

OLAP: adentrándose en el procesamiento analítico online

Más que un acrónimo extraño, OLAP (On-line Analytical Processing) representa la manera de guardar y generar nueva información, con el objetivo de agilizar el análisis de gran cantidad de datos.

La metodología OLAP se basa en la información contenida en las bases transaccionales que, normalmente, quedan registradas en el DataWarehouse de los sistemas de información de la empresa.



Las bases transaccionales poseen modelos de relación robustos, enfocados a la operatoria diaria, pero no han sido diseñadas para ejecutar grandes consultas sobre ellas sino que deben ser simples y, en general, unitarias. A este tipo de consultas se les denomina OLTP (*On-line Transaction Processing*) y responden a un nivel de preguntas del tipo: En mayo, Pepe Pérez compró un equipo de música en la tienda Refting por \$100.000.

Las bases transaccionales poseen una riqueza de información sobre la que posiblemente usted ni siquiera tiene total conocimiento. Esta información permitirá tomar decisiones informadas relacionadas con el negocio.

A partir de esta información se puede dar respuesta a preguntas de mayor complejidad como:

- ¿Cómo fueron las ventas en el Norte comparadas con las del Sur?
- ¿Cuál fue el ticket promedio de las compras de productos electrónicos?
- ¿Quién es cliente más rentable?
- ¿Cuántos clientes están abandonando?

Sin embargo, algunas empresas requieren de un ejército de analistas que extraen la información de las bases transaccionales para poder dar respuesta a este tipo de preguntas. Este esfuerzo en extraer la información se debe a que las bases transaccionales no están diseñadas para el análisis, sino que para **guardar** información. La solución para este problema está en una base OLAP, especialmente diseñada para el análisis.

¿Análisis sobre las bases transaccionales (OLTP)?

Efectivamente, es posible realizar el análisis directamente sobre estas bases. Sin embargo, las bases transaccionales son las que guardan las transacciones de una empresa. Estas bases insertarán, harán actualizaciones (*update*) y borrarán registros a una gran velocidad. No querríamos detener los procesos de la empresa por culpa de la rapidez con que se guarda la información. Esta capacidad de soportar la operatoria diaria es producto de un diseño orientado a este fin. Este diseño, que guarda transacciones rápida y precisamente, hace que sea muy complejo el análisis a partir de ellas por las siguientes razones:

1) Comprensión del modelo: Las bases OLTP contienen una gran cantidad de tablas, a veces cientos de ellas. Estas tablas poseen relaciones entre sí, por lo que se hace complejo comprender todo el modelo de datos del diseño (Ver Figura 1).

2) **Capacidad de extracción:** Para poder extraer información desde las bases OLTP necesitará crear y ejecutar procedimientos que poseen una secuencia de consultas. Estos procedimientos pueden demorar horas en ejecutarse y usar los recursos del servidor desde el cual se extrae la información, algo que no es aceptable en un sistema que aloja los procesos de negocio de la empresa.

3) **Datos variables en el tiempo:** Dado que las bases OLTP cambian constantemente la información, se hace imposible obtener análisis precisos con datos cambiantes.

4) **Datos individuales y no agregados:** las bases OLTP guardan registros individuales (En Mayo, Pepe Pérez compró un equipo de música en la tienda Refting por \$100.000) y normalmente los analistas usan información con niveles de agregación (sumas y conteos) según el tipo de análisis. Luego, el tener la información a nivel individual no ayuda para el objetivo del analista.

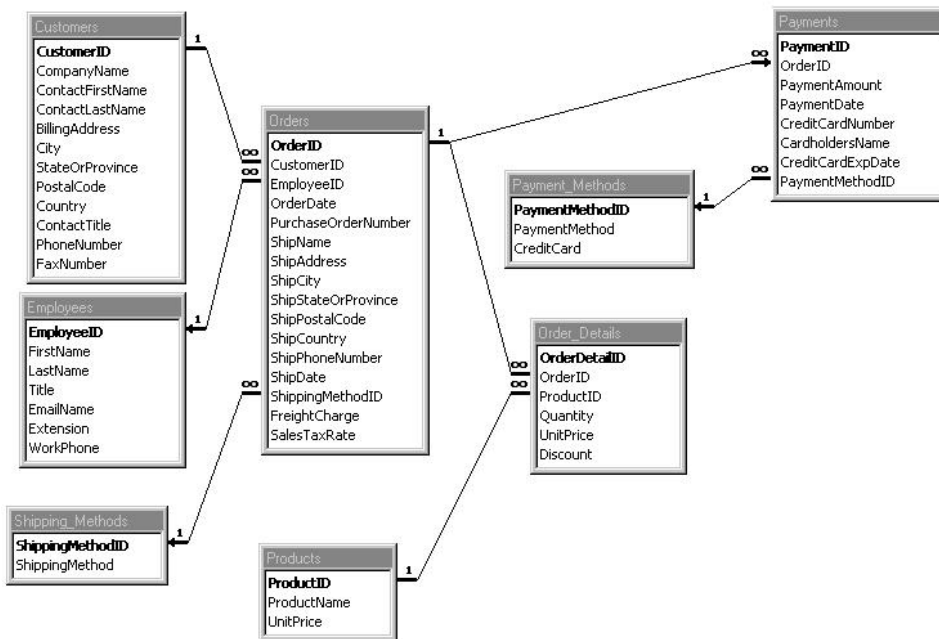


Figura 1 – Modelo de Base de Datos OLTP

La solución para todas las inconveniencias de realizar análisis sobre bases OLTP se resume en **On-line Analytical Processing (OLAP)**.

