quinn, Tom. "New and Unexpected Uses for Scoring Technology." *Credit Score Blog* 2011. Web. <http://blog.credit.com/2011/06/new-and-unexpected-uses-for-scoring-technology/>
- <sup>32</sup> Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal* 2013: Print. <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- <sup>33</sup> Sobrecarga de datos: En la actualidad, un médico clínico experimentado necesita casi 2 millones de datos para ejercer la medicina, y los doctores se suscriben a un promedio de siete revistas médicas, lo que representa más de 2500 artículos nuevos al año; así, resulta casi imposible mantenerse al corriente de la información

más actualizada acerca de diagnósticos, pronósticos, tratamientos y cuestiones médicas afines. "Clinical Decisions Support Systems: The Time Has Come..." Frost & Sullivan, 2009. Print. Market Insight. <http://www.frost.com/prod/servlet/cio/181298788>

- <sup>34</sup> "Clinical Decisions Support Systems: The Time Has Come..." Frost & Sullivan, 2009. Print. Market Insight. <http://www.frost.com/prod/servlet/cio/181298788>
- <sup>35</sup> "Kinect Sign Language Translator Expands Communication Possibilities." *Microsoft Research* 2013. Web. <http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx>
- <sup>36</sup> "Brats 2012 - Multimodal Brain Tumor Segmentation Challenge." *CodaLab*, 2012. Print. <https://www.codalab.org/competitions/191>
- <sup>37</sup> Grady, Denise. "3-D Mammography Test Appears To Improve Breast Cancer Detection Rate." *New York Times* 2014: p. A1 Print. [http://www.nytimes.com/2014/06/25/health/breast-cancer-3d-mammography-test-x-ray.html?emc=edit\\_th\\_20140625&nl=todaysheadlines&nid=435891&r=0](http://www.nytimes.com/2014/06/25/health/breast-cancer-3d-mammography-test-x-ray.html?emc=edit_th_20140625&nl=todaysheadlines&nid=435891&r=0)
- <sup>38</sup> "The game-changing technology that's transforming manufacturing." *Manufacturing Weekly*, January 31, 2014. <http://web.archive.org/web/20140131233544/http://www.manufacturingweekly.com/supercomputers/>
- <sup>39</sup> "The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend." *Cloud Platform News Bytes Blog* 2015. Web. <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- <sup>40</sup> Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- <sup>41</sup> Un aumento de la productividad del 1 por ciento puede parecer poco, pero, tal como lo expresa Jeff Immelt, Director Ejecutivo de GE: "tell an oil guy you can use software to save him one percent on something, and that guy will be your friend for life." Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. Web. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- <sup>42</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>43</sup> Gartner, "Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally To Support Big Data By 2015." 2012. Print. <http://www.gartner.com/newsroom/id/2207915>
- <sup>44</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>45</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>46</sup> According to ESG research, data managed per hospital is expected to increase from 168 terabytes in 2010 to 6 terabytes by 2015. "Digital Imaging in the Cloud." *There Magazine* 2012: 16. Print. [http://www.agfahealthcare.com/he/global/en/binaries/THERE\\_12\\_tcm541-95647.pdf](http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf)
- <sup>47</sup> Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. Print. [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
- <sup>48</sup> Los investigadores enseñaron un algoritmo de aprendizaje automático sobre datos de 133.000 pacientes. El modelo aún necesita que se realice más trabajo para reducir los falsos positivos. Rutkin, Aviva. "Machine Predicts Heart Attacks 4 Hours Before Doctors - New Scientist." *New Scientist*. 2014. Web. <http://www.newscientist.com/article/mg22329814.400-machine-predicts-heart-attacks-4-hours-before-doctors.html>
- <sup>49</sup> El modelo de fusión híbrido moderno de Ford genera hasta 25 GB de datos por hora. Hemsoth, Nicole. "How Ford Is Putting Hadoop Pedal To The Metal." *Datanami*. 2013. Web. [http://www.datanami.com/2013/03/16/how\\_ford\\_is\\_putting\\_hadoop\\_pedal\\_to\\_the\\_metal/](http://www.datanami.com/2013/03/16/how_ford_is_putting_hadoop_pedal_to_the_metal/)
- El Chevy Volt contiene más de 10 millones de líneas de códigos de software, y la profesión de desarrollador de software es una de las profesiones técnicas de más rápido crecimiento al sudeste de Michigan, una región con una larga reputación por su capacidad de producción. Trop, Jaclyn. "Detroit, Embracing New Auto Technologies, Seeks App Builders." *New York Times*. June 30, 2013. <http://www.nytimes.com/2013/07/01/technology/detroit-embracing-new-auto-technologies-seeks-app-builders.html>
- <sup>50</sup> Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. Print. <http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&r=0>
- <sup>51</sup> El 787 utiliza sensores de datos para reducir el uso de combustible y controlar los sistemas e, incluso, emplea acelerómetros en la parte delantera del avión para contrarrestar las turbulencias. Si los sensores registran

