

Del bit... al Big Data

El término **Big Data** se ha vuelto muy popular en los últimos meses, pero ¿qué es Big Data en realidad?
¡Te lo explicamos todo en esta guía!



Tabla de contenido

1. Big Data
2. Nuevas necesidades
3. Nuevos requerimientos
4. Big analytics
5. MAD- Nueva metodología
6. Hadoop
7. Cuatro megatendencias
8. Análisis en tiempo real
9. Implementar Big Data en su organización
10. Humanizando el Big Data

1. Big Data

¿En qué océano navega Big Data?

Actualmente, se generan muchos datos y las empresas exigen saber no sólo lo que sucede en la actualidad, sino también lo que va a pasar en un futuro, requiriendo niveles de servicio mucho más exigentes que hace unos años. Es precisamente en esta tesitura donde se encuentra Big Data hoy día.



Definamos entonces: ¿Qué es Big Data?

Para obtener una definición verdaderamente acertada de lo que significa Big Data debemos abrir la mente y romper con el estereotipo de que en “Big” esta la clave, ya que las exigencias actuales no siempre están basadas en el volumen, sino que éste es sólo una parte del rompecabezas, siendo muy diversos los parámetros que se tienen en cuenta en cada ocasión.

¿Por qué es tan importante?

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

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

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

Créditos foto: <http://blog.beeva.com/tag/big-data/>

2. Nuevas necesidades

Personalización máxima

Big Data nace de la exigencia de dar respuesta a toda una serie de necesidades del mercado actual, requerimientos que los avances han impulsado y a los que sólo se puede dar satisfacción por medios tecnológicos.

Hoy en día los clientes quieren ser tratados de forma totalmente personalizada. Es por esto que, ser capaces de detectar sus gustos, se ha convertido en una necesidad de primer nivel para poder aumentar el volumen de ventas, dirigiéndolas de forma mucho más directa para lograr el éxito esperado.

Además, hay que tener en cuenta que cada individuo se relaciona con otras personas, cuyo análisis permite tanto conocer más a fondo al propio cliente, como aumentar el número de clientes potenciales. Estos datos se transforman en un conocimiento de valor incalculable para el negocio.



De la transacción a la interacción

Hasta hace unos años lo más importante era la transacción de información pero, actualmente, esto está cambiando. Cada vez se da más importancia a la interacción constante con el mundo que nos rodea. Las empresas tienen una única opción: aprovecharse de toda la información que se genera.

En efecto, las empresas deben aprender a optimizar y conseguir la mayor información posible de sus clientes, creando, por ejemplo, nuevos canales de comunicación para poder interactuar con ellos.

Nuevos dispositivos

En la actualidad las personas están constantemente conectadas, **el teléfono móvil** ha dejado de ser un objeto de lujo y sus posibilidades van mucho más allá de la llamada, ofreciendo servicios basados en la geolocalización o permitiendo incluso detectar cuál ha sido la última interacción realizada y dónde se ha producido.

Lo mismo sucede con las **viviendas** que cada vez presentan mayores funcionalidades, inspiradas en la domótica o los **vehículos**, muchos de los cuales ya cuentan con varios sistemas de sensores, generando todos ellos información muy completa, perfecta para crear nuevos servicios de valor.

Big Data aprovecha todas estas nuevas fuentes de información provenientes de los dispositivos de última generación.

Créditos foto: <http://industriamusical.es/big-data-y-redes-sociales-al-servicios-del-ar/>

3. Nuevos requerimientos

Big Data requiere Big Analytics

Estamos acostumbrados a saber lo que pasó, pero hoy nos interesa más conocer lo que pasará: si la empresa, con su producto o servicio, seguirá siendo el gusto de los consumidores o si se hablará de ella bien o mal en las redes sociales; aspectos todos ellos que requieren de nuevos modelos de análisis mucho más complejos que lo que se podía experimentar hasta ahora.

