

Procesamiento Analítico con Minería de Datos

Analytical Processing with Data Mining

Angelino Feliciano Morales

Universidad Autónoma de Guerrero, México

afmorales@uagro.mx

René Edmundo Cuevas Valencia

Universidad Autónoma de Guerrero, México

reneecuevas@uagro.mx

José Mario Martínez Castro

Instituto Tecnológico de Chilpancingo, México

jmariomtz@yahoo.com

Resumen

Este trabajo describe la utilidad e importancia de la herramienta OLAP en Business Intelligence con el fin de recomendarla a los administradores de empresas para su toma de decisiones. La tecnología OLAP permite el rápido acceso a datos mediante data warehouse, agilizando el análisis de la información. Los cubos proveen de un rápido mecanismo de búsqueda de datos y de un tiempo de respuesta uniforme, independientemente de la cantidad de datos o de la complejidad del procedimiento de búsqueda. Tomando en cuenta su funcionamiento y estructura, el sistema OLAP se clasifica en tres categorías: ROLAP, MOLAP y HOLAP. Actualmente el sistema OLAP que más se utiliza es el denominado ROLAP.

Palabras clave: OLAP, Multidimensional, Cubo OLAP.

Abstract

This work describes the utility and importance of the OLAP Business Intelligence tool in order to recommend it to the administrators of companies for their decision making. OLAP technology enables faster access to data using a data warehouse, speeding up the information analysis. Cubes provide a fast mechanism of search data and a uniform response time, regardless of the amount of data or the complexity of the search procedure. Taking into account its operation and structure, the OLAP system is classified into three categories: ROLAP, MOLAP and HOLAP. Currently the OLAP system most widely used is the so-called ROLAP.

Key words: OLAP, Multidimensional, Cube OLAP.

Fecha recepción: Agosto 2015

Fecha aceptación: Diciembre 2015

Introducción

Este documento resalta la importancia de utilizar esta herramienta en bases de datos para agilizar la consulta de grandes cantidades de información, sobre todo en el mundo empresarial.

El logro de competitividad en la producción es una tarea primordial en Business Intelligence (Inteligencia de Negocios). Para ello es fundamental desarrollar en el sistema empresarial una mentalidad innovadora; con una estrategia de ciencia e innovación tecnológica se puede alcanzar un nivel adecuado de gestión tecnológica y mantener los estándares de calidad y eficiencia.

En el ámbito de las soluciones, las aplicaciones OLAP son una de las herramientas más utilizadas por las empresas, ya que han sido creadas en función a bases de datos multidimensionales que permiten procesar grandes volúmenes de información en campos bien definidos, y con un acceso inmediato a los datos para su consulta posterior.

Las herramientas OLAP proporcionan a las compañías un sistema confiable para procesar datos que luego serán utilizados para llevar a cabo análisis e informes que permiten mejorar las

operaciones productivas, tomar decisiones inteligentes y optimizar la competitividad. Además, dan soporte a las tecnologías de Data Warehouse. En general, estos sistemas OLAP deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos.

La principal característica de las herramientas OLAP es que son entornos especialmente diseñados para la ejecución de análisis multidimensional de los datos corporativos de cualquier usuario que soportan. Asimismo, brindan posibilidades de navegación, seleccionando información, permitiendo el análisis de datos segmentados que permiten ir reduciendo el conjunto de datos que se han reportado. Este tipo de selecciones se refleja en la visualización de la estructura multidimensional, mediante unos campos de selección que permiten elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de un dato en concreto.

La información es gestionada y procesada en grandes bloques organizativos, como pueden ser la estructura geográfica o la académica, llamados dimensiones. Dichas dimensiones de negocio se estructuran a su vez en distintos niveles de detalle.

En la actualidad su aplicación se ha extendido hacia todas las áreas empresariales y otros tipos de organizaciones que analizan volúmenes masivos de datos —incluyendo medianas empresas, academia, gobierno y demás instituciones públicas y privadas—, que requieren cada vez más de un análisis dinámico, potente y en línea para tomar decisiones adecuadas, generando así la demanda de este tipo de software.

METODOLOGÍA

La metodología utilizada en esta investigación fue el análisis y síntesis documental de la herramienta OLAP, utilizada por administradores de empresas para la mejor toma de decisiones. Asimismo, OLAP es un recurso que se usa en el área de la denominada inteligencia de negocios, cuyo objetivo es agilizar la consulta de grandes cantidades de datos, para lo cual recurre a estructuras multidimensionales o cubos OLAP que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales OLTP.

