



Sustento del uso justo
de **Materiales Protegidos**
derechos de autor para
fines educativos



UCI

Universidad para la
Cooperación Internacional



UCI

Sustento del uso justo de materiales protegidos por derechos de autor para fines educativos

El siguiente material ha sido reproducido, con fines estrictamente didácticos e ilustrativos de los temas en cuestión, se utilizan en el campus virtual de la Universidad para la Cooperación Internacional – UCI - para ser usados exclusivamente para la función docente y el estudio privado de los estudiantes en el curso **Administración de la Información para la toma de Decisiones** perteneciente al programa académico **MATI**.

La UCI desea dejar constancia de su estricto respeto a las legislaciones relacionadas con la propiedad intelectual. Todo material digital disponible para un curso y sus estudiantes tiene fines educativos y de investigación. No media en el uso de estos materiales fines de lucro, se entiende como casos especiales para fines educativos a distancia y en lugares donde no atenta contra la normal explotación de la obra y no afecta los intereses legítimos de ningún actor.

La UCI hace un USO JUSTO del material, sustentado en las excepciones a las leyes de derechos de autor establecidas en las siguientes normativas:

- a- Legislación costarricense: Ley sobre Derechos de Autor y Derechos Conexos, No.6683 de 14 de octubre de 1982 - artículo 73, la Ley sobre Procedimientos de Observancia de los Derechos de Propiedad Intelectual, No. 8039 – artículo 58, permiten el copiado parcial de obras para la ilustración educativa.
- b- Legislación Mexicana; Ley Federal de Derechos de Autor; artículo 147.
- c- Legislación de Estados Unidos de América: En referencia al uso justo, menciona: "está consagrado en el artículo 106 de la ley de derecho de autor de los Estados Unidos (U.S.Copyright - Act) y establece un uso libre y gratuito de las obras para fines de crítica, comentarios y noticias, reportajes y docencia (lo que incluye la realización de copias para su uso en clase)."
- d- Legislación Canadiense: Ley de derechos de autor C-11– Referidos a Excepciones para Educación a Distancia.
- e- OMPI: En el marco de la legislación internacional, según la Organización Mundial de Propiedad Intelectual lo previsto por los tratados internacionales sobre esta materia. El artículo 10(2) del Convenio de Berna, permite a los países miembros establecer limitaciones o excepciones respecto a la posibilidad de utilizar lícitamente las obras literarias o artísticas a título de ilustración de la enseñanza, por medio de publicaciones, emisiones de radio o grabaciones sonoras o visuales.

Además y por indicación de la UCI, los estudiantes del campus virtual tienen el deber de cumplir con lo que establezca la legislación correspondiente en materia de derechos de autor, en su país de residencia.

Finalmente, reiteramos que en UCI no lucramos con las obras de terceros, somos estrictos con respecto al plagio, y no restringimos de ninguna manera el que nuestros estudiantes, académicos e investigadores accedan comercialmente o adquieran los documentos disponibles en el mercado editorial sea directamente los documentos, o por medio de bases de datos científicas, pagando ellos mismos los costos asociados a dichos accesos.

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. 4 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 generamos muchos datos y las empresas, cada vez con más frecuencia, exigen saber; no sólo lo que sucede en la actualidad, sino también lo que va a pasar en un futuro solicitando niveles de servicio mucho más exigentes que hace unos años. Es precisamente en esta tesitura actual dónde se encuentra Big Data.



Definamos entonces: ¿Qué es Big Data?

Para realizar 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.

¿Porque 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 hay más valor. Hoy en día para muchas empresas puede llegar a ser más importante detectar el cliente que más influye al resto de posibles compradores, que propiamente el cliente al que más le venden.

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 Únicos pertenecen a un grupo

Big Data nace de la exigencia de dar respuesta a toda una serie de necesidades del mercado actual.

