

Propuesta de un método para el diseño y modelado de una bodega de datos

José Hernando Bahamón L.

Universidad Icesi
jbahamon@icesi.edu.co

Fecha de recepción: 15-4-2003

Fecha de aceptación: 25-8-2003

RESUMEN

El desarrollo de los Sistemas de Información Gerencial basados en tecnologías de Data Warehouse y Herramientas Olap, es relativamente reciente y, por lo tanto, no existe una propuesta metodológica universalmente válida y aceptada como tal, por la comunidad académica.

El presente artículo expone una propuesta metodológica para la realización del diseño de una bodega de datos, que utiliza como eje articulador la identificación de las necesidades de información por parte de la gerencia, para el soporte de los procesos de control y de toma de decisiones.

El método propuesto está compuesto de ocho pasos agrupados en tres fa-

ses. La primera fase comprende la identificación de las necesidades de información gerencial, desde la perspectiva del negocio. La segunda fase comprende todas las actividades relacionadas con la elaboración de un modelo lógico-conceptual de la estructura de la bodega de datos. La tercera fase incluye los pasos para realizar el diseño físico de la estructura de la bodega de datos.

PALABRAS CLAVES

Bodegas de datos, método de diseño de la estructura de una bodega de datos.

ABSTRACT

The development of Management Information Systems based on Ware-

house Data technologies and Olap tools is relatively new. Therefore, there is no valid methodological approach that is generally accepted as such by the academic community.

This article presents a methodological approach to the design of a data warehouse using the identification of management information requirements as a shaft that supports the control and decision-making processes. The suggested approach consists of eight steps grouped in three different stages. The first stage encompasses the identification of management

information requirements from a business perspective. The second one deals with all the activities associated with the preparation of a logical conceptual model for the data warehouse structure, and the third stage includes the steps to make the physical design of the data warehouse structure.

KEYWORDS

Data warehouses, approach to the design of a data warehouse structure.

Clasificación: A

INTRODUCCIÓN

El desarrollo de los Sistemas de Información Gerencial basados en tecnologías de Data Warehouse y herramientas Olap, es relativamente reciente y, por lo tanto, no existe una propuesta metodológica universalmente válida y aceptada como tal, por la comunidad académica.

Entre las propuestas más conocidas están: 1. Ralph Kimball,¹ con un esquema centrado en la identificación de los procesos de la empresa, como elemento clave para la definición de la estructura de variables y dimensiones; 2. W.H. Inmon,² con un esquema que parte de la construcción del modelo de datos corporativos, elaborado al más alto nivel de abstracción, para luego derivar la estructura del modelo de datos, para el diseño de la bodega; 3. Golfarelli Matteo, Maio Dario, Rizzi Stefano³ proponen un esquema que parte de los modelos E-R descriptivos de los sistemas transaccionales de la organización, para luego derivar el modelo E-R de la estructura, para la bodega de datos.

En este artículo se presenta una propuesta de sistematización del proceso de diseño de una bodega de datos, que se aparta de los esquemas de diseño referidos, y que utiliza como eje articulador, la identificación de la información gerencial, para el soporte de los procesos de control y de toma de decisiones en los niveles directivos de la organización.

MÉTODO PROPUESTO

El método de diseño propuesto está centrado en la identificación de la información clave y relevante para soportar los procesos de dirección y de toma de decisiones dentro de la organización. Este método utiliza, como punto de partida, la identificación y el modelado de: **qué es lo que el negocio está tratando de alcanzar**, para luego elaborar una estructura que apoye el proceso de gestión hacia el logro de las metas definidas.

Una vez que la información clave de apoyo a los procesos de gestión y control de la organización ha sido identificada, se inicia la elaboración del modelo lógico-conceptual de la estructura de la bodega de datos, que soportará las consultas y la exploración de los datos, a partir de los cuales se construirán los indicadores de gestión requeridos por los niveles directivos de la organización.

Para darle un orden a este proceso sistémico de diseño, los pasos del método propuesto, tal como se presentan en la Figura 1, se han agrupado en las siguientes fases:

- **Fase 1:** Identificación de las necesidades de información gerencial, desde la perspectiva del negocio.
- **Fase 2:** Elaboración del modelo lógico-conceptual de la estructura de la bodega de datos.

1. Kimball R. *The Data Warehouse Toolkit*. John Wiley & Sons, 1996.

2. Inmon W.H. *Building The Data Warehouse*. QED Press / John Wiley & Sons, 1992.

3. Golfarelli M., Maio D., Rizzi S. *Conceptual Design of Data Warehouse From E/R Schemes*. <http://www.csr.unib.it/~golfare/db.html>, 1998.

- **Fase 3:** Elaboración del modelo físico de la bodega de datos.

Fase 1:
Identificación de las necesidades de información gerencial, desde la perspectiva del negocio.

La primera fase, a partir de la cual se realiza el proceso de diseño de la estructura para una bodega de datos, comprende la identificación de las necesidades de información gerencial, lo que significa hacer explícitos los objetivos y los factores claves de éxito de la organización, o de un área del negocio.

Es bastante común empezar este proceso de identificación y modelado mediante entrevistas a los directivos, en las cuales la pregunta central es:

«¿Cuál es la información que desea obtener del sistema de información gerencial?». Este enfoque puede resultar muy peligroso, si el directivo no realiza un proceso sistemático y ordenado, para establecer sus necesidades de información, en relación con sus actividades de gestión y control.

Una forma ordenada y sistemática para realizar esta fase de identificación de las necesidades de información, que soporte los procesos de gestión y control gerencial, es la aplicación del enfoque de sistemas, para guiar el proceso de revisión o definición de: 1. Los objetivos estratégicos del negocio o del área; 2. Los factores clave para el logro de los objetivos definidos y 3. Los indicadores de control, tanto de los objetivos como de los factores clave.⁴

4. Véase Bahamón José H. *Construcción de indicadores de gestión bajo el enfoque de sistemas*. S&T Revista de la Facultad de Ingeniería, Universidad Icesi. 2003.

