



EL CAMINO HACIA BIG DATA

Análisis, desarrollo, entrega

Big data surge de la necesidad que tienen las organizaciones de sacar valor de su información —ya sea interna o externa, procedente de datos estructurados o no estructurados— para realizar consultas y conocer los beneficios que pueden obtener. Pero ¿qué es realmente *big data* y cuál es el camino adecuado para implementarlo en el contexto real en el que se encuentran las empresas?



ANDREA TOSCANINI
Manager Business Analytics



Son muchas las definiciones que podemos encontrar de *big data*. Para nosotros, surge desde el mismo momento en que una empresa tiene la necesidad de consultar varias bases de datos para obtener información contextualizada con el objetivo de utilizarla para mejorar las líneas estratégicas de la compañía.

Se trata de dar valor a la enorme cantidad de datos con los que cuentan las organizaciones. De hecho, *business intelligence* (BI) y *big data* están muy relacionados, son parientes muy cercanos. Un sistema de BI permite tomar decisiones

con la información estática interna, pero ahora las empresas quieren inyectar también muchos otros tipos de datos y hacer predicciones basadas en el histórico. Para ello es necesario tener en cuenta tres aspectos fundamentales: el almacenamiento, la capacidad de proceso y el *real-time*.

AMORTIZAR LA INVERSIÓN

El camino que nosotros definimos comienza con un análisis previo, que hacemos de forma conjunta con el cliente. Este proceso es fundamental, ya que las compañías cuentan con una infraestructura propia y han realizado ya inversiones en hardware y software, soluciones de BI, bases de datos... En este primer paso es importante asegurar que

están sacando todo el potencial de sus plataformas, que se están optimizando los recursos disponibles, que su infraestructura está siendo explotada al máximo nivel y se está sacando el mayor beneficio posible.

Por ejemplo, en vez de adquirir herramientas específicas para generar un árbol de decisiones, tal vez sea más inteligente utilizar predicciones a través de algoritmos en el *data mining* que tienen en su BI, que, como ocurre en la mayoría de las empresas, está infrautilizado. Otro ejemplo es el caso de SAP, que ofrece un importante número de opciones en el estándar que permiten sacar partido a los datos sin que sea necesario acudir a herramientas externas.

Es imposible que un proyecto de big data ofrezca un ROI positivo sin incorporar a los usuarios de TI con un nivel de protagonismo similar al de negocio

Desde ese momento estamos ahorrando costes en la inversión de lo que va a ser el futuro *big data* de la organización.

Los siguientes pasos consisten en detectar las necesidades —respecto a hardware, software o servicios— para cubrir el objetivo planteado.

Y el procedimiento vuelve a ser el mismo: cada paso que damos debe implicar necesariamente un análisis, en el que se decida también qué tipo de tecnología y recursos (internos o externos) serán necesarios.

Si hay que entrar en redes sociales es importante tener claro en cuáles de ellas y para analizar qué... Es posible que únicamente necesitemos capturar datos procedentes de *open data*, y eso se puede suplir con la utilización de un API y de un servidor de pequeño tamaño.

En este aspecto, hay que recordar que nosotros somos integradores y, como tales, agnósticos con respecto a la tecnología que se vaya a utilizar, aunque no con respecto a los procesos.

Una vez decidida la inversión, resulta fundamental realizar un estudio a largo plazo sobre lo que implicará en ámbitos como el coste de licencias, soporte y mantenimiento del hardware por un período razonable, el incremento que se puede producir teniendo en cuenta el crecimiento vegetativo de los datos almacenados...

Normalmente, en las pruebas de concepto o los pilotos se suelen planificar unos niveles de rendimiento y unos costes determinados, pero después, al llevarlo al entorno real y cargar todos los datos que se deben manejar, el precio para alcanzar ese rendimiento se multiplica.

La planificación que nosotros entregamos es siempre a cinco años, contemplando un escenario en el que incluimos todo lo que pueda afectar a esa puesta en marcha de *big data*. De hecho, podemos también hacernos cargo de ese mantenimiento y garantizar que los números van a cumplirse de aquí a cinco años. Dentro de nuestro servicio está el compromiso de que ese escenario va a ser así.

NEGOCIO Y TI

Un elemento clave en los proyectos de *big data* es el tema de la formación y de la implicación de los diferentes *key users*. Durante este camino a *big data*, el cliente está involucrado desde el minuto cero, acompañando a los consultores en todos los análisis que se realizan, así como en las decisiones que se toman.

Esto es especialmente importante, ya que se trata de tecnologías con el consiguiente reto derivado de un proceso de cambio. En nuestro camino, los *key user* saben lo que estamos haciendo y por qué, trabajan y definen el proyecto con nosotros. Llegamos de forma conjunta al final del camino y se sienten seguros de la tecnología elegida.

Además, para nosotros es tan importante la involucración del usuario de negocio como del de tecnología. TI es fundamental desde el primer momento, en el que buscamos conocer el nivel de optimización de los recursos hardware y software. Es imposible que un proyecto de *big data* ofrezca un ROI positivo si no hemos incorporado a los usuarios de TI con un nivel de protagonismo similar al de negocio.

De hecho, los departamentos de TI no tienen que sentir como una amenaza la llegada de proyectos de *big data*, sino más bien como una oportunidad de aprender, de hacer cosas nuevas y de sacar el máximo beneficio de los sistemas instalados.

A medida que se va descubriendo información se tienen perspectivas diferentes, que se traducen en nuevas ideas y en consultas basadas en los nuevos escenarios

RETROALIMENTACIÓN

Otro de los puntos importantes de “nuestro camino” es que los proyectos los hacemos utilizando metodologías ágiles, como XP o SCRUM, dividiendo el proyecto en pequeños desarrollos, con una duración máxima de tres semanas, para realizar entregables rápidos. Huimos de proyectos de largo alcance. Estas entregas cada poco tiempo permiten que el usuario vaya viendo resultados constantemente, ajustando el modelo y retroalimentándolo.

El objetivo de *big data* es descubrir datos y los resultados obtenidos pueden hacer que cambie el modelo, o la manera de investigar o lo que queremos buscar. Por esta razón es importante ir realimentando este ciclo constantemente a medida que las entregas rápidas van siendo validadas.

Cuando se descubre el dato se obtiene valor de negocio, pero, a medida que se va descubriendo información se van teniendo perspectivas que antes no se tenían, que se traducen en nuevas ideas y en consultas basadas en los nuevos escenarios. Haciendo entregas rápidas conseguimos muchas interacciones con los usuarios para ir acomodando las demandas de información a las nuevas perspectivas que se van abriendo.

Y esto no acaba nunca, volveríamos a comenzar el proceso en la búsqueda de nuevos retos, a descubrir los datos y a obtener valor para el negocio que redundaría en una nueva perspectiva... ●●●