BASES DE DATOS

Una base de datos es una colección de datos organizados y estructurados según un determinado modelo de información que refleja no solo los datos en sí mismos, sino también las relaciones que existen entre ellos. Una base de datos se diseña con un propósito específico y es organizada con una lógica coherente. Los datos podrán ser compartidos por distintos usuarios y aplicaciones, sin embargo, deben conservar su integridad y seguridad al margen de las interacciones de ambos. La definición y descripción de los datos deben ser únicas para minimizar la redundancia y maximizar la independencia en su utilización.

En una base de datos, las entidades y atributos del mundo real se convierten en registros y campos. Estas entidades pueden ser tanto objetos materiales como libros o fotografías, pero también personas e incluso conceptos e ideas abstractas. Las entidades poseen atributos y mantienen relaciones entre ellas. Las bases de datos pueden clasificarse según las características (figura 1).

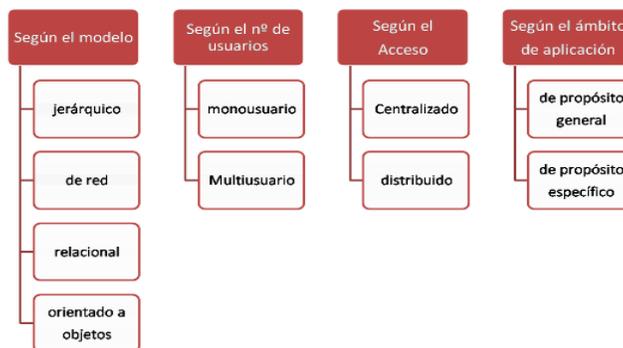


Figura 1. Clasificación de bases de datos

Una base de datos proporciona a los usuarios el acceso a la información, que pueden visualizar, ingresar o actualizar, en concordancia con los derechos de entrada que se les haya otorgado.

Una base de datos local puede ser utilizada por un solo usuario en un equipo o distribuir la información en equipos remotos y acceder a ella a través de una red.

La principal ventaja de utilizar bases de datos es que múltiples usuarios pueden acceder a ellas al mismo tiempo.

El manejo de datos requiere de un proceso para convertirse en información útil; algunas de las herramientas que necesita este tipo de procedimiento se presentan a continuación (figura 2).

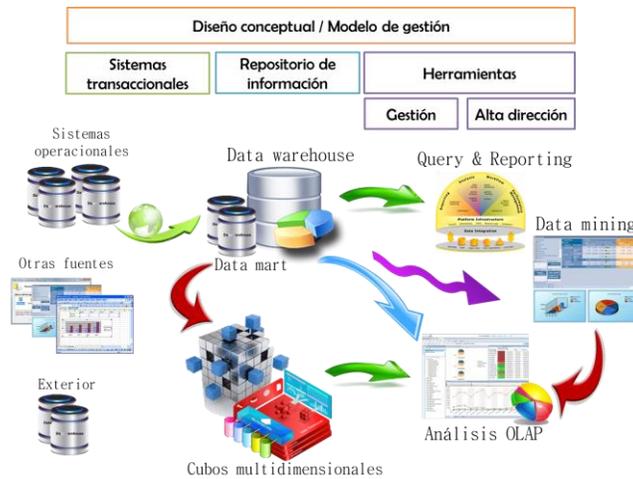


Figura 2. Proceso de los datos

Data Warehouse

Es una combinación de conceptos y tecnologías destinadas a satisfacer los requerimientos de una organización o empresa, en términos de mejorar la gestión con eficiencia y facilidad de acceso.

Es una base de datos relacional diseñada para la consulta y análisis en lugar del procesamiento de transacciones. Por lo general, contiene información histórica derivada de una transacción, pero puede incluir datos de otras fuentes. Separa el análisis de carga de trabajo de las transacciones y permite a una organización consolidar datos de varias fuentes.

Los almacenes de datos y sistemas OLTP (On-Line Transactional Processing) tienen necesidades muy diferentes. Los siguientes son algunos ejemplos de las diferencias entre los almacenes de datos típicos y sistemas OLTP:

- Carga de trabajo

De antemano, el usuario puede no conocer la carga de trabajo del almacén de datos, por lo que debe ser optimizado para propiciar un buen desempeño de variedad de posibles operaciones de consulta.

