

# La fiebre de los datos

Entre las últimas tendencias de las empresas se encuentra comenzar a sacar mayor provecho de sus almacenes de datos. Explotar todos esos datos que tienen almacenados de los clientes o que se han ido generando a lo largo de los años con la intención de poder generar mayor conocimiento... o eso es lo que intentan.

Y es una ardua tarea la que se plantean. Los costos de diseño e implantación son altos, ya que requiere una migración previa y una estrategia de escalado, ya que si la empresa crece el análisis requerirá de mayores recursos. Como no es de extrañar, muchas de esas empresas comienzan mal desde el principio. Sufren de la fiebre de la moda y no se llegan a plantear para qué quieren un sistema de BI (Business Intelligence). Y no todos los negocios requieren (o pueden disponer) de estas tecnologías debido al esfuerzo e inversión que es necesario realizar, o que el tamaño de la empresa no es suficientemente grande como para necesitar gestionar el conocimiento con este tipo de herramientas automáticas.

El primer paso para comenzar a implantar una solución de este tipo es decidir los requisitos de la empresa, lo que se quiere predecir y obtener como respuesta del sistema. Una vez se tiene el dominio de análisis, hay que realizar una revisión de los datos que ya existen en el sistema: tanto la estructura de estos datos, como el formato que tienen los datos y encontrar patrones en estos datos.

El siguiente paso es decidir qué forma se desea que tengan los datos. Unificarlos para que tengan una misma forma y patrón, de forma que se eviten duplicidades de datos, errores y falta de información. Durante esta fase se elimina toda la "basura" que existe en el sistema y se optimiza la base de datos.

Finalmente, el tercer paso de la migración se trata de

localizar fuentes de terceros para completar los datos almacenados en el sistema. Por ejemplo, se pueden emplear repositorios de OpenData para tener datos de una fuente fiable, con datos validados y con libertad de poder usarlos. De esta forma si existen direcciones sin completar o con fallos ortográficos, se puede corregir.

Una vez diseñada la solución de migración se debe realizar el despliegue sobre el nuevo sistema, donde podrá ser explotado con herramientas dotadas de inteligencia artificial para generar predicciones. El conocimiento generado debe ser formateado correctamente para mostrarse en un sencillo panel de mandos. Es importante que el diseño del panel sea sencillo ya que si se distrae la vista con demasiada información puede resultar abrumante. Esto puede provocar que la persona que lo tiene que analizar no reciba correctamente la información.

Un requisito que debiera estudiarse detenidamente es evitar la corrupción de esta nueva base de datos. Como ya he mencionado todo este proceso es realmente costoso tanto en tiempo como en dinero, para ello hay que evitar que los nuevos datos que llegan acaben con la consistencia que se obtiene mediante estos procesos.

Para evitar esto he pasado a realizar varias practicas en los proyectos que realizo. Durante el diseño del esquema de datos, los campos que varían levemente a lo largo del tiempo (país, dirección, títulos de educación, niveles de idiomas...), los añado directamente a la base de datos como *fixtures*. Después, cuando realizo los formularios que los usuarios deben rellenar, doy opción solo a que se puedan emplear las opciones que existen en la base de datos.

De esta forma se corrigen varios problemas: por un lado la consistencia de la base de datos no va a ser comprometida, y por el otro los formularios pasan a ser mucho más simples. Los usuarios pueden rellenar rápidamente un formulario sin necesidad de perder el tiempo. Como contra, requiere tener una

fase de mantenimiento de estos datos, ya que campos como las direcciones o los estudios están cambiando constantemente. Se pueden tener campos de libre edición, pero deben ser monitorizados y purificados de manera que se reduzca el impacto que pueda provocar al resto de conjunto de datos.

Llegar a un punto de equilibrio no es fácil, pero con una buena fase de diseño antes de meterse en harina puede reducir los problemas a largo plazo. Un buen conjunto de datos dará a un negocio la ventaja competitiva que ofrecen las soluciones de BI.