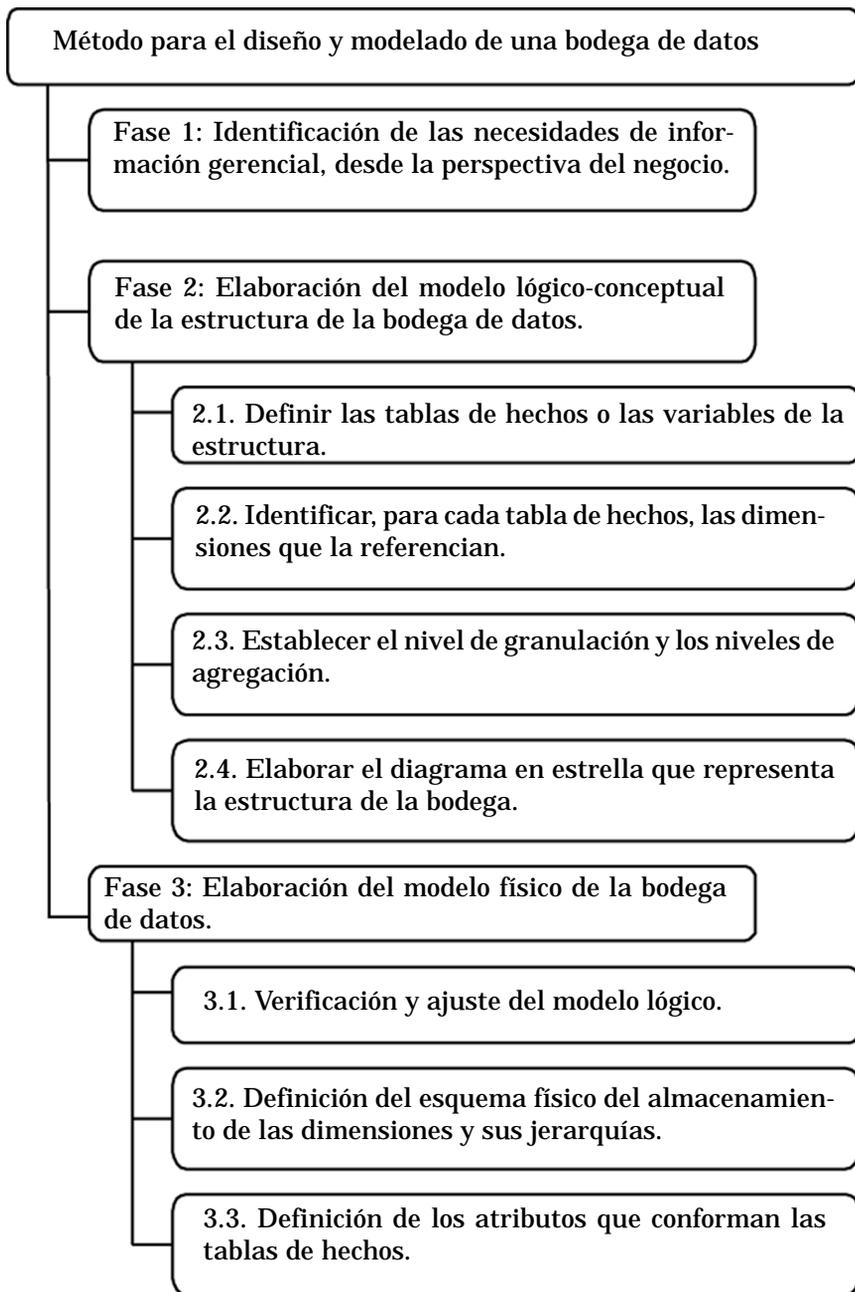


Figura 1. Método para el diseño y modelado de una bodega.

Como resultado de esta fase, se tendrá una visión del negocio y de la información requerida para la dirección y el control gerencial, representada fundamentalmente por: Los objetivos del negocio; los factores clave de éxito y, en especial, un conjunto de indicadores clave de la gestión.

Fase 2:
Elaboración del modelo lógico-conceptual de la estructura de la bodega.

En esta fase se elabora el modelo lógico de la estructura de la bodega, que soportará las consultas, mediante las cuales se obtendrá la información requerida por los niveles directivos como apoyo a sus procesos de gestión y de toma de decisiones.

La elaboración de este modelo lógico comienza con los indicadores de ges-

tión (necesidades de información gerencial), identificados en la fase anterior, y termina con la construcción de una representación multidimensional de las variables que conforman cada indicador. En esta representación multidimensional, cada variable es modelada mediante un arreglo dimensional (multidimensional) de celdas, como se presenta en la Figura 2.

Para facilitar el proceso de elaboración del modelo lógico, se utiliza una representación gráfica denominada diagrama tipo estrella, donde el elemento central del esquema es la Variable o Tabla de Hechos («Fact»), la cual es referenciada por un conjunto de ejes, denominados Dimensiones, a través de los cuales se seleccionan los valores contenidos en la tabla de hechos. En la Figura 3, se esquematiza el modelo de un diagrama en estrella.

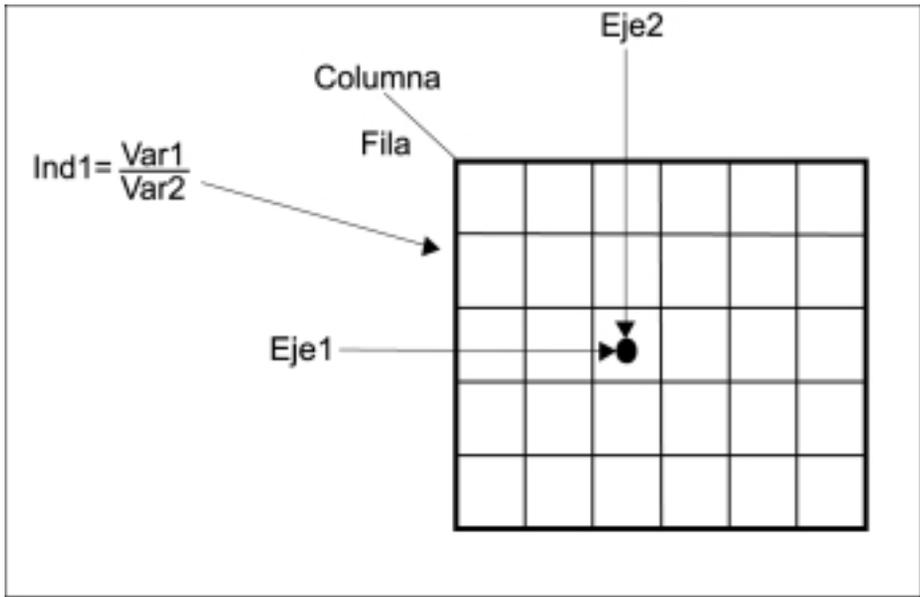


Figura 2. Vista multidimensional de una de las variables que conforman un indicador.

