

## MODELO MULTIDIMENSIONAL

### Resumen / Abstract

Data Warehousing, es una tecnología para el almacenamiento de grandes volúmenes de datos en una amplia perspectiva de tiempo para el soporte a la toma de decisiones. Debido a su orientación analítica, impone un procesamiento distinto al de los sistemas operacionales y requiere de un diseño de base de datos más cercano a la visión de los usuarios finales, permitiendo que sea más fácil la recuperación de información y la navegación. Este diseño de base de datos se conoce como modelo multidimensional, este artículo, abordará sus características principales.

*Data Warehousing is a technology that allows us to storage a huge amount of data in a wide time frame in order to help us with decision making problems. Due to its analytical perspective it imposes a different kind of processing from regular Operational Systems and it also requires a database design based more on the final users point of view, making the retrieving and surfing of the information easier. This database design is known as Multidimensional Model and this article will approach its principle characteristics.*

### Palabras clave / Key words

Sistemas de soporte a la toma de decisiones, almacén de datos, mercado de datos, modelo multidimensional, tablas de hechos, tablas de dimensiones, esquemas multidimensionales, esquema estrella, esquema copo de nieve, niveles, jerarquías, base de datos

*Decision support system, DSS, data warehouse, data mart, multidimensional model, fact tables, dimension tables, multidimensional schemes, star scheme, snowflake scheme, levels, hierarchy, database*

## INTRODUCCIÓN

Data Warehouse, es una tecnología para el manejo de la información, que soporta el procesamiento informático y provee una plataforma sólida que permite realizar análisis a partir de datos históricos y actuales. Su función esencial es ser la base de un sistema de información. Facilita la integración de sistemas de aplicación no integrados proveniente de fuentes de datos heterogéneas (bases corporativas, bases propias, de sistemas externos, ficheros, etc.), brinda una visión integrada de dicha información, especialmente enfocada hacia la toma de decisiones por parte del personal de la organización. Organiza y almacena los datos que se necesitan para el procesamiento analítico sobre una amplia perspectiva de tiempo.

Es un sitio donde se almacena fundamentalmente la información que se extrae de los sistemas de gestión que registran la operatoria diaria de la organización. Además, se almacenan datos

*Alexis Cedeño Trujillo, Ingeniero Informático, Centro de Estudios de Ingeniería de Sistemas (CEIS), Instituto Superior Politécnico José Antonio Echeverría Cujae, Ciudad de La Habana, Cuba*

e-mail: alexis@ceis.cujae.edu.cu  
alexis.cedeno@gmail.com

Recibido: octubre del 2005  
Aprobado: diciembre del 2005

estratégicos y tácticos que pueden ser de gran ayuda para aplicar sobre los mismos, técnicas de análisis de datos encaminadas a obtener información oculta (minería de datos).<sup>1</sup>

## DIFERENCIAS ENTRE EL MODELO ENTIDAD-RELACIÓN Y EL MULTIDIMENSIONAL

El modelo entidad-relación (MER) es una técnica poderosa para el diseño de sistemas transaccionales en el entorno de las bases de datos relacionales. Permite la normalización de la estructura de datos física, obteniéndose un diseño sin redundancias en los datos y ocupándose el menor espacio de almacenamiento. Sin embargo, no contribuye en la habilidad del usuario en el momento de consultar la base de datos. Una técnica mucho más poderosa para la interrogación de los datos, es el modelo dimensional o multidimensional.<sup>2</sup>

El modelo multidimensional, es mucho menos riguroso en cuanto a organización; le permite a analistas y diseñadores más flexibilidad en el diseño, para lograr un mayor desempeño y optimizar la recuperación de la información, desde un punto de vista más cercano al usuario final.

No es una tarea fácil, transformar un modelo entidad-relación en un modelo multidimensional aún cuando los datos que se modelan son los mismos. Ambos enfoques tienen diferentes puntos de partida, se emplean técnicas distintas y producen resultados de diseño de bases de datos desiguales.

