

Modelamiento Dimensional

Por

Elizabeth León Guzmán, Ph.D.

Profesora

Ingeniería de Sistemas

Grupo de Investigación MIDAS

Agenda

- Modelo Dimensional
 - Definición
 - Componentes
 - Ejemplos
 - Comparación con E/R
 - Proceso del diseño dimensional
 - Modelos de arquitectura
 - Arquitectura en bus

Modelamiento dimensional

Definición

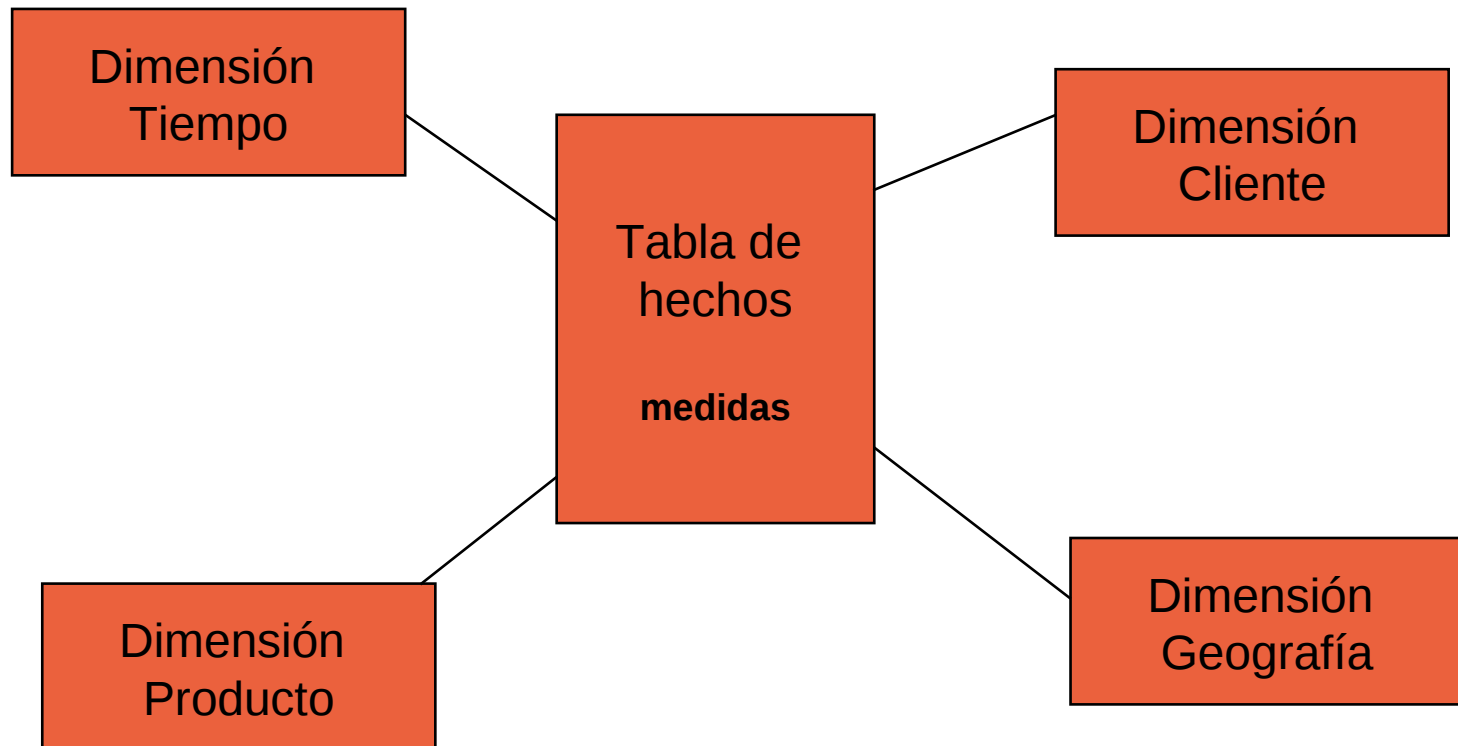
- Una técnica para **diseñar el modelo lógico** de la bodega de datos
- Permite **alto rendimiento** en el momento de acceder a los datos (orientado a consultas)
- Dimensional (orientado al negocio)
- Usa algunos conceptos del modelo entidad/relación
- Diferente del modelo entidad/relación

Modelo dimensional

- Define un nivel mínimo de detalle (granularidad)
- Se compone de:
 - **Hechos**
 - **Medidas**
 - **Dimensiones**
 - **Atributos**
 - Elementos
 - Jerarquías
 - **Relaciones**

-
- **Hechos**: algo que ocurre en el tiempo (operación o actividad)
 - **Medidas: Valores** numéricos que **describen** el **hecho** que se está analizando
 - **Dimensiones**: Son categorías que describen el **contexto** en el cual se analizan las medidas

Modelo dimensional



The diagram illustrates a hierarchical database schema structure. It is organized into a tree-like format with multiple levels of branching. Each node is represented by a rectangular box containing text. The text within each box typically follows a pattern: a table name in all caps, followed by a list of column names in all caps, separated by commas. The columns are often grouped into categories, such as personal information, contact information, and identification details. The diagram uses a color-coded system where different colors represent different types of data or relationships. The overall structure is highly detailed and covers a wide range of data elements, from basic personal data to more complex organizational and operational information.

Comparación

Operacionales	BI (Dimensional)
Enfocado a la actualización: Enfocado a eliminar redundancia, coordinar actualizaciones y repetir el mismo tipo de operaciones muchas veces en el día	Enfoque a la consulta
Altamente normalizadas para soportar actualizaciones consistentes y mantenimiento de la integridad referencial	Altamente desnormalizada ya que se requiere disminución de tiempos en la obtención de grandes cantidades de datos
Tiempos de respuesta en segundos o inferior	Tiempos de respuesta aceptables pueden ser segundos, minutos, horas
Almacenan pocos datos derivados	Gran cantidad de datos derivados (redundancia)
Pocos datos agregados	Agregación: Varios niveles de datos precalculados

Medidas

- Son las **variables** que ayudarán a medir el **desempeño del negocio**.
- Las medidas básicas, existen físicamente en el warehouse junto a los atributos que los caracterizan
- Las medidas derivadas o calculadas se construyen a partir de las medidas básicas y pueden o no estar almacenados físicamente en el data warehouse

Hechos y medidas

- Una fila en una tabla de hechos corresponde a una o varias medidas
- Todas las medidas en la tabla de hechos deben tener el mismo nivel de granularidad
- Las medidas mas usadas son numéricas y aditivas
- Las tablas de hechos representan la relación muchos a muchos que existe entre las dimensiones!

Dimensiones

- Son las **áreas temáticas o sujetos del negocio**.
- Proveen un método general para organizar la información corporativa
- Se definen como un grupo de uno o más atributos. Las dimensiones no comparten atributos
- Las dimensiones definen los niveles de análisis (**Jerarquías**). Esto permite obtener agregación o sumarización.

Atributos

- Los atributos son una **agrupación de elementos** o items dentro de una dimensión.
- Representan **categorías** o clases de elementos que tienen el mismo nivel lógico dentro de una dimensión.
- La finalidad de los atributos es ver la información de cada dimensión a **diferentes niveles de detalle** y **agrupar** los datos para ser analizados.

Elementos

- Son las **instancias** o valores de los atributos que, como componentes atómicos del modelo.

Relaciones

- Los atributos **dentro** de una dimensión están directamente relacionados uno con otros a través de los diferentes tipos de **relaciones** antes definidos.
- Los atributos en las **diferentes** dimensiones están relacionados otros a través de las **medidas** o variables del negocio definidas como intersección de las dimensiones.