Los sistemas OLTP apoyan las operaciones predefinidas. Sus aplicaciones pueden ser sintonizadas o diseñadas para soportar estas operaciones específicamente.

- Modificación de datos

Un almacén de datos se actualiza de forma regular por el proceso ETL (ejecutar por noche o cada semana) utilizando técnicas de modificación de la información a granel. Los usuarios finales de un almacén de datos no se actualizan directamente del almacén.

En sistemas OLTP, los usuarios finales emiten rutinariamente instrucciones de modificación de información individual a la base de datos. La base de datos OLTP está siempre al día, y refleja el estado actual de cada transacción comercial.

- Diseño del esquema

Los almacenes de datos a menudo usan esquemas que no están normalizados o no normalizados parcialmente (como un esquema en estrella) para optimizar el rendimiento de las consultas.

Los sistemas OLTP suelen utilizar esquemas totalmente normalizados para optimizar actualización / insertar / borrar rendimiento, y para garantizar la coherencia de datos.

- Las operaciones típicas

Una consulta típica de almacenamiento de datos escanea miles o millones de filas. Por ejemplo, "ubicar las ventas totales para todos los clientes el mes pasado".

Una operación típica OLTP accede solo a un puñado de registros. Por ejemplo, "recuperar el orden actual para este cliente".

- Los datos históricos

Los almacenes de datos suelen almacenar varios meses o años de información. Esto es para apoyar el análisis histórico.

Los sistemas OLTP suelen almacenar datos de unas pocas semanas o meses. El sistema OLTP almacena únicos datos históricos necesarios para afrontar con éxito los requisitos de la transacción actual.

Además de una base de datos relacional, un Data Warehouse incluye una solución de extracción, transporte, transformación y carga (ETL), un procesamiento analítico en línea (OLAP) del motor, herramientas de análisis de clientes y otras aplicaciones que gestionan el proceso de recopilación de información y la entrega a los usuarios de negocios (figura 3).

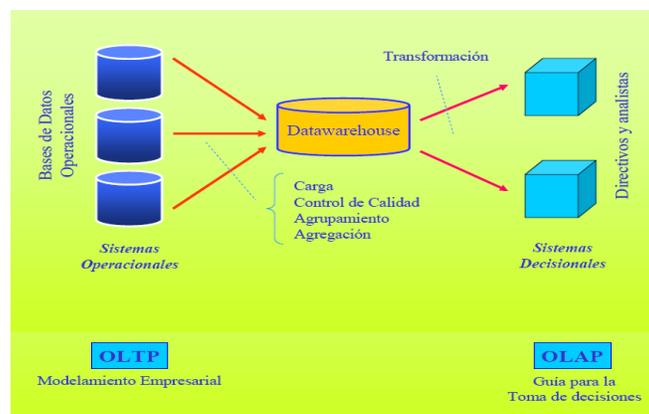


Figura 3. Procedimiento del DATA WAREHOUSE

Datamart

Es una base de datos departamental, especializada en almacenar información de un negocio específico. Se caracteriza por disponer una estructura óptima de datos para analizar información desde varias perspectivas que afecten los procesos de dicho departamento. Un Datamart puede ser alimentado desde los datos de un Data Warehouse o integrar por sí mismo un compendio de distintas fuentes de información.

Datamart OLAP

Se basan en los cubos OLAP populares, que se construyen según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es heterogéneo, en función de la herramienta final que se utilice.

Datamart OLTP

Las bases de datos OLTP pueden basarse en un simple extracto del data warehouse, no obstante, lo común es introducir mejoras en su rendimiento (los agregados y filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras comunes son las tablas report, que vienen a ser *fact-tables* reducidas (que agregan dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, aun con el objetivo de explotar la reescritura de queries (aunque solo es posible en algunos SGBD avanzados, como Oracle).

Los Datamart dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información

Proceso ETL

Los procesos ETL son una parte de la integración de datos, pero es un elemento importante cuya función completa el resultado de todo el desarrollo de la cohesión de aplicaciones y sistemas.

La palabra ETL corresponde a las siglas en inglés de:

- Extraer: extract.
- Transformar: transform.
- Cargar: load.

