

UNA APROXIMACIÓN A BIG DATA

AN APPROACH TO BIG DATA¹

JAVIER PUYOL MORENO

Magistrado (En Excedencia). Abogado. Doctorando UNED

Resumen: Big Data puede ser considerada como una tendencia en el avance de la tecnología que ha abierto la puerta a un nuevo enfoque para la comprensión y la toma de decisiones, que se utiliza para describir las enormes cantidades de datos (estructurados, no estructurados y semi-estructurados) que sería demasiado largo y costoso para cargar una base de datos relacional para su análisis. Así, el concepto de Big Data se aplica a toda la información que no puede ser procesada o analizada utilizando herramientas o procesos tradicionales. En términos generales, Big Data y los procesos que dicha técnica representa tiene un amplio espectro de aplicaciones potenciales. El mayor desafío para la inversión en Big Datas se produce con relación a los proyectos vinculados a la toma de decisiones sobre una gran cantidad de datos en la toma de decisiones, definición de estrategias y la obtención de mejores experiencias sobre los actos de consumo de las personas. El desafío de Big Data consiste en capturar, almacenar, buscar, compartir y agregar valor a los datos poco utilizados o inaccesibles hasta la fecha. No es relevante el volumen de datos o su naturaleza. Lo que importa es su valor potencial, que sólo las nuevas tecnologías especializadas en Big Data pueden explotar. En última

¹ Este estudio ha sido realizado como parte integrante del Trabajo de Tesis Doctoral con el título «Estudio y propuestas de regulación normativa del Cloud Computing» dirigido por el Profesor Dr. Federico Fernández de Buján. Dicho Proyecto de Tesis Doctoral se elabora en la Escuela de Doctorado de la Facultad de Derecho de la Universidad Nacional de Educación a Distancia (UNED).

instancia, el objetivo de esta tecnología es aportar y descubrir un conocimiento oculto a partir de grandes volúmenes de datos.

Palabras clave: Privacidad, Big Data, Dato de carácter personal.

Abstract: Overall Big Data can be seen as a trend in the advancement of technology has opened the door to a new approach to understanding and decision making, which is used to describe huge amounts of data (structured, unstructured and semi-structured) it would take too long and would be costly to load a relational database for analysis. So that, the concept of Big Data applies to all information that can't be processed or analyzed using traditional tools or processes. Broadly, Big Data has a broad spectrum of potential applications. The biggest challenge for investment in Big Data is provided with adequate and trained to execute these projects and make a large amount of data into decisions, strategies and best consumer experiences people. The Challenge of Big Data is to capture, store, search, share, and add value to underutilized or inaccessible data to date and one of the interesting aspects of Big Data is that businesses and this kind of data stored on a key that is not yet been given the appropriate interpretation. No data volume or nature that matters, but their potential value, which only art technologies of Big Data may explode. Ultimately, the aim with this technology is to discover hidden knowledge from large volumes of data.

Keywords: Privacy, Big Data, Personal character data.

Recepción original: 27/03/2014

Aceptación original: 22/04/2014

La digitalización en la que está inmersa la sociedad actual, según la Consultora Mackinsey actual afecta a todos los niveles. No hay vuelta atrás: la fiebre por el uso de diversos dispositivos tecnológicos prolifera, el auge de las redes sociales es un hecho, las ciudades se empiezan a dotar de sensores y, como fruto de todo esto la información de la que disponen individuos y empresas se multiplica sin precedentes. Estamos, sin duda, en la era marcada por el llamado Big Data².

Vivimos en un mundo de grandes volúmenes de datos. Sólo este año, más de un billón de gigabytes de datos nuevos se creará a nivel mundial. Big Data representa una nueva frontera para la innovación, la competitividad y la productividad³, y además, supone un gran desa-

² TicBEAT BIG DATA. «Big Data Spain 2013 calienta motores». 18 de octubre de 2013. <http://bigdata.ticbeat.com/big-data-spain-calienta-motores/>

³ Cfr. Consultoría McKinsey. Rizomática. Blog personal de José López Ponce. <http://www.rizomatica.net/big-data-la-fuente-para-el-conocimiento-en-el-siglo-xxi/>

fío –pero también nuevas oportunidades– para las empresas para elevarse por encima de la competencia⁴.

Cada día en el mundo se generan más de 2.5 exabytes de datos. Esto equivale a 1.000.000 de terabytes⁵. La generación de datos no solo crece, explota. El crecimiento exponencial es tan grande, que el 90% de los datos guardados en la actualidad, han sido creados en los dos últimos años.

Cada segundo sensores, tabletas, teléfonos y sistemas inteligentes generan cantidades de datos que crecen exponencialmente. Actualmente, la mayoría de los datos almacenados por las nuevas tecnologías no tienen más de dos años. Muchos de estos datos no se procesan porque los sistemas tradicionales de computación no son capaces de procesarlos y muchas empresas no tienen una solución unificada para recogerlos y analizarlos.

⁴ Cfr.: «Big Data Transforms Business». An EMC E-book. Big Data for The Predictive Organization. <http://www.emc.com/microsites/ebook/index.htm#/slide-intro>

⁵ En términos generales el Big Data puede ser considerado como a la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis. De tal manera que, el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Sin embargo, Big Data no se refiere a alguna cantidad en específico, ya que es usualmente utilizado cuando se habla en términos de petabytes y exabytes de datos. Entonces cabe plantearse ¿Cuánto es demasiada información de manera que sea elegible para ser procesada y analizada utilizando Big Data? Ello debe analizarse primeramente en términos de bytes:

$$\begin{aligned} \text{Gigabyte} &= 10^9 = 1,000,000,000 \\ \text{Terabyte} &= 10^{12} = 1,000,000,000,000 \\ \text{Petabyte} &= 10^{15} = 1,000,000,000,000,000 \\ \text{Exabyte} &= 10^{18} = 1,000,000,000,000,000,000 \end{aligned}$$

Además del gran *volumen* de información, esta existe en una gran *variedad* de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, video, sistemas GPS, incontables sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la *velocidad* de respuesta sea lo demasiado rápida para lograr obtener la información correcta en el momento preciso. Estas son las características principales de una oportunidad para Big Data. Finalmente señala este autor que es importante entender que las bases de datos convencionales son una parte importante y relevante para una solución analítica. De hecho, se vuelve mucho más vital cuando se usa en conjunto con la plataforma de Big Data. Cfr.: BARRANCO FRAGOSO. IBM Software Mexico. 18 de junio de 2012. <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

En líneas generales, según indica Girardotti, Big Data tiene un gran espectro de posibles aplicaciones⁶. Estas son:

- Sensores inteligentes aplicados a diferentes verticales de la industria, que almacenan continuamente datos de las líneas de producción que son luego analizados para, por ejemplo, mejorar procesos industriales.
- Horas de video grabadas para vigilancia u otros fines.
- Miles de pagos con tarjeta de crédito cada segundo alrededor del mundo.
- Millones de tweets por día. Miles de tweets por segundo.
- Numerosos comentarios en las páginas corporativas de las redes sociales.
- Gigas de archivos de documentos, planos, formularios, y muchos otros tipos de datos desestructurados que son digitalizados para hacer más eficiente su almacenamiento.
- Información de transacciones en la bolsa, cotizaciones de commodities.
- Movimiento de vehículos, carga, seguimiento por GPS. Información del clima: temperatura, presión, humedad, vientos, precipitaciones.

Alvarez Valle señala que estos datos son una oportunidad de negocio millonaria para la mayoría de las empresas. Saber recoger y utilizar grandes cantidades de datos permite a las empresas tomar mejores decisiones y ser más competitivas: esto es el Big Data⁷.

En este sentido, no hay que olvidar una serie de puntos de partidas, trascendentes a los efectos del presente análisis.

Según una encuesta de Gartner⁸ en las prioridades de negocio y de tecnología en general, así como de estrategia, se presenta el siguiente escenario:

⁶ Cfr.: GIRARDOTTI, Alejandro. Es el Regional Product Account Manager Data Services Level 3 Latin America. «Big Data es suelo fértil para nuevos modelos de negocio». Mercado. 5 de junio de 2.013. Obra citada. <http://www.mercado.com.ar/notas/tecnologiacutea-%7C/8012717/%3Cem%3EBig%20Data-%3Cem%3Ees-suelo-frtil-para-nuevos-modelos-de-negocio>

⁷ ALVAREZ VALLE, Javier. «El Big Data: Una Gran Oportunidad». La Nueva España. 11 de noviembre de 2.013. <http://www.lne.es/opinion/2013/11/10/big-data-una-gran-oportunidad/1497242.html>

⁸ Emerging Market Analysis: IT. Mexico, 2012 and beyond Gartner. Julio de 2012.

- Las cinco principales prioridades de negocio son:
 1. Incrementar el crecimiento de la empresa
 2. La atracción y retención de clientes
 3. La creación de nuevos productos y servicios (innovación)
 4. Mejorar la efectividad de marketing y ventas
 5. El mejoramiento del liderazgo empresarial
- Las cinco principales estrategias que fueron mencionadas son:
 1. El desarrollo de soluciones de negocio
 2. Implementación de soluciones en la nube (SaaS, PaaS, IaaS)
 3. Reducir el costo de entrega de TI para la organización
 4. Expandir el uso de información y *analytics*
 5. La implementación de mejoras en los procesos de negocio

Esta actividad comercial, indica López Ponce⁹, implica que los datos provengan de multitud de fuentes distintas vienen de diferentes

⁹ Lo que es cierto es que el Big Data, abre nuevos horizontes para el conocimiento, nunca jamás imaginado y una fuente de innovación hasta ahora desconocida. Erik Brynjolfsson, economista de la Sloan School of Management y responsable del grupo de Productividad digital en el Massachusetts Institute of Technology y Andrew McAfee, en un artículo publicado hace unos días en la revista *The Atlantic*, *The Big Data Boom Is the Innovation Story of Our Time*, para señalar la importancia del Big Data recurren a la siguiente historia (traducción libre del inglés):

«En 1670, en Delfos, Holanda, un científico llamado Anton van Leeuwenhoek hizo algo que muchos científicos llevaban haciendo desde hacía 100 años. Construyó un microscopio. Este microscopio era diferente de los otros, pero no tenía nada de extraordinario. Al igual que muchos inventores, que se había basado en el ingenio de sus predecesores. Pero cuando miró a través de su microscopio, encontró cosas que le parecían extraordinarias. Las denominó «animálculos» ... Eran de los microbios que estaban en las gotas de agua o en la sangre humana y que fueron el origen de la teoría de la microbiología que inspiró a muchos tratamientos médicos.