Relaciones

- Son **asociaciones lógicas de atributos** dentro de una jerarquía definida por las instancias de los atributos y transitivas dentro de una jerarquía:
 - Uno-a-uno (1:1)
 - Uno-a-muchos (1:M)
 - Muchos-a-uno (M:1)
 - Muchos-a-muchos (M:N)

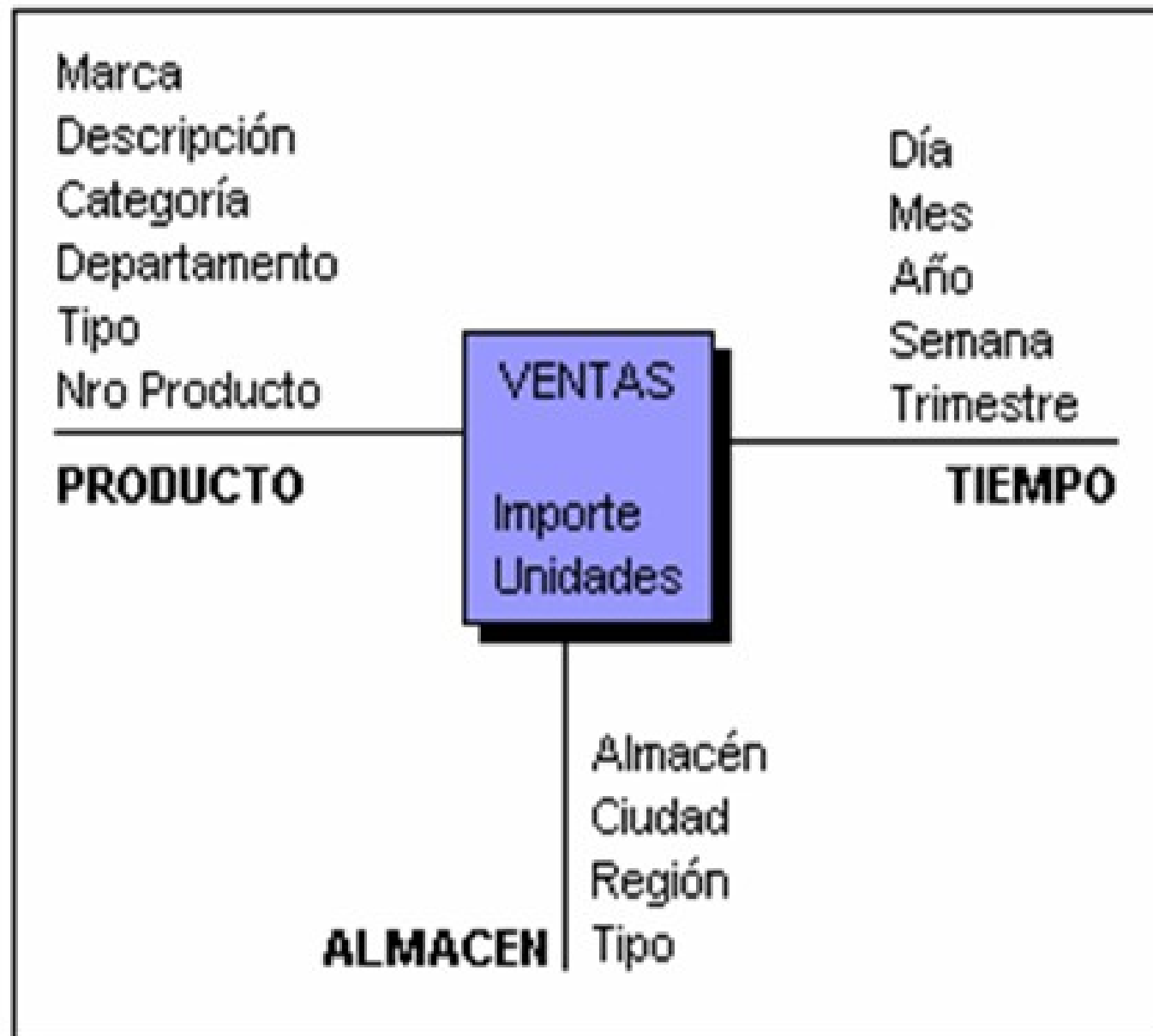
Jerarquías

- Representadas por un **ordenamiento lógico** dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión.
- Como convenciones del modelado, la jerarquía principal se dibuja verticalmente desde el atributo más agregado (arriba) hasta el más atómico (abajo) y las jerarquías características se adicionan por los costados.

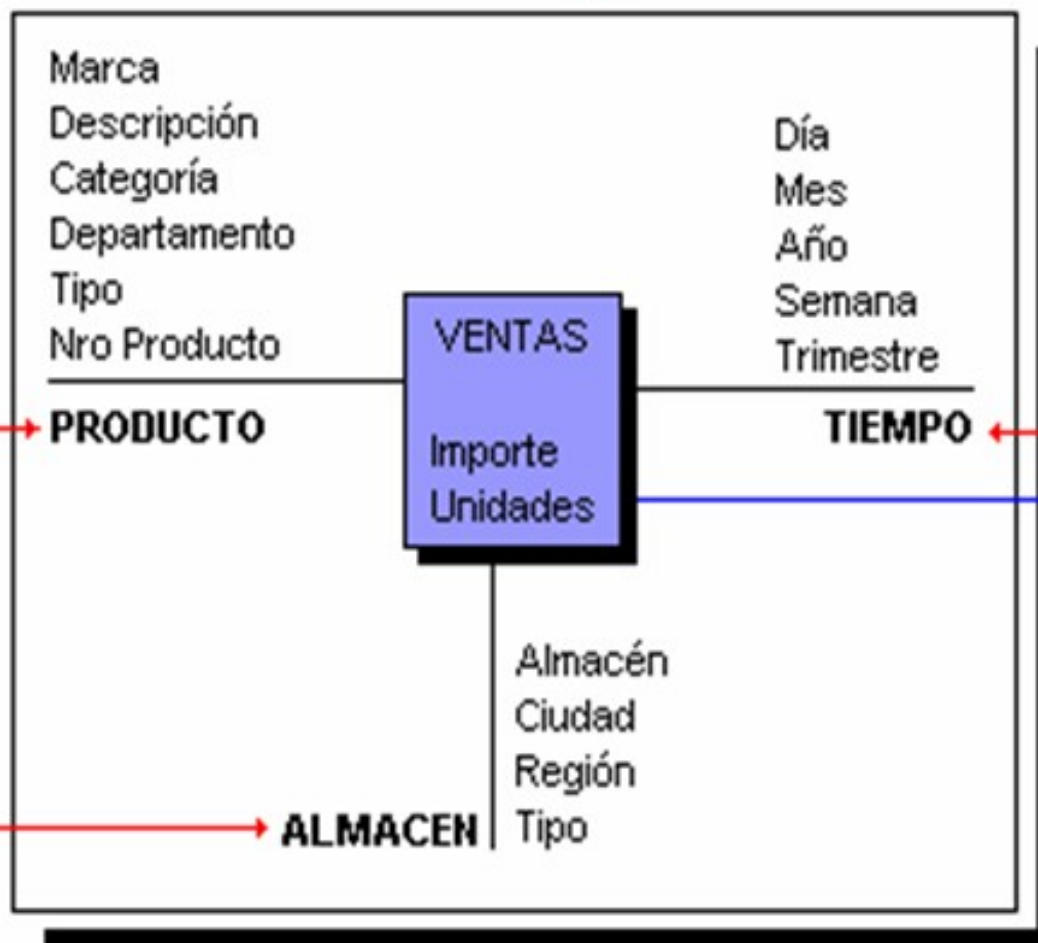
Ejemplo

- **Empresa:** Cadena de supermercados
- **Actividad objeto de análisis:** ventas de productos
- **Información registrada sobre una venta:**
"Se vendió 5 unidades del producto A, en el almacén número 1, el día 20/01/2007 por un total de 250.000 pesos"

Para hacer el análisis no interesa la venta individual realizada por un cliente, si no las ventas diarias de productos en los distintos almacenes de cadena



**Dimensiones (Puntos de Vista)
desde los que se puede analizar la
actividad**



**Actividad que es objeto de
análisis con indicadores
que intenta analizar**

Mitos del modelo dimensional

Modelos dimensionales y data marts

- Solo para datos sumarizados
- Son contruidos por departamentos
- Son no escalables
- No pueden ser integrados

Proceso de diseño dimensional

Pasos

1. Seleccionar el proceso del negocio a modelar
2. Definir el nivel de granularidad del proceso del negocio
3. Escoger las dimensiones que aplican en cada fila de la tabla de hechos
4. Identificar los hechos numéricos que poblaran la tabla de hechos

1. Seleccionar el proceso de negocio a modelar

- El primer modelo dimensional debe ser uno con gran impacto (debe responder la mayoría de preguntas del negocio y estar accesible para la extracción de los datos)
- Decidir el proceso(s) del negocio a modelar combinando y entendiendo los requerimientos del negocio con los datos disponibles

2. Definir el nivel de granularidad del proceso

- Preferiblemente, desarrollar el modelo dimensional con los datos mas atómicos capturados del proceso de negocio
- Datos atómicos son los mas detallados
- Una bodega casi siempre requiere los datos expresados al mas bajo nivel de granuralidad posible en cada dimensión.