- una caída abrupta, le dicen de inmediato a los alerones que se ajusten (en nanosegundos) y, al hacer eso, lo que solía ser una caída de 9 pies en un avión más antiguo se puede reducir a tan solo 3 pies en el 787, para lograr un vuelo mucho más cómodo. Gosling, Kevin. "E-Enabled Capabilities of the 787 Dreamliner." *Aero Quarterly* 2009: 22-24. [http://www.boeing.com/commercial/aeromagazine/articles/qtr\\_01\\_09/pdfs/AERO\\_Q109\\_article05.pdf](http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf)
- <sup>52</sup> El fabricante de motores de aviones para pasajeros, GE, afirma que los datos del motor le permiten detectar aspectos como posibles desperfectos 2000 veces más rápido que antes. Hardy, Quentin. "What Cars Did for Today's World, Data May Do for Tomorrow's?" *New York Times* 2014: B7. Print. [http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?\\_php=true&\\_type=blogs&\\_php=true&\\_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&\\_r=1&](http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs&region=Body&_r=1&)
- <sup>53</sup> Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. [https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy\\_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf](https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf)
- <sup>54</sup> El sector de producción almacenó alrededor de 2 exabytes de datos nuevos solo en 2010. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. Print. [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)
- <sup>55</sup> Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. Print. [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)
- <sup>56</sup> "Data Smart' Strategies for Customers Are Yielding 'Early But Impressive Returns.'" *Microsoft Research the Fire Hose* 2014. Web. <http://blogs.microsoft.com/firehose/2014/05/22/data-smart-strategies-for-customers-are-yielding-early-but-impressive-returns/>
- <sup>57</sup> Somers, Dan. "Manufacturing 4.0 – From Industrialization to Data-Driven Product Lifecycle." *Citizenekkk*. 2013. Web. <http://citizenekkk.com/2013/11/05/manufacturing-4-0-industrialisation-data-driven-product-lifecycle/>
- <sup>58</sup> McKinsey informa que al usar estas técnicas de diseño habilitadas por los datos, Toyota pudo eliminar el 80 por ciento de los defectos antes de construir el primer prototipo físico. Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. Print. [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)
- <sup>59</sup> Findings of the New Intelligent Enterprise Study. IBM 2010. *New Intelligent Enterprise Global Executive Study*. 2010. Print.
- <sup>60</sup> Geron, Tomio. "Cows in the Cloud: The Hot Startup Moving Farmers into the Cloud." *Forbes* 2012. Web; Helmer, Jodi. "Get Ready For Robot Farmers." Yahoo 2014. Web. <https://www.yahoo.com/tech/get-ready-for-robot-farmers-100613764059.html>
- <sup>61</sup> De-Identification of Personally Identifiable Information, National Institute of Science and Technology, DRAFT NISTIR 8053 (April 2015).
- <sup>62</sup> Cavoukian, Ph.D., Ann, and El Emam, Ph.D., Khaled, Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy, Information and Privacy Commissioner of Ontario, (June 2011); Cavoukian, Ph.D., Ann, and Daniel Castro Castro. *Big Data And Innovation, Setting The Record Straight: De-Identification Does Work*. ITIF, 2014. Print. <http://www2.itif.org/2014-big-data-deidentification.pdf>
- <sup>63</sup> See for example, Microsoft's add on protecting privacy as their priority [https://www.youtube.com/watch?feature=player\\_embedded&v=bt51MWll1oY](https://www.youtube.com/watch?feature=player_embedded&v=bt51MWll1oY)
- <sup>64</sup> Apple, en Solicitudes de información del gobierno, observa que la empresa ha incorporado cifrado de vanguardia en el sistema operativo de su iPhone, de modo que: "your personal data such as photos, messages (including attachments), email, contacts, call history, iTunes content, notes, and reminders is placed under the protection of your passcode," at [www.apple.com/privacy/government-information-requests/](http://www.apple.com/privacy/government-information-requests/)
- <sup>65</sup> Hachman, Mark, "Microsoft's updated privacy policy makes it clear it's not selling ads against your words," *PCWorld*, June 11, 2014, <http://www.pcworld.com/article/2362130/microsofts-updated-privacy-policy-makes-it-clear-its-not-selling-ads-against-your-words.html>,