Es cierto, *«el descubrimiento de Leewenhoek fue crucial para nuestra comprensión de la innovación, no sólo porque cambió la faz de la bioquímica, sino también porque representa un paradigma fundamental en el descubrimiento: los avances en la innovación a menudo se basan en los avances en la medición (yo añadiría en el cambio de dimensión en la observación y el análisis).»*

Brynjolfsson y McAfee ponen el énfasis en la oportunidad que tienen las empresas de medir su actividad y su relación con los clientes con una precisión sin precedentes, sobre todo en la economía digital, donde los clics ofrecen unas perspectivas muy precisas y en tiempo real sobre el comportamiento del consumidor. A cambio, los consumidores están actuando como consultores involuntarios de las empresas. Porque nuestras compras, nuestras búsquedas y nuestra actividad online son trazas para mejorar los procesos. Asimismo, señalan el potencial de la *reality mining* (la minería

fuentes: vídeos, imágenes, redes sociales, transacciones de comercio electrónico, búsquedas en internet, comunicaciones 3G, 4G y GPS, y sensores del internet de las cosas, por mencionar algunos ejemplos.

Estos datos ya son Big Data¹⁰.

de la realidad) para analizar los miles de millones de datos que están generando la actividad de las personas y los dispositivos en las redes facilitando la toma de decisiones de una forma nunca imaginada y que permitirá el desarrollo de nuevas tecnologías de análisis de datos para reemplazar a los expertos humanos que se basan más en la intuición. Aunque discrepo en la generalización de este último punto, porque el dejar en manos de los algoritmos complejos la toma de decisiones puede conducir a efectos no deseados (Véase: Las TIC y la volatilidad de los mercados financieros).

Sin embargo, el Big Data es una potente herramienta para la experimentación, el análisis y la toma de decisiones. Es una oportunidad para experimentar en tiempo real rompiendo con las barreras de los costes y el tiempo requerido en obtener los datos, porque éstos están ahí, de forma masiva, para su explotación. Desde los comportamientos de los consumidores, tal como se ha señalado, hasta los temas reales que preocupan a los ciudadanos en diversos ámbitos, o los comportamientos de todos los agentes que intervienen en los procesos de negocios. Todo depende de los criterios que vayamos a utilizar para explotar unos datos que se duplican cada dos años.

Cfr. Rizomática. Blog personal de José López Ponce. <http://www.rizomatica.net/big-data-la-fuente-para-el-conocimiento-en-el-siglo-xxi/>

En este mismo sentido, por ejemplo cfr.» «Big data»: los datos como valor de negocio». TICbeat. 21 de octubre de 2.013, donde se señala que *Big data* es información, proceso y almacenamiento, pero no solo eso. El diferencial de esta nueva tendencia tecnológica reside en su capacidad como herramienta de apoyo a la toma de decisiones a partir del análisis de los grandes volúmenes de datos estructurados y no estructurados que entran, salen y giran alrededor de las organizaciones; es decir, en su potencial para convertir los datos en valor de negocio. Esta es su naturaleza revolucionaria llamada a transformar la actividad de prácticamente todos los sectores, y donde citando a IDG se señala, además que *Big Data* como un nuevo valor económico basado en la toma de decisiones a partir del análisis de grandes volúmenes de datos procedentes de una amplia variedad de fuentes, desde las aplicaciones empresariales convencionales a los datos móviles, los medios sociales y el Internet de las Cosas, un campo aún incipiente donde en 2020 convivirán 212.000 millones de dispositivos conectados. Se estima que el 23% (643 exabytes) de todo este universo digital sería susceptible de ser aprovechado por *Big Data*; sin embargo, hoy sólo el 3% de los datos potencialmente útiles está clasificado, y menor es aún la tasa de los que están siendo analizados. El potencial de esta nueva revolución, como ya consideran *Big data* tanto expertos en tecnología como los más visionarios economistas, es pues enorme. Sus ventajas para el negocio en áreas como la gestión de las relaciones con el cliente, el desarrollo de nuevos productos, la detección del fraude o la predicción del comportamiento de los consumidores permiten a las compañías obtener resultados financieros un 20% por encima de sus competidores, según estimaciones de Gartner. <http://bigdata.ticbeat.com/big-data-los-datos-como-valor-de-negocio/>

¹⁰ ¿Qué es el Big Data?. Innoquant. <http://innoquant.com/big-data-es>

Tascon afirma que Big data¹¹, que en español significa literalmente «grandes datos», es un término que se ha acuñado para referirse a la manipulación de gran cantidad de datos. El volumen masivo, variedad y velocidad que ahora toma la información hace imprescindible capturar, almacenar y analizar todo este complejo engranaje¹².

En el ámbito anglosajón nacieron como *Big Data* y son el resultado de una carrera tecnológica sin precedentes en la que nosotros, casi sin saberlo, también participamos. Hablamos de millones y millones de bytes de información sobre toda clase de fenómenos y actividades –desde las variaciones atmosféricas hasta nuestros patrones diarios de consumo– que se producen, difunden o almacenan a través de medios tan variopintos como un teléfono móvil, una red social o la memoria de una máquina expendedora de refrescos.

En torno a este concepto¹³, de acuerdo con el criterio establecido por Elliott, giran múltiples tecnologías que contribuyen a organizar, almacenar y, sobre todo, analizar en profundidad los datos.

¹¹ En al ámbito empresarial y el mundo de los negocios, durante la última década del pasado siglo y los primeros años de este se hablaba de *Business Intelligence* (BI) para hacer referencia al conjunto de estrategias y herramientas que una empresa tenía a su disposición para poder analizar los datos de su organización. Con el BI se hacían previsiones y análisis.

Big Data también está emparentado con lo que se ha conocido como minería de datos, un campo de las Ciencias de la Computación que intenta descubrir patrones en grandes volúmenes de datos. La minería de datos (parte de BI), al igual que el *Big Data*, utiliza los métodos de la *Inteligencia Artificial* (IA) y la Estadística para analizar los patrones en las bases de datos con las que trabaja.

Cfr.: TASCÓN, Mario. «Introducción al Big Data. Pasado, Presente y Futuro». Revista Telos. Revista de Pensamiento sobre Comunicación, Tecnología y Sociedad. Número 96: El documento Digital Junio-Septiembre de 2013.

http://sociedadinformacion.fundacion.telefonica.com/seccion=1266&idioma=es_ES&id=2013062110090002&activo=6.do

¹² Cfr.: BELICOVE, Mikal E. «Cómo aprovechar el Big Data en tu empresa». En SoyEntrepreneur.com. 22 de octubre de 2013. <http://www.soyentrepreneur.com/25933-como-aprovechar-el-big-data-en-tu-empresa.html>

¹³ Frente a la confusión actual sobre el término «Big Data», algunas de las principales definiciones que se pueden obtener de Big Data, son las siguientes: Lo primero a destacar es que - a pesar de lo que dice Wikipedia - todo el mundo en la industria en general está de acuerdo en que los grandes datos no es sólo acerca de tener más datos.

(1) El Big Data original. Big Data como los tres Vs: Volumen, Velocidad y Variedad. Esta es la definición más venerable y bien conocida, acuñada por Doug Laney de la consultora Gartner durante doce años atrás. Desde entonces, muchos otros han tratado de llevarlo a 11 con Vs adicional, incluyendo la validez, la veracidad, valor y visibilidad.

(2) Big Data como Tecnología ¿Por qué un término de 12 años de edad, de repente hacer un zoom en el punto de mira? No fue simplemente porque de hecho ahora tenemos mucho más volumen, la velocidad, y la variedad que hace una década. En cambio, se vio impulsado por las nuevas tecnologías, y en particular el aumento

Además, los convierten en información inteligente para la empresa. Unas tecnologías en las que, según los últimos datos de Gartner, el 64% de las organizaciones a escala mundial ya está invirtiendo o espera hacerlo este año 2013 y que conformarán un mercado que,

rápido de las tecnologías de código abierto como Hadoop y otras NoSQL formas de almacenar y manipular datos. Los usuarios de estas nuevas herramientas necesitaban un término que los diferenciaba de las tecnologías anteriores, y de alguna manera terminó -de decidirse por el término Big Data.

(3) Big Data como datos o Distinciones El problema con -Big Data- como -la tecnología es que: (a) es lo suficientemente imprecisa de que todos los proveedores en la industria saltó a reclamarlo para sí mismos y (b) a todo el mundo «sabía» que se suponía que debían elevar el debate y hablar de algo más negocio y rentabilidad. Aquí hay dos buenos intentos de ayudar a las organizaciones a entender por qué los grandes datos ahora es diferente de los meros datos grandes en el pasado:

- Las transacciones, interacciones y observaciones. Éste concepto procede de Shaun Connolly de Hortonworks. Transacciones constituyen la mayor parte de lo que hemos recogido, almacenado y analizado en el pasado. Las interacciones son datos que viene de cosas como que hacen clic en las páginas web. Las observaciones son datos recogidos automáticamente.

- Proceso mediante datos, Human-Citas de la Información, y generadas por la máquina de Datos. Este concepto es aportado por Barry Devlin, quien co-escribió el primer artículo sobre el almacenamiento de datos. Es básicamente similar al anterior, pero con una terminología más precisa.

(4) Los datos del tamaño de Señales. Este es otro enfoque de negocio-y que divide al mundo por la intención y el momento en lugar del tipo de datos, por corte-sía de SAP Steve Lucas. El «viejo mundo» es acerca de las transacciones, y para cuando se registran estas operaciones, ya es demasiado tarde para hacer algo al respecto: las empresas están constantemente «manejando datos desde el espejo retrovisor». En el «nuevo mundo», en lugar de las empresas pueden utilizar los nuevos datos «señal» para anticipar lo que va a pasar, e intervenir para mejorar la situación. Los ejemplos incluyen el seguimiento de la confianza de la marca en las redes sociales (si caen sus «gustos» de un acantilado, sus ventas seguramente seguirá) y el mantenimiento predictivo (algoritmos complejos a determinar cuándo es necesario sustituir una parte de la aeronave, antes de que el avión llega costosamente atascado en la pista).

(5) Big Data como Oportunidad. Este es un concepto acuñado por Matt Aslett y define Big Data como «análisis de datos que fue ignorado previamente debido a las limitaciones de la tecnología.

(6) Big Data como metáfora. En su libro «El Rostro Humano de Big Data», el periodista Rick Smolan señaló que Big Data es «el proceso de ayudar al planeta crezca un sistema nervioso, en el que somos más que otro, humano, el tipo de sensor.»

(7) Big Data como nuevo término para viejas cosas. Este es el uso caracterizado por el hecho de que los proyectos que eran posibles con la tecnología anterior, y se habrían denominado BI o analytics en el pasado han sido rebautizados en un intento de asumir el empuje del Big Data. La conclusión: lo que los desacuerdos sobre la definición, todo el mundo está de acuerdo en una cosa: los grandes datos es una gran cosa, y dará lugar a nuevas y enormes oportunidades en los próximos años. Cfr: ELLIOTT, Timo. «7 Definiciones de los grandes datos que debe saber sobre Big Data». Business Analytics. 5 de julio 2013 <http://timoelliott.com/blog/2013/07/7-definitions-of-big-data-you-should-know-about.html>

según los datos de IDC, alcanzará los 16.900 millones de dólares en 2015¹⁴.

