



Sistemas de Bases de Datos 2

2024

Ing. Luis Alberto Arias Solórzano

Unidad 7



Esquemas dimensionales

- Cada vez es más necesario apoyar la búsqueda con el acceso directo a la base de datos desde varias perspectivas diferentes.

Por ejemplo:

- Tener la posibilidad de encontrar todas las transacciones de negocios que involucran a un producto en particular o todas las transacciones de negocios que suceden dentro de un periodo en particular o todas las transacciones de negocios que se refieren a un cliente en particular.
- Un método de organización que soporta este tipo de acceso fue llamado base de datos "multicatálogo". Para continuar con nuestro ejemplo, dicha base de datos consistiría en un gran archivo central de datos que contuviera los datos de las transacciones de negocios, junto con tres archivos de "catálogo" individuales para productos, periodos y clientes, respectivamente.



Esquemas dimensionales

- Las relaciones de catálogo se parecen a los índices dado que contienen apuntadores hacia los registros que están en el archivo de datos; sin embargo, (a) las entradas pueden ser colocadas en las tablas explícitamente por el usuario y (b) los archivos pueden contener información suplementaria (por ejemplo, la dirección del cliente) que luego puede ser eliminada del archivo de datos.
- Por lo general las tablas de catálogo son pequeños en comparación con la tabla de datos. Esta organización es más eficiente en términos de espacio y E/S que el diseño original (el cual involucra un solo archivo de datos).
- En el ejemplo: la información sobre productos, periodos y clientes del archivo central de datos ahora se reduce a sólo *identificadores* de producto, periodo y cliente.



Esquemas dimensionales

- Cuando se emula este enfoque en una base de datos relacional, las tablas de datos y los catálogos se convierten en relaciones; los apuntadores que están en los datos de catálogo se convierten en claves primarias en las tablas que son imagen de estos datos, y los identificadores que están en la tabla datos se convierten en claves externas en la tabla que es imagen del archivo de datos.
- Por lo general, estas claves primarias y externas están indexadas. Bajo tal arreglo, a la imagen del archivo de datos se le llama **tabla de hechos** y a las imágenes de los archivos de catálogo se les llama **tablas de dimensión**.
- Al diseño general se le menciona como **esquema dimensional, o de estrella**, debido a la forma en que aparece cuando es trazado como diagrama de entidad-vínculo.



Modelo Estrella

- Es un modelo de datos que tiene una tabla de hechos (o fact table) que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones.

Características:

- 1. Consultar una base de datos con esquema de estrella involucra el uso de las tablas de dimensión para encontrar todas las combinaciones de clave externa que son de interés, y luego el uso de esas combinaciones para acceder a la tabla de hechos.



Modelo Estrella

- 2. Los esquemas de estrella en realidad son físicos y no lógicos, aunque se habla de ellos como si fueran lógicos. El problema es que en realidad en el enfoque del esquema de estrella no hay concepto de diseño lógico para distinguirlo con respecto al diseño físico.
- 3. El enfoque del esquema de estrella no siempre da como resultado un diseño físico legítimo (es decir, uno que conserve toda la información en un diseño lógico relacionalmente correcto). Esta limitación se hace más aparente conforme el esquema es más complejo.
- 4. Debido a que hay muy poca disciplina, los diseñadores incluyen a menudo varios tipos de hechos diferentes en la tabla de hechos. Como consecuencia, las filas y columnas de la tabla de hechos no tienen una interpretación uniforme. Lo que es más, por lo general ciertas columnas sólo se aplican a determinados tipos de hechos, lo que implica que las columnas en cuestión deben permitir nulos. Y conforme son incluidos más y más tipos de hechos, la tabla se hace cada vez más difícil de mantener y comprender, y el acceso se hace cada vez menos eficiente.