Timberg, Craig. "Newest Androids Will Join iPhone In Offering Default Encryption, Blocking Police." *Washington Post* 2014: Print. <http://www.washingtonpost.com/blogs/the-switch/wp/2014/09/18/newest-androids-will-join-iphones-in-offering-default-encryption-blocking-police/>

- <sup>66</sup> El agente de datos Acxiom inaugura un sitio web de datos enfocados en el consumidor, ofrece la opción de exclusión: <http://cir.ca/news/acxiom-gives-consumers-data-peek>
- <sup>67</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>68</sup> McKinsey Global Institute. *Internet Matters: The Net's Sweeping Impact On Growth, Jobs, And Prosperity*. McKinsey & Co., 2011. Print.
- <sup>69</sup> Manyika, James et al. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. Print. [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)
- <sup>70</sup> According to Salaries of Data Scientists, an April 2014 study from Burtch Works.
- <sup>71</sup> Bagley, Rebecca. "How The Cloud And Big Data Are Changing Small Business." *Forbes* 2014. Web. <http://www.forbes.com/sites/rebeccabagley/2014/07/15/how-the-cloud-and-big-data-are-changing-small-business/>
- <sup>72</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>73</sup> Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- <sup>74</sup> Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- <sup>75</sup> Gerbis, Nicholas. "10 Correlations That Are Not Causations." *HowStuffWorks*. 2015. Web. <http://science.howstuffworks.com/innovation/science-questions/10-correlations-that-are-not-causations.htm>
- <sup>76</sup> Vesset, Dan, Henry D. Morris, and John F. Gantz. *Capturing the \$1.6 Trillion Data Dividend*. IDC, 2014. Print. IDC White Paper.
- <sup>77</sup> Westerman, George, Didier Bonnet, and Andrew McAfee, "The Advantages of Digital Maturity." November 2012, MIT Sloan.
- <sup>78</sup> "Meeting the Big Data Challenge: Don't Be Objective." *Forbes* 2013. Web. <http://www.forbes.com/sites/darden/2013/02/01/meeting-the-big-data-challenge-dont-be-objective/>
- <sup>79</sup> IDG Enterprise's 2014 Big Data research. IDG. *CEOs Call for Big Data and IT Continues To Lead Investment Decisions*. 2014. Print. <http://www.idgenterprise.com/press/ceos-call-for-big-data-and-it-continues-to-lead-investment-decisions>
- <sup>80</sup> Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. Print. <http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&r=0>
- <sup>81</sup> Clemens, Samuel. "7 Facts about Data Quality [Infographic]." *InsightSquared*. January 3, 2012. Web. <http://www.insightsquared.com/2012/01/7-facts-about-data-quality-infographic/>
- <sup>82</sup> Economist Intelligence Unit. *Big Data Harnessing a Game-Changing Asset*. SAS, 2011. Web. [http://www.sas.com/resources/asset/SAS\\_BigData\\_final.pdf](http://www.sas.com/resources/asset/SAS_BigData_final.pdf)
- <sup>83</sup> "The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend." *Blog de noticias de la plataforma en la nube* 2015. Web. <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- <sup>84</sup> BSA/IPSOS Global Data Analytics Poll, November 2014, [www.bsa.org/datasurvey](http://www.bsa.org/datasurvey)
- <sup>85</sup> Las inversiones en TI en toda la economía de los EE. UU., incluidas las ventas minoristas, durante el alto crecimiento de la década del noventa le sumaron entre un 1 y 2 por ciento a la tasa compuesta de crecimiento anual de la productividad de los EE. UU. Farrell, Diana et al. *How IT Enables Productivity Growth*. San Francisco: McKinsey Global Institute High Tech Practice, 2002. Print. [http://www.mckinsey.com/insights/business\\_technology/how\\_it\\_enables\\_productivity\\_growth](http://www.mckinsey.com/insights/business_technology/how_it_enables_productivity_growth)
- <sup>86</sup> National Bureau of Economic Research. NBER Working Paper No. 18315: *Is U.S. Economic Growth Over? Faltering Innovation Confronts The Six Headwinds*. 2012. Print. <http://www.nber.org/papers/w18315>
- <sup>87</sup> Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. Web. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>