OLAP al servicio del analista

Las bases OLAP sólo reciben datos históricos del negocio y éstos nunca se modifican. Los usuarios nunca realizan ingresos o edición de datos. Todo lo que pueden hacer es consultas sobre las bases.

Las bases OLAP usan pocas tablas y poseen un tipo de organización distinto. El objetivo es minimizar el número de JOINS a realizar sobre las bases a partir de un modelo estrella (ver Figura 2).

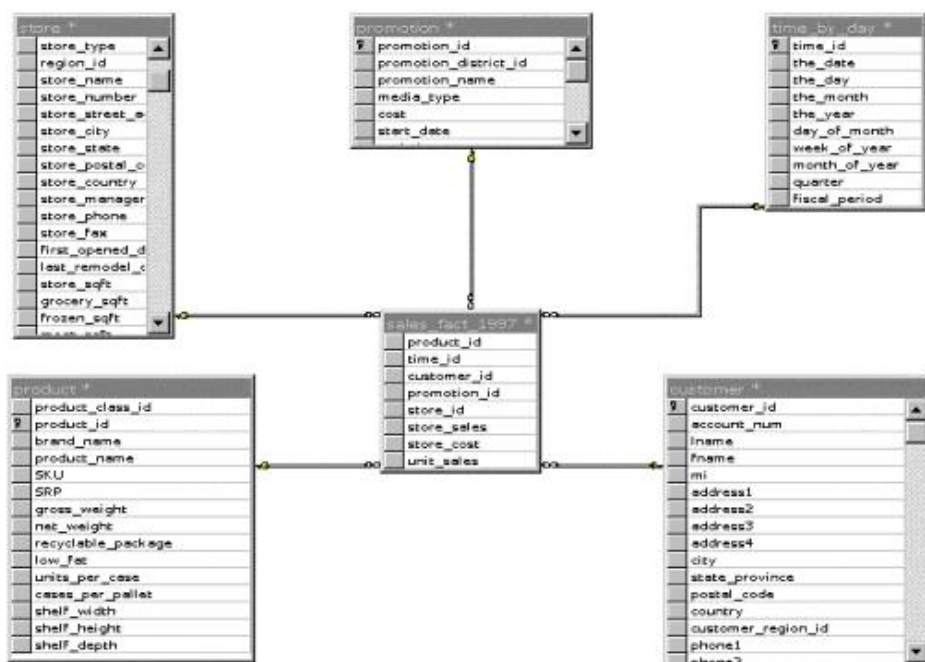


Figura 2 – Modelo de Base de Datos OLAP

El modelo estrella

La **Figura 2** muestra un modelo estrella en el cual existe una tabla central (tabla “de hecho” o *facttable*) y tablas relacionadas a ésta (tabla de dimensiones). La tabla de hecho contiene, en general, datos numéricos que se relacionan con las dimensiones (códigos postales, códigos de comuna, ID de cliente, ID de tienda, ID de producto, etc.) y datos aditivos que resumen de manera agregada la transaccionalidad (volumen de compras, costos, número de visitas, etc.).

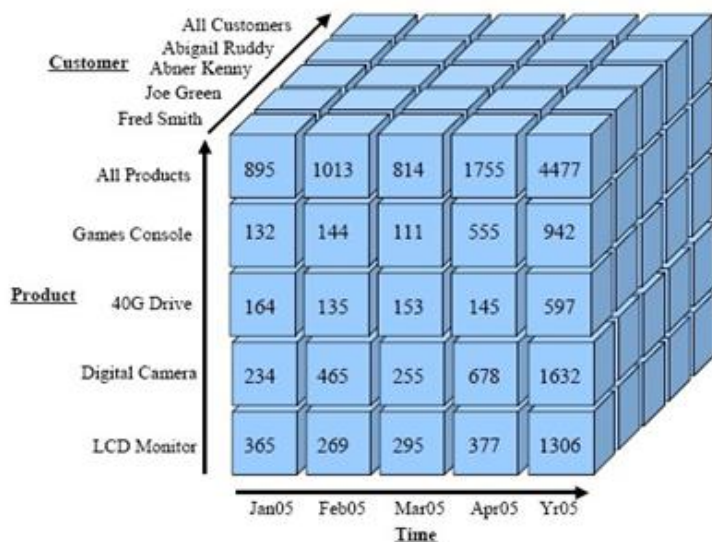
Tal como hemos visto anteriormente, los datos numéricos en una tabla utilizan menos espacio que un carácter (tipo *string*) y funcionan mucho mejor al momento de crear [claves primarias o índices](#). Luego, el usar enteros como claves para el modelo estrella mejorará el rendimiento y capacidad de almacenamiento de los datos.