3. Escoger las dimensiones

- La granularidad determina la dimensionalidad de la tabla de hechos
- Es posible adicionar mas dimensiones a la granularidad básica de la tabla de hechos:
 - Toman solo un valor bajo cada combinación de las dimensiones
- Si la dimensión adicional viola la granularidad causando mas filas en la tabla de hechos, la granularidad debe ser revisada para acomodar la dimensión.

4. Identificar los hechos

- Determinar las medidas que aparecerán en la tabla de hechos
- La granularidad define el nivel de detalle de las medidas



Errores comunes - prevenir

Error 10: Permanecer “enamorado” con la tecnología y los datos en lugar de enfocarse en los requerimientos y metas del negocio

Errores comunes - prevenir

Error 9: No apadrinarse de un administrador visionario, influyente, accesible y razonable y metas del negocio que cumpla las veces de patrocinador del datawarehouse.

Errores comunes - prevenir

Error 8: Embarcarse en un proyecto muy galáctico y milenario antes que perseguir un proyecto más manejable con procesos iterativos.

Errores comunes - prevenir

Error 7: Asignar energía para construir un modelo de datos normalizado, buscando *ahorrar recursos físicos* en vez de optar por la facilidad de entendimiento que se le puede dar al usuario final.

Errores comunes - prevenir

Error 6: Poner mas atención en el rendimiento operacional y facilidad de desarrollo del “backroom” que en el rendimiento y facilidad de uso del “front-room”

Errores comunes - prevenir

Error 5: Los diseñadores de base de datos que prefieren una presentación mas compleja posteriormente gastan una enorme cantidad de tiempo dando soporte a los usuarios.

Errores comunes - prevenir

Error 4: Poblar los modelos dimensionales sobre una base aislada sin considerar una arquitectura de datos que opere unida, utilizando dimensiones compartidas.

Errores comunes - prevenir

Error 3: Cargar únicamente datos
sumarizados en las estructuras
dimensionales del área de presentación.

Errores comunes - prevenir

Error 2: Presumir que el negocio, sus requerimientos y analíticas, así como los datos subyacentes y la tecnología que lo soportan, son estáticos.

Errores comunes - prevenir

Error 1: Negarse a reconocer que el éxito de un datawarehouse esta ligado directamente a la aceptación de los usuarios.

Caso de estudio

- Cadena de almacenes
- 100 almacenes en cinco (5) regiones
- Cada almacén tiene varios departamentos

Realidad de
los datos



Modelo dimensional

1. Proceso del negocio
2. Granularidad
3. Dimensiones
4. Hechos



Requerimientos
del
negocio

Caso de estudio

1. Seleccionar el proceso de negocio a modelar

Administrador desea

Entender las compras de los clientes desde el POS (punto de venta)




Proceso del negocio

Ventas realizadas en el POS

Analizar que productos se están vendiendo, en que almacenes, en que días, y en que condiciones de promoción

Caso de estudio

2. Definir el nivel de granularidad del proceso del negocio

- ¿Cuál es el nivel de detalle de los datos? 
- Posibles análisis:
 - Diferencia de ventas entre Lunes y Viernes
 - Existencia de productos en ciertos almacenes, ej: cereales
 - Entender por que ciertos compradores tomaron la promoción del “shampoo”

“Requieren datos detallados en diferente forma”

“No es posible responder con datos sumariados”

«A nivel de producto»

Caso de estudio

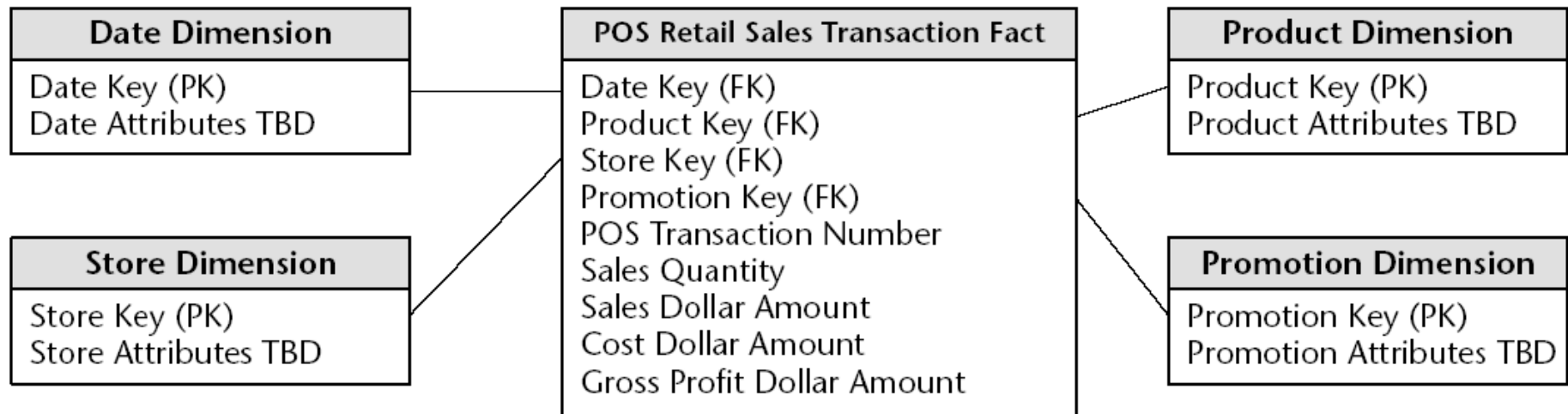
3. Escoger las dimensiones

- La fecha, producto y almacén (dimensiones primarias)
- ¿otras dimensiones?
 - Promoción
- ¿Nueva dimensión viola la granularidad? (nuevas filas en la tabla de hechos)



Caso de estudio

Modelo preliminar



Caso de estudio

4. Identificar los hechos

- ¿Qué más debe aparecer en la tabla de hechos?
- Hechos que correspondan con la granularidad: Transacción de venta

- Precio por unidad

- Cantidad vendida

- Venta total

- Costo del producto

- Calcular la ganancia ($\text{Venta total} - \text{Costo total}$)

aditivas

Dilema: ¿Almacenar calculado o calcular con herramienta de consulta?

Caso de estudio

4. Identificar los hechos

- Porcentaje de ganancia (calculada al dividir la ganancia por el total de la venta)

No puede ser sumariada por ser **no aditiva** con respecto a la dimensión!

Los porcentajes y proporciones no son aditivos. El numerador y el denominador deben ser almacenados en la tabla de hechos. La proporción puede ser calculado a tiempo de consulta de acuerdo con la situación.

“La proporción de las sumas != La suma de las proporciones”

Caso de estudio

4. Identificar los hechos

- Estimar el número de filas que se almacenarán en la tabla de hechos
 - Caso de estudio: Hablar con la persona experta para conocer cuantas transacciones de venta pueden ocurrir en un periodo de tiempo
- Estimar el número de filas que se almacenarán anualmente a la tabla de hechos



Determinar si es razonable

Atributos de las dimensiones

La selección de los atributos de cada dimensión es un paso muy importante ya que este permitirá a los analistas responder muchas mas preguntas del tipo ¿Por qué?