¿Qué es lo interesante de estos gigantescos conjuntos de datos? Pues según Curto, que su estudio ya no es un privilegio de los grandes laboratorios de supercomputación; que pueden mejorar nuestra comprensión sobre lo que hacemos, lo que somos... y lo que compramos; que pueden ayudarnos a predecir lo que ocurrirá dentro de unos minutos, a paliar los efectos de una catástrofe natural... o, quién sabe, a resolver una crisis económica como la presente.¹⁵

Hoy son 2,5 zetabytes de información. Esa es la capacidad de almacenamiento de la Tierra o, mejor dicho, la de todos los soportes de datos disponibles en nuestras ciudades, hogares y centros de trabajo, ya sean memorias de ordenador o consolas de videojuegos. Si dispusiésemos ese volumen en unidades de DVD, tendríamos una pila de discos tan larga que cubriría 1,5 veces la distancia hasta la Luna. Pero lo más espectacular no es eso. Lo sorprendente para Rodríguez, es que esa cifra, generada principalmente en Europa y Norteamérica, se ha producido en apenas los dos últimos años y que cada anualidad, además, crece a un ritmo aproximado del 50 por ciento¹⁶.

Por ello, según señala Soraya Paniagua, la aparición de Big Data constituye una oportunidad para lograr ventajas comerciales sin precedentes. Sin embargo, se requiere un nuevo enfoque¹⁷ de la infraes-

¹⁴ TICbeat BIG DATA. Obra citada.

¹⁵ Para CURTO, Josep, Big Data va a suponer un cambio transgresor y disruptivo en las organizaciones. Una nueva realidad que ya está transformando las operaciones y los negocios. Big Data va a situar el mercado de business analytics en cifras espectaculares, las previsiones hablan de 50,7 billones de dólares para 2016.

Cfr.: En Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 23 de julio de 2.012. <http://www.sorayapaniagua.com/2012/07/23/josep-curto-de-icd-big-data-es-un-cambio-transgresor-y-disruptivo/>

¹⁶ Cfr.: RODRIGUEZ, José Luis. «Big Data: La Galaxia de los Datos». 9 de julio de 2012. <http://www.territoriocreativo.es/etc/2012/07/big-data-la-galaxia-de-los-datos.html>

¹⁷ No se puede hacer Big Data si no hay una idea, un modelo de uso, un concepto, en definitiva si no hay un objetivo. No se puede hacer Big Data sin gastar dinero. No se puede extraer valor sin invertir... En banca hacemos Big Data desde hace muchos años. La gente no lo sabe pero cada vez que pagamos con tarjeta, un sistema anti-fraude que analiza en tiempo real una gran cantidad de datos: patrón de pagos, importes, el comercio, la localización, los hábitos de compra, etc. Esto permite generar un ratio de peligrosidad o de propensión al fraude. También se utiliza en análisis de riesgo, es una calibración estadística que funciona bien en situaciones de normalidad.

Cfr.: Banco de Sabadell, de Social Media a Big Data. En Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 5 de junio de 2012. <http://www.sorayapaniagua.com/2012/06/05/banco-sabadell-de-social-media-a-biga-data/>

estructura de TI y una manera diferente de pensar sobre la manera de hacer negocios y de TI trabajen juntos¹⁸.

El inconveniente hasta ahora del Big Data era que requería de una gran inversión para lograr la capacidad de computación necesaria. Sin embargo, gracias al «Cloud Computing» («computación en la nube»), este problema ha desaparecido. El «Cloud Computing», como servicio, permite que cualquiera pueda acceder a centros de datos repartidos por todo el mundo con un sistema de pago por uso.

El mayor reto, como más adelante se verá, para la inversión en Big Data es disponer de personas adecuadas y formadas para ejecutar estos proyectos y convertir una gran cantidad de datos en decisiones, estrategias y mejores experiencias para los consumidores. Por ello, el hecho de tener la información adecuada es la clave para un futuro sostenible. En este sentido, tal y como afirmó Benjamín Franklin, *«invertir en conocimientos produce siempre los mejores beneficios»*.

En definitiva, según expone Alvarez Valle, es necesario partir de la idea de que los datos nunca mienten, y que tomar las decisiones basadas en los datos nos aleja sustancialmente de la especulación.¹⁹

En este mismo sentido, indica Schmidt, que debe tenerse en cuenta que *«desde el amanecer de la civilización hasta el 2003, se crearon más de 5 Exabytes de información. En la actualidad, esta cantidad se está generando cada 2 días»*²⁰.

El Big Data fue un problema para los negocios en el pasado porque es costoso de guardar y, para que sea útil, requiere muchos conocimientos sobre informática. Al hilo de ello, Molina²¹, señala que la década de los 90 Data Mining²² tenía ciertos problemas, además de los tecnológicos:

No había una metodología consolidada y completa.

¹⁸ Cfr.: «Big Data Transforms Business». An EMC E-book. Big Data for The Predictive Organization. Obra citada.

¹⁹ ÁLVAREZ VALLE, Javier. Obra citada. <http://www.lne.es/opinion/2013/11/10/big-data-una-gran-oportunidad/1497242.html>

²⁰ Cfr.: SCHMIDT, Eric, es CEO de Google. 2.010. <http://innoquant.com/big-data-es>, página web citada.

²¹ Cfr.: MOLINA FELIX, Luis Carlos. «Del Data Mining al Big Data» Director de Power Builders». http://media.wix.com/ugd/d3e6e7_872c1dd6fcd93442d4c1894c252ad58f.pdf

²² La Minería de Datos o «Data Mining» es el proceso de detectar la información procesable de los conjuntos de grandes datos. Utiliza el análisis matemático para deducir los patrones y las tendencias que existen en los datos. Normalmente estos patrones no se pueden detectar mediante la exploración tradicional de los datos por-

Se le vinculaba mucho como una etapa posterior y dependiente del Data Warehouse.

Los proyectos tenía una fuerte dependencia del «gurú» que lo había desarrollado y solo él sabía lo que había hecho, ya que muchos proyectos no contaban con documentación detallada.

Muchos de los proyectos realizados solo con estadística, pronóstico o inclusive Data Warehouse se vendía como de Data Mining.

El especialista de Data Mining no tenía experiencia en mejoramiento de la calidad de datos.

El especialista en Data Mining tenía poca experiencia en temas de negocio.

Al momento de recibir los resultados, algunos directivos los hacían más como casos anecdóticos, que como verdaderos elementos de apoyo a la toma de decisiones.

Pero con el tiempo se ha vuelto más barato guardar el Big Data y han aparecido recursos de computación que permiten que el análisis de los datos no sea un reto financiero²³.

Las empresas punteras en comercio electrónico (E-Commerce) utilizan el Big Data para conocer mejor el mercado y ofrecer mejores productos y servicios a través de canales digitales²⁴.

El Big Data hace referencias al conjunto de las tecnologías y herramientas que permiten resolver diferentes tipos de limitaciones

que las relaciones son demasiado complejas o porque hay demasiados datos. Cfr.: Power Builders. Minería de datos. <http://www.powerbuilders.com.mx/>

El *data mining* surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos. De forma general, los *datos* son la materia prima bruta. En el momento que el usuario les atribuye algún significado especial pasan a convertirse en *información*. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación del confronto entre la información y ese modelo represente un valor agregado, entonces nos referimos al *conocimiento*.

Cfr.: MOLINA FELIX, Luis Carlos. «Data mining: torturando a los datos hasta que confiesen». Coordinador del programa de Data Mining (UOC). http://media.wix.com/ugd//d3e6e7_8c76150e4369bc3b4b6c7ef4b717c686.pdf

²³ Cfr.: RUBENS, Paul. « La ciencia de encontrar perlas en el Big Data». BBC Mundo. 16 de marzo de 2013. http://www.bbc.co.uk/mundo/noticias/2013/03/130310_tecnologia_big_data_datos_informacion_perla_dp.shtml

²⁴ Cfr.: ROVIRA, Pere. «Las redes sociales ofrecen Big Data para conocer y satisfacer las demandas de los clientes». BDIGITAL GLOBAL CONGRESS, 22 de diciembre de 2013. <http://www.bdigitalglobalcongress.com/2013/las-redes-sociales-ofrecen-big-data-para-conocer-y-satisfacer-las-demandas-de-los-clientes-2/>

asociadas al tratamiento de grandes volúmenes de datos²⁵. En este sentido, y en el ámbito de las llamadas Tecnologías de la Información, el concepto de «Big Data» se define como *«un conjunto de datos de gran volumen y complejidad. Estas características dificultan el proceso de manejar, almacenar y extraer información de valor a partir de ellos»*.

Del mismo modo, se puede afirmar que Big Data son grandes volúmenes de datos en cantidad, se capturan a un índice rápido, y son estructurados o no estructurados²⁶.

²⁵ Cfr.: Pragsis. Tecnología e >Innovación. «Bidoop». http://bigdata-hadoop.pragsis.com/pages/2/big_data_hadoop_bidoop

IDC define el Big Data como *«la colección de tecnologías y estrategias capaces de capturar y analizar, de forma económica, grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad»*. Pero esta definición varía según las características de las empresas. Para unas empresas prima la volumetría, para otras la velocidad, para otras la variabilidad de las fuentes. En cada una la cadena de valor va a ser sutilmente diferente. Las empresas con mucha volumetría van a estar interesadas en capturar la información, guardarla, analizarla e incorporarla en sus procesos de negocio. Pero hay empresas que, aunque tienen mucho volumen, no necesitan almacenar, necesitan trabajar en tiempo real y a gran velocidad. Un ejemplo clásico son los sistemas de recomendación, sistemas que en tiempo real capturan la información de lo que está haciendo el usuario en la web, la combinan con la información histórica de ventas y lanza, en tiempo real, las recomendaciones. Otras empresas tienen otro tipo de retos como fuentes de datos heterogéneas y lo que quieren es combinarla. La captura es más compleja ya que hay que combinarlas en un mismo sitio y analizarla.

Del mismo modo, Big Data supone un proceso de cambio en la organización, no sólo desde la perspectiva tecnológica sino principalmente desde la de negocio.

La explotación de datos anteriormente inabordable habilita no sólo una mayor comprensión del negocio, sino también proporciona la capacidad de reimaginar los servicios que se ofrecen.

De hecho, Big Data representa una oportunidad para que desde el departamento TI se impacte en el negocio de forma realmente significativa. El CIO debe liderar la adopción de un conjunto de tecnologías – como Big Data, Analytics o tecnologías del lenguaje – y de mejores prácticas para extraer valor de los datos, para lo que será necesario redefinir nuevas funciones y realizar un rediseño organizacional. En particular, Big Data necesita de un nuevo rol que es el científico del dato. La falta de talento es uno de los obstáculos de las organizaciones que van a tener de apoyarse en los servicios de terceros.

Cfr. IDG en Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 23 de julio de 2012. <http://www.sorayapaniagua.com/2012/07/23/josep-curto-de-idc-big-data-es-un-cambio-transgresor-y-disruptivo/>

²⁶ Desde la perspectiva médica, por ejemplo un dato estructurado es un dato que puede ser almacenado, consultado, analizado y manipulado por máquinas. Un dato desestructurado es todo lo contrario. Por ejemplo datos no estructurados son las recetas de papel, los registros médicos, las notas manuscritas de médicos y enfermeras, las grabaciones de voz, las radiografías, resonancias magnéticas, TAC y otras imágenes. Los datos estructurados y semiestructurados incluyen archivos electrónicos de contabilidad, datos de actuario o datos clínicos.