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 nuestras ventas, dirigiéndolas de forma mucho más directa para lograr el éxito esperado. Además, éstos están relacionados con otras personas que nos permiten tanto conocer más a nuestro cliente como aumentar el número de clientes potenciales, conocimiento que aporta un valor incalculable a nuestra empresa.



De la transacción a la interacción

Hasta hace unos años siempre lo más importante era la transacción, pero actualmente esto está cambiando, ya que cada vez se da más importancia a la interacción constante con el mundo que nos rodea y las empresas deben aprovecharse de toda esta información que generamos.

En efecto, las empresas deben aprender a optimizar y a conseguir la mayor información posible de

sus clientes, creando por ejemplo un nuevo canal 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 capacidades van mucho más allá de la llamada, pudiendo realizar servicios basados en la geoposición en la que estoy o incluso detectar cuál ha sido la última interacción realizada y dónde se ha producido, y lo mismo sucede con las **casas** que cada vez tienen más domótica inteligente o los **coches**, muchos ya con varios sistemas de sensores, generando todos ellos información muy valiosa perfecta para crear nuevos servicios de valor.

Big Data aprovecha todas estas nuevas fuentes de información provenientes de estos nuevos dispositivos.

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 nos prefieren, qué dicen de nosotros, aspectos todos ellos que requieren de nuevos modelos mucho más complejos.

Esas nuevas necesidades han llevado a nuevos requerimientos por parte de las empresas y, precisamente para poder analizar toda esa información, lo que anteriormente se conocía como Business Intelligence actualmente requiere de un nuevo modelo de análisis, **Big analytics**, que permita dar forma al Big Data.



Esta intervención necesaria se debe al 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 utilización de procesos matemáticos complejos, trabajando con toda la data al nivel de granularidad más bajo de que disponemos, la data en crudo, y 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.

En efecto, con el sistema tradicional los procesos necesarios para poder cargar unos modelos, por ejemplo, eran muy largos, lo que conllevaba con frecuencia que al finalizar dicho proceso el negocio

ya había evolucionado por lo que la utilidad del modelo se veía reducida. Realizar los cambios pertinentes para alienarlo de nuevo con el negocio era costoso, largo, poco ágil e ineficaz, al no poder además aplicar los algoritmos que quizás ahora nos gustaría usar 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 análisis a tiempo real tan deseados. Esto a llevado a la inevitable necesidad de unas nuevas infraestructuras capaces de soportar ese nuevo paradigma de datos que nos están llegando.

Además, en una empresa es importante saber lo que me va a costar la infraestructura actual con lo que es muy importante en Big Data que esa arquitectura sea escalable y que pueda ser predecible para así saber el coste de mi crecimiento. De este modo, permite poder proyectar y determinar la inversión necesaria para seguir creciendo, lo cual no sucedía con las infraestructuras actuales en 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 se basaban en unos modelos que habían sido predefinidos muchas veces 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, permite a esos datawarehouse evolucionar y además proporciona unas capacidades ilimitadas tanto para incorporar todo tipo de datos como también para el procesamiento de esos datos.

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

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 tenga que sustituir a los datawarehouse actuales, ya que cada uno de ellos **cubre necesidades distintas, y usan infraestructuras y tecnologías muy diversas.**



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 mediante el uso de Analytics

información muy valiosa escondida detrás de los propios datos y hasta el momento desapercibida por la empresa.

Cómo integramos estos dos mundos?

Se puede usar un **modelo híbrido**, es decir, podemos seguir manteniendo para un cierto tipo de datos el data warehousing actual, nutriéndose con la data estructura, y para todo el resto podemos incorporar el nuevo modelo de analytics data warehousing.

Ciertamente, los analytics data warehousing pueden convivir con los actuales peromantiéndolos para realizar análisis y funciones diferentes. Usaremos herramientas de trabajo distintas, sin embargo podremos explotarlos desde las actuales herramientas de visualización del BI tradicional,

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

5. MAD- Nueva Metodología

La existencia de nuevas necesidades en el mercado han requerido que nuestra empresa haya tenido que adaptarse e incorporar nuevas infraestructuras y requerimientos. 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, ya sean datos antiguos, nuevos, estructurados o no-estructurados etc. y se deben captar de forma muy rápida sin perder tiempo en pre-procesos costosos.