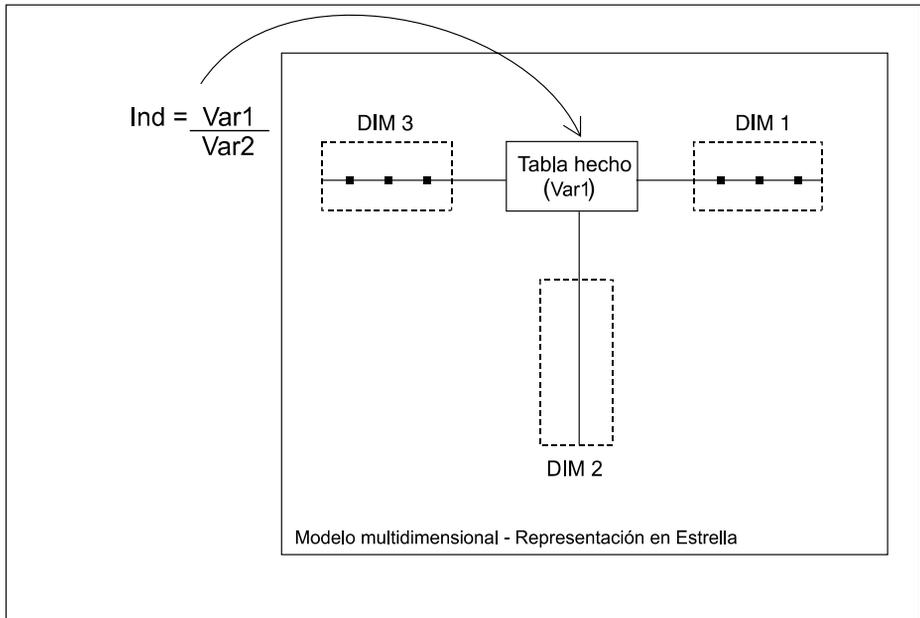


Figura 3. Diagrama en estrella de una estructura multidimensional.

Antes de presentar los pasos propuestos por el método para la elaboración del modelo lógico, es pertinente precisar algunos de los términos utilizados en el método propuesto. En el

Cuadro 1 se presentan las definiciones adoptadas para los diferentes conceptos utilizados en el método propuesto.

Cuadro 1: Definición de conceptos básicos.

Gráfica	<p>Una gráfica es una red de nodos interconectados. Una gráfica direccional es aquella en la cual la conexión entre dos nodos tiene una dirección específica. Un modelo E-R puede ser considerado una gráfica direccional.</p>
Trayectorias cíclicas y acíclicas	<p>En una gráfica, una trayectoria acíclica es aquella que sólo tiene una forma de recorrido (en un solo sentido). Una trayectoria cíclica es aquella que se puede recorrer en dos o más secuencias diferentes.</p>
Tabla de hechos	<p>Es la tabla central de la estructura de la bodega. Esta tabla contiene los datos de interés para el negocio, es decir, los valores para la construcción de los indicadores claves del negocio.</p> <p>Técnicamente, la tabla de hechos es una entidad de intersección cuya llave primaria está compuesta por la unión de los dominios de las diferentes dimensiones que la referencian.</p>
Dimensión	<p>Las dimensiones corresponden a los ejes con los cuales se construye la vista multidimensional de la información clave del negocio, almacenada en la tabla de hechos.</p> <p>Las atributos almacenados en las dimensiones determinan la granulación adoptada para el modelo.</p> <p>Las dimensiones pueden ser:</p> <ul style="list-style-type: none"> - Propias: Cuando el conjunto de entidades que conforman la dimensión se encuentran unidas a la tabla de hechos, en una trayectoria acíclica. - Impropias: Cuando el conjunto de entidades que conforman la dimensión se encuentran unidas a la tabla de hechos, en una trayectoria cíclica. - De Información: Cuando los atributos contenidos en la dimensión definen qué tipo de datos se encuentran almacenados en la tabla de hechos.
Jerarquías	<p>Determinan cómo las instancias de la tabla de hechos pueden ser agregadas. Las jerarquías permiten las operaciones de «drill-down» o «rollup», en los procesos de consulta.</p> <p>Una jerarquía está conformada por el conjunto de entidades que constituyen la dimensión.</p>

A continuación se presentan los cuatro pasos propuestos para la sistematización del proceso de elaboración del modelo lógico:

Paso No. 1:

Definir las tablas de hechos o las variables de la estructura.

Este paso se realiza a partir del conjunto de los indicadores de gestión, definidos en la fase de identificación de las necesidades de información gerencial, desde la perspectiva del negocio. El paso se inicia con la evaluación de las variables (divisores y dividendos) de cada indicador, para determinar cuáles de éstas pueden ser almacenadas en una tabla de hechos, y cuáles no.

En el Cuadro 2 se presenta, a manera de ejemplo, la información obtenida, al aplicar los pasos de la fase 1 al área de ventas de una organización. A partir de estos resultados se identifican las variables o tablas de hechos, como lo establece el paso 1 de esta fase.

Aplicación del paso 1: Identificación de las tablas de hechos

- El indicador definido para el monitoreo del objetivo puede ser construido con una sola tabla de hechos: Ventas. Se toma una sola variable, por cuanto las ventas del año y las ventas del año anterior, que son las dos variables que conforman el indicador, se pueden almacenar en la misma tabla de hechos.