Pero los avances tecnológicos están generando nuevas cascadas de datos (tanto estructurados como no estructurados), son los que provienen de dispositivos para

Pero también es importante comprender que además de los datos estructurados, aquellos otros que provienen de fuentes de información conocidas y que, por tanto, son fáciles de medir y analizar a través de los sistemas tradicionales, empezamos a poder y querer manejar datos no estructurados²⁷: los que llegan de la Web, de las cámaras de los móviles y vídeos, redes sociales, sensores de las ciudades y edificios... La variedad de su origen, además de la rapidez con la que se incrementa su volumen, son algunos de los factores que habían dificultado su análisis hasta ahora. El nuevo *software* y los nuevos modelos permiten la incorporación a los estudios tanto de un tipo como de otro. Los avances en análisis semántico también permiten estructurar mínimamente parte de los textos escritos por personas de forma automática.

La característica de «variedad» utilizada para definir Big Data lleva muchas veces a pensar, tal como lo hace Ambriz, que solo es apropiado para el almacenamiento y análisis de información no estructurada²⁸. La expresión «información no-estructurada» se refiere típicamente a aquellos datos que no están organizados bajo el Modelo de Datos Relacional, definido por Edgar Codd²⁹ en 1970. Algunos

fitness (sensores), de los medios sociales, de App en smartphones o de la genética y genómica.

Cfr.: «Big Data en sanidad para predecir, prevenir y personalizar». Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 23 de julio de 2012. <http://www.sorayapaniagua.com/2012/11/12/big-data-en-sanidad-para-predecir-prevenir-y-personalizar/>

²⁷ TASCÓN. Mario. «Introducción al Big Data. Pasado, Presente y Futuro». Revista Telos. Revista de Pensamiento sobre Comunicación, Tecnología y Sociedad. Número 96: El documento Digital. Junio-Septiembre de 2.013. Obra citada.

http://sociedadinformacion.fundacion.telefonica.com/seccion=1266&idioma=es_ES&id=2013062110090002&activo=6.do

²⁸ AMBRIZ llega a la conclusión de que conforme las soluciones de Big Data basadas en Hadoop, se han descubierto casos de uso que van más allá de las aplicaciones originalmente consideradas para Big Data. Del procesamiento de datos no-estructurados se ha evolucionado a la capacidad de administrar datos estructurados en soluciones basadas en Hadoop. Uno de los casos de uso que mejor demuestran la sinergia que se puede lograr al unir el paradigma del modelo relacional con Hadoop es la extensión a Data Warehouses, en particular para optimizar el almacenamiento y análisis de los datos accedidos con poca frecuencia.

Cfr. AMBRIZ RIVAS, M. Sc. «Almacenamiento de datos estructurados en Big Data. Como puede ayudar Big Data a extender su infraestructura actual de administración y análisis de datos». IBM México. 30 de septiembre de 2013. <https://www.ibm.com/developerworks/ssa/library/bd-almacenamiento-datos/>

²⁹ Edgar Frank Codd (*Ted Codd*), fue un científico informático inglés (19 de agosto de 1923 - 18 de abril de 2003), conocido por sus aportes a la teoría de bases de datos relacionales. En las décadas de los sesenta y los setenta trabajó en sus teorías sobre modelado de datos, publicando su trabajo «*Un modelo relacional de datos para grandes bancos de datos compartidos*» («*A Relational Model of Data for Large Shared Data*

ejemplos comunes de información no estructurada son los archivos de texto, documentos (PDF, Word), imágenes, audio y video, entre otros.

No obstante, casi desde su concepción, se consideró el uso de Big Data para almacenar y analizar datos estructurados. Google fue el primero en implementar una tecnología con tal fin: Bigtable³⁰.

Bigtable se utiliza para guardar datos de varios proyectos en Google, como por ejemplo los índices de búsqueda Google Earth y Google Finance. Asimismo, es capaz de escalar hasta petabytes, soporta desde procesamiento batch hasta acceso en tiempo real con baja latencia. Bigtable utiliza Google File System (precursor de Hadoop Distributed File System) para almacenar datos y bitácoras.

El reto del Big Data consiste en capturar, almacenar, buscar, compartir, y poner en valor datos hasta fecha infrautilizados o inaccesibles y uno de sus aspectos interesantes del Big Data es que las empresas ya almacenan este tipo de datos claves sólo que aún no se les ha

Banks) en 1970. Para su descontento, IBM no se apresuró a explotar sus sugerencias hasta que no empezaron a ser puestas en práctica por rivales comerciales. Por ejemplo, Larry Ellison diseñó la base de datos Oracle basándose en las ideas de Codd.

Codd continuó expandiendo y desarrollando su modelo relacional, en ocasiones en colaboración con Chris Date. También trabajó el área de los autómatas celulares, sobre la que versó su tesis doctoral.

Codd definió las tres primeras Formas Normales que se aplican para la normalización de sistemas de bases de datos. Además, la Forma normal de Boyce-Codd lleva el nombre en su honor.

También acuñó el término OLAP y redactó las doce leyes del procesamiento analítico informático.

Cfr.: Edgar Frank Codd. Wikipedia. http://es.wikipedia.org/wiki/Edgar_Frank_Codd

³⁰ BigTable es un sistema de gestión de base de datos creado por Google con las características de ser: distribuido, de alta eficiencia y propietario. Está construido sobre GFS (Google File System), Chubby Lock Service, y algunos otros servicios y programas de Google, y funciona sobre «commodity hardware» (sencillos y baratos PCs con procesadores Intel). BigTable comenzó a ser desarrollado a principios de 2004.

BigTable almacena la información en tablas multidimensionales cuyas celdas están, en su mayoría, sin utilizar. Además, estas celdas disponen de versiones temporales de sus valores, con lo que se puede hacer un seguimiento de los valores que han tomado históricamente.

Para poder manejar la información, las tablas se dividen por columnas, y son almacenadas como «tabletas» de unos 100-200 Mbytes cada una. Cada máquina almacena 100 tabletas, mediante el sistema «Google File System». La disposición permite un sistema de balanceo de carga (si una tableta está recibiendo un montón de peticiones, la máquina puede desprenderse del resto de las tabletas o trasladar la tableta en cuestión a otra máquina) y una rápida recomposición del sistema si una máquina «se cae».

Cfr.: Concepto de Bigtable. Wikipedia. <http://es.wikipedia.org/wiki/BigTable>

dado la interpretación adecuada. No es el volumen de datos o su naturaleza lo que importa, sino su valor potencial, que sólo tecnologías de última generación de Big Data pueden explotar³¹. En definitiva, lo que se pretende con esta tecnología es descubrir conocimiento oculto a partir de grandes volúmenes de datos.

En los últimos años, debido al desarrollo tecnológico a niveles exponenciales tanto en el área de cómputo como en la de transmisión de datos, de acuerdo con lo expuesto por Molina³², ha sido posible que se gestionen de una mejor manera el manejo y almacenamiento de la información. Sin duda existen cuatro factores importantes que nos han llevado a este suceso:

1. El abaratamiento de los sistemas de almacenamiento tanto temporal como permanente.
2. El incremento de las velocidades de cómputo en los procesadores.
3. Las mejoras en la confiabilidad y aumento de la velocidad en la transmisión de datos.
4. El desarrollo de sistemas administradores de bases de datos más poderosos.

Nuestra capacidad para almacenar datos, concluye Molina, ha crecido en los últimos años a velocidades exponenciales. En contrapartida, nuestra capacidad para procesarlos y utilizarlos no ha ido a la par. Por este motivo, el *data mining* se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido usando grandes volúmenes de datos.

Descubrir nuevos caminos que nos ayuden en la identificación de interesantes estructuras en los datos es una de las tareas fundamentales en el *data mining*.

En el ámbito comercial, resulta interesante encontrar patrones ocultos de consumo de los clientes para poder explorar nuevos horizontes. Saber que un vehículo deportivo corre un riesgo de accidente casi igual al de un vehículo normal cuando su dueño tiene un segun-

³¹ «Big Data, el gran y desconocido aliado de la economía». En ABC Tecnología. 5 de noviembre de 2.013. <http://www.abc.es/tecnologia/informatica-soluciones/20131101/abci-data-spain-beneficios-espana-201310312135.html>

³² Cfr.: MOLINA FELIX, Luis Carlos. «Data mining: torturando a los datos hasta que confiesen». Coordinador del programa de Data Mining (UOC). Obra citada. http://media.wix.com/ugd//d3e6e7_8c76150e4369bc3b4b6c7ef4b717c686.pdf

do vehículo en casa ayuda a crear nuevas estrategias comerciales para ese grupo de clientes.

Asimismo, predecir el comportamiento de un futuro cliente, basándose en los datos históricos de clientes que presentaron el mismo perfil, ayuda a poder retenerlo durante el mayor tiempo posible.

Las herramientas comerciales de *data mining* que existen actualmente en el mercado son variadas y excelentes. Las hay orientadas al estudio del web o al análisis de documentos o de clientes de supermercado, mientras que otras son de uso más general. Su correcta elección depende de la necesidad de la empresa y de los objetivos a corto y largo plazo que pretenda alcanzar. La decisión de seleccionar una solución de data mining no es una tarea simple. Es necesario consultar a expertos en el área con vista a seleccionar la más adecuada para el problema de la empresa.

Por ello, el data mining se presenta como una tecnología emergente, con varias ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa y abre nuevas oportunidades de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sinnúmero de detalles debido a que el producto final involucra «toma de decisiones».

Complementariamente a ello, y según IDC³³, Big Data es una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico de grandes volúmenes de datos heterogéneos habilitando una captura, identificación y/o análisis a alta velocidad. Big Data se caracteriza por tener cuatro dimensiones: volumen, variedad, velocidad y valor.

Volumen: hace referencia a la necesidad de procesamiento intensivo y complejo de subconjuntos de datos de gran tamaño que contengan información de valor para una organización mediante tecnologías de Big Data.

Es decir, la cantidad de datos³⁴. Siendo quizá la característica que se asocia con mayor frecuencia a big data, el volumen hace referencia

³³ IDG Analyze the Future. White Paper. Resumen Ejecutivo. «Big Data: Un mercado emergente. José Curto. Junio de 2012. <http://www.diarioabierto.es/wp-content/uploads/2012/06/Resumen-Ejecutivo-IDC-Big-Data.pdf>

³⁴ Cfr.: SCHROECK, Michael, SHOCKLEY, Rebecca, ROMERO-MORALES, Dolores, TUFANO, Peter. «Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos». Informe ejecutivo. IBM Global Business Services Business Analytics and Optimisation y la Escuela de Negocios Saïd en

a las cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones en toda la empresa. Los volúmenes de datos continúan aumentando a un ritmo sin precedentes. No obstante, lo que constituye un volumen verdaderamente «alto» varía en función del sector e incluso de la ubicación geográfica y es más pequeño que los petabytes y zetabytes a los que a menudo se hace referencia.