Esta nueva realidad ha motivado nuevos requerimientos por parte de las empresas en relación con el análisis de datos. Y, precisamente, para poder analizar toda esa información de que hoy se dispone, lo que anteriormente se conocía como Business Intelligence actualmente requiere de un nuevo modelo de análisis: **Big analytics**, el único que permite dar forma al Big Data.



Esta necesaria aparición tiene que ver con el hecho de que, para poder hacer un análisis predictivo o una “clusterización” de los perfiles de clientes, se requiere de algo más que la propia inteligencia de negocio, y ese algo más se concreta precisamente en Big Analytics.

Analytics permite la aplicación de procesos matemáticos complejos. Sus técnicas hacen posible trabajar con los datos al nivel de granularidad más bajo disponible, en crudo, y empleando para ello modelos mucho más ágiles que los actuales modelos de BI.

Las limitaciones del Business Intelligence tradicional

El Business Intelligence con el que las empresas han crecido y se han externalizado, tiene unas limitaciones bastante claras, basadas en la falta de agilidad en sus procesos, que hoy se ponen en

evidencia.

En efecto, con el sistema tradicional, los procesos necesarios para el procesamiento y análisis de información carecían de dinamismo. Por ejemplo, la inversión de tiempo necesaria para proceder a la carga de modelos era excesiva y provocaba que, con frecuencia, al finalizar dicho proceso hubiese perdido utilidad, por haber evolucionado el negocio. Realizar los cambios pertinentes para alinearlo de nuevo con las necesidades corporativas era una tarea costosa, larga, poco ágil e ineficaz, al no poder tampoco aplicarse los algoritmos que hoy día se utilizan para el análisis.

La llegada de esta cantidad de datos ha provocado que las infraestructuras actuales no sean capaces de soportar esas volumetrías, variabilidad en los datos, ni de hacer los deseados análisis en tiempo real. Esto ha llevado a la inevitable búsqueda de infraestructuras capaces de soportar ese nuevo paradigma de datos al que las empresas se enfrentan.

Además, en cualquier compañía es importante poder calcular los costes en materia de infraestructura. Cuando la interacción con Big Data se realiza en términos de escalabilidad, la predicción del gasto se simplifica y se hace accesible la estimación del coste asociado al crecimiento de negocio. De este modo, es posible proyectar y determinar la inversión necesaria para seguir creciendo, algo que resulta impensable en términos de infraestructuras actuales y en un entorno de business intelligence tradicional.

Créditos foto: <http://www.ebanking.cl/tendencias/que-es-big-data-que-oportunidades-abre-en-la-banca-retail-0013613>

4. Big Analytics

¿Qué es un Analytic Data Warehouse?

Hasta ahora, las estructuras usadas para la analítica de datos y su procesamiento se basaban en unos modelos que habían sido predefinidos según las verticales correspondientes, caracterizándose por su gran rigidez, lo que provocaba que la empresa no se pudiera adaptar a las nuevas necesidades.

Con la llegada de Big Analytics este panorama cambia: los data warehouse pueden evolucionar a la medida de las necesidades, sin perjuicio de la posibilidad de disfrutar de unas capacidades ilimitadas tanto para incorporar datos de cualquier tipo, como para llevar a cabo su procesamiento.

Esta evolución tiene su razón de ser en que no es lo mismo aplicar determinados algoritmos complejos sobre grandes volúmenes de datos, que aplicar estos mismos sobre un conjunto reducido de ellos.

Evolución del DWH al ADW

No obstante, la llegada de Analytics no representa en absoluto un cambio radical en las aplicaciones existentes hasta el momento. En efecto, la aplicación de los Analytics data warehouse no significa que se tengan que sustituir los data warehouse actuales, ya que cada uno de ellos **cubre necesidades distintas, y usan infraestructuras y tecnologías muy diferentes.**



Actualmente, los datos esconden muchísima información y de lo que se trata es de ser capaces de descifrarla. Es precisamente por

ello que, si se ofrece esta información a un perfil nuevo dentro de la empresa, los llamados Data Science, éstos serán capaces de descubrir información muy valiosa escondida dentro de los propios datos y hasta el momento desconocida por la empresa, mediante el uso de Analytics.