- El indicador 1 del F.C.E1 puede ser construido con dos tablas de hechos que son: ventas por vendedor y cuota de ventas de cada vendedor.
- El indicador 2 del F.C.E1 puede ser construido con dos tablas de hechos que son: número de visitas realizadas por cada vendedor, y número de visitas presupuestadas por cada vendedor.
- El indicador 1 del F.C.E2 puede ser construido con una tabla de hechos: número de clientes nuevos en la base de datos. En este caso, el denominador del indicador se asume como un único valor y, por lo tanto, no tiene sentido almacenarlo en otra tabla de hechos.
- Los demás indicadores se analizan de igual manera.

En suma, al realizar el análisis de todos los indicadores, obtenemos las siguientes tablas de hecho:

- Ventas.
- Ventas por vendedor.
- Cuota de ventas de cada vendedor.
- Número de visitas realizadas por cada vendedor.
- Número de visitas presupuestadas por cada vendedor.
- Número de clientes nuevos en la base de datos.
- Número de vendedores capacitados que aprobaron los cursos.

Cuadro 2: Información gerencial del área de ventas,
obtenida al realizar la fase 1.

Área del Negocio - Descripción. Se trabaja con el área de ventas de una organización dedicada a la producción de recipientes elaborados en plástico.

Objetivo del Área: para propósitos del ejemplo, se toma el siguiente objetivo:

Lograr al final del año un incremento del 15% en las ventas totales de la compañía, con respecto a las ventas del año anterior:

Factores claves de éxito. Luego de realizado el análisis de las acciones y las condiciones necesarias para garantizar el logro del objetivo planteado, se identificaron los siguientes F.C.E:

- **F.C.E.1:** Planeación y control de la fuerza de ventas.
- **F.C.E.2:** Búsqueda de nuevos clientes rentables para la organización.
- **F.C.E.3:** Capacitación y entrenamiento de la fuerza de ventas.

Indicadores claves de gestión. Para el control y seguimiento de los F.C.E y los objetivos, se proponen los siguientes indicadores:

$$I_{obj}: \frac{\text{Ventas del año}}{\text{Ventas del año anterior}} - 1$$

$$I1_FCE1: \frac{\text{Ventas del vendedor}}{\text{Cuota de ventas}}$$

$$I2_FCE1: \frac{\text{Número de visitas de venta realizadas}}{\text{Número de visitas presupuestadas}}$$

$$I1_FCE2: \frac{\text{Número de clientes nuevos en la base de datos}}{\text{Número de clientes nuevos presupuestados}}$$

$$I1_FCE3: \frac{\text{Número de vendedores capacitados que aprobaron los cursos}}{\text{Número presupuestado de vendedores capacitados}}$$

Paso No. 2:
Identificar, para cada tabla de hechos, las dimensiones que la referencian.

Para cada variable o tabla de hechos se identifican, con la colaboración del usuario líder del área de negocio, los ejes de visualización multidimensional los cuales constituyen las dimensiones de la variable.

En este paso, se espera que el usuario visualice cada variable, como un conjunto de valores almacenados en una estructura de varias dimensiones, donde los valores almacenados son referenciados por la combinación de los valores definidos para cada eje (dominio de la dimensión), tal como se esquematiza en la Figura 4.

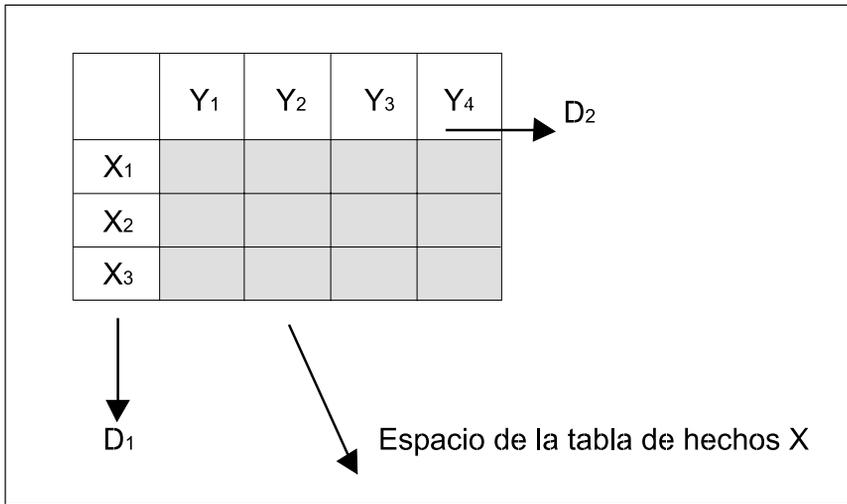


Figura 4: Esquema de una vista multidimensional de una tabla de hechos.

Paso 3:
Establecer el nivel de granulación y los niveles de agregación de cada dimensión.

Una vez que las dimensiones han sido identificadas se debe establecer, para cada una de ellas, el menor nivel de granulación, el cual corresponde al conjunto de atributos que referencian el mayor nivel de detalle deseado para la variable o tabla de hechos.

A manera de ejemplo, se aplican los dos pasos anteriores para la Tabla de

Hechos sobre Ventas, definida en el ejemplo anterior.

Aplicación del paso 2: Identificación de las dimensiones.

Supongamos que el gerente de ventas expresa su interés por visualizar la información de ventas organizada de la siguiente manera: primero, por cada producto de la compañía; en segundo término, por cada lugar en donde se venden los productos y, finalmente, por cada semana. Podemos establecer la necesidad de utilizar

tres ejes para elaborar la vista multidimensional (dimensiones) de la Tabla de Hechos - Ventas:

- Dim1: Producto
- Dim2: Lugar de venta
- Dim3: Tiempo.

Aplicación del paso 3: Definición del nivel de granulación.

De acuerdo con la solicitud del gerente, se establece para cada dimensión la siguiente granulación:

- **Dim. Producto:** El menor nivel de granulación es el *Tipo de Productos*. Podemos establecer otros niveles como: *Línea de Productos*, que tiene un nivel de granulación mayor, pero un menor nivel de detalle en la variable ventas; o *Referencias de Productos*, que tiene un menor nivel de granulación, pero un mayor nivel de detalle.
- **Dim. Lugar:** El nivel de granulación requerido es la *Ciudad*. Se habrían podido seleccionar otros niveles, como *Almacén*, que tiene un menor nivel, o *Región*, que tiene uno mayor.
- **Dim. Tiempo:** El menor nivel de granulación requerido es la *Semana*. Se habrían podido seleccionar otros niveles, como *el Día*, que tiene un menor nivel, o el *Mes*, que tiene uno mayor.