Modelo Estrella

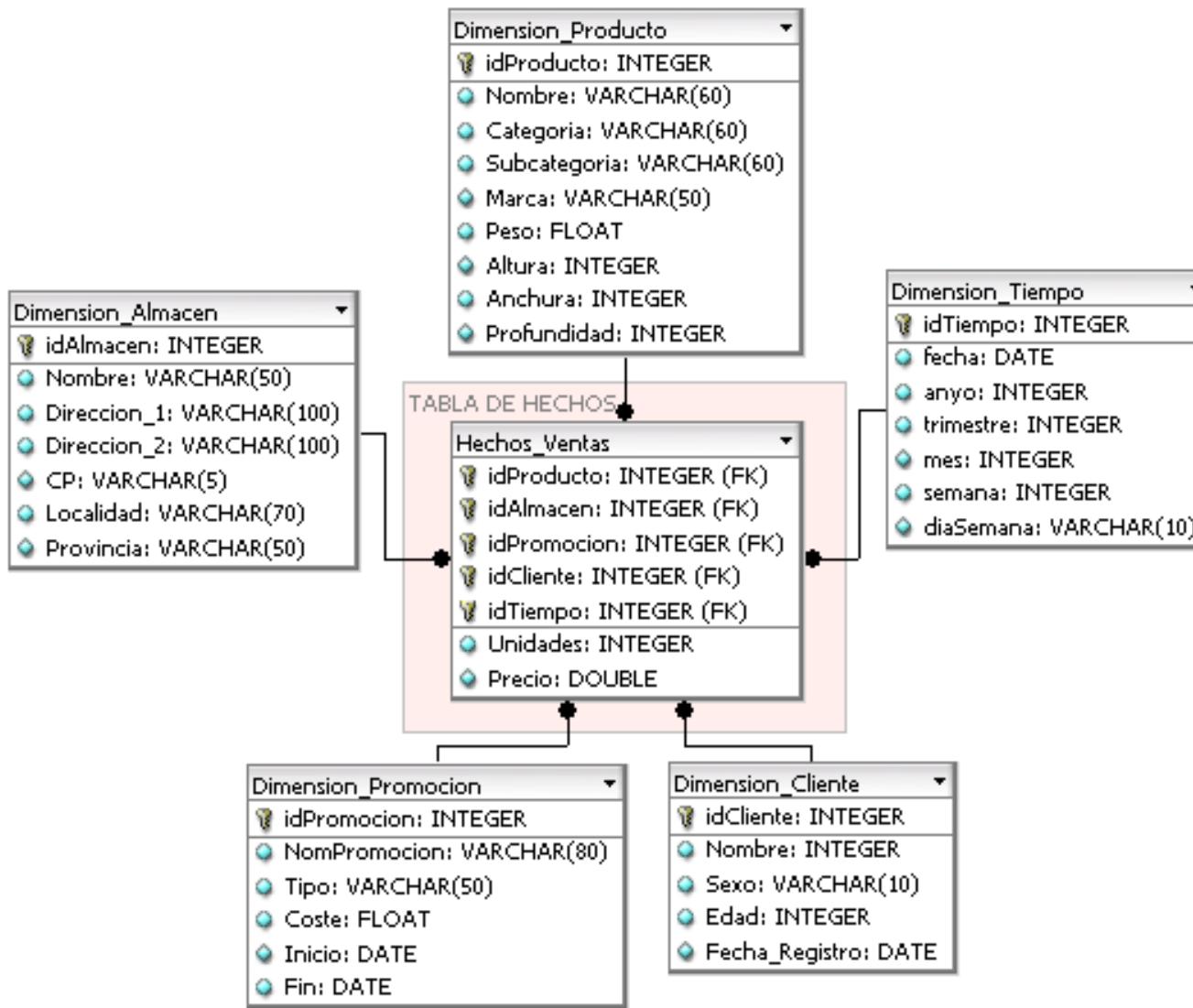
- 5. De nuevo, debido a la falta de disciplina, las tablas de dimensión también pueden llegar a ser no uniformes. Por lo general este error sucede cuando la tabla de hechos es usada para mantener datos que se refieren a diferentes niveles de agregación. Por ejemplo, podríamos (erróneamente) añadir filas a la tabla de envíos para que mostraran las cantidades totales de partes para cada día, cada mes, cada trimestre, cada año e incluso, el gran total a la fecha.