Dimensión **Tiempo**

- Está presente en casi todos los esquemas multidimensionales.
- Es la dimensión esencial.
- Casi siempre se usa como el primer criterio de un ordenamiento y/o filtro



Dimensión **Tiempo**

- Por qué usar una dimensión de tiempo si con el tipo de dato Datetime y las funciones de SQL puedo obtener diferentes atributos?
- Por que hace más fácil de entender y manejar el modelo.
- Depende del negocio se presentar atributos particulares por ejemplo:
 - Calendarios fiscales diferentes.
 - Días laborales o festivos

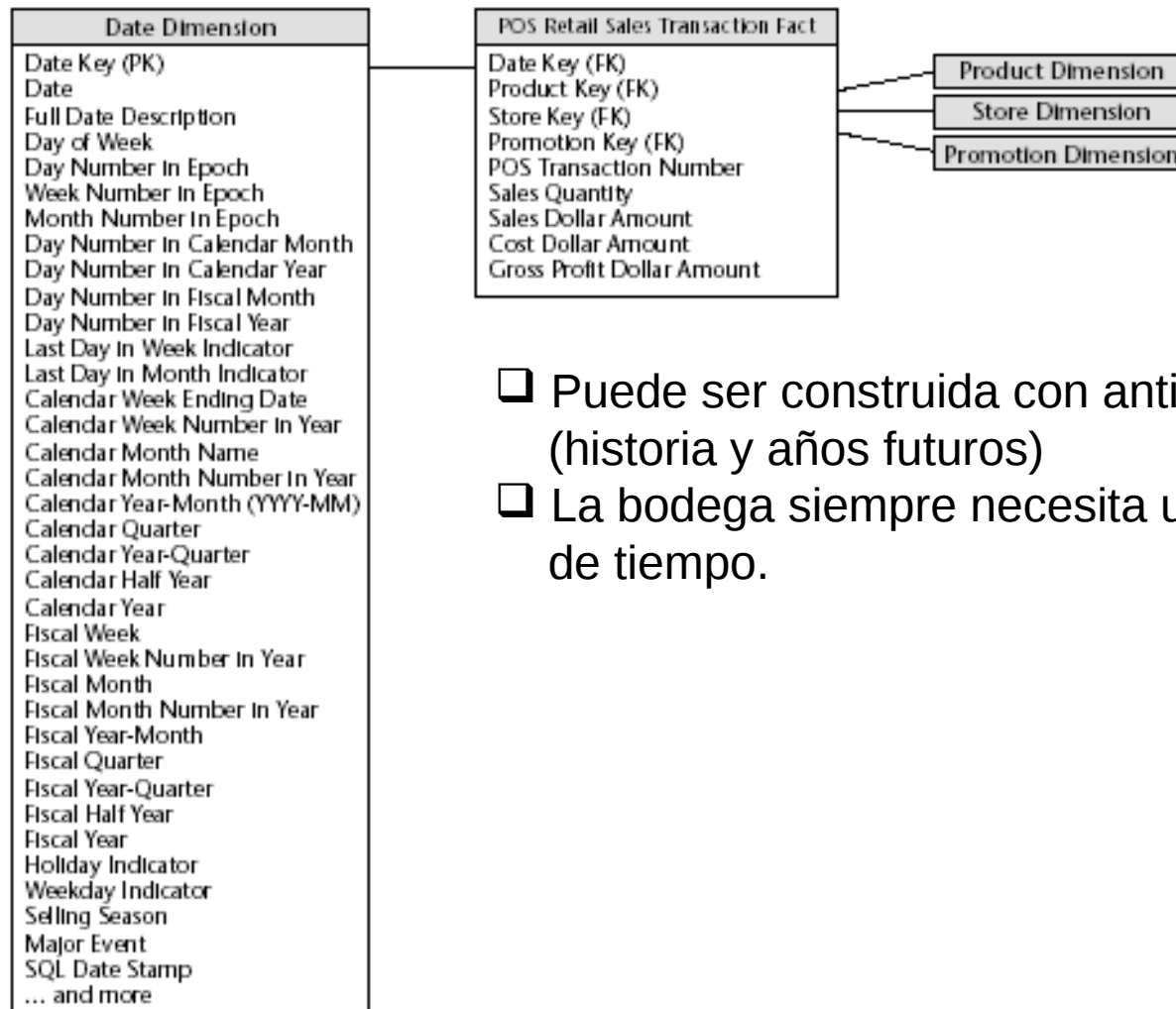
Dimensión **Tiempo**

Date Dimension
Date Key (PK)
Date
Full Date Description
Day of Week
Day Number in Epoch
Week Number in Epoch
Month Number in Epoch
Day Number in Calendar Month
Day Number in Calendar Year
Day Number in Fiscal Month
Day Number in Fiscal Year
Last Day in Week Indicator
Last Day in Month Indicator
Calendar Week Ending Date
Calendar Week Number in Year
Calendar Month Name
Calendar Month Number in Year
Calendar Year-Month (YYYY-MM)
Calendar Quarter
Calendar Year-Quarter
Calendar Half Year
Calendar Year
Fiscal Week
Fiscal Week Number in Year
Fiscal Month
Fiscal Month Number in Year
Fiscal Year-Month
Fiscal Quarter
Fiscal Year-Quarter
Fiscal Half Year
Fiscal Year
Holiday Indicator
Weekday Indicator
Selling Season
Major Event
SQL Date Stamp
... and more

Caso de estudio

Atributos de las entidades

- Dimensión Tiempo



- ❑ Puede ser construida con anticipación (historia y años futuros)
- ❑ La bodega siempre necesita una dimensión de tiempo.

Caso de estudio

Atributos de las entidades- Dimensión Tiempo

- La Dimensión Tiempo en detalle

Date Key	Date	Full Date Description	Day of Week	Calendar Month	Calendar Year	Fiscal Year-Month	Holiday Indicator	Weekday Indicator
1	01/01/2002	January 1, 2002	Tuesday	January	2002	F2002-01	Holiday	Weekday
2	01/02/2002	January 2, 2002	Wednesday	January	2002	F2002-01	Non-Holiday	Weekday
3	01/03/2002	January 3, 2002	Thursday	January	2002	F2002-01	Non-Holiday	Weekday
4	01/04/2002	January 4, 2002	Friday	January	2002	F2002-01	Non-Holiday	Weekday
5	01/05/2002	January 5, 2002	Saturday	January	2002	F2002-01	Non-Holiday	Weekend
6	01/06/2002	January 6, 2002	Sunday	January	2002	F2002-01	Non-Holiday	Weekend
7	01/07/2002	January 7, 2002	Monday	January	2002	F2002-01	Non-Holiday	Weekday
8	01/08/2002	January 8, 2002	Tuesday	January	2002	F2002-01	Non-Holiday	Weekday

Caso de estudio

Atributos de las entidades- Dimensión Tiempo

- Si se desea acceder al tiempo de la transacción por partes del día (Ej: tardes)
 - Conviene una Dimensión por horas
 - Es mejor tener las mismas 3.650 filas y una dimensión con 1.440 filas correspondientes a los minutos del día
 - Si se maneja en una dimensión a nivel de Fecha y hora (HH:MM), se tendría una dimensión con demasiados registros:
 $365 \text{ días} \times 10 \text{ años} \times 24 \text{ horas} \times 60 \text{ minutos}$
 $= 5.256.000 \text{ filas}$