Una vez se han definido los menores niveles de granulación para cada dimensión, se identifican los niveles de agregación requeridos para los valores almacenados en la tabla de he-

chos, por cada dimensión. Estos niveles de agregación representan la jerarquía de cada dimensión.

- **Jerarquía en la Dim. Producto:** Las ventas por productos pueden ser agregadas por grupos de productos, por líneas de productos y, por el total de la venta. De esta manera, los niveles de agregación de la dimensión producto son:
 - Por grupos de productos.
 - Por líneas de productos.
 - Total.
- **Jerarquía en la Dim. Lugar:** Las ventas por lugar pueden ser agregadas por regiones y por el total del país.
- **Jerarquía en la Dim. Tiempo:** Las ventas por tiempo pueden ser agregadas por mes, por trimestre, por semestre, por año.

Paso 4:

Elaborar el diagrama en estrella que representa la estructura de la bodega.

Luego de identificar los elementos que conforman la estructura de la vista multidimensional, de la información gerencial requerida por la organización, se pasa a la elaboración de una representación gráfica, en forma de estrella; para ello se puede utilizar la notación simplificada de los diagramas E-R, o la notación denominada «Dot modeling».⁵

5. Todman, Chris. *Designing a Data Warehouse: Supporting Customer Relationship*. Prentice Hall, 2001.

Notación tipo E-R:

En esta notación, el diagrama en estrella está conformado por una entidad central asociativa, que corresponde a la tabla de hechos, y por un con-

junto de trayectorias de entidades y relaciones de uno a muchos, que corresponde a las dimensiones y a sus jerarquías. En la Figura 5 se presenta un diagrama en estrella con esta notación.

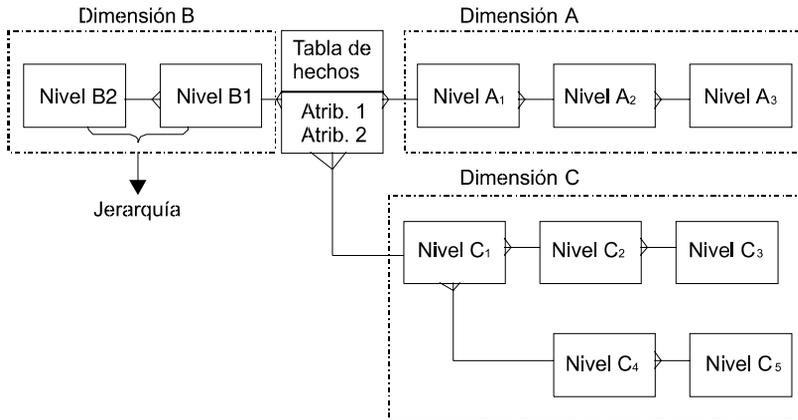


Figura 5. Representación de un diagrama en estrella, mediante la notación tipo E-R.

Notación «Dot Modeling»

En esta notación, el diagrama en estrella está conformado por una entidad central que corresponde a la Tabla de Hechos, y por un conjunto de

trayectorias compuestas por puntos («dots»), que representan las dimensiones y sus jerarquías. En la Figura 6 se presenta un diagrama en estrella con esta notación.

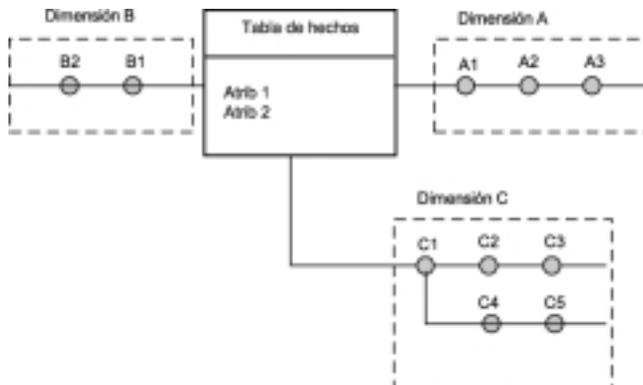


Figura 6. Representación de un diagrama en estrella, mediante la notación «Dot Modeling».

A manera de ejemplo se presenta en la Figura 7, el diagrama en estrella, con notación «Dot Modeling», para la Tabla de Hechos y para las dimensiones identificadas en el ejemplo

anterior. En la Figura 8 se representan los mismos elementos de la estructura de la bodega, pero con notación tipo E-R.

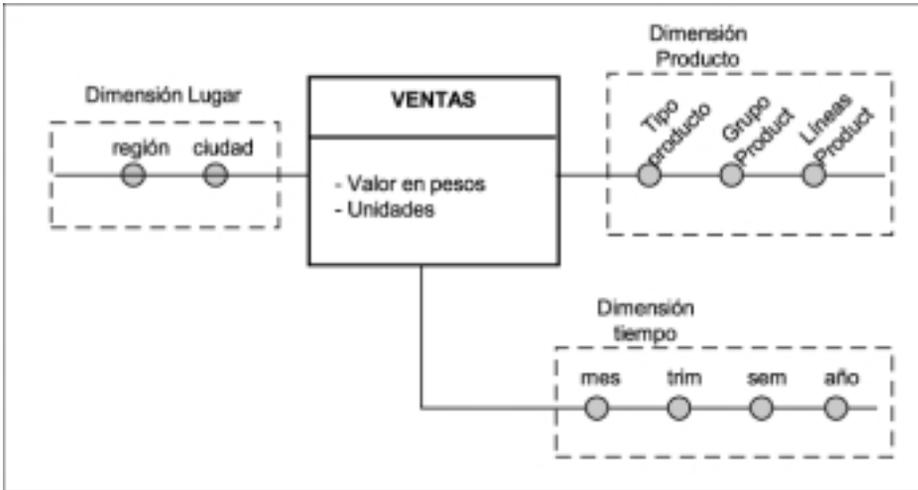


Figura 7. Representación, mediante la notación «Dot Modeling», de la estructura para la bodega de datos del ejemplo anterior.