El modelo multidimensional produce una base de datos que es simple de navegar y encuestar. Existen menor cantidad de tablas y relaciones en este, que en el modelo entidad-relación, el cual tiene cientos de tablas relacionadas entre sí y hay diferentes caminos para obtener una misma información, desde la perspectiva del usuario final resulta prácticamente inusable.<sup>2</sup>

## MODELO MULTIDIMENSIONAL

El modelo multidimensional dentro del entorno de las bases de datos, es una disciplina de diseño que se sustenta en el modelo entidad-relación y en las realidades de la ingeniería de texto y datos numéricos.<sup>3</sup>

Modela las particularidades de los procesos que ocurren en una organización, dividiéndolos en mediciones y entorno. Las medidas son en su mayoría, medidas numéricas, y se les denomina hechos. Alrededor de estos hechos existe un contexto que describe en qué condiciones y en qué momento se registró este hecho. Aunque el entorno se ve como un todo, existen registros lógicos de diferentes características que describen un hecho, por ejemplo, si el hecho referido, es la venta de un producto en una cadena de tiendas, se podría dividir el entorno que rodea al hecho de la **cantidad** vendida, en el **producto** vendido, el **cliente** que lo compró, la **tienda** y la **fecha** en que se realizó la venta. A estas divisiones se le denomina **dimensiones** y a diferencia de los hechos que son numéricos, estos son fundamentalmente textos descriptivos.

Las medidas, como se expresó anteriormente, se registran en las tablas de hechos, siendo la llave de esta tabla, la combinación de las múltiples llaves foráneas que hacen referencia a las dimensiones que describen la ocurrencia de este hecho, en otras palabras, cada una de las llaves extranjeras en las tablas de hecho se corresponden con la llave primaria de una dimensión.

## TABLAS DE HECHOS

Las tablas de hechos, representan los procesos que ocurren en la organización, son independientes entre sí (no se relacionan unas con otras). En estas, se almacenan las medidas numéricas de la organización. Cada medida, se corresponde con una intersección de valores de las dimensiones y generalmente se trata de cantidades numéricas, continuamente evaluadas y aditivas. La razón de estas características, es que facilita que los miles de registros que involucran una consulta, sean comprimidos más fácilmente y se pueda dar respuesta con rapidez, a una solicitud que abarque gran cantidad de información.

La llave de la tabla de hechos, es una llave compuesta, debido a que se forma de la composición de las llaves primarias de las tablas dimensionales a las que está unida. se pueden distinguir dos tipos de columnas en una tabla de hechos, columnas de hechos y columnas llaves. Las columnas de hechos almacenan las medidas del negocio que se quieren controlar y las columnas llaves, forman parte de la llave de la tabla.

Existen tablas de hechos que no contienen medidas, a estas tablas se les denomina tablas de hechos sin hechos.<sup>4</sup> La semántica de la relación entre las dimensiones que definen la llave de esta tabla de hechos, implica por sí sola la ocurrencia de un evento, por ejemplo, si se quiere representar el hecho de que un estudiante matriculó en una universidad, la combinación de las siguientes dimensiones definiría este suceso: el estudiante matriculado, la carrera en que matriculó, la fecha de matrícula, el tipo de curso que va a cursar, etcétera.

## TABLAS DE DIMENSIONES

Una tabla de dimensión contiene, por lo general, una llave simple y un conjunto de atributos que describen la dimensión. En dependencia del esquema multidimensional que se siga, pueden existir atributos que representen llaves foráneas de otras tablas de dimensión, es decir, que establecen una relación de esta tabla con otra dimensión.

Las tablas de dimensión, son las que alimentan a las tablas de hechos, como se expresó anteriormente, la llave de un hecho es la composición de las llaves de las dimensiones que están conectados a esta, por tanto, los atributos que conforman las tablas de dimensiones también describen el hecho.

Los atributos dimensionales son fundamentalmente textos descriptivos, estos desempeñan un papel determinante, son la fuente de gran parte de todas las necesidades que deben cubrirse, además, sirven de restricciones en la mayoría de las consultas que realizan los usuarios. Esto significa, que la calidad del modelo

multidimensional, dependerá en gran parte de cuán descriptivos y manejables sean los atributos dimensionales escogidos.<sup>5</sup>

Las tablas de dimensión en general, son mucho más pequeñas que las tablas de hechos en cuanto a cantidad de registros. En cuanto a cantidad de atributos, una tabla de hechos bien descriptiva puede tener un gran número de estos.