Variedad: a medida que se multiplican los canales de interacción con clientes, empleados, proveedores y procesos de negocio, la información de valor es cada vez más el resultado de la combinación de datos de múltiple origen y tipología que puede estar en forma estructurada, semiestructurada o no estructurada.

Implica, en definitiva la existencia de diferentes tipos y fuentes de datos³⁵. La variedad tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los datos estructurados, semiestructurados y no estructurados. Las organizaciones necesitan integrar y analizar datos de un complejo abanico de fuentes de información tanto tradicional como no tradicional procedentes tanto de dentro como de fuera de la empresa.

Con la profusión de sensores, dispositivos inteligentes y tecnologías de colaboración social, los datos que se generan presentan innumerables formas entre las que se incluyen texto, datos web, tuits, datos de sensores, audio, vídeo, secuencias de clic, archivos de registro y mucho más.

Velocidad: aunque los ciclos de negocio se han acelerado, no todos los datos de una organización tienen la misma urgencia de análisis asociada. La clave para entender en qué punto del espectro de la velocidad es necesario trabajar (desde el procesado en lote hasta el flujo de datos continuo) está asociada a los requerimientos de los procesos y los usuarios.

Ello lleva consigo la perspectiva de los datos en movimiento³⁶. La velocidad a la que se crean, procesan y analizan los datos continúa

la Universidad de Oxford. http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

³⁵ Cfr.: SCHROECK, Michael, SHOCKLEY, Rebecca, ROMERO-MORALES, Dolores, TUFANO, Peter. «Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos». Informe ejecutivo. IBM Global Business Services Business Analytics and Optimisation y la Escuela de Negocios Saïd en la Universidad de Oxford. Obra citada. http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

³⁶ Cfr. SCHROECK, Michael, SHOCKLEY, Rebecca, ROMERO-MORALES, Dolores, TUFANO, Peter. «Analytics: el uso de big data en el mundo real. Cómo las empresas más

aumentando. Contribuir a una mayor velocidad es la naturaleza en tiempo real de la creación de datos, así como la necesidad de incorporar datos en streaming a los procesos de negocio y la toma de decisiones. La velocidad afecta a la latencia: el tiempo de espera entre el momento en el que se crean los datos, el momento en el que se captan y el momento en el que están accesibles. Hoy en día, los datos se generan de forma continua a una velocidad a la que a los sistemas tradicionales les resulta imposible captarlos, almacenarlos y analizarlos. Para los procesos en los que el tiempo resulta fundamental, tales como la detección de fraude en tiempo real o el marketing «instantáneo» multicanal, ciertos tipos de datos deben analizarse en tiempo real para que resulten útiles para el negocio.

Valor: en el contexto de Big Data, valor hacer referencia a los beneficios que se desprenden del uso de Big Data (reducción de costes, eficiencia operativa, mejoras de negocio).

A estas características puede serle añadida otra distinta, denominada «veracidad»³⁷.

Veracidad: la incertidumbre de los datos. La veracidad hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos. Esforzarse por conseguir unos datos de alta calidad es un requisito importante y un reto fundamental de big data, pero incluso los mejores métodos de limpieza de datos no pueden eliminar la imprevisibilidad inherente de algunos datos, como el tiempo, la economía o las futuras decisiones de compra de un cliente. La necesidad de reconocer y planificar la incertidumbre es una dimensión de Big Data que surge a medida que los directivos intentan comprender mejor el mundo incierto que les rodea.

En definitiva, Big Data es una combinación de estas características que crea una oportunidad para que las empresas puedan obtener una ventaja competitiva en el actual mercado digitalizado³⁸. Permite

innovadoras extraen valor de datos inciertos». Informe ejecutivo. IBM Global Business Services Business Analytics and Optimisation y la Escuela de Negocios Saïd en la Universidad de Oxford. Obra citada. http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

³⁷ Cfr. SCHROECK, Michael, SHOCKLEY, Rebecca, ROMERO-MORALES, Dolores, TUFANO, Peter. «Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos». Informe ejecutivo. IBM Global Business Services Business Analytics and Optimisation y la Escuela de Negocios Saïd en la Universidad de Oxford. Obra citada. http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

³⁸ Cfr. SCHROECK, Michael, SHOCKLEY, Rebecca, ROMERO-MORALES, Dolores, TUFANO, Peter. «Analytics: el uso de big data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos». Informe ejecutivo. IBM Global Busi-

a las empresas transformar la forma en la que interactúan con sus clientes y les prestan servicio, y posibilita la transformación de las mismas e incluso de sectores enteros. No todas las organizaciones adoptarán el mismo enfoque con respecto al desarrollo y la creación de sus capacidades de Big Data. Sin embargo, en todos los sectores existe la posibilidad de utilizar las nuevas tecnologías y analíticas de Big Data para mejorar la toma de decisiones y el rendimiento.

Estos factores hacen que los Big Data sean difíciles de capturar, extraer, y gestionar usando métodos tradicionales. Pero debe tenerse presente que el uso de la tecnología Big Data no se restringe sólo a grandes volúmenes de datos. Los Big Data pueden hacer referencia al concepto de dato de carácter estructurados como no estructurados. Por ello, a través de la extensibilidad, también se gestionan datos no estructurados. La tecnología Hadoop³⁹, a la que más adelante se hará especial referencia, trae técnicas de programación nuevas y más accesibles para trabajar en almacenamientos de datos masivos con datos tanto estructurados como no estructurados⁴⁰.

ness Services Business Analytics and Optimisation y la Escuela de Negocios Saïd en la Universidad de Oxford. Obra citada. http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

³⁹ Apache Hadoop es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop se inspiró en los documentos Google para MapReduce y Google File System (GFS). Hadoop es un proyecto de alto nivel Apache que está siendo construido y usado por una comunidad global de contribuidores, mediante el lenguaje de programación Java. Yahoo! ha sido el mayor contribuyente al proyecto, y usa Hadoop extensivamente en su negocio.

Hadoop consiste básicamente en el *Hadoop Common*, que proporciona acceso a los sistemas de archivos soportados por *Hadoop*. El paquete de software *The Hadoop Common* contiene los archivos.jar y los scripts necesarios para hacer correr *Hadoop*. El paquete también proporciona código fuente, documentación, y una sección de contribución que incluye proyectos de la Comunidad *Hadoop*. Una funcionalidad clave es que para la programación efectiva de trabajo, cada sistema de archivos debe conocer y proporcionar su ubicación: el nombre del rack (más precisamente, del switch) donde está el nodo trabajador. Las aplicaciones *Hadoop* pueden usar esta información para ejecutar trabajo en el nodo donde están los datos y, en su defecto, en el mismo rack/switch, reduciendo así el tráfico de red troncal (*backbone traffic*). El sistema de archivos HDFS usa esto cuando replica datos, para intentar conservar copias diferentes de los datos en racks diferentes. El objetivo es reducir el impacto de un corte de energía de rack o de fallo de interruptor de modo que incluso si se producen estos eventos, los datos todavía puedan ser legibles.

Cfr.: Wikipedia. Concepto de Hadoop. <http://es.wikipedia.org/wiki/Hadoop>

⁴⁰ Cfr.: «Big data de código abierto para el impaciente, Parte 1: Tutorial Hadoop: Hello World con Java, Pig, Hive, Flume, Fuse, Oozie, y Sqoop con Informix, DB2, y MySQL». <http://www.ibm.com/developerworks/ssa/data/library/techarticle/dm-1209hadoopbigdata/>

Los desafíos actuales incluyen resolver problemas de captura, transformación, almacenamiento, búsqueda, visualización y análisis de estos datos. Analizar un gran conjunto de datos y relacionarlos entre sí, es una metodología mucho más eficaz que analizar por separado de solo un pequeño conjunto. De esta manera se pueden detectar tendencias en los mercados, determinar la calidad de una investigación, entender estructuras que ayuden a la prevención de enfermedades, a combatir el delito y/o determinar las condiciones de tráfico de una carretera en tiempo real⁴¹.

Baquero⁴² señala que el Big Data es pura tecnología pero sobre todo es análisis, el cómo explotar la información. Tener datos por tener datos no sirve de nada. Lo que pasa ahora es que Big Data democratiza la tecnología, permite que empresas muy pequeñas recojan mucha información. Con muy poca inversión empresas pequeñas pueden dar mucho valor a los datos. Y continúa señalando que, Big Data es una solución para tiempos de crisis al igual que las soluciones «*open source*»⁴³. Un proyecto que antes era inabordable en términos presupuestarios ahora se puede hacer incluso de manera incremental, invirtiendo poco a poco. Ahora, con la revolución de las redes sociales el CIO se enfrenta a retos muy grandes, retos a la hora de poder procesar toda esa información y dotarla de valor. Ahí necesita el apoyo del Cloud Computing y del Big Data.

Muchos son los cambios que se producen con el advenimiento del Big Data, pero Curto⁴⁴, ha señalado que los Departamentos de Tecno-

⁴¹ Cfr.: <http://innoquant.com/big-data-es>. Dirección web citada.

⁴² BAQUERO, Carlos de GMV: «Big Data es pura tecnología». En Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 16 de julio de 2012. <http://www.sorayapaniagua.com/2012/07/16/carlos-baquero-de-gmv-big-data-es-pura-tecnologia/>

⁴³ «*Open Source*» o «Código abierto» es la expresión con la que se conoce al software distribuido y desarrollado libremente. Se focaliza en los beneficios prácticos (acceso al código fuente) que en cuestiones éticas o de libertad que tanto se destacan en el software libre. En la actualidad el código abierto se utiliza para definir un movimiento nuevo de software (la Iniciativa Open Source), diferente al movimiento del software libre, incompatible con este último desde el punto de vista filosófico, y completamente equivalente desde el punto de vista práctico, de hecho, ambos movimientos trabajan juntos en el desarrollo práctico de proyectos. La idea bajo el concepto de *código abierto* es sencilla: cuando los programadores (en Internet) pueden leer, modificar y redistribuir el código fuente de un programa, éste evoluciona, se desarrolla y mejora. Los usuarios lo adaptan a sus necesidades, corrigen sus errores a una velocidad impresionante, mayor a la aplicada en el desarrollo de software convencional o cerrado, dando como resultado la producción de un mejor software.

Cfr. Concepto de «*Open Source*». Wikipedia. http://es.wikipedia.org/wiki/C%C3%B3digo_abierto

⁴⁴ Curto, Josep. Obra citada.

logías de la Información⁴⁵ están viviendo una profunda transformación. No sólo por el gran volumen de información o la necesidad de digestión de esa información, sino también por Cloud Computing. Cloud presupone que el departamento Tecnologías de la Información pasa de ser un gestor de infraestructuras a ser un gestor de servicios, de requerimientos.