Con ello, se puede decir que todo proceso ETL consta precisamente de estas tres fases: extracción, transformación y carga. A continuación se define en qué consiste cada una de ellas.

Fase extracción

Para realizar de manera correcta el proceso de extracción se deben seguir los siguientes pasos:

- Extraer los datos desde los sistemas de origen.
- Analizar los datos extraídos obteniendo un chequeo.
- Interpretar este chequeo para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no fuese así, los datos deberían ser rechazados.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación.

Además, una de las prevenciones más importantes a tomar en cuenta durante el proceso de extracción sería el exigir siempre que esta tarea cause un impacto mínimo en el sistema de origen. Este requisito se basa en la práctica, ya que si los datos a extraer son muchos el sistema de origen se podría ralentizar e incluso colapsar, provocando que no pudiera volver a ser utilizado con normalidad para su uso cotidiano.

Fase de transformación

La fase de transformación de un proceso de ETL aplica una serie de reglas de negocio o funciones sobre los datos extraídos para convertirlos en datos que serán cargados. Estas directrices pueden ser declarativas, basarse en excepciones o restricciones, pero para potenciar su pragmatismo y eficacia es preciso asegurarse de que sean:

- Declarativas
- Independientes
- Claras
- Inteligibles
- Con una finalidad útil para el negocio

Proceso de carga

En esta etapa, los datos procedentes de la fase anterior (fase de transformación) son cargados en el sistema de destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

Existen dos formas básicas de desarrollar el proceso de carga:

- Acumulación simple: esta manera de cargar los datos consiste en realizar un resumen de todas las transacciones comprendidas en el periodo seleccionado y transportar el resultado como una única transacción hacia el data warehouse, almacenando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada. Es la forma más sencilla y común de llevar a cabo el proceso de carga.
- Rolling: este proceso sería el recomendable en los casos donde se busque mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondiendo a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etcétera).

Cualquiera que sea la manera de desarrollar este proceso, hay que tener en cuenta que esta fase interactúa directamente con la base de datos de destino, y por eso al realizar esta operación se aplicarán todas las restricciones que se hayan definido en esta. Si están bien definidas, la calidad de los datos en el proceso ETL estará garantizada (figura 4).

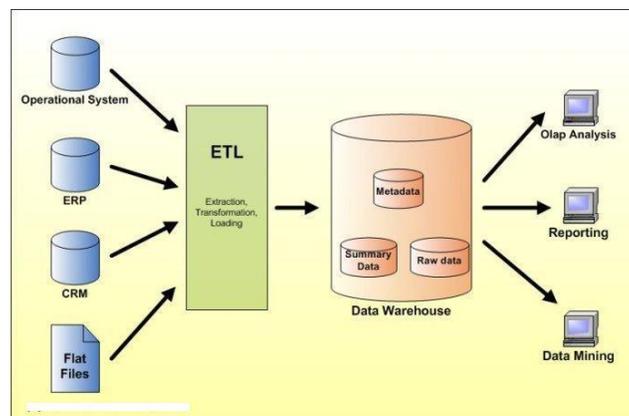


Figura 4. Proceso ETL

¿QUÉ ES OLAP?

OLAP es el acrónimo en inglés de procesamiento analítico en línea. Es una solución utilizada en el campo de la inteligencia empresarial (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o cubos OLAP) que contiene información resumida de grandes bases de datos. Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares

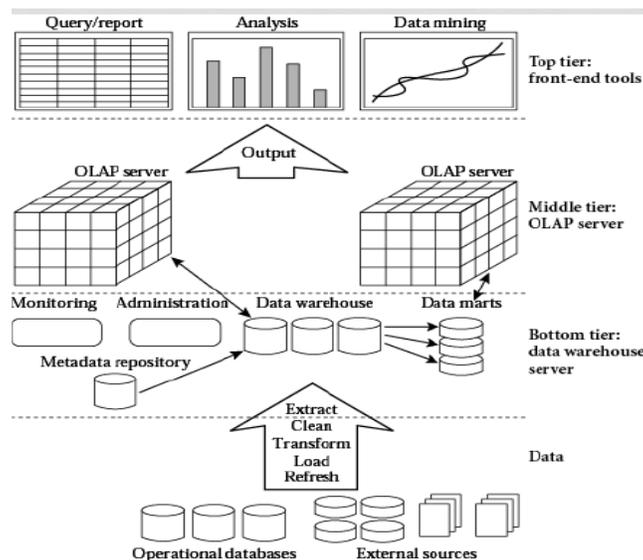


Figura 5. Estructura de OLAP

La herramienta utiliza estructuras de datos **multidimensionales** o cubos OLAP, que son bases **multidimensionales** en la cual el almacenamiento físico de la información se realiza en un vector **multidimensional**. Los cubos OLAP se pueden considerar como una ampliación de las 2 dimensiones de una hoja de cálculo, estos contienen información resumida de grandes bases de datos o Sistemas Transaccionales.