- <sup>88</sup> Gertner, Joey. "GE for Making the 'Internet of Things' Real." *Fast Company* 2014. Web. <http://www.fastcompany.com/most-innovative-companies/2014/ge>
- <sup>89</sup> Evans, Peter C., and Marco Annunziata. *Pushing the Boundaries of Minds and Machines*. GE, 2012. Web. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- <sup>90</sup> "City Of Barcelona Realizes Vision of Innovative City Governance with Cloud, Devices, and Apps." Customers. microsoft.com. 2014. Web. <https://customers.microsoft.com/Pages/Home.aspx>
- <sup>91</sup> "Autodesk the Gallery Masdar Headquarters Positive Energy Building." *Autodesk.com*. 2015. Web. <http://www.autodesk.com/gallery/exhibits/currently-on-display/adrian-smith-gordon-gill-architecture-masdar-headquarters>
- <sup>92</sup> Bunge, Jacob. "Big Data Comes To The Farm, Sowing Mistrust." *Wall Street Journal* 2014. Print. <http://online.wsj.com/news/articles/SB10001424052702304450904579369283869192124>
- "Supply Chain Management Solution for Fast Moving Consumer Goods & Food Industries - Farm to Fork Tech Mahindra." *Techmahindra*. 2015. Web. [http://www.techmahindra.com/en-US/wwd/solutions/Pages/Enterprises/retail\\_farm\\_fork.aspx](http://www.techmahindra.com/en-US/wwd/solutions/Pages/Enterprises/retail_farm_fork.aspx)
- <sup>93</sup> Entre 2013 y 2020, la división del universo digital entre los mercados establecidos y los emergentes (p. ej., China) cambiará —de un 60 por ciento representado por los mercados establecidos a un 60 por ciento de los datos del universo digital proveniente de los mercados emergentes. EMC Digital Universe. *Executive Summary Data Growth, Business Opportunities, and the IT Imperatives*. IDC, 2014. Print. <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>
- <sup>94</sup> Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. [https://www.accenture.com/us-en/~/\\_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy\\_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf](https://www.accenture.com/us-en/~/_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf)
- <sup>95</sup> Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. [https://www.accenture.com/us-en/~/\\_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy\\_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf](https://www.accenture.com/us-en/~/_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf)
- <sup>96</sup> Levy, Stephen. "Bill Gates and President Bill Clinton on the NSA, Safe Sex, and American Exceptionalism." *Wired* 2013: Print. <http://www.wired.com/2013/11/bill-gates-bill-clinton-wired/2/>
- Chhachhar, Abdul Razaque, and Siti Zobidah Omar. "Use of Mobile Phone among Fishermen for Marketing and Weather Information." *Archives Des Sciences* 65.8 (2012): 107-119. Print. [http://www.academia.edu/4592505/Use\\_of\\_Mobile\\_Phone\\_among\\_Fishermen\\_for\\_Marketing\\_and\\_weather\\_information](http://www.academia.edu/4592505/Use_of_Mobile_Phone_among_Fishermen_for_Marketing_and_weather_information)