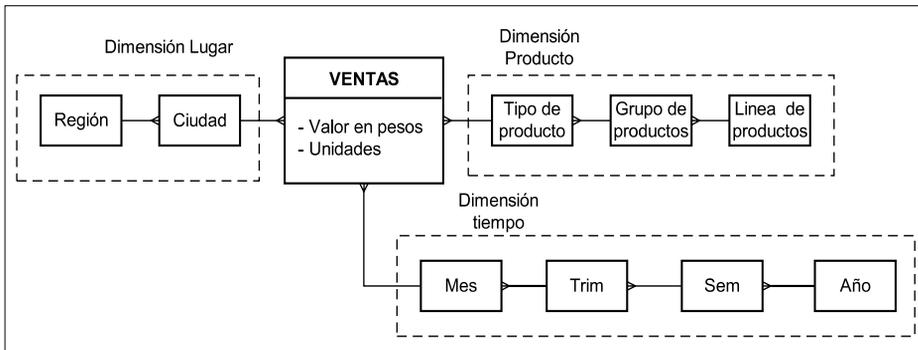


Figura 8. Representación, mediante la notación E-R, de la estructura para la bodega de datos del ejemplo anterior.

Fase 3:

Elaboración de la estructura física de la bodega

Durante esta fase, se realiza la transformación del modelo lógico conceptual en la estructura física, que posteriormente será implementada en alguna herramienta de «Data Warehouse».

Este proceso de transformación se realiza mediante los siguientes pasos:

1. Verificación y refinamiento del modelo lógico para determinar su consistencia.
2. Definición del esquema físico de almacenamiento de las estructuras jerárquicas de las dimensiones.
3. Identificación de los atributos que conforman las tablas de hechos y las dimensiones.

Paso 1:

Verificación y ajuste del modelo lógico.

Durante este paso, se realiza la verificación del modelo lógico, obtenido en la fase anterior, para garantizar que el modelo, además de soportar todas las consultas requeridas por los niveles ejecutivos, siempre retorne información confiable.

Para iniciar este proceso de verificación se debe elaborar una matriz de cruce, entre los requerimientos de información gerencial, definidos en la fase inicial, y las estructuras (estrellas), definidas en la fase anterior. En la matriz de cruce se confirma si el requerimiento está completamente

soportado. Si esta verificación no es correcta, se debe retornar a la fase anterior, para incorporar las estructuras que soporten los requerimientos faltantes de información.

Terminada la revisión anterior, el proceso continúa con la evaluación de la estructura, para asegurar la validez de todas las consultas de información realizadas sobre dicha estructura.

Para realizar este proceso de comprobación de validez de la estructura, recurrimos a la teoría de grafos, según la cual una estructura de consulta es válida cuando está conformada por trayectorias acíclicas. Al aplicar esta teoría, se puede afirmar que cualquier diseño para una bodega de datos permitirá siempre consultas correctas, si la estructura propuesta está conformada únicamente por dimensiones propias, es decir, por trayectorias acíclicas.

Si al realizar la comprobación de la estructura se encuentran trayectorias acíclicas, éstas deben ser transformadas, para asegurar la confiabilidad de las consultas. Las posibles transformaciones son:⁶

1. Ajuste para los casos de trayectorias cíclicas simples

Este caso ocurre cuando la trayectoria de una dimensión presenta una trayectoria alterna que tiene dos entidades comunes. En la Figura 9, se esquematiza una trayectoria cíclica simple.

6. Mcguff, F. *Designing the perfect Data Warehouse*. 1998.

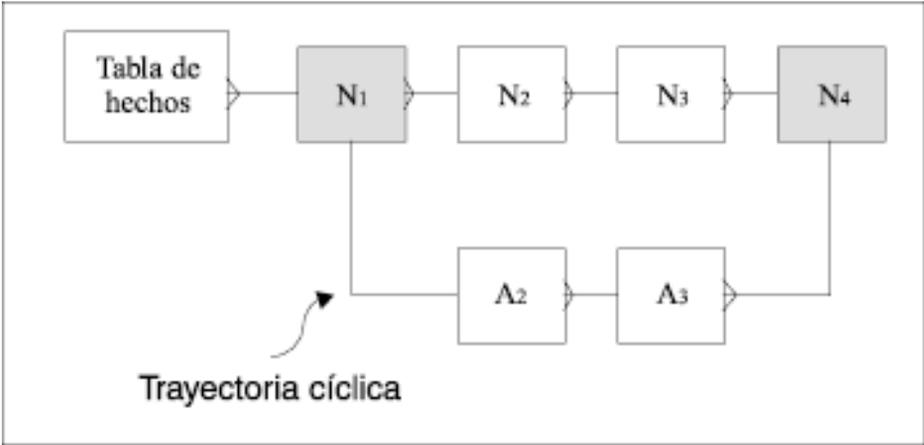


Figura 9. Dimensión con trayectoria cíclica.

Las opciones de transformación para esta clase de trayectorias son:

- Tratar cada trayectoria como una nueva dimensión, lo cual significa redibujar el diagrama, eliminando las relaciones N1-A2 y A3-N4, para luego crear la relación: Tabla de Hechos - A2.
- Convertir la trayectoria cíclica en una trayectoria alterna, eliminando la relación A3-N4.

2. *Ajuste para los casos de trayectorias alternas, mezcladas con trayectorias cíclicas.*

Se presenta cuando la trayectoria de una dimensión está conformada por una trayectoria alterna, más una trayectoria cíclica, tal como se esquematiza en la Figura 10.

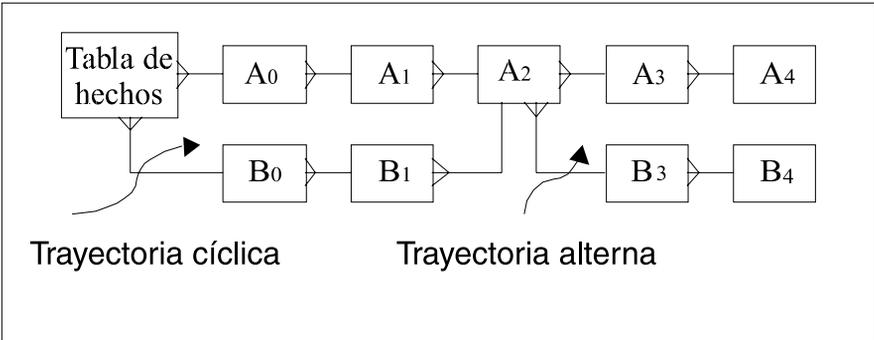


Figura 10. Dimensión con trayectoria alterna, más trayectoria cíclica.

