

Introducción a los *Datawarehouses**

Diana Graciela Chuc Durán †

Universidad Juárez Autónoma de Tabasco, DACB

Vivimos en la era de la información y esto implica que las empresas requieren de esta para ser más competitivas y tomar decisiones acertadas en el momento oportuno. La construcción de un Datawarehouse está dirigida fundamentalmente a responder de manera rápida a las solicitudes de análisis de los directivos de las empresas. Conocer como diseñar un Datawarehouse es importante hoy en día pues las empresas requieren de esta tecnología para tener ventajas en el mercado.

We live in information era and this implies enterprises requiring of informations for be more competitive and take assert decitions in the righth time. Building a datawarehouse

Palabras clave: Datawarehouse, DataMarts, Dimensiones.

Keywords: Datawarehouse, DataMarts, Dimensions.

1. Introducción

El mundo avanza y cada día las empresas tienen mayor número de aplicaciones automatizadas, almacenan la información diaria en grandes bases de datos, y pueden conocer al momento su stock de inventarios, su volumen de ventas del día, tener sus precios actualizados, etc.

Poco a poco las empresas fueron almacenando un gran número de información en diferentes fuentes de datos (archivos, documentos de texto, bases de datos, etc.), y los directivos de las empresas se dieron cuenta de que ésta, podría ser útil pues reflejaba la mayoría de las operaciones diarias del negocio.

A su vez en los mercados empresariales existe mayor competencia, por lo que las empresas requieren mayor rapidez y eficiencia en sus procesos, y precisión en la información para tomar decisiones adecuadas. Por eso, se pensó que lo ideal sería unificar las diferentes fuentes de información de las cuales disponían, almacenándolas en un único lugar, de tal forma que solo se le incorporara información relevante. Este nuevo repositorio de datos debería tener una estructura organizada, integrada, lógica y dinámicamente, y además ser de fácil explotación. La respuesta a esto fueron los Almacenes de datos o Data Warehouses (DW)][2].

En 1994 el 90% de las empresas, según la revista Fortune 2000, planeaba implementar un data warehouse entre 1994 y 1996. En 1996 el 90% de las grandes corporaciones consideraba adoptar la tecnología del datawarehouse. Hill Hostian, de la empresa Gartner1, estima que para el 2007, 50% de los proyectos de 2007, 50% de los proyectos de inteligencia de negocio (data warehouse), requerirán de un proveedor de servicios para librar los obstáculos debidos a falta de personal capacitado y recursos.

*Recibido el 13 de febrero de 2007 y aceptado el 30 de marzo de 2007

†**Dirección postal:** Carr. Cunduacán-Jalpa Km 1, Cunduacán Tabasco, México. A.P. 24 C.P. 86690. Tel.(+52)914 336-0928. **Correo electrónico:** diana.chuc@basicas.ujat.mx

En la página del e-semanal Publicado el 31 de Julio del 2006 [1] se dice que las soluciones de Data warehouse (DW) son un nicho de la industria que continúa en franco crecimiento, ya que durante 2005 se pudo observar un crecimiento de 14% en lo que se refiere a aquellas empresas que comercializan soluciones completas de DW, crecimiento que se deriva principalmente de la actividad en sectores como el financiero, retail, petroquímica, gobierno y consumo, los cuales se han dado a la tarea de adoptar este tipo de soluciones buscando un mayor control del negocio mediante el uso adecuado y eficiente de la información.

2. Data warehouse

Existen diversas definiciones de Data warehouse, por ejemplo:

Manuel de la Herran Gastón, define el DataWarehouse como un *almacén de datos que reúne la información histórica generada por todos los distintos departamentos de una organización, orientada a consultas complejas y de alto rendimiento*.

Susan Osterfeldt en 1993 publica una definición que sin duda acierta en la clave del DW:

“Yo considero al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico”.

Sin embargo, la definición más conocida para el DataWarehouse, fue propuesta por William H. Inmon [4] ¹ en 1992:

“Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”.

Al datawarehouse también se le puede llamar Bussines Inteligent o decisión support [3].

3. Características

Según, Bill Inmon, existen generalmente cuatro características que describen un almacén de datos o datawarehouse:

1. Orientado al sujeto: los datos se organizan de acuerdo al sujeto en vez de la aplicación, por ejemplo, una compañía de seguros usando un almacén de datos podría organizar sus datos por cliente, premios, y reclamaciones, en lugar de por diferentes productos (automóviles, vida, etc.). Los datos organizados por sujetos contienen solo la información necesaria para los procesos de soporte para la toma de decisiones.

¹Inmon, William H. es considerado el padre de las Bases de Datos

2. Integrados: cuando los datos residen en muchas aplicaciones separados por los distintos entornos operacionales, la descodificación de los datos es a menudo inconsistente. Por ejemplo, en una aplicación, la palabra gender podría codificarse como “m” y “f” en otra como “0” y “1”. Cuando los datos fluyen de un entorno operacional a un entorno de almacén de datos o de data warehouse, ellos asumen una codificación consistente, por ejemplo gender siempre se transformaría a “m” y “f”.
3. Variación-temporal: el almacén de datos contiene un lugar para guardar datos con una antigüedad de 5 a diez años, o incluso más antiguos, para poder ser usados en comparaciones, tendencias y previsiones. Estos datos no se modificarán.
4. No son inestables: los datos no serán modificados o cambiados de ninguna manera una vez ellos han sido introducidos en el almacén de datos, solamente podrán ser cargados, leídos y/o accedidos.