Un Data Warehouse debe aceptar la responsabilidad de describir el pasado con precisión, en un sistema operacional normalmente si se produce un cambio, se sobrescribe el nuevo valor en el registro y se pierde el anterior. Cuando se quieren mantener las trazas de lo que ha ocurrido en el entorno multidimensional, por lo general ocurren cambios en la dimensiones, esto implica que se debe mantener almacenado el valor anterior y registrar el nuevo valor; a este fenómeno se le conoce como pequeños cambios en la dimensiones.<sup>6</sup> En estos casos se debe prestar especial atención pues las tablas de dimensiones pueden crecer en el número de registros y afectar el rendimiento.

### DIMENSIÓN TIEMPO

La dimensión más importante de un Data Warehouse, es la dimensión tiempo. Cada hechos que se registra en un tabla de hecho, tiene asociada una marca de tiempo, es decir, en qué momento ocurrió este hecho. Esto permite el almacenamiento y análisis histórico de la información. La inserción de datos, en la base de datos multidimensional, se hace por intervalos de tiempo, lo cual asegura un orden implícito.<sup>5</sup>

### ESQUEMAS MÁS COMUNES PARA EL MODELO MULTIDIMENSIONAL

Existen varios esquemas para el modelado de los datos en un Data Warehouse, los esquemas más usados son:

**Esquema estrella:** Es un paradigma de modelado que tiene un solo objeto en el medio conectado con varios objetos de

manera radial (figura 1) El objeto en el centro de la estrella es la **tabla de hechos** y los objetos conectados a ella son las **tablas de dimensiones**. Las tablas de dimensiones solo se relacionan con la tabla de hechos, es decir, no existen relaciones entre estas.<sup>7</sup>

**Esquema copo de nieve:** El esquema copo de nieve es una extensión del esquema estrella en donde cada una de las puntas de la estrella puede dividirse en más puntas. En esta forma de esquema, las tablas de dimensión pueden tener relación con otras tablas de dimensión.

La ventaja fundamental que proporciona este esquema, es que se ocupa menor espacio de almacenamiento, sin embargo, aumenta el número de tablas con las que el usuario debe interactuar e incrementa la complejidad de las consultas a realizar. El esquema estrella proporciona mayor compresión, navegabilidad, es más cercano a cómo el usuario final refleja la visión de una consulta empresarial. Se recomienda, de ser posible, que se emplee un esquema estrella antes que un copo de nieve (figura 2), justificándose la utilización de mayor espacio de almacenamiento, en la disminución del tiempo de obtención de la información que se necesita.

### JERARQUÍAS Y NIVELES

Un nivel se corresponde con un atributo dentro de una dimensión, este representa un nivel particular de agregación.

Una **jerarquía**, es una estructura arbórea lógica, que está compuesta por uno o varios niveles e implica una organización de estos dentro de una dimensión.<sup>8</sup> Cada nivel consta de un nivel superior y cero o varios niveles inferiores. Se tiene una relación de uno a mucho entre objetos de nivel superior e inferior (un objeto de un nivel superior agrupa uno o muchos objetos del nivel inferior). Las jerarquías definen cómo los datos son agregados desde los niveles más bajos hacia los más altos. A partir de una dimensión se puede definir varias jerarquías.