En estos casos, el problema ocurre con la trayectoria cíclica; por lo tanto, la transformación se maneja como se explicó en el caso anterior.

**Paso 2:
Definición del esquema físico del almacenamiento de las dimensiones y sus jerarquías.**

El modelo en estrella que conforma la estructura lógica propuesta para

la bodega de datos, debe ser convertido en una estructura totalmente desnormalizada, tal como se presenta en la Figura 11. Este modelo físico está conformado por una tabla de hechos, y por las entidades en las cuales se almacenarán los dominios de las dimensiones con sus correspondientes niveles jerárquicos.

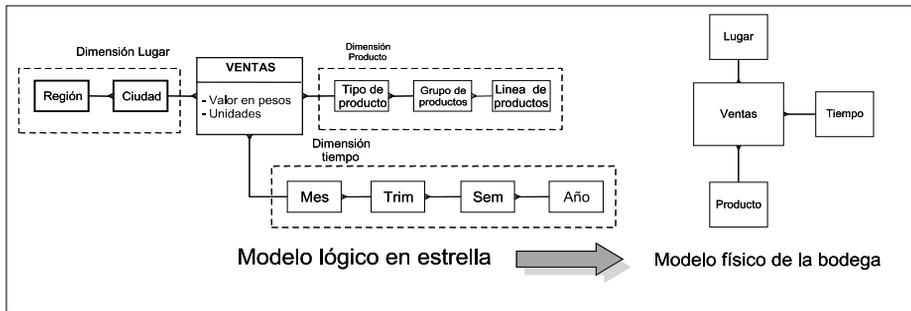


Figura 11. Modelo lógico en estrella, y modelo físico de la bodega.

Para el proceso de conversión de cada una de las trayectorias que conforman el modelo en estrella, en entidades desnormalizadas, se puede utilizar uno de los siguientes esquemas de conversión.⁷

1. Conversión vertical o recursiva

En esta conversión, se utiliza una llave primaria única, para cada dimensión. El dominio de esta llave primaria se obtiene mediante la unión de

todos los dominios de las entidades que conforman la trayectoria de la dimensión, es decir, si los dominios de las entidades que conforman la trayectoria son: {enero, febrero, marzo, abril}; {1er_trim, 2º_trim, 3er_trim, 4º_trim}; {1er_sem, 2º_sem}, el dominio de la llave primaria será: {enero, febrero, marzo, abril, ..., 1er_trim, 2º_trim, 3er_trim, 4º_trim; 1er_sem,2º_sem}. En la Figura 12 se presenta, de manera gráfica, este esquema de conversión.

7. Mcguff, F. *Designing the perfect Data Warehouse*. 1998.

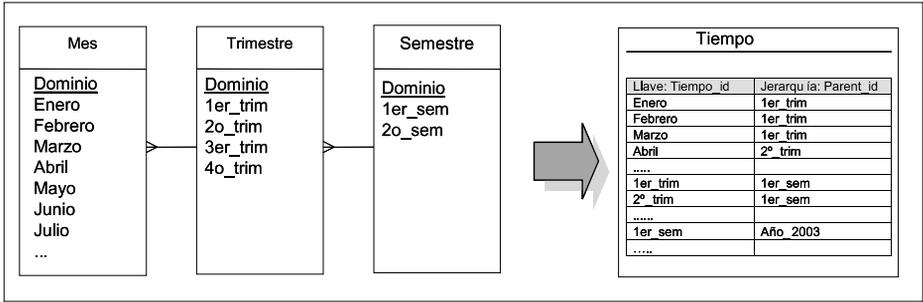


Figura 12. Conversión vertical de la trayectoria de una dimensión.

Adicionalmente, en este esquema de conversión a cada valor del dominio se le asocia un valor padre, el cual también pertenece al dominio; de esta manera se implementa la jerarquía definida en la trayectoria, representada en la dimensión, dentro del modelo en estrella.

Este esquema para el manejo de las jerarquías (id_dimensión, id_padre) permite implementar fácilmente la operación de desenrolle («drill-down»), cuando se realizan consultas a la bodega de datos. Sin embargo, esta estructura es eficiente, si las agregaciones para cada nivel jerár-

quico son precalculadas y almacenadas en la bodega.

Este esquema de conversión es el más recomendado para implementar la estructura física de una bodega, cuando las dimensiones están compuestas por jerarquías desbalanceadas.

2. Conversión horizontal

En esta conversión, la llave primaria de la dimensión se conforma como una llave compuesta por las llaves de cada una de las entidades que conforman la trayectoria de la dimensión. En la Figura 13 se presenta, de manera gráfica, este esquema de conversión.

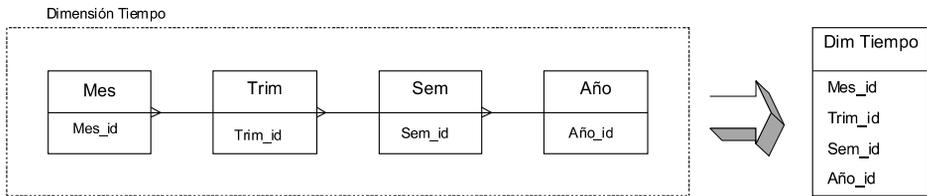


Figura 13: Conversión horizontal de la trayectoria de una dimensión.

Este esquema de conversión es el más recomendable, si las agregaciones de datos se realizan de manera dinámica.

**Paso 3:
Definición de los atributos que conforman las tablas de hechos y las dimensiones del modelo.**

En este paso final, se identifican para cada tabla de hechos y cada dimensión las características de los atributos

que conforman cada estructura. Una vez asignados todos los atributos, se realiza un análisis cruzado entre la tabla de hechos y las dimensiones, para establecer los tipos de cálculo matemático que pueden ser realizados, sobre la tabla de hechos.

La especificación de los atributos que conforman la tabla de hechos se debe realizar siguiendo el formato que aparece en el Cuadro 3.