Modelo Estrella

- 6. Con frecuencia, las tablas de dimensión no están normalizadas por completo.* El deseo de evitar juntas conduce frecuentemente a los diseñadores a agrupar en esas tablas información distinta, que sería mejor mantener separada. En el caso extremo, las columnas a las que simplemente se tiene acceso en conjunto, son mantenidas juntas en la misma tabla de dimensión. Debe quedar claro que seguir tal "disciplina" extrema, y no relacional, conducirá con seguridad a una redundancia sin control y probablemente incontrolable.

Modelo Estrella

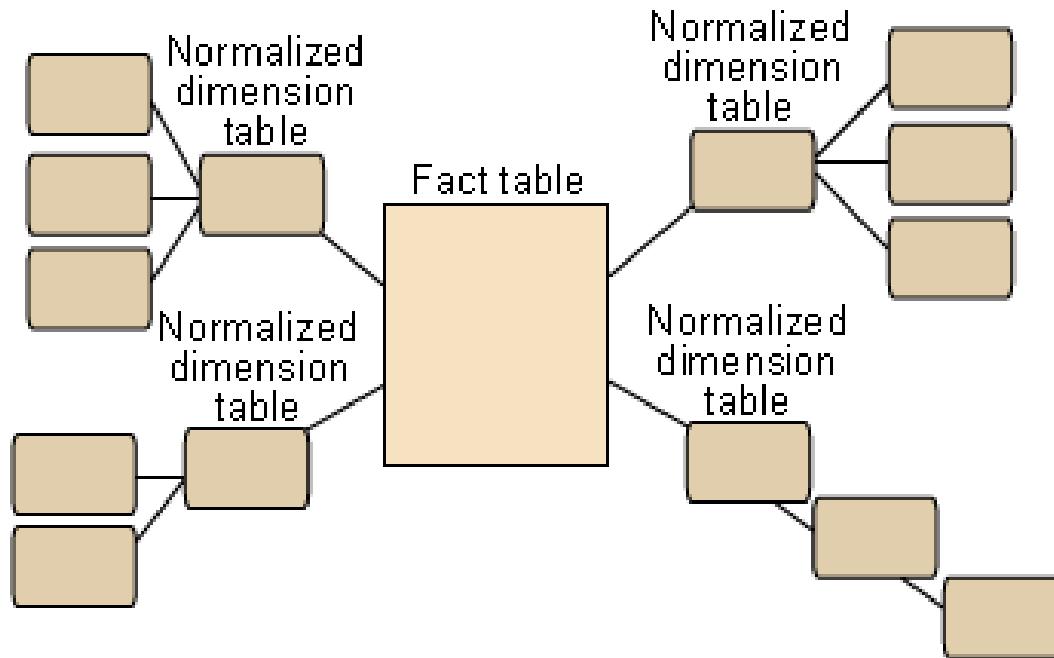


Modelo Estrella - Desventajas

- Con mucha frecuencia, los "modelos estrella" (también conocidos como esquemas *dimensionales*) son el resultado de intentar "tomar atajos" en una técnica adecuada de diseño. Es poco lo que se puede ganar con esos atajos.
- Con frecuencia afectan el rendimiento y la flexibilidad conforme crece la base de datos, y la resolución de tales dificultades por medio del rediseño físico fuerza también a hacer cambios en las aplicaciones (ya que en realidad los esquemas de estrella son esquemas *físicos*, aunque estén expuestos a las aplicaciones).
- Los diseñadores tratan frecuentemente de ahorrar espacio permitiendo nulos en las columnas (este truco *puede* funcionar si la columna en cuestión es de algún tipo de dato de longitud variable y el producto en cuestión representa a los nulos en dichas columnas por medio de cadenas vacías en el nivel físico). Sin embargo, por lo general dichos intentos son erróneos.