- <sup>97</sup> Neuman, William, and Ralph Blumenthal. "New to the Archaeologist's Tool Kit: The Drone." *New York Times* 2014. Print. [http://mobile.nytimes.com/2014/08/14/arts/design/drones-are-used-to-patrol-endangered-archaeological-sites.html?\\_r=1&referrer](http://mobile.nytimes.com/2014/08/14/arts/design/drones-are-used-to-patrol-endangered-archaeological-sites.html?_r=1&referrer)
- <sup>98</sup> "Forty Thousand Relics to Be Digitized In Five Years." *Thanhniem News*. 2010. Web. <http://www.thanhniemnews.com/entertainment/forty-thousand-relics-to-be-digitized-in-five-years-22816.html>
- <sup>99</sup> Long, Jessica, and William Brindley. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. Print. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. [https://www.accenture.com/us-en/~/\\_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy\\_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf](https://www.accenture.com/us-en/~/_media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf)
- <sup>100</sup> Future of Privacy Forum. *Big Data: A Tool for Fighting Discrimination and Empowering Groups*. Future of Privacy Forum and Anti-Defamation League, 2014. Print. <http://www.futureofprivacy.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf>
- <sup>101</sup> Wladawsky-Berger, Irving. "The Changing Nature of Globalization in Our Hyperconnected, Knowledge-Intensive Economy." *Wall Street Journal* 2014. Print. [http://blogs.wsj.com/cio/2014/06/20/the-changing-nature-of-globalization-in-our-hyperconnected-knowledge-intensive-economy/?mod=wsj\\_ciohome\\_cioreport](http://blogs.wsj.com/cio/2014/06/20/the-changing-nature-of-globalization-in-our-hyperconnected-knowledge-intensive-economy/?mod=wsj_ciohome_cioreport)
- <sup>102</sup> Por ejemplo, Argentina, Australia, Brasil, Canadá, Chile, China, Colombia, Costa Rica, Grecia, Hong Kong, India, Indonesia, Corea, México, Perú, Rusia, Suiza y Vietnam han adoptado o propuesto normas que prohíben o restringen considerablemente la posibilidad de que las empresas transfieran información personal fuera de sus territorios domésticos respectivos.

## ACERCA DE BSA | THE SOFTWARE ALLIANCE

BSA | The Software Alliance ([www.bsa.org](http://www.bsa.org)) es la principal defensora de la industria del software a escala mundial ante los gobiernos y el mercado internacional. Sus miembros se encuentran entre las empresas más innovadoras del mundo; las soluciones de software que crean impulsan la economía y mejoran la vida moderna.

Con su sede central en Washington, DC y operaciones en más de 60 países, BSA es pionera en los programas de cumplimiento normativo que promueven el uso de software legítimo y es partidaria de las políticas públicas que fomentan la innovación tecnológica e impulsan el crecimiento de la economía digital.



[www.bsa.org](http://www.bsa.org)

### BSA Sede Central Mundial

20 F Street, NW  
Suite 800  
Washington, DC 20001

T: +1.202.872.5500  
F: +1.202.872.5501

### BSA Asia-Pacífico

300 Beach Road  
#25-08 The Concourse  
Singapur 199555

T: +65.6292.2072  
F: +65.6292.6369

### BSA Europa, Oriente Medio y África

2 Queen Anne's Gate Buildings  
Dartmouth Street  
Londres, SW1H 9BP  
Reino Unido

T: +44.207.340.6080  
F: +44.207.340.6090