¿Cómo se pueden integrar ambos mundos?

Se puede usar un **modelo híbrido**, es decir, es posible seguir manteniendo el data warehousing actual para un cierto tipo de datos, nutriéndose con la data estructura, y, para todo el resto, incorporar el nuevo modelo de analytics data warehousing. Ciertamente, los analytics data warehousing pueden convivir con los data warehouse actuales pero siempre que se ocupen de realizar análisis y funciones diferentes.

Créditos foto: <http://blog.prabasiva.com/>

5. MAD- Nueva Metodología

El dinamismo de los mercados actuales ha provocado que las empresas tengan que adaptarse a sus exigencias, incorporando nuevas infraestructuras. No obstante, para que dicha adaptación sea completa también se necesita de una **nueva metodología**, que desde Powerdata la llamamos **MAD**, esto es: Magnética, Ágil y Detallada.



Magnético

Debe tener la capacidad de captar toda la información disponible: ya sean datos antiguos, nuevos, estructurados o no-estructurados etc. Además, esta captación se debe poder llevar a cabo de forma muy rápida, sin perder tiempo en pre-procesos costosos.

Ágil

La nueva metodología tiene que poder ingerir, digerir, producir y acomodar los datos de una manera rápida y eficaz. A modo de ejemplo, debería ajustarse a los nuevos requerimientos, implementar los controles de cambio, realizar un análisis predictivo de forma rápida, visual y sencilla. La metodología, por tanto, debe ser adaptable, rápida y eficiente.

Detallado

La metodología conlleva la utilización de **sofisticados métodos estadísticos** para poder percatarse incluso de los detalles más minuciosos. Hasta ahora, y con frecuencia, se trabajaba a nivel de agrupaciones, al ser imposible el trabajar a un nivel más preciso de detalle. Este hecho no acarrea ningún inconveniente a la hora de practicar análisis a nivel genérico, pero si se quería profundizar más, resultaba del todo insuficiente.

Usando la nueva metodología es posible llegar a un nivel de detalle muy preciso, por lo que las decisiones que se tomen a nivel empresarial resultan mucho menos arriesgadas y más acertadas que con la metodología tradicional.

Créditos foto: <http://www.tech-wd.com/wd/2013/07/24/what-is-big-data/>

6. Hadoop

Historia y business drivers que lo impulsan aparecer

Hadoop nace en el momento en que Google se ve incapaz de poder indexar la web al nivel que exige el mercado y, por ello, decide buscar una solución.

Esta solución se basa en un **sistema de archivos distribuido**, que hace suyo el lema "divide y vencerás". Es decir, se basa en pequeños ordenadores pero en gran cantidad, procesando cada uno de ellos una porción de información pero actuando como uno solo.



En 2006, la comunidad Open Source basándose en las enseñanzas de Google, desarrolla la implementación que se conoce como Hadoop. A partir de ahí, Yahoo toma el relevo y lo impulsa, facilitando que grandes empresas, como Facebook, comiencen a usarlo y a sumarse a la comunidad Open Source en su labor de desarrollo.

¿Qué es Hadoop?

Hadoop es un **sistema de código abierto** que se utiliza para almacenar, procesar y analizar grandes volúmenes de datos, aislando a los desarrolladores de todas las dificultades presentes en la programación paralela.

Hadoop cuenta con todo un **ecosistema** de ayuda que, además de distribuir el fichero en sus nodos (que no son más que ordenadores con commodity-hardware), hace posible ejecutar procesos en paralelo; disponiendo también de módulos de control, monitoreo y consultas.

Empiezan entonces aparecer distintos **add-ons** que ayudan a poder trabajar, manipular y monitorizar la información que se está guardada sobre Hadoop.

Los componentes básicos de Hadoop son los siguientes:

HDFS

Es el sistema de archivo distribuido que permite que el fichero de datos no se guarde en una única máquina, sino que la información se distribuya en distintas.