Caso de estudio

Dimensión **Producto**

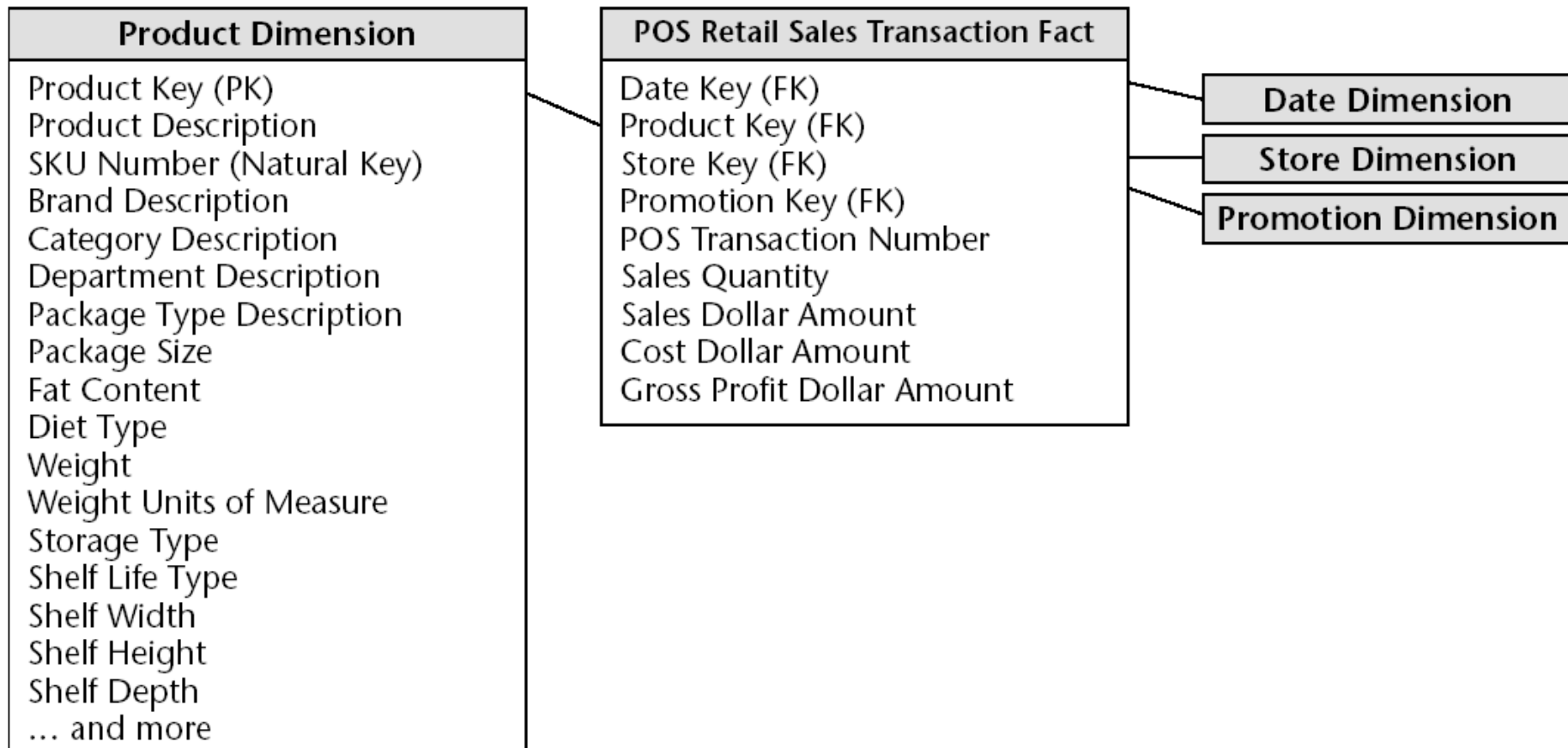
Su definición debe ser lo mas amplia posible y sus atributos van a dar respuestas a las diferentes áreas funcionales de la empresa.

- Ejemplo: Compras se preguntará por las marcas mas Vendidas
- Logística por la presentación más vendida
- Mercadeo por el color de los productos más vendidos.

Caso de estudio

Atributos de las entidades

■ Dimensión Producto



Caso de estudio

Atributos de las entidades- dimensión producto

Product Key	Product Description	Brand Description	Category Description	Department Description	Fat Content
1	Baked Well Light Sourdough Fresh Bread	Baked Well	Bread	Bakery	Reduced Fat
2	Fluffy Sliced Whole Wheat	Fluffy	Bread	Bakery	Regular Fat
3	Fluffy Light Sliced Whole Wheat	Fluffy	Bread	Bakery	Reduced Fat
4	Fat Free Mini Cinnamon Rolls	Light	Sweeten Bread	Bakery	Non-Fat
5	Diet Lovers Vanilla 2 Gallon	Coldpack	Frozen Desserts	Frozen Foods	Non-Fat
6	Light and Creamy Butter Pecan 1 Pint	Freshlike	Frozen Desserts	Frozen Foods	Reduced Fat
7	Chocolate Lovers 1/2 Gallon	Frigid	Frozen Desserts	Frozen Foods	Regular Fat
8	Strawberry Ice Creamy 1 Pint	Icy	Frozen Desserts	Frozen Foods	Regular Fat
9	Icy Ice Cream Sandwiches	Icy	Frozen Desserts	Frozen Foods	Regular Fat

Caso de estudio

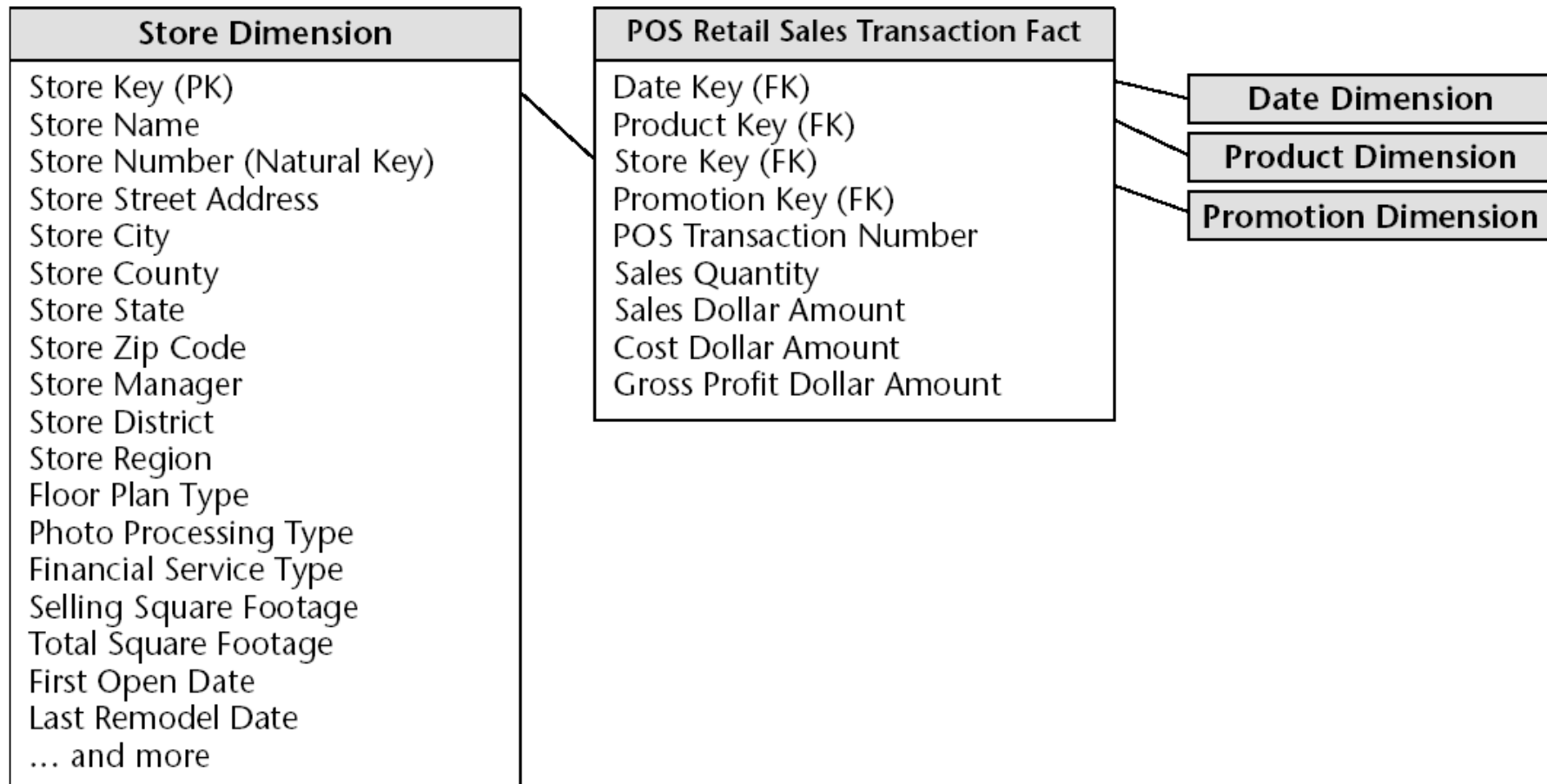
Dimensión Almacén

- Describe cada almacén de la cadena del retail.
- Posiblemente no existe un archivo maestro de tiendas. Por lo tanto en ocasiones debe construirse manualmente.
- Contiene atributos geográficos de análisis