Ágil

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

Detallado

La metodología conlleva utilizar **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 incapaces de trabajar a un nivel más preciso de detalle, lo cual para extraer un análisis a nivel genérico no conllevaba problemas pero si se quería profundizar más esta metodología era insuficiente.

Usando, pues, esta metodología que te permite llegar a un nivel de detalle muy precioso las decisiones que se tomen a nivel empresarial serán mucho menos riesgosas y seguramente 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**, haciendo suyo el lema "divide y vencerás", es decir, se basa en pequeños ordenadores pero en gran cantidad, procesando cada uno de ellos un trozo de información pero actuando como uno solo.



En 2006 Google saca un *paper* explicando sobre ello, lo toma la comunidad Open Source y de ahí sacan esta implementación a la que denominan Hadoop. A partir de ahí, Yahoo lo toma y lo impulsa, grandes empresas como Facebook empiezan a usarlo y a desarrollar junto la comunidad Open Source.

¿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 un **ecosistema** para ayudarte ya que, aparte de distribuir el fichero en los nodos (que no son más que ordenadores con commodity-hardware) es capaz de ejecutar procesos en paralelo y también tienen módulos de control, monitoreo y para poder realizar 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 que son el corazón de Hadoop son los siguientes:

HDFS

Se trata del sistema de archivo distribuido que permite que el fichero de datos no se guarde en una única máquina sino que se distribuya la información en distintas máquinas.

MAPREDUCE

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

La gran ventaja es la siguiente: permitirte poder usar el lenguaje o las herramientas más adecuadas para la tarea concreta que deba realizar. Un ejemplo, para la manipulación de cadenas usar Perl para su tratamiento.

¿Para qué le puede servir?

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 aquella información que antes las empresas no podían procesar por la metodología existente hoy puedan hacerlo con Hadoop, pudiendo de este modo sacar información nueva y a su vez descubrir y aplicar otro tipo de análisis como por ejemplo, una regresión lineal, sobre millones de registros de su histórico.

Es precisamente por ello que se está expandiendo tanto su uso entre las Empresas porque ven que con un costo relativamente bajo y un retorno rápido de la inversión pueden afrontar nuevos retos y problemáticas que antes no podían afrontar.

A su vez, para minimizar los riesgos de su aplicación, existen en el mercado distintas distribuciones de Hadoop con soporte 24/7 y no depender de la comunidad Open Source para solucionar un bug que pudieran encontrar. Esto ha hecho que se haya impulsado mucho su adopción ya en entornos productivos

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

7.4 Megatendencias

Las megatendencias que actualmente tiene el mercado de TI se podrían concretar en las siguientes básicamente:

Movilidad

Actualmente el acceso, interacción o transacción en relación con los datos de una empresa es algo que se puede realizar desde cualquier lugar y desde dispositivos propios que no pertenecen a la empresa, lo cual supone un reto añadido para éstas, las cuales antes tenían acotado el terreno a controlar, pero actualmente con la llegada de los móviles, ipads, etc. ven dificultado su ámbito 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 correo. De este modo, si la empresa crece y tiene que contratar a más personal podrá saber exáctamente qué inversión va a tener que hacer para ofrecer el servicio de correo que la empresa necesita.



Y ello a diferencia de lo que sucedía antes, cuando las empresas tenían que comprar como activos toda su infraestructura, , con lo cual resultaba muy costoso de mantener por la inversión, y costes de administración entre otros aspectos.

Con Cloud además, se asegura un mejor (más alto) nivel de servicio, pues los proveedores disponen de sistemas avanzados de alta disponibilidad del servicio.

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, mostrando quienes son nuestros amigos, lo que nos gusta y lo que no nos gusta. Todo esto es información profundamente valiosa para las empresas, que pueden orientar sus estrategias de captación, fidelización y producto a estos medios.