Clasificación y comparación de sistemas OLAP

De acuerdo al funcionamiento y estructura, los sistemas OLAP han sido clasificados en distintas categorías, como ROLAP, MOLAP, HOLAP. El sistema OLAP más utilizado hoy en día es el llamado ROLAP.

A continuación se describen los distintos Sistemas OLAP.

ROLAP

Significa Procesamiento Analítico en Línea Relacional, es una herramienta OLAP construido sobre una base de datos relacional. En este sistema tiene importancia la tabla de hechos, donde se almacena la historia de la información relevante para la empresa que requiere ser estudiada.

En la industria del OLAP, el sistema ROLAP es conocido por ser capaz de escalar grandes volúmenes de información, pero su rendimiento a la hora de ejecutar consultas es inestable comparado con otro procedimiento de la industria OLAP, MOLAP. Sin embargo, después de haberse realizado un estudio se llegó a la siguiente conclusión: que una compañía que utiliza ROLAP tiene menor rendimiento que aquellas que utilizan el sistema MOLAP (figura 6).

MOLAP

Significa Procesamiento Analítico Multidimensional en Línea, cuyo sistema guarda los datos en una matriz multidimensional de almacenamiento y requiere que el procesamiento y la acumulación de información estén contenidos en el cubo OLAP.

HOLAP

Significa Procesamiento Analítico en Línea Híbrido, es una combinación de los sistemas ROLAP y MOLAP permitiendo ordenar una parte de los datos en un MOLAP mientras que el resto lo hace como un ROLAP (figura 6).

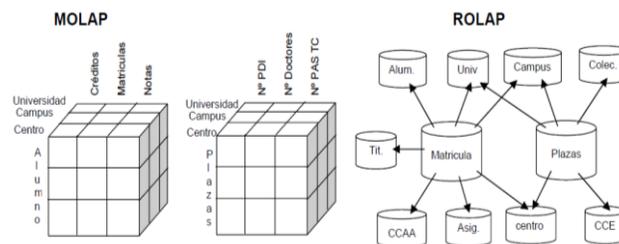


Figura 6. Vistas Molap y Rolap

Comparación de sistemas ROLAP y MOLAP

Cuando se comparan las dos arquitecturas se pueden realizar las siguientes observaciones:

- El ROLAP delega la negociación entre el tiempo de respuesta y el proceso batch al diseño del sistema, mientras que el MOLAP suele requerir que sus bases de datos se pre compilen para conseguir un rendimiento aceptable en las consultas, incrementando los requerimientos batch.
- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación ad-hoc. Los sistemas ROLAP soportan bien esta consolidación dinámica, mientras que los MOLAP están orientados hacia consolidaciones batch.
- Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes de datos controlados. El volumen de información con los que se trabaja son los que están implementados en el cubo de análisis, mientras que en los entornos ROLAP en todo momento se dispone de la totalidad de la base de datos.
- El entorno MOLAP almacena la información en estructuras con formato de matrices multidimensionales, mientras que los ROLAP gestionan la información mediante metadatos que *mapean* esquemas de base de datos, en estrella, en vistas multidimensionales.

De igual manera que las organizaciones utilizan una variedad de herramientas, para el trabajo cotidiano OLAP puede requerir diferentes tipos de recursos dependiendo del nivel o área de análisis. Los entornos de planificación, análisis financieros y localización de recursos pueden requerir entornos MDDB. Por su parte, los entornos de análisis de ventas o campañas de marketing que requieren datos con millones de cambios continuos, tanto de productos como de clientes o atributos, requieren entornos ROLAP.

- ROLAP

Los datos son introducidos directamente desde el Datawarehouse u otra fuente de datos relacional y no son almacenados por separado.

- **MOLAP**

La información es pre calculada y luego es almacenada en cubos de datos multidimensionales.