Tal como se ve en la **Figura 2**, por ejemplo, los productos están con un ID que seguramente corresponden a un entero. Si no existiera una dimensión de producto, sería necesario traer toda la información que caracteriza al producto (15 columnas x número de registros) en la tabla de hechos, afectando enormemente el rendimiento de la base. Además, si fuera necesario modificar un atributo del producto, sería necesario realizar un *update* sobre toda la tabla de hechos.

Generalmente las dimensiones poseen jerarquías que serán útiles al momento de realizar consultas sobre la base. Por ejemplo, en la tabla de productos, es posible observar un nivel de jerarquía mayor llamado *product_class_id* que permitirá subir un nivel en el análisis (ej: podremos saber las ventas y número de transacciones de artículos electrónicos como categoría). Calcular, por ejemplo, el número total de las ventas de una categoría de productos sobre bases OLTP podría involucrar docenas de tablas, haciendo que la consulta sea muy lenta y compleja.

Cubos

Con los cubos, en tanto, se puede completar el proceso de consolidación de información con foco en el análisis. Los cubos generarán los niveles de agregación para cada una de las dimensiones de la base OLAP. Estos niveles de agregación precalculados permiten obtener respuestas sobre los cubos a una velocidad que no sería posible bajo otro esquema de datos y agregación. Una vez que el cubo ha sido generado, es posible visualizarlo en las herramientas Office u otras de uso cotidiano y así permitir que los usuarios accedan a la información de manera simple y rápida tal como lo harían con una tabla dinámica de Excel.



Reporting y Balanced ScoreCard

Una vez generada la infraestructura de datos para el análisis a partir de bases OLAP y cubos, podemos abordar el desafío de distribuir la información a los que deben gestionar el negocio y tomar decisiones. Sin una buena infraestructura será imposible generar reportes de buen nivel.

Podríamos discutir ampliamente sobre las características de un cuadro de mando, de las herramientas que existen para su desarrollo y las técnicas de despliegue de los datos. Hay una infinidad de literatura que podría apoyar esta discusión. Sin embargo, en esta etapa del conocimiento del BI se estima conveniente hablar de la realidad y la experiencia del entorno empresarial actual en cuanto a las necesidades de información y cómo el especialista BI debe satisfacer esa necesidad y no morir en el intento.

El proceso de conocimiento es creciente y nunca termina

Según la experiencia de varios procesos de generación de conocimiento, hay algunas premisas que difícilmente encontrará en la literatura tradicional y que son importantes de considerar al momento de abordar un proyecto de reporting o cuadro de mando:

- **El usuario rara vez sabe lo que quiere:** aunque el usuario final sabe que necesita información y que ésta sea lo más amplia posible, es muy poco común que sepa exactamente qué y cómo ver al inicio de un proyecto de cuadro de mando.
- **Una vez que el cuadro de mando está terminado, el proyecto recién comienza:** es posible que hayamos puesto todo nuestro esfuerzo y tiempo en desarrollar el mejor cuadro de mando y que, sin embargo, una vez que haya finalizado el proyecto se requiera ampliar el nivel de información o modificar algunas cosas que, con el correr del tiempo, han sido detectadas.

El hecho de que el usuario no conozca bien lo que quiere y que requiera realizar modificaciones en el futuro no es una crítica al usuario, sino que establece una realidad sobre el entorno empresarial. Las necesidades de conocimiento son dinámicas. Lo que hoy era relevante, puede ser que mañana no lo sea, y viceversa. El especialista BI está al servicio del negocio, por lo que debe diseñar sus procesos e infraestructura de información de manera de resolver el problema de un entorno dinámico y no entorpecer el proceso de generación de conocimiento rigidizando las soluciones.

El desafío de la flexibilidad en la generación de conocimiento es algo que pocos especialistas BI logran abordar exitosamente. Para ello, es importante considerar dos aspectos fundamentales:

- **Desarrolle de menos a más:** no caiga en la trampa de embarcarse en costosos proyectos de desarrollo de cuadros de mando si aún no ha generado, durante un tiempo prudente, un proceso de generación de conocimiento que se ha ido

complejizando a través del tiempo.

- **Acompa e el proceso de conocimiento:** el especialista BI no debe estar ajeno a la necesidad del negocio; por lo tanto, debe estar en contacto cercano con  l. Debe agregar valor y detectar oportunidades de negocio. De esta forma, los cambios que deben realizarse sobre los reportes o cuadros de mando ser n naturales y no algo impuesto.

Siempre considere que todo el proceso de BI debe tener como foco el negocio. Cada decisi n, desde el modelamiento de datos hasta la generaci n de reportes, debe ser tomada considerando el objetivo de negocio.

Si se falla en esta visi n, el  rea de BI ser  mero un proveedor de datos y no aqu lla que est  integrada hacia el negocio con el objetivo de agregar un valor que antes no exist a.