Analytics

Esta megatendencia representa el análisis llevado a otro nivel, mucho más allá del BI tradicional. Va a tener en cuenta cualquier tipo de dato, sea de la índole que sea, y ello con el objetivo de poder sacar la máxima información posible y poder navegar por la información sin patrones prefijados de búsqueda.

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

8. Análisis en Tiempo Real

Actualmente, la tendencia del ser humano es que cuando quiere algo lo quiere en ese mismo instante, lo cual ha conllevado que el nivel de exigencia sea cada vez más elevado para las empresas



Esta exigencia, es un aspecto importante no sólo para los servicios que ofrecen las empresas, sino también va a impactar mucho en los análisis de los datos que se hagan en tiempo real.

Respecto a este análisis en tiempo real, que es lo que Big Data ofrece, existen básicamente 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 guarda toda la información obtenida.

CEP

Son las siglas pertenecientes a “*Complex Event Processing*”.

Se trata también de un análisis en tiempo real pero centrándose en los **eventos que puedan suceder en un determinado proceso**, es decir, su objetivo es buscar eventos pre-definidos, analizando los datos en tiempo real.

Como ejemplos de esta técnica, podríamos citar el supuesto de un banco en el que se analiza en tiempo real todo lo que está sucediendo con mis tarjetas de crédito en cajeros y comercios.

Otro ejemplo sería el caso de los sensores de maquinaria, donde se analiza en tiempo real la información que generan dichos sensores, para de este modo poder alertar antes de que falle la máquina correspondiente y proceder a la sustitución de una pieza, por ejemplo.

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 una determinada organización, lo primero que se debe hacer es detectar cuánto valor puede ofrecer el Big Data a la empresa, es decir, debemos observar qué cosas no estamos haciendo y qué nos gustaría hacer, qué información no está siendo ni captada, ni analizada y proceder a identificarla. Además, es necesario cuantificar el valor que se puede sacar de ello.



Una vez realizadas estas acciones, se podrá decidir cuál va a ser el proyecto de Big Data más adecuado para la empresa.

Elegir el Data Warehouse de Big Data correcto

Antes de empezar debemos entender qué tecnología es la que mejor se adecúa a nuestras necesidades y ello dependerá de cuáles son nuestros objetivos.

Por ejemplo, si lo que quiero es realizar un análisis de eventos en tiempo real lo que necesito será una tecnología del CEP para poder determinarlo, mientras que si mi objetivo es realizar un análisis predictivo necesito de bases de datos con capacidades de procesamiento MPP que puedan hacer un análisis de lo actual contra todo el histórico.

Es precisamente por ello que, en función de nuestros objetivos, se deberá elegir qué tecnología nos

resuelve nuestra problemática concreta.

Conocer (y avanzarse a) el negocio es imperativo

Toda iniciativa de Big Data necesita perfiles de negocio que conozcan el negocio en detalle, ya que van a aparecer nuevos roles, como es el caso de Data Scientist o Data Analyst, y se van a poder descubrir muchas cosas escondidas en los datos que con frecuencia ni tan siquiera sabemos que están allí.

Por ejemplo, para realizar un análisis completo de sentimiento de la percepción que se tiene de tu marca y de tu empresa, necesitamos conocer a la perfección el producto que ofrecemos, qué canales usamos, a qué perfiles nos dirigimos, o qué publicidad realizamos, con la intención de poder encontrar toda aquella información valiosa que hay detrás de los datos.

Dependiendo del negocio de que se trate habrá más o menos información en la que indagar, y de ahí precisamente la importancia de conocer hasta los máximos detalles del negocio en cuestión.