- **HOLAP**

Mantiene volúmenes de información más grandes en la base de datos relacional y las agregaciones en un MOLAP separada.

OLAP y OLTP

OLAP

Son aplicaciones que se encargan de analizar datos del negocio para generar información táctica y estratégica que sirve de soporte para la toma de decisiones. Mientras que las transacciones OLTP utilizan bases de datos relacionales u otro tipo de archivos, OLAP logra su máxima eficiencia y flexibilidad operando sobre bases de datos multidimensionales.

A continuación se enuncian las características más sobresalientes de estas aplicaciones.

- Estructura de datos transparente al usuario.
- Solo realizan consulta, trabajan sobre la información operacional generada por los sistemas OLTP.
- Consultas sobre grandes volúmenes de datos predecibles.
- Información histórica.
- Modo de actualización batch.
- Mayor renuncia de datos para facilitar la generación de consultas y obtener buenos tiempos de respuesta.
- Trabaja con resúmenes de miles de registros condensados en una sola respuesta.

OLTP (On-Line Transactional Processing)

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico, y puede involucrar operaciones de inserción,

modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura (por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).
- Los datos se estructuran según el nivel de aplicación (programa de gestión a medida, ERP o CRM implantado, sistema de información departamental, etcétera).
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad y la existencia de islas de datos).
- El historial de datos suele limitarse a los datos actuales o recientes.

Son aplicaciones que definen el comportamiento habitual de un entorno operacional de gestión y ejecutan las operaciones del día. Las características más comunes de este tipo de transacciones son:

- Altas/bajas/modificaciones
- Consultas rápidas, escuetas y predecibles
- Poco volumen de información y disgregada
- Transacciones rápidas
- Gran nivel de concurrencia
- Modo de actualización on-line
- Baja redundancia de datos

Algunos ejemplos de este tipo de aplicaciones son:

- Compras
- Ventas
- Inventario
- Sueldos

OLTP también se ha utilizado para referirse a la transformación en la que el sistema responde de inmediato a las peticiones del usuario. Un cajero de un banco es un ejemplo de una aplicación de procesamiento de transacciones comerciales.

La tecnología OLTP se utiliza en innumerables aplicaciones, como en banca electrónica, procesamiento de pedidos, comercio electrónico, supermercados o industria.

Conceptos de datos multidimensionales (cubos OLAP)

Para su funcionamiento, las aplicaciones OLAP utilizan un tipo de base de datos que posee la peculiaridad de ser multidimensional, denominada comúnmente cubo OLAP.

Básicamente, el cubo OLAP, cuyo nombre proviene de su característica multidimensional, es una base de datos que posee diversas dimensiones (figura 7).

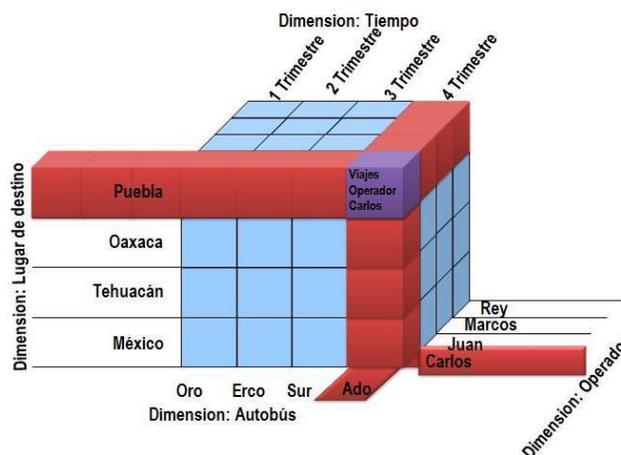


Figura 7. Cubo OLAP

En el análisis multidimensional, los datos se representan mediante dimensiones como destino, autobús, operador y tiempo. En general, las dimensiones se relacionan en jerarquías, por ejemplo, ciudad, estado, región, país y continente. El tiempo es también una dimensión estándar con sus propias jerarquías tales como: día, semana, mes, trimestre y año.

Para tener una idea más simple de la función de los cubos OLAP dentro de una base de datos multidimensional, cabe destacar que cada una de las dimensiones o escalas del cubo corresponde básicamente a una jerarquía de datos.

Pueden utilizarse otras dimensiones del cubo para recabar información referente a situaciones geográficas, clasificación de los productos por categorías, gastos realizados por la empresa, y demás.

