

Artículo 3. Using OLAP and multidimensional data for decision making

José María García García

12 de marzo de 2016

1. PREGUNTAS

Leer el artículo y obtener de él información acerca de los siguientes puntos.

- **¿Qué dos sistemas puede utilizar un gestor para obtener un mejor conocimiento de su empresa?**

Procesamiento Analítico En Línea (On-Line Analytical Processing, OLAP) y Bases de Datos de Múltiples Dimensiones (MultiDimensional DataBases, MDDB). Las bases se utilizan para guardar datos que se usarán en análisis (ya veremos porque estas bases son mejor que las relacionales para esta finalidad) y con los sistemas OLAP manipulamos la información de dichas bases. Los sistemas OLAP nos presentan la información resumida que han extraído de bases de múltiples dimensiones y nos permiten estudiar los datos que contienen, trocear los datos para estudiar un aspecto de los mismos, etc.

- **¿Qué diferencia hay entre los sistemas OLTP y OLAP?**

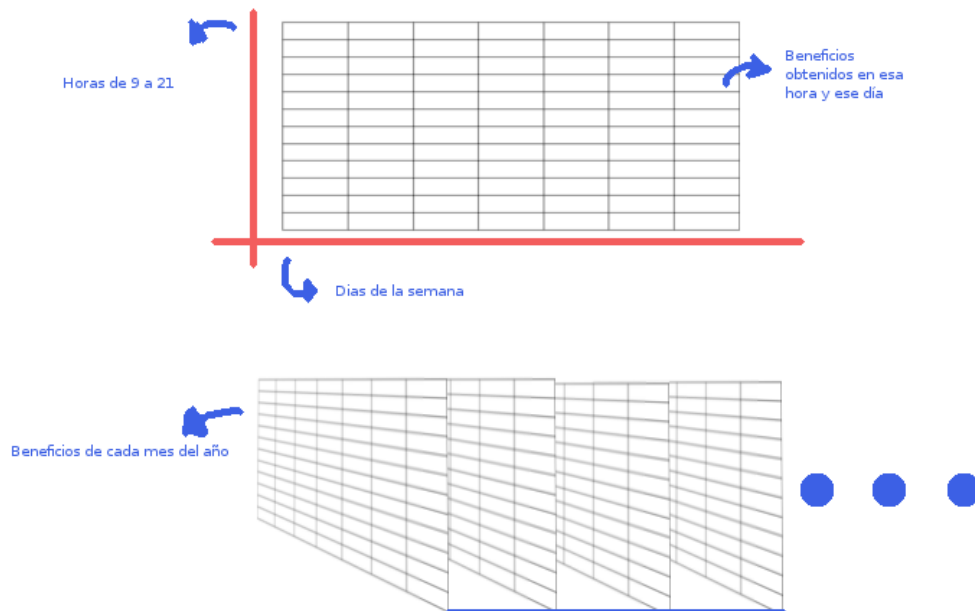
Los sistemas OLTP son Sistemas de Procesamiento de Transacciones En Línea (On-Line Transaction Processing), por lo que están centrados en las operaciones del día a día de la empresa, y no tanto en el análisis de los datos que generan esas transacciones. Aunque estos sistemas suelen incorporar mecanismos para generar informes según las actividades de la empresa, estos no sirven para tener una visión general ni para apreciar rápidamente puntos fuertes y puntos débiles. Además, son informes cuya lectura resulta tediosa y los gestores no suelen leerlos.

- **A la hora de obtener información sobre el funcionamiento general de la empresa, ¿qué características tiene el tipo de sistema preferido por los gestores?**

A diferencia del sistema comentado anteriormente, un sistema que usa OLAP y MDDB es intuitivo, fácil de usar y está orientado a la búsqueda de información de tal forma que permita, mediante un vistazo rápido, saber que está pasando en la empresa y formarse una idea general sobre la misma.

- **¿Cómo se representa la información cuando utilizamos una base de datos multidimensional?. Indica un ejemplo**

La información se representa mediante arrays (vectores, listas) de varias dimensiones. A la estructura que toman los datos en la base se la llama **hipercubo**, aunque esta puede no tener forma de cubo, esto es, si el array en cuestión tiene una dimensión 4, no podemos representar esa estructura como un tetraedro, ya que para ello el array debería ser de dimensión 3. Estas estructuras son preferidas a otras porque permiten “jugar” con la información para observarla desde muchos puntos de vista diferentes, centrar el análisis en un único aspecto de los datos, etc. El uso más simple de estas bases es el que podemos observar en el ejemplo.



En este caso, el dato a estudiar ha sido el beneficio que obtiene la empresa y las variables según las cuales lo hemos estudiado han sido la hora del día, el día de la semana y el mes del año. Esta representación permite ver como varían los datos según cada una de las variables, localizar los puntos donde se detecten pérdidas y arreglarlos, etc. Cabe también destacar que cada valor de beneficio está identificado de forma unívoca por sus coordenadas en el cubo (mes, día, hora).

- **¿Por qué las base de datos relacionales son adecuadas para almacenar información transaccional y no analítica?**

Porque estas bases permiten guardar información de las transacciones de forma rápida y sencilla, información que luego se puede estudiar con los informes autogenerados o a través de órdenes SQL. No obstante, dada la fragmentación y volumen que presentan los datos usados en analítica, manejar estas bases de datos dejaba de ser una trivialidad, por lo que para analítica se prefieren las MDDB.

- **¿Qué tipo de información suele almacenarse en las base de datos multidimensionales?**

Los datos comúnmente almacenados en estas bases de datos suelen ser numéricos y muestran el rendimiento que ha tenido la empresa u organización a lo largo de su historia. Estos datos provienen de otras bases de la empresa, que se vuelcan en la base multidimensional y esta limpia, resume y procesa la información para el uso de los gestores.