Encontrar las herramientas de análisis correctas

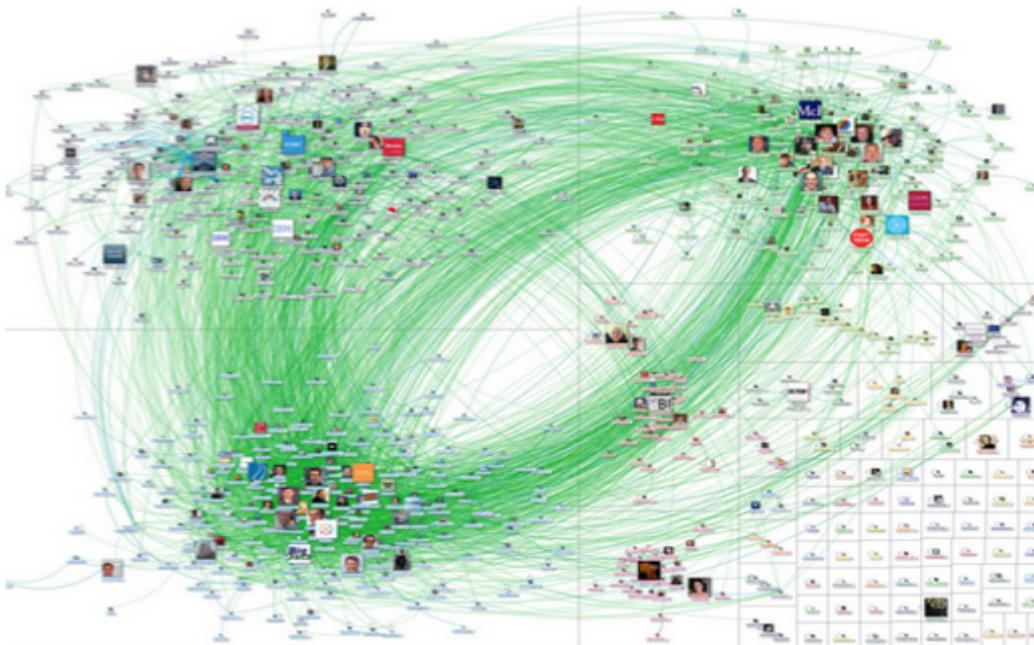
En función del tipo de análisis que se quiera realizar se elegirán unas herramientas u otras. No son las mismas las herramientas que se usarán al realizar un análisis predictivo que las usadas para realizar un análisis en tiempo real.

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

Como afecta Big Data en tu día a día

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



Muchas de las cosas que hacemos, las hacemos gracias a la utilización de Big Data. Ejemplos de ello los podemos encontrar en la compra a través de internet, existiendo detrás de esta acción multitud de datos a tener en cuenta, o en el uso de un auto mediante sensores, o en el simple hecho de ir al médico, donde todo tu historial está digitalizado, centralizado y comparado con otros pacientes.

Big Data permite Innovar

En una empresa, no es lo mismo copiar e ir detrás de otras empresas que ser el primero en ofrecer un servicio nuevo, y de ahí proviene la importancia de la innovación.

Vale la pena correr el riesgo de fracasar, para poder ofrecer un servicio totalmente nuevo, y ahí es donde entra en juego Big Data, pues te permite hacer volar tu imaginación al no tener ningún tipo

de restricción en la gestión de datos.

Seguridad

Esta cantidad de datos que somos capaces de analizar debe ser tratada teniendo en cuenta cuál es el nivel de privacidad que nuestros clientes estén dispuestos a aceptar.

Por ejemplo, en Google no todo es gratis; el precio que pagas es el de tu privacidad. Debes valorar si tu nivel de tolerancia es aceptable en relación con lo que Google puede saber de ti, y lo mismo sucede con Street View o Google Maps.

Las empresas deben balancear la cantidad y tipos de datos que son capaces de analizar de sus clientes y potenciales, respecto a los límites de tolerancia de éstos con respecto al acceso a su privacidad. Si sobrepasa estas barreras, las consecuencias serán que su nivel de servicio se verá afectado, por lo que hay que tenerlo muy en cuenta a la hora de usar esa cantidad de datos que permiten descubrir tendencias.

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