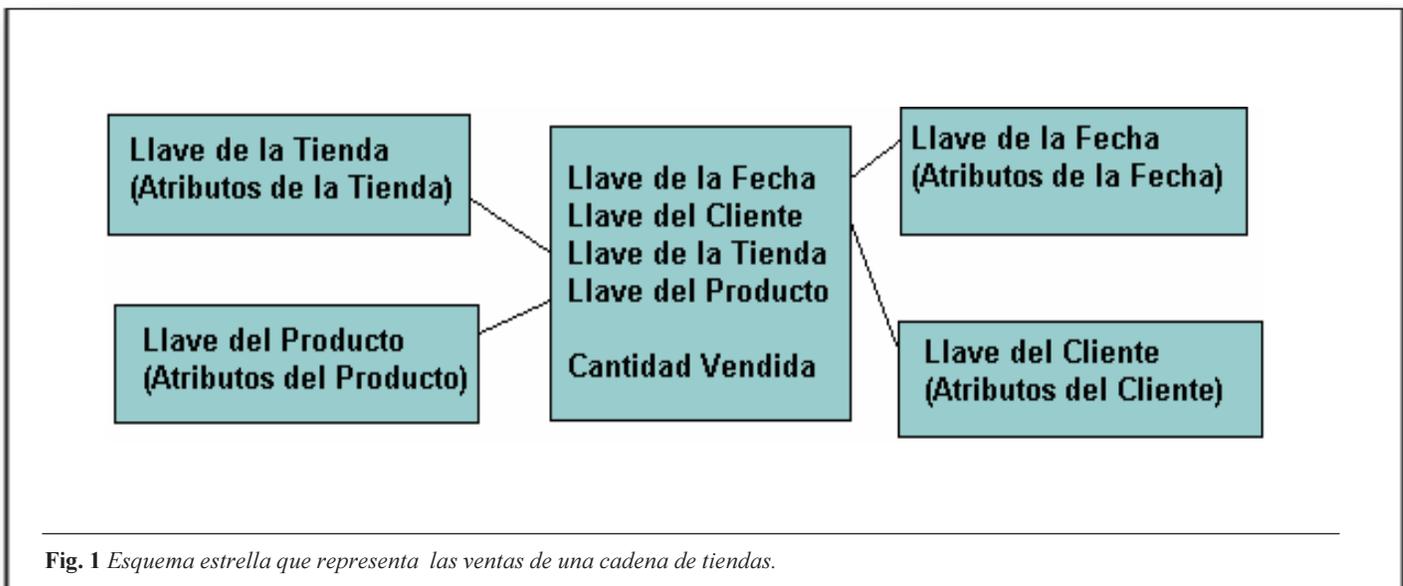
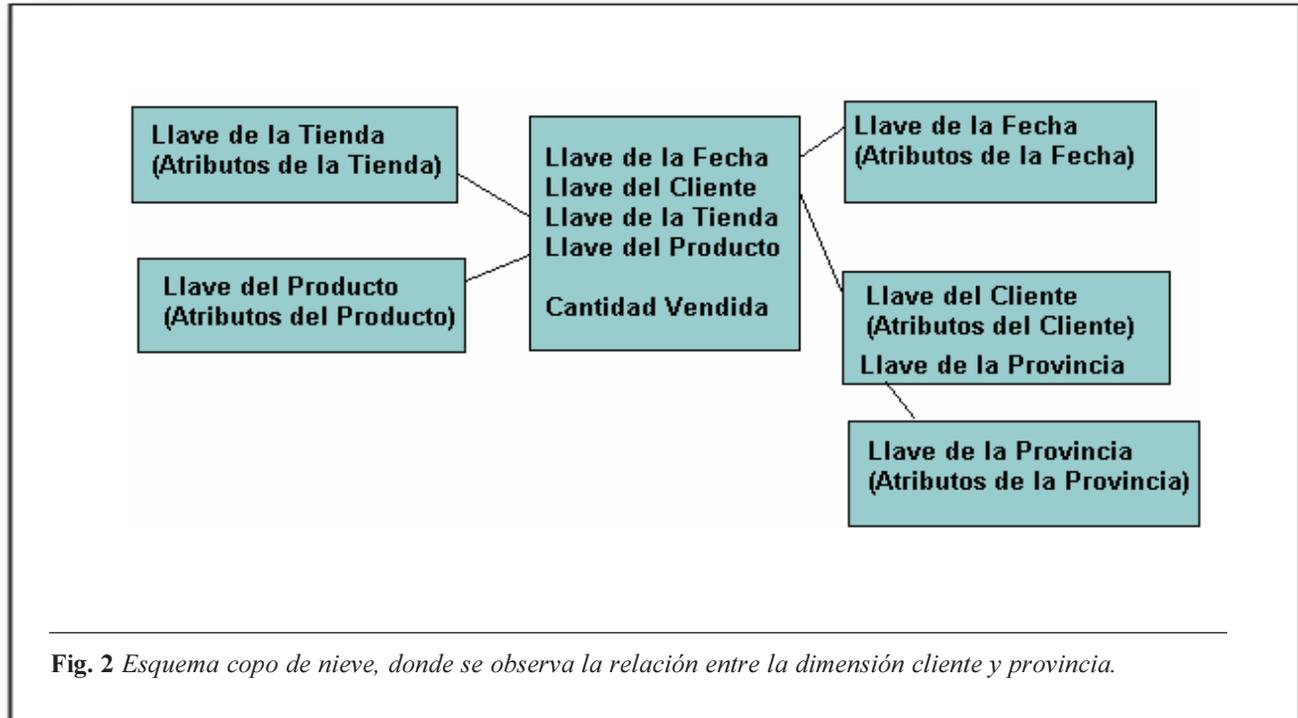