Caso de estudio

Atributos de las entidades

- Dimensión Almacén



Caso de estudio

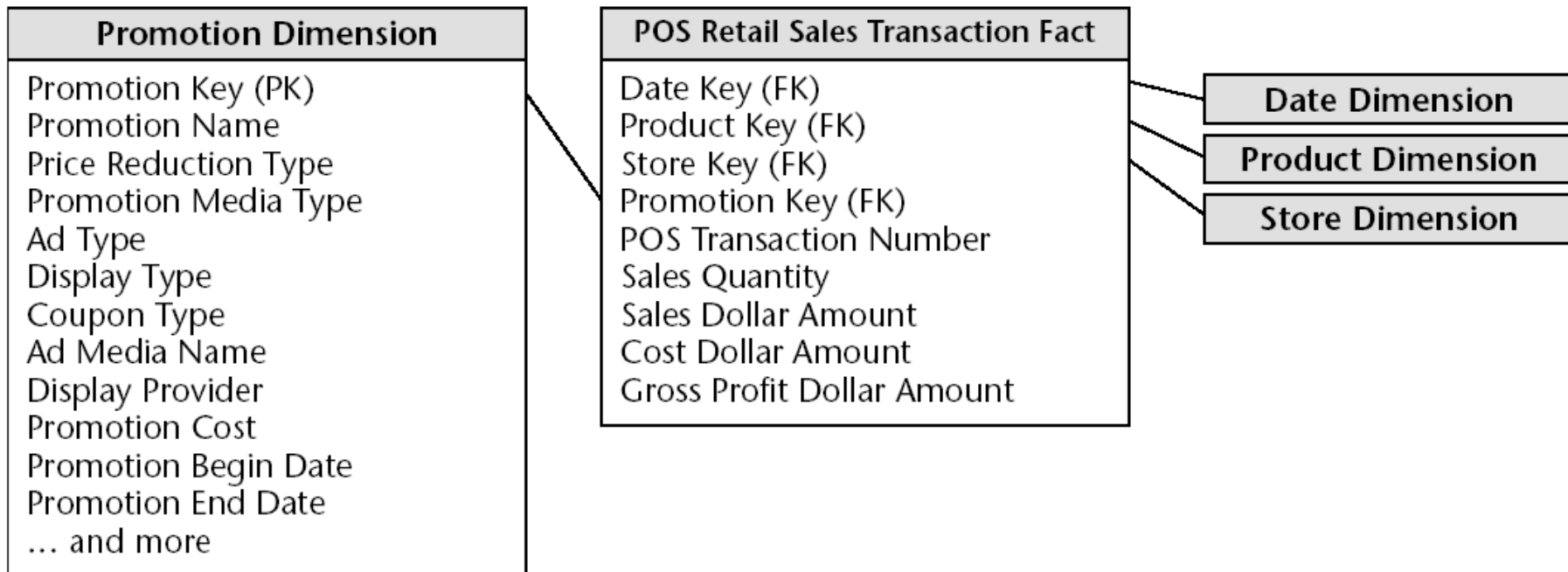
Dimensión Promoción

- Describe las condiciones de la promoción bajo las cuales se ha vendido un producto.
- Se clasifica como una dimensión casual.

Caso de estudio

Atributos de las entidades

- Dimensión promoción



Caso de estudio

Atributos de las entidades

- **Dimensión degenerada**

- Número de transacción del POS

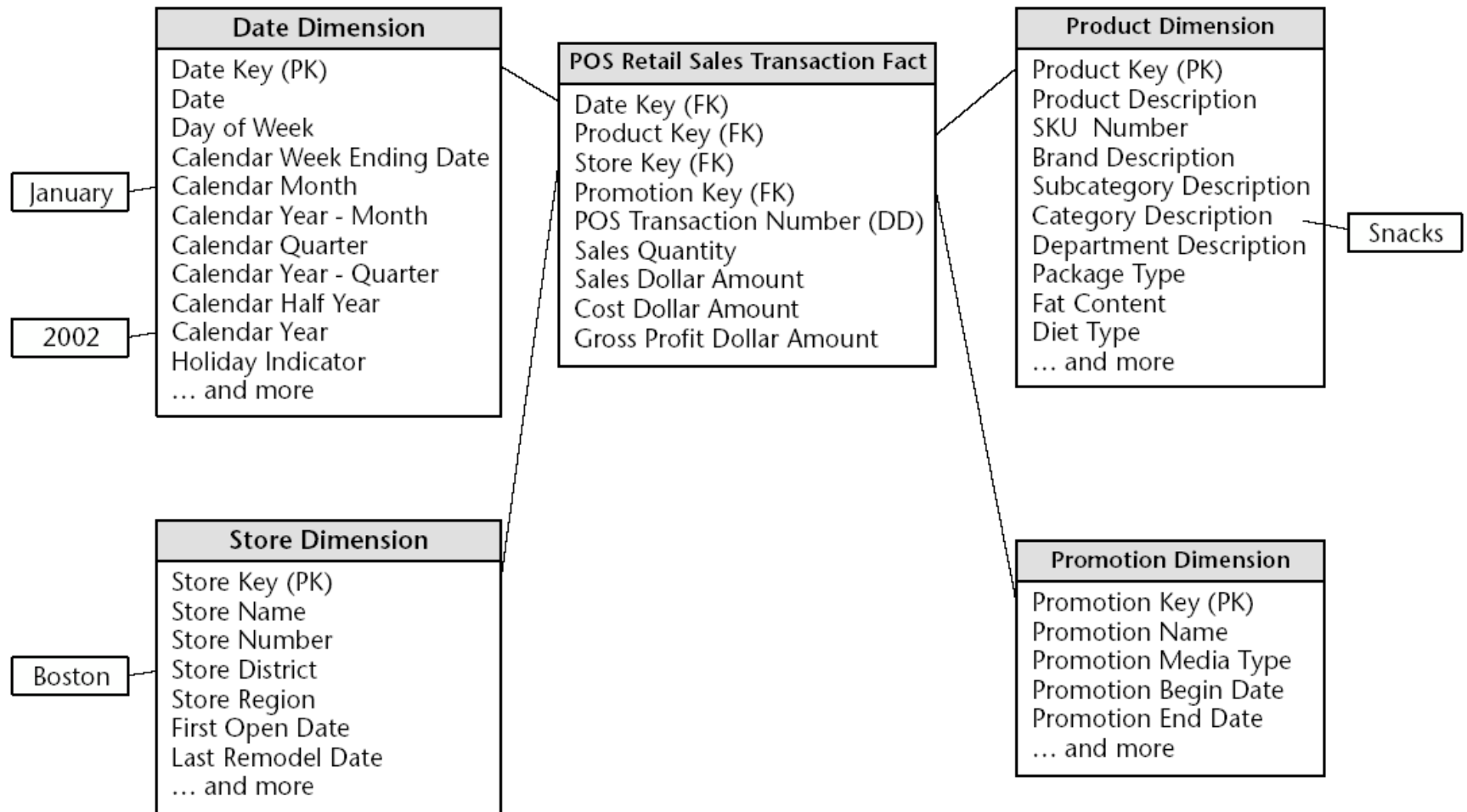
POS Retail Sales Transaction Fact	
Date Key (FK)	
Product Key (FK)	
Store Key (FK)	
Promotion Key (FK)	
POS Transaction Number	
Sales Quantity	
Sales Dollar Amount	
Cost Dollar Amount	
Gross Profit Dollar Amount	

Las dimensiones degeneradas son comunes cuando la granularidad de la tabla de hechos representa una simple transacción (Ej: número de ordenes, número de recibos, número de la factura, etc)