MAPREDUCE

Se trata de un framework de trabajo que permite aislar al programador de todas las tareas propias de la programación en paralelo. Es decir, hace posible que un programa escrito en los lenguajes de programación más comunes se pueda ejecutar en un cluster de Hadoop.

La gran ventaja es el poder usar el lenguaje o las herramientas más adecuadas para la tarea concreta que se ha de realizar en cada momento.

Utilidades de Hadoop

Hadoop es un sistema que se puede implementar sobre hardware a un costo relativamente bajo, siendo a su vez totalmente gratuito para software.

Ello ha comportado que, toda la información que antes las empresas no podían procesar por las limitaciones de la metodología existente, hoy pueda ser procesada gracias a Hadoop. de esta forma se puede, no sólo obtener información nueva, sino también descubrir y aplicar otro tipo de análisis como por ejemplo, una regresión lineal, sobre millones de registros de su histórico.

Éste es el principal motivo detrás de la rápida propagación de su uso entre las empresas, que ven que, con una inversión relativamente baja, pueden afrontar nuevos retos y problemáticas que antes no podían afrontar y con un ROI muy rápido.

A su vez, para minimizar los riesgos de su aplicación, existen en el mercado distintas distribuciones de Hadoop con soporte 24/7 que ayudan a no depender de la comunidad Open Source, lo que ha contribuido a impulsar su adopción en entornos productivos

Créditos foto: <http://ctovision.com/2013/07/now-we-can-reveal-centripetal-networks-a-2013-fintech-innovation-lab-company/>

7. Cuatro megatendencias

Las megatendencias que arrasan en el mercado de TI tienen que ver con ampliar horizontes, superar límites y alcanzar, directamente, otro nivel. Las cuatro corrientes que inspiran la innovación y el cambio, tanto a nivel tecnológico, como de negocio, tienen nombre propio:

Movilidad

Actualmente, el acceso, interacción o transacción en relación con los datos de una empresa es algo que ya se puede realizar desde cualquier lugar. Para conseguirlo, la otra novedad es que este intercambio de datos se produce desde dispositivos propios que no pertenecen a la empresa, algo que supone un reto añadido para éstas, que ven dificultada su capacidad de control.

Cloud

Cada vez con más frecuencia, las empresas tienden a contratar un servicio de correo en el que se paga una cierta cantidad al año por tener una cuenta de email. De este modo, si la empresa crece y tiene que contratar a más personal podrá saber exactamente qué inversión va a tener que hacer para ofrecer el servicio que necesita.



Lo mismo sucede con el almacenamiento de datos. La diferencia con el pasado es que, gracias a la nube, ya no es necesario comprar como activos toda la infraestructura, por lo que la inversión inicial, costes de mantenimiento y de administración se ven considerablemente reducidos.

Con Cloud, además, se asegura un mejor nivel de servicio, pues los proveedores disponen de sistemas avanzados que garantizan su alta disponibilidad.

Social Media

El ser humano es un ser social por naturaleza y es precisamente en estos medios donde realmente nos mostramos como somos y nos relacionamos de un modo natural. Las redes sociales son el medio idóneo para dejar constancia de quiénes son nuestros amigos, cuáles son nuestros gustos y cuáles nuestras preferencias. Todo esta información se traduce en datos de gran valor para las empresas, que pueden orientar sus estrategias de captación, fidelización y producto a estos medios, extrayendo el máximo partido posible de lo que se publica en la red.

Analytics

Esta megatendencia representa el análisis llevado a otro nivel, mucho más allá del alcanzado por el BI tradicional. Analytics tiene en cuenta cualquier tipo de dato, sea de la índole que sea, con el objetivo de extraer la máxima información posible, al permitir navegar por la información sin las restricciones que imponen unos patrones prefijados de búsqueda.