Fig. 1 Esquema estrella que representa las ventas de una cadena de tiendas.



Cada nivel sobre el nivel base, representa el total de los datos desde el nivel inferior, por ejemplo:

Si se considera una dimensión tiempo donde se ha definido una jerarquía de tres niveles: Mes, semestre y año, el nivel mes representa el nivel base de la jerarquía, el nivel semestre representa la suma de los totales por mes y el nivel año representa la suma de los totales para los semestres.

Los niveles de agregación, otorgan flexibilidad adicional a usuarios finales de aplicaciones para el análisis de los datos y se logra mayor velocidad de respuesta en las consultas.<sup>5</sup>

## CONCLUSIONES

El modelado multidimensional:

- Es una técnica de diseño de bases de datos que responde a las necesidades analíticas que impone la tecnología Data Warehousing. Se logra un diseño más cercano a la visión del usuario final de una consulta empresarial, la información es fácil de encuestar y navegar.
- Obtiene mayor velocidad de recuperación de las consultas en comparación con el modelo Entidad-Relación.
- Asegura un orden implícito y permite el almacenamiento y análisis histórico de la información, mediante la inserción de los datos por intervalos de tiempo.
- En dependencia de los datos que se necesiten modelar, se debe escoger el esquema multidimensional más adecuado. ☐

## REFERENCIAS

1. MÉNDEZ, ANDREAYARIEL MÁRTIRE: "Fundamentos de Data Warehouse, Centro de Actualización Permanente en Ingeniería del Software, Escuela de Postgrado, Instituto Tecnológico, Buenos Aires, Argentina, 2004.
2. KIMBALL, RALPH: *Is ER Modeling Hazardous to DSS*, October, 1995, <http://www.dbmsmag.com/9510d05.html>, August, 2005.
3. \_\_\_\_\_: *Fact Tables and Dimension Tables*, January, 2003, [http://www.intelligententerprise.com/030101/602warehouse1\\_1.jhtml](http://www.intelligententerprise.com/030101/602warehouse1_1.jhtml), July, 2005.
4. \_\_\_\_\_: *Factless Fact Tables*, September, 1996, <http://www.dbmsmag.com/9609d05.html>, August, 2005, Wolf Gloria Carmen, "Modelamiento multidimensional", agosto 2002 <http://www.inf.udec.cl/revista/>, abril 2005, August, 2005.
5. WOLF, GLORIA CARMEN: *Modelamiento multidimensional*, Agosto, 2002 <http://www.inf.udec.cl/revista/>, abril 2005, September, 2005.
6. KIMBALL, RALPH: *Slowly Changing Dimensions*, April, 1996, <http://www.dbmsmag.com/9604d05.html>, September, 2005.
7. GILL, HARJINDER S.Y PRAKASH C. RAO: *Data Warehousing: La integración de la información para la mejor toma de decisiones*, Prentice Hall, 1996
8. *Microsoft SQL-Server, SQL-Server Books on Line*, Microsoft Corporation, 2000.