Cuadro 3: Formato para la definición de los atributos de una tabla de hechos.

Nombre estructura de la bodega			
Tabla de hechos			
Atributos	Tipo	Pk	Descripción

Igualmente, para la especificación de los atributos que conforman las dimensiones se debe utilizar el formato que aparece en el Cuadro 4.

Cuadro 4: Formato para la definición de los atributos de una dimensión.

Nombre estructura de la bodega			
Dimensión			
Atributos	Tipo	Pk	Fuente

Finalmente, se deben establecer los tipos de cálculos matemáticos como suma, conteo, promedio, mínimo, máximo, que pueden ser aplicados a los valores almacenados en las tablas

de hechos. El resultado de esta revisión debe quedar consignado en una matriz de cruce, como la presentada en el Cuadro 5.

Cuadro 5: Operaciones matemáticas para cada atributo de la tabla de hechos.

Tabla de hechos	<input style="width: 100%;" type="text"/>				
Atributo 1	<input style="width: 100%;" type="text"/>				
Dimensiones	Suma	Conteo	Prom.	Mín.	Máy.
1. Dimensión a	✓		✓	✓	✓
2. Dimensión b	✓				
3. Dimensión c		✓			
....					
Atributo 2	<input style="width: 100%;" type="text"/>				
Dimensiones	Suma	Conteo	Prom.	Mín.	Máy.
1. Dimensión a					
2. Dimensión b					
3. Dimensión c					
....					
Atributo 3	<input style="width: 100%;" type="text"/>				
Dimensiones	Suma	Conteo	Prom.	Mín.	Máy.
1. Dimensión a					
2. Dimensión b					
3. Dimensión c					
....					

A manera de ejemplo, se presenta en los siguientes cuadros la definición de atributos para la tabla de hechos y para las dimensiones definidas en el ejemplo anterior, y esquematizadas en la Figura 7.

Cuadro 6: Definición de los atributos de la tabla de hechos sobre ventas.

Nombre de la estructura de la bodega		Área de ventas	
Tabla de hechos		Ventas	
Atributos	tipo	Pk	Descripción
Id lugar	C(35)	Si	Identif. de la dimensión lugar
Id tiempo	C(12)	Si	Identif. de la dimensión tiempo
id producto	C(30)	Si	Identif. de la dimensión producto
Unidades vendidas	N(8,0)		Valor 1 de la tabla de hechos
Pesos-venta	N(10,2)		Valor 2 de la tabla de hechos

Cuadro 7: Definición de los atributos de las dimensiones, para la estructura de ventas.

Nombre de la estructura de la bodega		Área de ventas / ventas	
Dimensión		Lugar	
Atributos	tipo	Pk	Fuente
id_lugar	C(35)	Si	
Jerarquía_lugar_1	C(35)		

Nombre de la estructura de la bodega		Área de ventas / ventas	
Dimensión		Tiempo	
Atributos	tipo	Pk	Fuente
id_tiempo	C(12)	Si	
Jerarquía_tiempo	C(12)		

Nombre de la estructura de la bodega		Área de ventas / ventas	
Dimensión		Producto	
Atributos	tipo	Pk	Fuente
id_producto	C(30)	Si	
Jerarquía_producto	C(30)		

Cuadro 8: Operaciones matemáticas para cada atributo de la tabla de hechos sobre ventas.

Tabla de hechos	Ventas				
Atributo 1	Unidades vendidas				
Dimensiones	Suma	Conteo	Prom	Mín	Máx
Lugar	✓		✓	✓	✓
Tiempo	✓		✓	✓	✓
Producto	✓		✓	✓	✓
Atributo 2	Pesos-venta				
Dimensiones	Suma	Conteo	Prom	Mín	Máx
Lugar	✓		✓	✓	✓
Tiempo	✓		✓	✓	✓
Producto	✓		✓	✓	✓

CONCLUSIÓN

Mediante la aplicación del enfoque de Sistemas para la definición de los indicadores claves de gestión de la organización, se ha logrado articular una propuesta para modelar, de manera ordenada y sistémica, las estructuras de las bodegas de datos que servirán de soporte a la implementación de sistemas de información gerencial, hechos a la medida de las necesidades de información de la gerencia. Esta propuesta facilita, ordena y sistematiza un proceso que en algunas

organizaciones se realiza de manera intuitiva y, en otras mediante la utilización de estructuras de bodegas que han sido definidas para otras organizaciones. El modelo propuesto, que se aparta de muchos de los enfoques presentados por los investigadores en este campo, se convierte en una opción válida para el diseño de sistemas de información gerencial, en particular para el diseño de bodegas de datos departamentalizadas («Data Marts»).

BIBLIOGRAFÍA

- McGuff, F. *Designing the perfect Data Warehouse*. 1998. <http://members.aol.com/fmcguff/dwmodel/index.htm>
- Kimball R. *The Data Warehouse Toolkit*. John Wiley & Sons, 1996.
- Inmon W.H. *Building The Data Warehouse*. QED Press /Jhon Wiley, 1992.
- Golfarelli, M; Maio, D; Rizzi, S. *Conceptual Design of Data Warehouse From E/R schemes*. <http://www.csr.unib.it/~golfare/db.html>, 1998.
- Bahamón, J. H. *Construcción de indicadores de gestión bajo el enfoque de sistemas*. S&T Revista de la Facultad de Ingeniería, Universidad Icesi. 2003.
- Chuck, B; Dick, H; Don, S; Rhonda; Eunsaeng, K.; Ann, V. *Data*

Modeling Techniques for Data Warehouse. IBM. 1998.

- Todman, C. *Designing a Data Warehouse: Supporting Customer Relationship*. Prentice Hall. 2001

CURRÍCULO

José Hernando Bahamón L. Ingeniero Electrónico de la Universidad del Cauca, especialista en Administración de la Universidad Icesi y magíster en Dirección Universitaria de la Universidad de los Andes. Profesor investigador de la Universidad Icesi. Vinculado a la Universidad Icesi desde 1988. Ha sido jefe del Departamento Académico de Sistemas (1988-1998), Director del programa de Ingeniería de Sistemas (1998-2000), y en la actualidad es el Director Académico de la Universidad. 