Esta confluencia de la información permite llevar a cabo un análisis completo de diversas situaciones, para obtener las soluciones correctas a los problemas de negocios.

Mediante la incorporación de estos vectores o cubos, se han ampliado las posibilidades de las bases de datos relacionales, permitiendo el procesamiento de importantes volúmenes de información, de lo contrario sería imposible realizar dicha actividad. Cada una de las dimensiones que posee la base de datos incorpora un campo determinado para cierto tipo específico, que luego podrá ser comparado con la información contenida en el resto de dimensiones, para hacer posible la evaluación y posteriores informes de las actividades relevantes para una compañía.

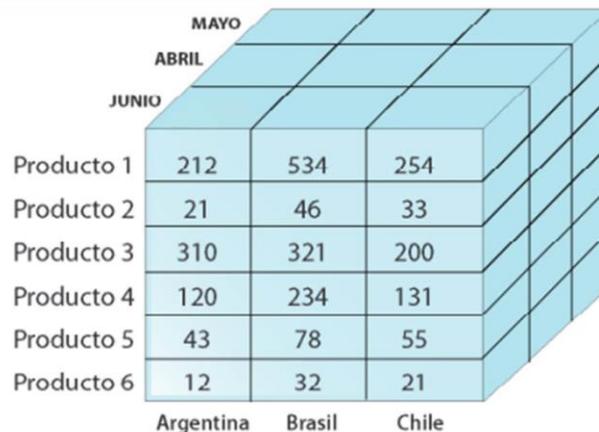


Figura 8. Estructura multidimensional

SMBD CON OLAP

Las aplicaciones OLAP son uno de los pilares de cualquier solución de Inteligencia de Negocios, debido a que provee a los que toman las decisiones de acceso a información sumariada mediante métodos convenientes de navegación que les permitan analizar y mantener una conversación fluida con los datos de la organización, en óptimos tiempos de respuesta.

Los servidores de bases de datos OLAP utilizan estructuras multidimensionales para almacenar información y las relaciones existentes entre los mismos. La mejor forma de visualizar las

estructuras multidimensionales es en cubos. Cada lado de un cubo es una dimensión. El cubo puede expandirse para incluir otra dimensión, por ejemplo, familia de productos. El cubo también soporta la aritmética de matrices.

La tecnología OLAP soporta las operaciones analíticas comunes, por ejemplo, la consolidación, la profundización y la navegación.

La consolidación implica agregar datos, por ejemplo, totalizaciones simples o expresiones complejas que impliquen datos interrelacionados.

La profundización es la operación inversa de la consolidación e implica mostrar información detallada comprendida en los datos consolidados.

La navegación hace referencia a la capacidad de examinar los datos desde diferentes puntos de vista. La navegación suele realizarse a menudo a lo largo de un eje temporal, con el fin de analizar tendencias y ubicar patrones.

A continuación se muestra un ejemplo de cómo funciona el sistema OLAP, en cuanto a las consultas (figura 9).

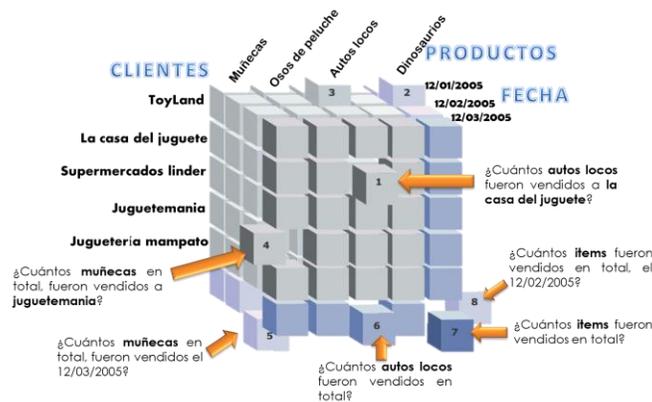


Figura 9. Ejemplo de un análisis con un cubo OLAP

Pentaho

Pentaho es una herramienta de Business Intelligence, que cuenta con la incorporación de las principales herramientas del mercado Open Source. Actualmente es la más completa y extendida. Cuenta con una gran comunidad de desarrollo que realiza constantes mejoras y extensiones en la Plataforma.