Créditos foto: <http://www.channelbiz.es/2013/04/25/emc-crece-confia-negocio/>

8. Análisis en tiempo real

Hoy día, la paciencia no es una virtud demasiado extendida. Dicho de otra forma, el valor del tiempo es realmente apreciado y, por eso mismo, cuando una persona quiere algo, lo quiere en ese mismo



instante. Aplicar este principio al mundo de los negocios impone elevar el nivel de exigencia a que las empresas han de someterse en

sus intentos por dar satisfacción a la demanda.

Esta imposición es un aspecto importante, no sólo de cara a los servicios que se ofrecen, sino también en lo relativo al análisis de los datos, que se ha de llevar a cabo en tiempo real. Para lograrlo en el entorno Big Data, existen dos técnicas a tener en cuenta:

Data Stream

La velocidad es su máxima preocupación y lo que permite es realizar un **análisis mientras los datos están en movimiento**, antes de que lleguen a su destino final, momento en el que ya se está en disposición de almacenar toda la información obtenida.

CEP

Son las siglas correspondientes a “*Complex Event Processing*”. Se trata también de un análisis en tiempo real pero, en este caso, con una peculiaridad: que se ocupa de buscar eventos predefinidos.

Ejemplos de esta técnica serían una entidad bancaria, que tiene la capacidad de analizar en tiempo real todo lo que sucede con las tarjetas de crédito de un determinado usuario en cajeros y comercios; o el caso de los sensores de maquinaria, que permiten conocer en tiempo real la información generada por los equipos para, de este modo, poder enviar una señal de alerta antes

de la aparición de un fallo, haciendo posible tomar acción a tiempo.

Créditos foto: <http://bigdata.ticbeat.com/espana-pisa-acelerador-big-data/>

9. Implementar Big Data en su organización

Cuando el objetivo es la implementación de Big Data en la organización, lo primero que se debe hacer es conocer y cuantificar cuánto valor puede ofrecer al negocio. Es necesario llevar a cabo una autoevaluación que permita conocer cuáles son los límites que impiden avanzar, o acelerar el progreso, qué expectativas se tienen a medio y largo plazo o qué restricciones afectan a la recogida, procesamiento y análisis de la información.



Una vez respondidas todas estas cuestiones, se está en posición de decidir cuál es el proyecto de Big Data más adecuado para la empresa.

Elegir el data warehouse más adecuado para Big Data

El primer paso es la elección de tecnología. Hay que seleccionar la que mejor se adapta a las necesidades corporativas, en otras palabras, la que facilitará la consecución de los objetivos de negocio. Son las metas las que determinan la tecnología, y no al revés.

Por ejemplo, si se quiere realizar un análisis de eventos en tiempo real, hará falta contar con una tecnología del CEP para poder determinarlo; mientras que, si el objetivo es realizar un análisis predictivo, entonces se necesitan bases de datos con capacidades de procesamiento MPP que puedan hacer un análisis de lo actual contra todo el histórico.

Conocer y anticiparse al negocio es imperativo

Toda iniciativa de Big Data necesita perfiles de negocio que conozcan la empresa y el sector en detalle, ya que van a aparecer nuevos roles, como el Data Scientist o Data Analyst, que harán posible, en base a esa información, acceder a un nivel de conocimiento superior, inalcanzable hasta el momento.

Por ejemplo, para realizar un análisis completo de sentimiento de la percepción que se tiene de una tu marca y la empresa que hay detrás, es necesario conocer a la perfección el producto que se ofrece, los canales que se emplean, los perfiles a que se dirige o el tipo de campañas de marketing que se lanzan. Sólo así es posible encontrar toda la valiosa información que hay detrás de los datos.