Big Data presupone que el Departamento Tecnologías de la Información se convierte en un departamento de innovación donde se van a incorporar nuevos perfiles como el llamado «Científico de Datos», al que más adelante haremos referencia y que se trata de un profesional que se va a preocupar de formular las preguntas adecuadas para que los procesos de negocio sean más óptimos y eficientes. Hay una gran oportunidad de reposicionar el departamento Tecnologías de la Información, pero a la vez las personas son reticentes al cambio. Estamos viviendo lo que IDC denomina «*La nueva normalidad*», una realidad donde cada vez hay menos dinero pero el negocio debe seguir funcionando con los mismos niveles de calidad. En esa nueva normalidad Big Data y Cloud juegan un papel protagonista.

⁴⁵ Debe tenerse presente que las bases de datos y los sistemas de administración de datos han jugado un papel primordial en el crecimiento y éxito de las organizaciones corporativas, los cambios en la economía y la globalización de las empresas han aumentado su importancia. Esto ha dado origen a que los departamentos de tecnología de la información inicien la búsqueda y explotación de nuevas herramientas y paradigmas para el procesamiento de datos y de la información generada.

Con ello inicia el concepto de Inteligencia de Negocios permitiendo a las organizaciones realizar análisis sobre un conjunto de datos con el objetivo de obtener y mejorar el conocimiento de las operaciones de la empresa y ayudar en la toma de decisiones, incrementando o manteniendo la competitividad de la organización.

Existe un crecimiento en el uso de técnicas y enfoques tradicionales para hacer uso de bases de datos empresariales y herramientas de análisis de datos, por lo que en los diferentes mercados y tipos de organizaciones se ha requerido un mayor estudio y entendimiento sobre estas herramientas, modelos, procesos, algoritmos, etc., referentes al descubrimiento de conocimiento y que gracias a la integración de tecnologías como Minería de Datos con Inteligencia de Negocios es posible ofrecer de manera automatizada e inteligente información de suma importancia para ejecutar acciones para la solución de problemas organizacionales.

La Minería de Datos es un proceso fundamental en Inteligencia de Negocios, ya que permite obtener conocimiento a partir de los datos almacenados, mediante un proceso no trivial de extracción de información implícita, previamente desconocida y potencialmente útil. Descubrir conocimiento implica buscar patrones de comportamiento aún no conocidos en los datos. El conocimiento se puede manifestar como: patrones, reglas de conocimiento, asociaciones, grupos, restricciones, tendencias, etc. Los dominios de aplicación prácticamente abarcan cualquier campo y tipo de empresas (comercialización, producción, finanzas, energía, gobierno, tecnología, etc.).

Cfr.: MOLINA FELIX, Luis Carlos, «Del Data Mining al Big Data», Director de Power Builders. <http://em.fis.unam.mx/mochan/eventos/pdf05Rys9Jn2t.pdf>.

Por tanto, debe indicarse que adoptar la filosofía Big Data en una empresa no es una cuestión puramente tecnológica⁴⁶. Requiere de una visión clara de los beneficios que genera basar tu negocio en datos y del valor y el conocimiento que se puede extraer integrando los datos internos y externos. Otro factor importante es contar con profesionales que sepan romper la barrera entre los aspectos más técnicos y los de negocio. En ese sentido cobra especial importancia la figura del científico de datos. Owen⁴⁷ la definió como «*una persona que entiende la estadística mejor que un ingeniero software y es mejor en ingeniería software que cualquier estadístico*». Sin duda a estas habilidades hay que añadir el conocimiento del negocio y la capacidad para plantear las preguntas adecuadas. Son, según Macías, en definitiva, unos profesionales que deben saber cómo analizar todos los datos que manejan las organizaciones y convertirlos en valioso conocimiento de modo que éstas puedan ser más competitivas y responder con mayor eficiencia a las necesidades no solo de los clientes sino también de los socios y de los propios empleados.⁴⁸

Complementariamente a ello es interesante traer a colación la opinión de Deloitte⁴⁹ con relación a los primeros pasos hacia una evolu-

⁴⁶ Así se indicó en la Segunda Edición de Big Data Spain uno de los eventos pioneros en las tecnologías y aplicaciones del procesamiento masivo de datos se celebró el 7 y el 8 de Noviembre de 2013 en Madrid. El evento que consiguió atraer a más de 400 asistentes, el doble que el año anterior, refleja el creciente interés en estas tecnologías también en España.

⁴⁷ OWEN, Sean, es Director de Data Science de la empresa Cloudera, y sus opiniones fueron vertidas en el seno de la Segunda edición del Foto Big Data Spain, citado.

⁴⁸ No hay que confundir al científico de datos con el tradicional analista de datos o de negocio. De hecho, se podría decir que el primero es un rol resultante de la evolución del segundo. Obviamente la formación del científico de los datos debe ser similar a la que tiene el analista de datos. Según Gartner, ambos deben tener fuertes conocimientos en informática (sobre todo en herramientas y lenguajes como Hadoop, Pig, Python y Java), estadística, análisis y matemáticas.

La diferencia es que el científico de los datos debe tener ciertas capacidades de las que el anterior perfil carecía: un sólido conocimiento del negocio y, sobre todo, la capacidad de comunicarse tanto con las áreas de tecnología como las de negocio, dentro de su organización.

Cfr.: MACIAS, Esther. «Big data propicia el auge de los científicos de datos en las organizaciones». TICbeat. 28 de mayo de 2.013. <http://bigdata.ticbeat.com/big-data-propicia-auge-cientificos-datos-organizaciones/>

Con relación a las oportunidades laborales que en el ámbito técnico y empresarial va a generar el Big Data, es interesante el artículo de Miguel Galera «Big Data y oportunidades laborales». TICbeat, de 7 de marzo de 2012. <http://bigdata.ticbeat.com/big-data-oportunidades-laborales/>

⁴⁹ En este sentido, debemos remitirnos al Informe de la Consultora Deloitte México titulado: «Qué hacer para evolucionar hacia Big Data en México. Cómo entrenar a tu BI, para que evolucione hacia Big Data». [http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx\(es-mx\)BigData_2013.pdf](http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx(es-mx)BigData_2013.pdf)

ción de Big Data, el orden debe ser conforme al orden que se propone, y que como se pueden evidenciar de su simple lectura, excede de lo que modernamente constituye un Departamento Tecnologías de la Información, para abarcar todo el conjunto de actividad que representa una empresa. Dichos elementos propuestos por la meritada Consultora, son los que se describen a continuación:

- Negocio
- Establecimiento de la estrategia de aproximación, lo cual incluye: identificación de las áreas estratégicas del negocio y cómo pueden ser éstas auxiliadas por el análisis predictivo
- Profesionales capaces de analizar, interpretar y detectar comportamientos con base en la presente información
- Información interna útil al negocio
- Fuentes estructuradas, como son *data warehouses* o bases de datos transaccionales
- Fuentes no estructuradas
- Información externa útil al negocio
- Económica
- Análisis de percepción o sentimiento
- Marketing
- Infraestructura
- Almacenamiento
- Procesamiento
- Niveles de servicio

La tendencia con Big Data⁵⁰ no es enfocarse en lo que sucedió en el pasado, sino en aprender de lo sucedido y anticipar posibles situa-

⁵⁰ En este sentido, de acuerdo con el citado Informe de la Consultora Deloitte México titulado: «Qué hacer para evolucionar hacia Big Data en México. Cómo entrenar a tu BI, para que evolucione hacia Big Data», donde en el mismo se describen los pasos a seguir para la evolución hacia el Big Data, señalando a tal efecto que en el establecimiento de la estrategia de aproximación, deberán determinarse qué áreas estratégicas de la organización serán las beneficiadas con dicha iniciativa entre el establecimiento del *road map* y los modelos de madurez presentes dentro de la organización.

Como segundo paso se tiene la adquisición o identificación de profesionales que a través de algoritmos de análisis de tendencias, como *market basket*, puedan encontrar patrones de comportamiento dentro de la información, así como la identificación de variables externas e internas que influyen en ellas; esto en la jerga informática es conocido como los «mineros de datos». En conjunto con los dueños estratégicos del

ciones para el futuro, y con esto realizar una planeación estratégica de riesgos más precisa, ya que en estos tiempos la única constante es el cambio. Todo ello, lleva a la consecuencia que la mejor forma de empezar a «hacer Big Data» es ir poco a poco, y, al mismo tiempo, ir abordando proyectos con objetivos de negocio bien definidos. En otros casos, sin embargo, la necesidad viene por el lado de innovar, bien mediante la integración de datos externos o el diseño de productos basados en los datos.

En cuanto a los requisitos necesarios para la operatividad del Big Data, debe tenerse presente, tal como pone de manifiesto Innoquant⁵¹, que al evaluar una solución de Big Data para la empresa vale la pena considerar algunas de las cuestiones fundamentales que se plantean en los desafíos del Big Data. Para comenzar con el proceso de evaluación de los proveedores de tecnologías Big Data, desde dicha Empresa se recomienda responder algunas preguntas clave y comprobar si el grado de madurez y ajuste de la respuesta se adecúa a las necesidades:

- ¿La solución es realmente escalable a grandes volúmenes de datos, mientras se mantiene baja la latencia?
- ¿La solución permite el procesamiento de flujos de datos, y cálculo incremental de métricas?
- ¿La solución usa procesamiento paralelo y aprovecha las ventajas del cómputo distribuido?
- ¿La solución realiza agregación o indexación resumida para acelerar las consultas en grandes conjuntos de datos?
- ¿Los entornos de exploración de datos y lenguaje de consultas permiten realizar un rápido e interactivo análisis exploratorio de los datos?
- ¿La solución provee visualización de datos y está preparada para integrarse con piezas de otros proveedores de tecnología?

Conseguir la respuesta a estas preguntas, es un punto fundamental para el éxito de una iniciativa Big Data.

negocio, ellos son los encargados de identificar la información útil a nivel interno y externo, que podría ayudar a sentar los cimientos para los análisis de tipo predictivo, pronósticos y de regresión.

En el siguiente paso dentro de la estrategia de aproximación, se obtendrá la identificación de la información necesaria para cumplir con tal cometido. [http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx\(es-mx\)BigData_2013.pdf](http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx(es-mx)BigData_2013.pdf)

⁵¹ INNOQUANT. Dirección web citada.

En este sentido, debe traerse a colación un interesante estudio llevado a cabo por IDG⁵², donde se señala que los mercados y las empresas están viviendo una transformación de base tecnológica y social cuya principal derivada es el crecimiento exponencial de datos tanto dentro como fuera de los sistemas empresariales. De hecho, según IDC el volumen de datos digitales alcanzará los 35 Zbytes, 44 veces más que en 2009. Este crecimiento se caracteriza principalmente por estar formado en un 80% de datos no estructurados.

Las organizaciones se encuentran ante el reto de capturar, transformar, analizar y almacenar datos con sistemas tradicionales que no son capaces de resolver esta nueva problemática adecuadamente.

Según IDC, este nuevo escenario está lleno de retos y oportunidades. Por un lado es marco perfecto para el despliegue de iniciativas basadas en Big Data. Por otro, es necesario reducir las barreras de adopción en el seno de la organización.