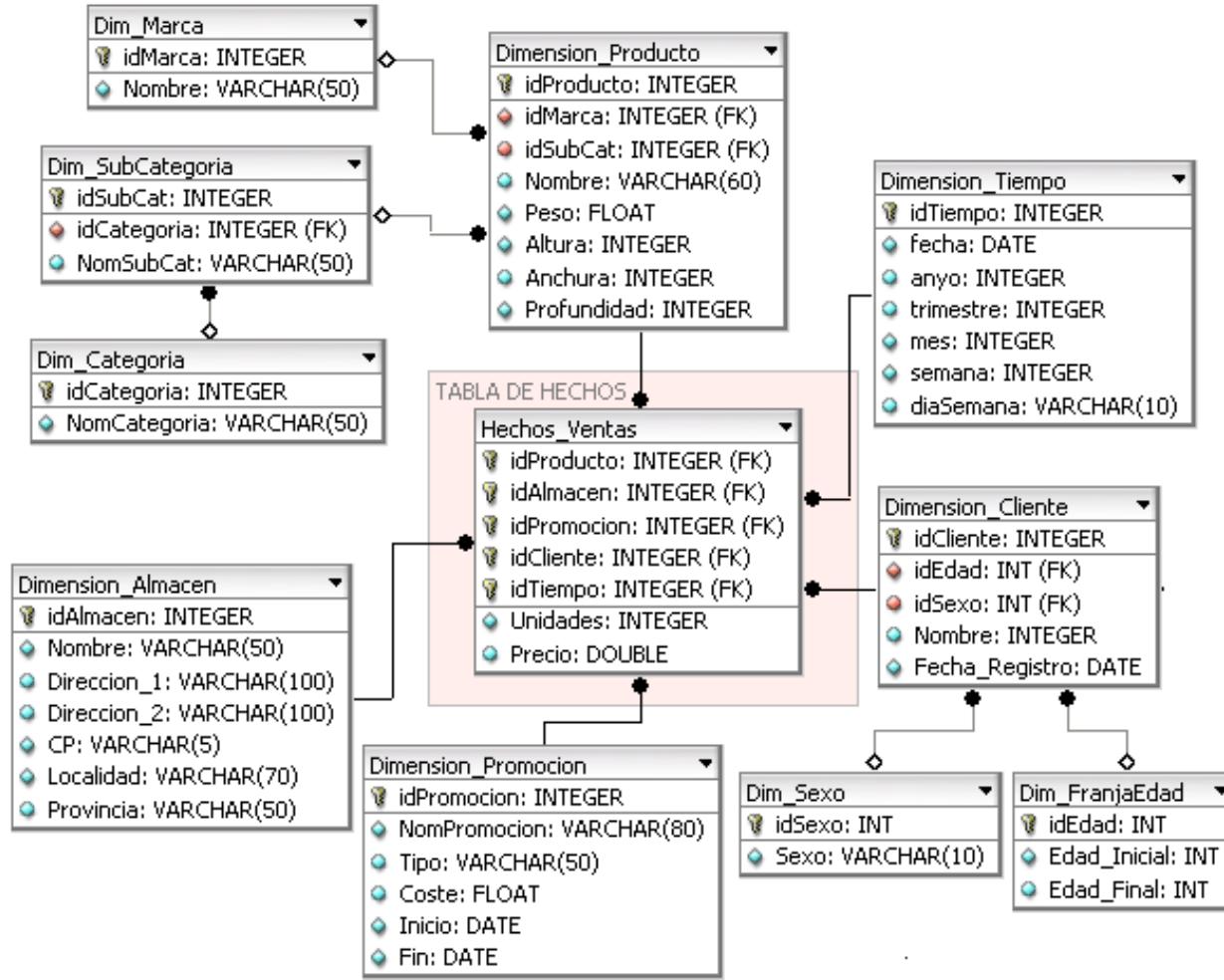
Modelo Copo de Nieve



Modelo Copo de Nieve

- El modelo de copo de nieve es una estructura algo más compleja que el esquema en estrella. Se da cuando alguna de las dimensiones se implementa con más de una tabla de datos. La finalidad es normalizar las tablas y así reducir el espacio de almacenamiento al eliminar la redundancia de datos; pero tiene la contrapartida de generar peores rendimientos al tener que crear más tablas de dimensiones y más relaciones entre las tablas (JOINS) lo que tiene un impacto directo sobre el rendimiento.
- El único argumento a favor de los esquemas en copo de nieve es que al estar normalizadas las tablas de dimensiones, se evita la redundancia de datos y con ello se ahorra espacio. Pero si tenemos en cuenta que hoy en día, el espacio en disco no suele ser un problema, y sí el rendimiento, implica tener que realizar código más complejo para realizar una consulta que a su vez se ejecutará en un tiempo mayor.

Modelo Copo de Nieve





Gracias