Encontrar las herramientas de análisis correctas

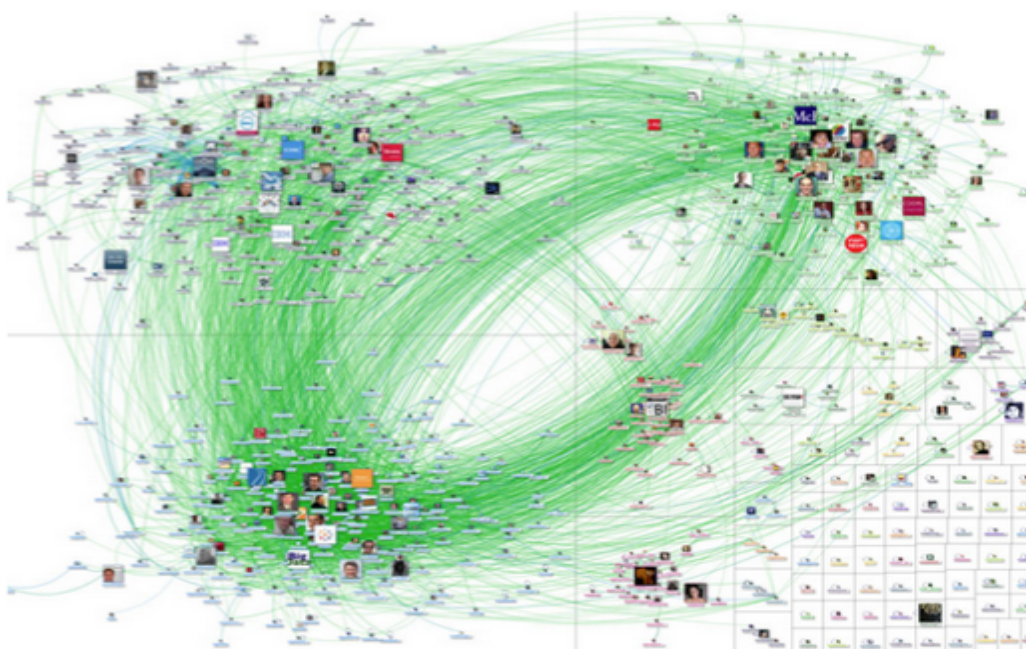
En función del tipo de análisis que se quiera realizar se elegirán unas herramientas u otras. No son iguales las herramientas que se emplean al realizar un análisis predictivo que las usadas para llevar a cabo un análisis en tiempo real.

Créditos foto: <http://www.v3.co.uk/v3-uk/the-frontline-blog/2260731/using-big-data-and-creativity-for-social-good>

10. Humanizando el Big Data

Cómo Big Data impacta en tu día a día

Existe un estudio llamado *"The Human face of Big Data"* que muestra que Big Data es algo más habitual en el día a día de lo que se cree.



Muchas de las decisiones que se toman, derivan en acciones que se hacen realidad gracias a la utilización de Big Data. Ejemplos de ello son las compras a través de internet, tras cada una de las cuales se esconden multitud de datos a tener en cuenta; la conducción de un vehículo mediante sensores; o el simple hecho de acudir al médico, donde el historial de cada paciente está digitalizado, centralizado y permite ser comparado con el de otros pacientes.

Big Data permite innovar

En una empresa, no es lo mismo copiar e ir detrás de los competidores que ser el primero en ofrecer un servicio nuevo. De ahí proviene la importancia de la innovación que Big Data propicia,

que permite hacer volar la imaginación sin límites al eliminar por completo cualquier tipo de restricción en la gestión de datos.

Seguridad

La cantidad de datos que hoy día se puede analizar debe ser tratada teniendo en cuenta el nivel de privacidad que los clientes y usuarios de negocio están dispuestos a aceptar.

Por ejemplo, en Google no todo es gratis. El precio que se paga es el de la privacidad. Al interactuar con el motor de búsqueda, éste recopila información personal que ayuda a completar el perfil individual de cada usuario.

Las empresas deben encontrar el equilibrio entre la cantidad y tipos de datos que son capaces de analizar de sus clientes y los límites de tolerancia de éstos con respecto al acceso a su privacidad. Descubrir tendencias puede no interesar tanto si para hacerlo hay que sobrepasar determinadas barreras.

Créditos foto: <http://www.datacenterjournal.com/it/thoughts-downsizing-big-data/>