Los principales puntos que este estudio destaca son:

- Big Data habilita la extracción de valor para las organizaciones a partir de grandes volumetrías de datos con una alta variabilidad mediante la captura, la extracción y el análisis de datos a alta velocidad.
- El mercado español de Big Data está en eclosión, con una adopción del 4,8% que se acelerará en los próximos años.
- Existe aún un desconocimiento importante del concepto de Big Data y de las diferencias respecto a otras tecnologías como Business Intelligence.
- Big Data habilita una transformación de las organizaciones a todos los niveles acentuando la generación de ventajas competitivas y la creación de servicios basados en datos. Las primeras aplicaciones de estas tecnologías se centran en usos genéricos como la reducción de costes aunque los enfoques verticales van adquiriendo importancia.
- Existe una falta de talento en las organizaciones. El científico de datos, rol fundamental en la generación de valor para las iniciativas de Big Data, es una barrera para la adopción de las

⁵² IDG Analyze the Future. White Paper. Resumen Ejecutivo. «Big Data: Un mercado emergente. José Curto. Junio de 2012. <http://www.diarioabierto.es/wp-content/uploads/2012/06/Resumen-Ejecutivo-IDC-Big-Data.pdf>

tecnologías de Big Data que debe suplirse mediante el apoyo en servicios de terceros⁵³.

Tal como pone de manifiesto De la Prida⁵⁴ las organizaciones españolas son ahora conscientes del beneficio que las tecnologías de Big Data pueden representar en su proceso de transformación. Resulta significativo observar que el fenómeno de Big Data no solo es aceptado como un motor de transformación del negocio, sino como elemento clave en las estrategias de seguridad. Sin duda, Big Data es una parte fundamental de la transformación de las empresas e instituciones españolas, que ya están inmersas en procesos de virtualización, Cloud Computing e infraestructuras definidas por software.

Cabe cuestionarse con EMC⁵⁵ el hecho consistente en ¿cómo aprovechar los grandes datos, maximizar su valor, y obtener una visión más profunda, sin aumentar los costos?

⁵³ Según una encuesta llevada a cabo por EMC España realizado entre 510 profesionales de TI en España, sobre los desafíos y oportunidades que las tecnologías Big Data suponen para sus organizaciones. Los datos cuentan que las empresas españolas comienzan a ser conscientes de la importancia de la aplicación de Big Data en sus procesos de negocio y de gestión. Un 29% de los encuestados manifiesta que sus organizaciones ya han obtenido ventajas competitivas, y un 55% cree que las empresas e instituciones que utilicen estas tecnologías lograrán mayor éxito. Una amplia mayoría, 77%, considera que Big Data mejorará la toma de decisiones en sus organizaciones. Además, el 46%, considera que sus equipos directivos confían en Big Data como impulsor de decisiones de negocio innovadoras. Pero, frente a estos buenos deseos, los datos también nos aportan tres estadísticas para la reflexión:

- El 69% de los encuestados considera que el factor más importante detrás de la toma de decisiones empresariales es el instinto y la experiencia de los directivos senior (y no en la objetividad de los datos).

- La mayoría cita la falta de las habilidades como la principal barrera para la adopción de Big Data.

- Un 74% cree que será un desafío mantener la evolución de las capacidades de los equipos de TI al ritmo de la innovación tecnológica que se producirá en los próximos tres años.

Cfr.: «La falta de habilidades, principal barrera para la adopción de Big Data en España». En Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social. 16 de octubre de 2.013. <http://www.sorayapaniagua.com/2013/10/16/la-falta-de-habilidades-principal-barrera-para-la-adopcion-de-big-data-en-espana/>

⁵⁴ DE LA PRIDA, Fernando. Es Director General de EMC España. Cfr.: <http://spain.emc.com/index.htm>

Y en «La falta de habilidades, principal barrera para la adopción de Big Data en España». En Soraya Paniagua. Datos, IoT, Makers, Medios Sociales y Aprendizaje Social.

⁵⁵ EMC, Obra citada. <http://www.emc.com/microsites/ebook/index.htm#/slide-overview>

Y para ello, se proponen a estos efectos, las soluciones que se citan a continuación:

1. Es necesario construir sobre una base de almacenamiento que puede escalar en minutos y proporcionar un rendimiento lineal.
2. Se hace preciso añadir una plataforma de análisis unificado que le ayuda a acceder rápidamente, analizar y colaborar en todos los datos resultantes en una visión más profunda.
3. Y también se exige incorporar los conocimientos sobre aplicaciones de actuar sobre las decisiones más rápido.

Ante esta serie de cuestiones, cabe plantearse ¿Qué hacer con toda esta información? ¿Cómo procesarla para convertirla en reportes útiles que ayuden a la toma de decisiones en el negocio? Para Anukoll⁵⁶, la respuesta es justamente la promesa de saber con precisión lo que ocurre en el presente. Si una empresa se puede beneficiar de los sensores de registro de datos, toda esa información se traduce en decisiones precisas en el momento correcto. Hoy, los datos ya no se utilizan para saber qué sucedió en el pasado; en cambio, son útiles para conocer el acontecer en tiempo real. Esto permite brindar una dirección útil al marketing, mejorar los servicios, reducir costos y ahorrar tiempo. En fin, las posibilidades son ilimitadas.

Tal como sugiere Belicove⁵⁷, el primer paso que es necesario afrontar, consiste en analizar de qué datos se dispone y cuáles son los que se pretenden obtener. Esto incluye desde la información generada por el software de servicio al cliente, hojas de Excel, información de los socios, y recibos de venta, hasta encuestas y datos oficiales. Todo lo que permita alimentar el sistema para generar una estrategia que genere beneficios para la Sociedad que pretende hacer ese tratamiento de datos masivos. El siguiente paso a llevar a cabo consiste en encontrar las maneras de conectar esta información y ordenarla por paquetes para obtener algo más contundente y vigente. Una forma fácil de lograrlo es empezar por un problema que se pretenda solventar. En todo caso, continua este autor, la clave consiste en relacionar toda

⁵⁶ Cfr.: «Cómo aprovechar el Big Data en tu empresa». Obra citada. Anukool Lakhina, es el fundador y CEO de Guavus, una compañía que crea software para análisis de datos. <http://www.soyentrepreneur.com/25933-como-aprovechar-el-big-data-en-tu-empresa.html>

⁵⁷ Cfr.: BELICOVE, Mikal E. «Cómo aprovechar el Big Data en tu empresa». Obra citada. <http://www.soyentrepreneur.com/25933-como-aprovechar-el-big-data-en-tu-empresa.html>

esta información en un programa de fácil acceso para todos los miembros de tu equipo de trabajo.

Con relación al tipo de dato que se han de ser tratados aptos para ser tratados en un proyecto de Big Data, podemos establecer la siguiente clasificación siguiendo a Barranco⁵⁸, de la siguiente forma:

1. *Web and Social Media*: Incluye contenido web e información que es obtenida de las redes sociales como Facebook, Twitter, LinkedIn, etc., blogs.
2. *Machine-to-Machine (M2M)*: M2M se refiere a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (velocidad, temperatura, presión, variables meteorológicas, variables químicas como la salinidad, etc.) los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.
3. *Big Transaction Data*: Incluye registros de facturación, en telecomunicaciones registros detallados de las llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.
4. *Biometrics*: Información biométrica en la que se incluye huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc. En el área de seguridad e inteligencia, los datos biométricos han sido información importante para las agencias de investigación.
5. *Human Generated*: Las personas generan diversas cantidades de datos como la información que guarda un «*Call Center*» al establecer una llamada telefónica, notas de voz, correos electrónicos, documentos electrónicos, estudios médicos, etcétera.

No obstante todo lo anterior, es cierto que Big Data presenta también evidentes dificultades técnicas en algunas de las ramas de actividad descritas, y un ejemplo significativo de ello lo encontramos

⁵⁸ Cfr.: BARRANCO FRAGOSO. IBM Software Mexico. 18 de junio de 2.012. Obra citada. <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

en lo afirmado por Neff⁵⁹⁶⁰, donde señala, con relación al ámbito

⁵⁹ NEFF, Gina es Profesora asociada de la Universidad de Washington, y por la misma se señala que uno de los hallazgos de esta investigación parece a primera vista tan obvio, que es difícil de creer, que ha sido pasado por alto en el diseño e implementación de tecnologías de innovación para el cuidado de la salud. Es decir, la gente se imagina *los datos* de maneras muy diferentes. La comprensión de este hecho clave sobre los datos nos ayuda a entender por qué las llamadas soluciones «Big Data» a la atención de salud son muy difíciles de implementar en la práctica. Los médicos, los pacientes y los empresarios de la salud tienen perspectivas de los *valores* de datos de maneras muy diferentes. Un médico se limitó a decir: «Yo no necesito más datos: necesito más recursos». Decir esto en Silicon Valley o en TEDMED sería equivalente a la herejía. Lo mismo ocurre con aquellos de nosotros que trabajamos en la investigación y llevamos a cabo la obtención, el tratamiento, la gestión, el análisis y la interpretación de los datos. Desde la perspectiva del médico, sin embargo, los datos requieren sin excepción una acción interpretativa adicional, de carácter administrativo, laboral y empresarial. Esta es la perspectiva sobre los «Datos», al menos en lo que respecta a la práctica clínica actual, en que los datos consumen más recursos que los beneficios que proporcionan. En otras palabras, la mayoría de los médicos creen que la innovación obtenida por medio de los datos significa más trabajo para ellos, y no menos, y, además, les quita tiempo de lo que ven como sus principales prioridades en la prestación de una atención de calidad».

Cfr.: NEFF, Gina. «Why big data won't cure us». DOI: 10.1089/big.2013.0029. Mary Liebert INC. Volumen 1, número 3.º Septiembre de 2013.

⁶⁰ En este sentido, Enrique Dans señala que una de las tendencias que me ha llamado la atención revisando proyectos es la aplicación de *Big data* al mundo de la salud: los hospitales, a pesar de la creciente sofisticación de sus sistemas, suelen vivir en lo que Seth Godin llama «la fase pre-digital», a pesar de que la incorporación de inteligencia analítica en este sentido puede ser de importancia crítica. El entorno médico-hospitalario está cada día más invadido por máquinas de todo tipo que generan torrentes de datos acerca de los pacientes a los que están conectadas. Datos que, sin embargo, suelen simplemente no almacenarse – son usados para análisis coyuntural y asociado a un momento específico – o ser impresos y recogidos de manera rudimentaria en una carpeta. En el plano personal, Google Health, uno de los proyectos recientemente clausurados de Google, intentaba dar un soporte a la información de salud y facilitar que fuese compartida con terceros: introduce los resultados de tus análisis, tus prescripciones, tus medicaciones, etc. en un archivo y compártelo con tu médico o con hospitales, facilitando incluso el que éstos puedan acceder a tu archivo para la inclusión de más información. Una idea con posibilidades, pero cuyo bajo nivel de adopción no permitió su supervivencia. Se calcula que un paciente medio genera alrededor de dos gigas de información, que crecen rápidamente en el caso de determinados tratamientos. ¿De qué tipo de información hablamos? De todo un poco: desde información perfectamente tabulada, como en el caso de resultados de analíticas, hasta datos no estructurados, como imágenes de todo tipo o lecturas de parámetros variados. Toda ella es información digitalizable, pero en muy pocos casos se digitaliza y almacena adecuadamente. Sin duda, un campo perfecto para la aplicación de técnicas de *Big data*, no solo por una cuestión de aplicación al paciente, sino también – y con grandes posibilidades – al tratamiento de la información agregada.