4. Ventajas del Data Warehouse

Los objetivos del datawarehouse se construyen tomando en cuenta las sugerencias de los administradores y gerentes de las empresas, inevitablemente las cuestiones que ellos desean resolver son:

1. Tenemos un mundo de información almacenada pero no podemos consultarla.
2. Tenemos que copiar y pegar la información resumida que arroja el sistema cada vez que la requerimos para análisis.
3. Yo creo que sería mas fácil que los mismos usuarios pudieran tomar la información directamente
4. Me vuelve loco tener la información de dos personas con los mismos reportes pero diferentes resultados.
5. Quiero que mi personal utilice la información para soportar sus decisiones.

Los DWH están en la categoría de los sistemas para el soporte de decisiones (DSS) que tienen como objetivos medir y controlar el desarrollo de las variables importantes del negocio, buscando identificar, proyectar y predecir tendencias a partir de los datos acumulados.

Los datos que se manejan en el DWH son informativos , esto significa que son datos resumidos y periódicos a diferencia de los datos operacionales.

Algunos ejemplos de consultas típicas en este sistema para la toma de decisiones de la empresa serían;

- ¿Cuál ha sido el total de ingresos de los clientes por meses y años?
- ¿Cuáles han sido los pagos realizados por los clientes y por qué conceptos?
- ¿Cuáles han sido las principales cadenas hoteleras que aportan a sus Casas Matrices?

El DWH se soporta sobre el modelo dimensional a diferencia de los sistemas de bases de datos que están basados en el modelo Entidad-Relación. Este modelo contiene la misma información que el modelo E/R pero empaqueta los datos en un

formato simétrico cuyo objetivo es ganar una mayor comprensión del usuario y garantizar la ejecución rápida y eficiente de las consultas. A diferencia del modelo E/R, el modelo dimensional no necesita anticipar las consultas que se van a realizar y es muy elástico a los cambios que se produzcan en los patrones de los usuarios.

El modelo dimensional divide el mundo de los datos en dos grandes tipos: las medidas y las descripciones del entorno de estas medidas. Las medidas, que generalmente son numéricas, se almacenan en las tablas de hechos y las descripciones de los entornos que son textuales se almacenan en las tablas de dimensiones. Las tablas de hechos son las tablas primarias en el modelo dimensional y contiene los valores del negocio. Los hechos más comunes son valores numéricos. Cada tabla representa una interrelación muchos-muchos y contiene dos o más llaves extranjeras que acoplan con sus respectivas tablas de dimensiones.

Las tablas de dimensiones son las compañeras de las tablas de hechos. Cada dimensión se define por su llave primaria que sirve para mantener la integridad referencial en la tabla de hechos a la que se acopla. Los atributos de estas tablas sirven de base a las solicitudes que se hacen al DWH.

Las tablas de dimensiones contienen información jerárquica que permitirán la realización de las agregaciones o las profundizaciones.

5. Desventajas

El implementar un datawarehouse conlleva muchas ventajas para la empresa, pero también existen problemas que se deben solucionar. Una de las principales desventajas de un proyecto de datawarehouse es el costo, pues requiere una inversión de 2 a 10 millones de dólares en tres años[3]. Como cada empresa tiene sus propias características, no es fácil estandarizar este tipo de proyectos, por lo que se vuelve complejo su desarrollo[3]. Aunado a esto en Septiembre de 2005, Hostian, empleado de la empresa consultora Gartner, explicó que muchas empresas fallan en la implementación de almacenes de datos (data warehouse) debido a que la falta de calidad en los datos se detecta hasta el final del proyecto, y además resalto otro problema importante, "A los usuarios se les enseña a usar las herramientas, pero muchos de ellos no entienden los resultados" [5].

6. Empresas que utilizan Data Warehouse

Walmart	Procter & Gamble
Whirpol	3M
Coca Cola	Walt Disney
Tv Azteca	Banorte
Banco de Mexico	Banamex
(http://www.computerworld.com.mx/soluciones/otras/apuestabanco.htm)	
Nike	Baxter
	GNP
	Warner Lambert

7. Empresas Consultoras en México

INFOMEDIA, es una empresa mexicana formada por más de 50 consultores, que proporciona diversos servicios, entre los cuales se encuentran:

1. Definición de estrategia tecnológica del cliente
2. Definición de arquitectura
3. Implantación de nuevas tecnologías
4. Definición de normatividad
5. ETLs
6. Data Marts
7. Data Warehouse
8. Bases de Datos Multidimensionales (OLAP)
9. Tableros de control
10. Minería de Datos

Su sitio oficial en Internet es <http://www.infomedia.com.mx/>

Referencias

- [1] eSemanal en Internet. Publicación de negocios en formato HTML, disponible en la dirección <http://www.esemanal.com.mx>. Artículo revisado publicado el 31 de julio de 2006 URL http://www.esemanal.com.mx/articulos.php?id_sec=12&id_art=4032 revisado el 10 de febrero de 2007.
- [2] Padrón Torres, Liudmila. *Almacenes de datos: importancia de la estandarización de las direcciones para las empresas de hoy en día*. Monografía disponible en Internet en URL <http://www.monografias.com/trabajos31/almacenes-datos/almacenes-datos.shtml>. Creado en 1 de enero de 2006 y visitado el 10 de febrero de 2007.
- [3] Cruz Chávez, Marco Antonio. *Programacion en computadoras*. Recurso didactico disponible en Internet en URL <http://www.uaem.mx/posgrado/mcruz/cursos/miic/datawarehouse2.ppt>, diapositiva 1. (ago 2006)
- [4] Inmon, William H. *Data Warehouse - a perspective of data over time*. 370/390 Data Base Management, Feb 1992.
- [5] Gartner. URL <http://www.gartner.com/> Sitio oficial de la empresa consultora.