- **¿Qué tipo de interactividad es permitida sobre los datos de una base de datos multidimensional? ¿Qué tipo de aplicaciones permiten esta interactividad?**

Siguiendo la metáfora del cubo respecto a la representación de la información, el cubo puede **hacerse rodajas**, para observar los valores de los datos respecto a un valor concreto de una de las coordenadas de los mismos. Estas rodajas pueden **trocearse** además para observar la información enmarcada en un rango de valores. Podemos observar cuadrados concretos del cubo o subconjuntos del mismo. Podemos **comprimir** el cubo para observar un resumen de todos los datos, o al contrario, podemos **aumentar** una sección del cubo para estudiar datos muy localizados. Podemos pasar el cubo por un **filtro** para quedarnos con los bloques que satisfagan ese filtro. Normalmente todo esto se hace mediante interfaces gráficas, lo que hace el análisis más ameno que la lectura de un informe. Al conjunto de herramientas que se usan para esa manipulación del cubo se les llama OLAP.

- **¿Qué opciones tenemos a la hora de implantar una MDDB? ¿En qué se centra cada una de ellas?**

Existen dos enfoques a la hora de implementar una MDDB:

1. Enfoque top-down (de arriba a abajo): centrado en la planificación del negocio.
2. Enfoque bottom-up (de abajo a arriba): centrado en bases y sistemas existentes.
3. Enfoque middle-out (mezcla de los dos anteriores): centrado al mismo tiempo en lo que necesita la empresa y las tecnologías en las que se apoya.

Desarrollaremos los pasos de cada uno de estos enfoques en apartados posteriores.

- **¿En qué consiste el “síndrome de los datos disponibles”?**

Es un síndrome en el que se puede caer al usar como enfoque de implementación el bottom-up, que implica que al partir de datos y sistemas que ya existen, se intenta crear una base que lo contenga todo y lo tenga disponible para todos, cuando realmente eso no es necesario.

- **¿Qué pasos se dan en una implantación top-down?**

En este proceso de implementación se empieza desde arriba, es decir, partimos de lo que la empresa quiere y para qué lo quiere.

- Se lleva a cabo una **elicitación de requisitos**, proveídos por el gestor de la empresa o negocio.
- El desarrollador determina las **medidas** de la MDDB en base a esos requisitos.
- Se prepara un **prototipo** de MDDB que se muestra a la empresa haciendo uso de un software OLAP.

Este enfoque asegura que la base que se desarrolle se ajustará a las necesidades de la empresa, lo que no siempre es bueno, ya que la base se desarrollará exclusivamente para un fin concreto, además de que se hará sin tener en cuenta los sistemas sobre los que se apoya, lo que puede dar lugar a problemas si no se prevén posibles modificaciones que puedan sufrir las bases subyacentes durante el desarrollo de la MDDB.

- **¿Qué pasos se dan en una implantación bottom-up?**

En este tipo de implementación partimos de unos datos que ya existen en unas bases de datos de una tecnología concreta, es decir, empezamos la implementación por abajo. En este enfoque, aunque nos guiamos por requisitos también, empezamos la implementación fijándonos en el origen de los datos, de manera que estos determinarán como será la MDDB. Los pasos que se dan en este enfoque son los siguientes:

- El primer paso es **analizar dichos datos y tomar medidas** para la confección de la base de datos multidimensional. Esta toma de medidas da lugar a lo que se conoce como diagrama en estrella, que contiene dos tipos de tablas:
 - **Tabla de hecho**, formada por datos numéricos que existen dentro de la base de datos.
 - **Tablas de dimensiones**, formadas por datos más descriptivos que se pueden asociar a las dimensiones de la base dentro de la empresa.
- Haciendo uso de este diagrama, se **traduce** la base de datos relacional a una MDDB, ya que permite concretar el tema central del análisis. Esta traducción es también una simplificación y puede incluso automatizarse.

El problema que tiene este enfoque es que hacer esa traducción es una tarea de lo más complicada, incluso para expertos cualificados (además durante la ejecución de esta actividad se puede caer en el síndrome mencionado anteriormente).

▪ **¿Qué enfoque alternativo puede plantearse? ¿En qué consiste?**

Existe un enfoque alternativo a estos dos que es realmente una media de ambos, el enfoque **middle-out**. Dado que podríamos considerar el enfoque top-down como un enfoque centrado en el usuario y el bottom-up como un enfoque centrado en el desarrollador, podríamos considerar este enfoque como uno orientado a ambas figuras y que consta de los siguientes pasos.

1. Elegir un gestor apto para ejercer el rol de representante del proyecto (en el lado de la empresa). Extraer de este representante las necesidades de la empresa.
2. Localizar las fuentes de datos necesarias para tomar las medidas de la MDDB e identificar un subconjunto de los datos a modo de muestra con el que se pueda trabajar fácilmente.
3. Identificar los datos que el representante no haya identificado inicialmente y comentarlos con él (quizá suponga una ventaja incluir esos datos en la base y el representante simplemente olvidó comentarlo).
4. Desechar datos que no sean esenciales, para evitar sufrir el síndrome de los datos disponibles.
5. Construir un prototipo de la MDDB y que lo revisen tanto el representante del proyecto como los gestores de las bases de datos actuales.
6. Si el gestor mencionado lo considera, modificar las fuentes de datos para facilitar la extracción a la MDDB.

Este enfoque tiene partes tanto top-down, en las que se determina que información nos interesa controlar y analizar, como bottom-up, en fases en las cuales los datos y sus fuentes determinan los datos que estarán disponibles así como la forma en la que se modelará la información. A pesar de ser un enfoque intermedio, no es tampoco la opción perfecta en todas las situaciones.