Los primeros proyectos están centrándose en cuestiones relacionadas con el ahorro y la gestión hospitalaria, en los que es posible llevar a cabo una estimación de impacto económico objetiva o una mejor asignación de recursos. Pero sin duda, hay un enorme potencial para eso que empieza a resultar cada vez más paradójico que en un entorno como el que vivimos no esté disponible aún: el almacenamiento de los

sanitario y asistencial, que la recopilación de tal magnitud de datos puede resultar una tarea difícil de realizar si se hace de forma manual, por lo que habrá que implementar estrategias o mecanismos que se encarguen de la recogida, haciendo que ello no suponga una carga adicional de trabajo para los profesionales. Y ello además se justifica en la razón de que Big Data no consiste sólo en recopilar un gran número de datos y cruzarlos a la espera de grandes resultados, sino en formular las preguntas adecuadas, para conseguir así, una adecuada explotación, y que de verdad sirvan para mejorar la salud de las personas.

En cuanto a las perspectivas de futuro, puede afirmarse que Big Data continuará dominando la inteligencia empresarial en 2013⁶¹, cuestión que es proyectable en los años venideros, pues constituye una tecnología que no ha hecho más que empezar. Los expertos están empezando a ver una creciente sofisticación en cómo las empresas lo utilizan.

En este sentido, se cree que Big Data seguirá siendo la mayor tendencia en la inteligencia de negocio para el futuro previsible pues se vislumbra como la herramienta competitiva del futuro. Así las principales tendencias de evolución que previsiblemente se seguirán son las que se enuncian a continuación.

1. Herramienta competitiva

La generación de datos está creciendo a tasas astronómicas en casi todos los mercados verticales, y las personas que analizan los negocios se están dando cuenta de que no es sólo un producto más económico o un método escalable para hacer Data Warehousing y Business Intelligence (BI), sino que es un arma competitiva para esta década.

datos de una persona de tal manera que permita un tratamiento y análisis centralizado en los momentos que realmente hace falta.

Cfr. Dans, Enrique. «Big Data e información de salud». Periódico Asus. http://asus.usal.es/index.php?option=com_content&view=article&id=12895: alias&catid=528: el-blog-de-enrique-dans&Itemid=273

⁶¹ Cfr.: Big Data: 10 tendencias para 2013. Enterprise Apps Today. Traducido y adaptado por División consultoría de Evaluando Software. [Evaluandosoftware.com](http://www.evaluandosoftware.com/nota-3136-Big-data-10-tendencias-para-2013.html). 9 de abril de 2.013. <http://www.evaluandosoftware.com/nota-3136-Big-data-10-tendencias-para-2013.html>

2. La segunda ola de Big Data

A pesar de la publicidad que rodea a Big Data, algunos creen que sólo se está en las primeras etapas de su potencial analítico. Hay conciencia sobre el hecho de que hay grandes volúmenes de datos y que deben ser extraídos, para definir con mayor claridad la forma en que se puede aprovechar y obtener una ventaja. En el 2012 se vió la primera ola de Big Data. Las empresas están empezando a ver que mediante el aprovechamiento de estas tecnologías, la transformación del negocio podría ser más barata y más rápida, lo que les permite a los ejecutivos centrarse en lograr una ventaja competitiva en el mercado frente a sólo mantener la empresa en funcionamiento.

3. Análisis más sofisticados

Como parte de esta relación cada vez mayor entre Big Data y el valor de negocio, las empresas prevén desarrolladores cada vez más sofisticados en sus intentos de recoger el valor de vastos repositorios de datos no estructurados.

En lugar de análisis derivados de las correlaciones estadísticas básicas, las tecnologías de Big Data va a evolucionar y aprovechar algoritmos altamente sofisticados que permiten decidir en la situación «*qué pasa si*» (what-if).

4. Análisis avanzados diferenciales

El responsable de información de un laboratorio internacional, cree que la llegada de Big Data ha puesto de manifiesto la diferencia entre la inteligencia de negocios tradicional y análisis avanzados que profundizan en los datos.

«En cierto sentido, el BI tradicional se ha convertido en el generador de preguntas, y la práctica analítica se está convirtiendo en el» contestador automático», dijo el ejecutivo, que agregó: «Con los informes y cuadros de mando ejecutando grandes conjuntos de datos, con mayores necesidades que sea en tiempo real, usted puede obtener una instantánea de su empresa y determinar dónde es necesario prestar atención a las ventas, el marketing de apoyo, y las finanzas.»

Una vez que sepa que las ventas han bajado, o llamadas de los clientes han aumentado, o los gastos de publicidad son demasiado

altos, se pregunta ¿qué hacer al respecto? Y a continuación empieza a hacerse más preguntas.

¿Por qué son las ventas bajan en una región, pero no en otra?

¿Qué nos dicen los clientes?

¿Cómo podemos reducir los costos de comercialización sin afectar los ingresos?

«Estos son los tipos de preguntas que la práctica analítica avanzada es capaz de responder, y hay una creciente presión para que las organizaciones van más allá de simples cuadros de mando e informes».

5. Herramientas simplificadas Big Data

El ejecutivo citado piensa que todavía estamos en las primeras etapas de la determinación de la forma de descubrir conocimientos con los Big Data. Si lo hace en la actualidad requiere de herramientas complejas que trabajan en pequeñas muestras de datos, manejados por consultores caros o difíciles para emplear científicos de datos.

«Si usted fue a alguna de las conferencias sobre estructuras de datos, podría haber pensado que todo el mundo estaba usando hermosas herramientas visuales para extraer sus datos y obtener grandes joyas ocultas».

«Pero la realidad es que la mayoría de la gente está luchando para obtener el máximo potencial de todas las nuevas tecnologías Big Data. Creo que 2013 es el año en que Big Data tiene que alcanzar un nivel de madurez con las aplicaciones que nos dan la profundidad, el rendimiento y la facilidad de uso que hemos llegado a esperar de las herramientas de Business Intelligence tradicionales que se ejecutan en fuentes de datos tradicionales».

6. Más allá de la ciencia de los datos

En las reuniones cumbres sobre ciencia de datos que se celebran en todo el mundo, la gente discute la falta de profesionales capacitados para dar sentido a las corrientes sociales y otros datos. Pero como sucede en otras prácticas, en la tecnología, finalmente reinará la sencillez. Así como la inteligencia empresarial tradicional ha sido tomado de las manos de unos pocos y dado a los muchos, lo mismo es probable que ocurra con el análisis de grandes datos.

«Vamos a ver pasar grandes volúmenes de datos más allá de las ciencias de datos en 2013», dijo el cofundador de una firma de análisis de

datos. «Big Data va a fundir cada vez más con BI tradicionales, y proporcionará información para gente de negocios que pueden acoplar información de fuentes no estructuradas con los estructuradas. Como analista de negocios, usted no tiene que poseer un doctorado universitario, para empezar a tomar ventaja de todo lo que Big Data tiene para ofrecer».

7. Surgimiento del analista de datos

Hay un futuro alternativo. En lugar de eliminar la necesidad de especialistas de datos, haciendo análisis de grandes volúmenes de datos de manera sencilla, surgirá un nuevo tipo de analista. Las cifras de Gartner Research apoya esta tendencia. En 2015, se crearán 4,4 millones puestos de trabajo de TI a nivel mundial para apoyar el manejo de grandes volúmenes de datos (Big Data). Sin embargo, la limitada aplicación de grandes volúmenes de datos en las universidades y colegios de hoy contribuirá a una brecha de habilidades.

«Debido a la ubicuidad de los datos, las empresas requieren un nuevo conjunto de especialistas que pueden cerrar la brecha entre TI y el negocio por la evolución de su pensamiento básico y el enfoque de análisis de toma de decisiones», dijo un ejecutivo de la industria quien los llama «los artesanos de datos». Además predice una generación emergente de los analistas que pueden proporcionar respuestas a las preguntas empresariales complejas en un corto período de tiempo, utilizando todos los datos que se requiere.

8. Nuevos actores, nuevas innovaciones

En 2011 la tendencia en materia de tratamiento de datos, fue la implementación de software de Business Intelligence (BI) tradicionales de mega-proveedores (SAP Business Objects, IBM Cognos, Oracle y MicroStrategy OBIEE). Se cree que el mercado de inteligencia de negocios es ahora más diverso que, con muchas nuevas empresas que compiten en el espacio y aportando ideas innovadoras.

«Cuando nos fijamos en los resultados recientes de los proveedores existentes de BI, se ve que el crecimiento se ha mitigado enormemente y que el ingreso neto está viniendo de las áreas auxiliares como las licencias de bases de datos», dijo el presidente de una firma implementadora de BI. «El mercado de BI pasó por una consolidación, pero ahora está pasando por una segunda ola de innovación y la diversidad».

9. Big Data, consumismo y «la nube»

Se prevé la monetización de grandes volúmenes de datos, gracias a la creciente utilización de plataformas en «la nube». Las aplicaciones analíticas empresariales recibirán una renovación en el 2013, cuando las organizaciones comiencen a adoptar fáciles de utilizar modelos amigables para el análisis. Las plataformas basadas en la nube darán a los usuarios las herramientas necesarias para acceder, intercambiar y seleccionar aplicaciones de análisis.

10. Contexto por medio de conectores de datos

El mantenimiento de almacenes de grandes volúmenes de datos no aportará un valor añadido sin contexto. En ese contexto se incluye la estrategia de negocio, los cambios competitivos, la percepción del mercado y más. Las empresas estarán mejor equipados para aprovechar los datos que antes no estaba disponible a través de conectores a los orígenes como las aplicaciones on-line (Ejemplo. Salesforce, Zoho) u of-line (Teradata). Al combinar estos datos con otros de los medios internos o incluso sociales, las empresas deberán obtener el contexto que necesitan para tomar las decisiones correctas.

En la implementación de Big Data, han de darse una serie de ítems o procesos⁶², que se concretan en las siguientes pautas:

a) **Recolección**

Obtención de datos de diferentes fuentes.

b) **Procesamiento**

Uso de Altas tecnologías de paralelismo para que el procesamiento de los datos sea mas entendible, y la consiguiente determinación de los resultados.

⁶² Cfr.: PUENTES, Juan Carlos. Es Country Manager SAS Colombia. «Big Data Analytics for Development». SAS

c) Gerenciamiento de los Datos

Se refiere a las labores de depuración técnica y jurídica de los datos, su transformación, así como las labores correspondientes a las funciones de auditoría, y de seguridad.

d) Medición

Que consiste en la integración o la obtención de correlaciones entre datos, basados en métricas inspiradas en el propio negocio.

e) Consumición

Se corresponde con el uso final de los datos resultantes

f) Almacenamiento

Hace referencia a las Soluciones de Storage

g) Gobernanación

Implica la facultad de disposición o gobierno de datos desde la perspectiva de Negocio