Se compone de un entorno al que podrá acceder vía web de forma segura. Cada usuario visualiza todos los elementos habilitados para su perfil, el cual incluye informes, análisis OLAP y cuadros de mando con indicadores y tablas. El usuario podrá generar nuevas vistas de análisis y nuevos informes y guardarlos para consultarlos más tarde, así como exportarlos a Excel, PDF o realizar su impresión en papel.

Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de BI. Ofrece soluciones para la gestión y análisis de la información, incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario.

La plataforma ha sido desarrollada bajo el lenguaje de programación Java y tiene un ambiente de implementación basada en la misma herramienta, haciendo que Pentaho sea una solución flexible al cubrir una alta gama de necesidades empresariales. Con Pentaho se puede tener una visión más clara de los datos (figura 10).



Figura 10. OLAP con Pentaho

CONCLUSIÓN

Después de haber realizado la investigación respecto al procesamiento analítico en línea y a la herramienta utilizada para dar solución en el campo de la inteligencia empresarial, se concluye lo siguiente.

- Implementar de forma adecuada un sistema OLAP permite obtener beneficios de mayor productividad para los usuarios finales de la organización
- El acceso controlado y oportuno de la información de carácter estratégico permite realizar un proceso en la toma de decisiones con mayor efectividad.
- Se obtienen mayores ingresos y beneficios potenciales, al permitirle a la organización responder rápidamente a las demandas del mercado.
- Utilizando OLAP se garantiza una respuesta rápida en la consulta y una visión mejor comprendida por el usuario para la toma de decisiones.
- Pueden aprovecharse las ventajas de los informes multidimensionales en forma de cubos OLAP generados por el motor.
- Las tablas dinámicas aportan gran dinamismo a la plataforma, facilitando a los usuarios las herramientas necesarias para la creación de cubos personalizados.

Bibliografía

- Cervantes M., Ballesteros B.B., y Hernández F. (2012). "Programas para elevar la competitividad de las empresas en México" en Observatorio de la Economía Latinoamericana. Sitio web: <http://www.eumed.net/cursecon/ecolat/mx/2012/vlc.html>
- Gonzalo H- Morales, Álvaro (2000). "Migrator OLAP", en Facultad de Ingeniería, Universidad de la República Oriental del Uruguay. Sitio web: <http://www.fing.edu.uy/~ruggia/T5s/Migrador-OLAP.pdf>
- Hernández, E. N., y Mosquera, C. I. I. S. (2012). "Pentaho: software líder de Inteligencia de Negocio de código abierto", en Revista Telem@tica. Sitio web: <http://revistatelematica.cujae.edu.cu/index.php/tele/article/view/44>
- Ibarra M, A. (2006). "Procesamiento Analítico en Línea (OLAP)" en Universidad Nacional del Noreste. Sitio web: <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/OLAPMonog.pdf>
- Informática Hoy (2007-2012). "Cubo OLAP; una base de datos multidimensional". Sitio web: <http://www.informatica-hoy.com.ar/telefonos-celulares/Cubo-OLAP-una-base-de-datos-multidimensional.php>
- Núñez F. D. J. (2011). "Sistemas ROLAP y MOLAP". Sitio web: <http://repository.uaeh.edu.mx/bitstream/handle/123456789/14495>
- Pentaho (2013). "Pentaho BI - Software Libre para la Toma de Decisiones". Sitio web: <http://gravitar.biz/pentaho/>
- Romucho P. (2010). "Bases de datos y sistemas de información", en Monografías. Sitio web: <http://www.monografias.com/trabajos77/base-datos-sistema-informacion/base-datos-sistema-informacion.shtml>
- Sinnexus (2012). "Datamart", en Sinnexus. Sitio web: http://www.sinnexus.com/business_intelligence/datamart.aspx

Tamayo, M., y Moreno, F. J. (2006). "Comparing the MOLAP the ROLAP storage models", en Ingeniería e Investigación. Sitio web: http://www.scielo.org.co/scielo.php?pid=S0120-56092006000300016&script=sci_arttext&tlng=pt

Urquizo P. (2011). "¿Qué es OLAP?", de Business Intelligence. Sitio web: <http://www.businessintelligence.info/definiciones/que-es-olap.html>

Wikipedia (2013). "OLTP", Enciclopedia libre. Sitio web: <http://es.wikipedia.org/w/index.php?title=OLTP&oldid=70248939>.