Algunas veces la dimensión degenerada es llave primaria de la tabla de hechos

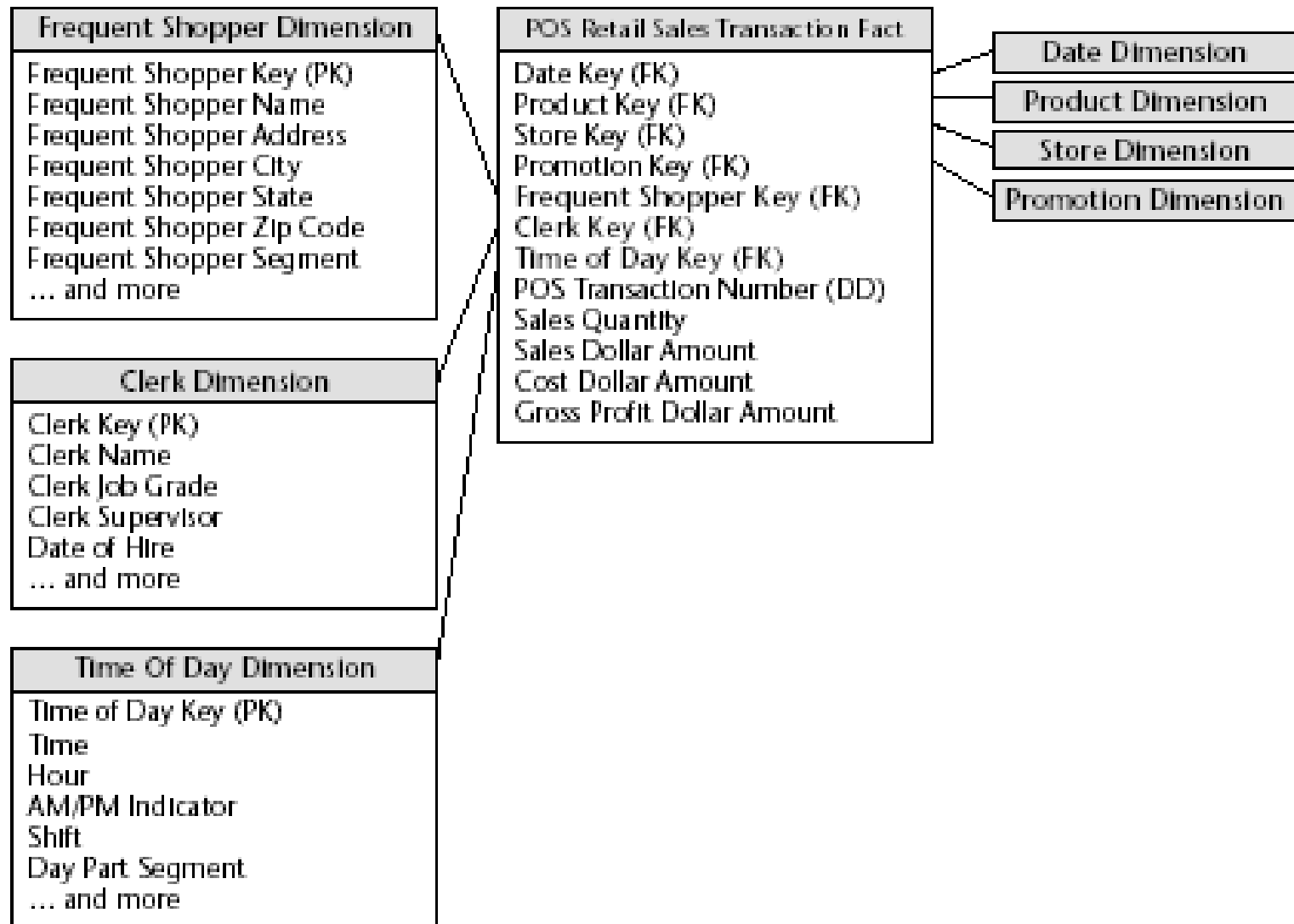
Caso de estudio

Modelo



Caso de estudio

Modelo extendido



Bodegas de Datos

Modelos de Arquitectura

Dos arquitecturas de acuerdo con la normalización de sus dimensiones:

- Estrella
 - Desnormalizado
- Copo de Nieve
 - Normalizado

Estrella vs Copo de nieve

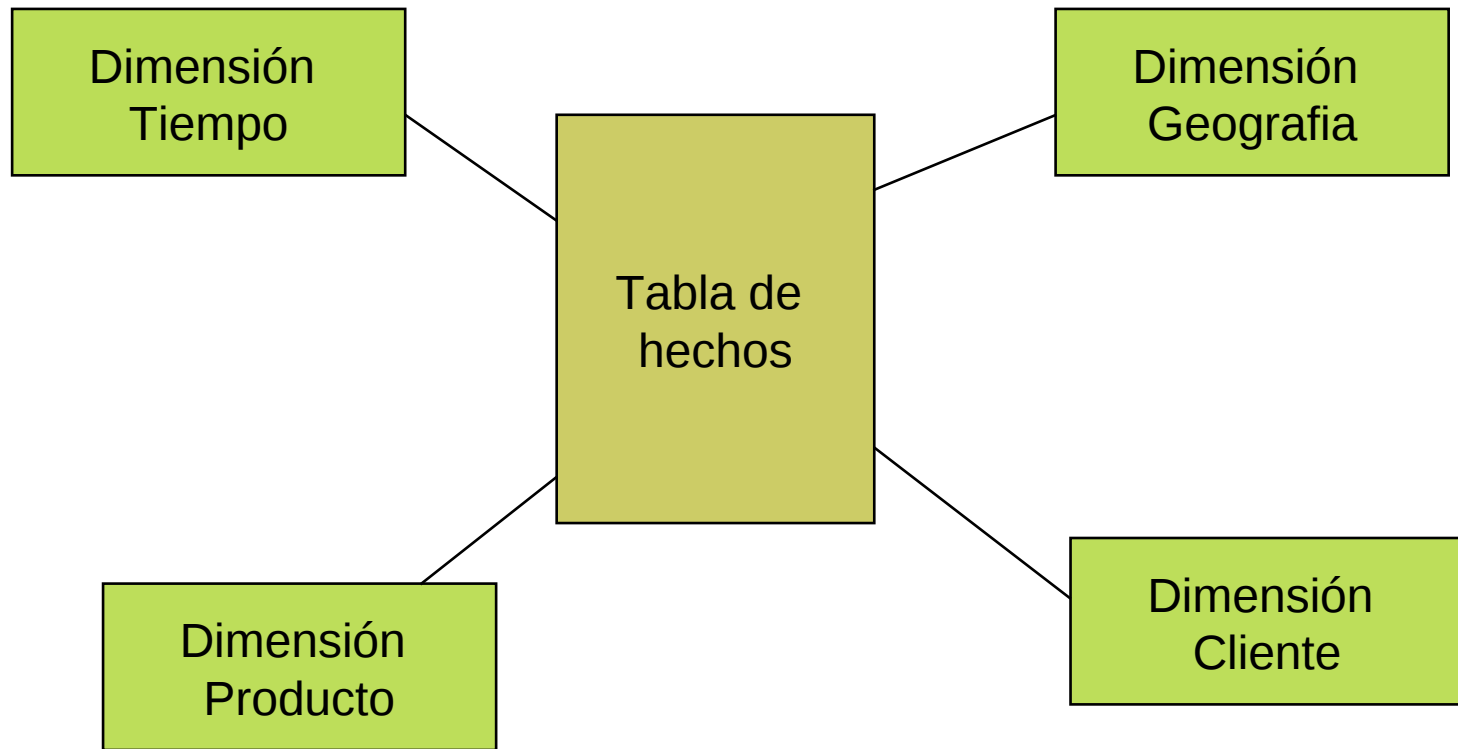
- **Estrella**

- Desnormalizado
- Habilidad para análisis dimensional

- **Copo de nieve**

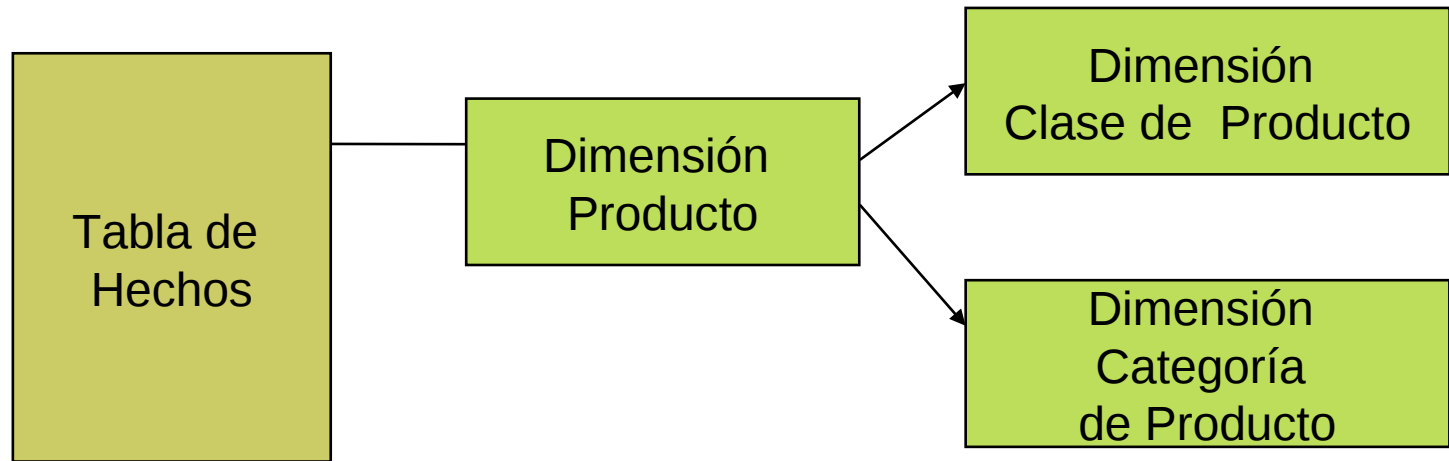
- Variación del modelo estrella
- Forma normalizada de las dimensiones (solo las dimensiones primarias están enlazadas con la tabla de hechos)
- Rompe el análisis dimensional
- Se usa cuando no se puede implementar un modelo estrella

Modelo Estrella



Modelo Copo de nieve

- Dimensiones desnormalizadas (valores repetidos) puede llevar a la normalización de una dimensión

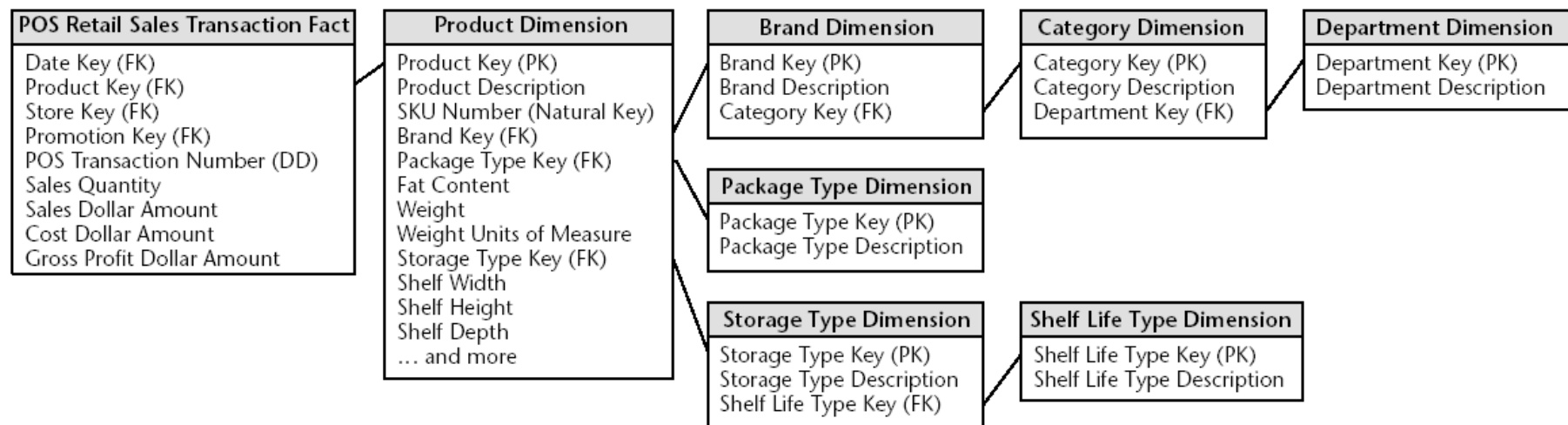


Modelo Copo de nieve

- Atributos redundantes son removidos de la entidad y creando una segunda dimensión.
- Reduce espacio
- Fácil de mantener (si la descripción de un departamento cambia, solo se necesita actualizar una ocurrencia y no alrededor de 3.000)
- Extensión del modelo dimensional (válido)

Caso de estudio

Dimensión producto (parcial) en copo de nieve



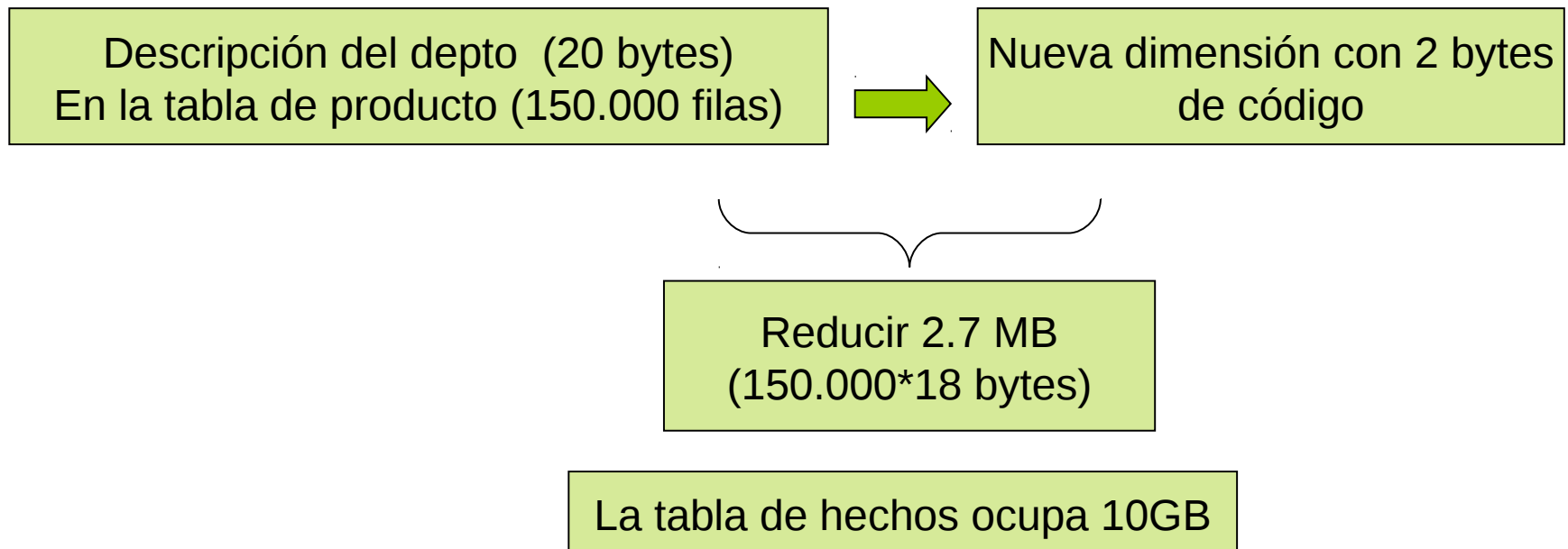
Normalizado en 3FN

Modelo Copo de nieve

- En el diseño del modelo se debe tener en cuenta: **uso** y **desempeño**
- Consideraciones de este modelo
 - Múltiples tablas aumentan la complejidad de uso
 - Mas tablas y joins afectan el desempeño de las consultas
 - Navegar a través de las dimensiones puede ser más lento (cruce de dimensiones)
 - No se pueden aprovechar los índices Bitmap que son útiles para indexar campos con baja cardinalidad

Modelo Copo de nieve

- El espacio salvado en las dimensiones es insignificante comparado con el tamaño de la tabla de hechos.



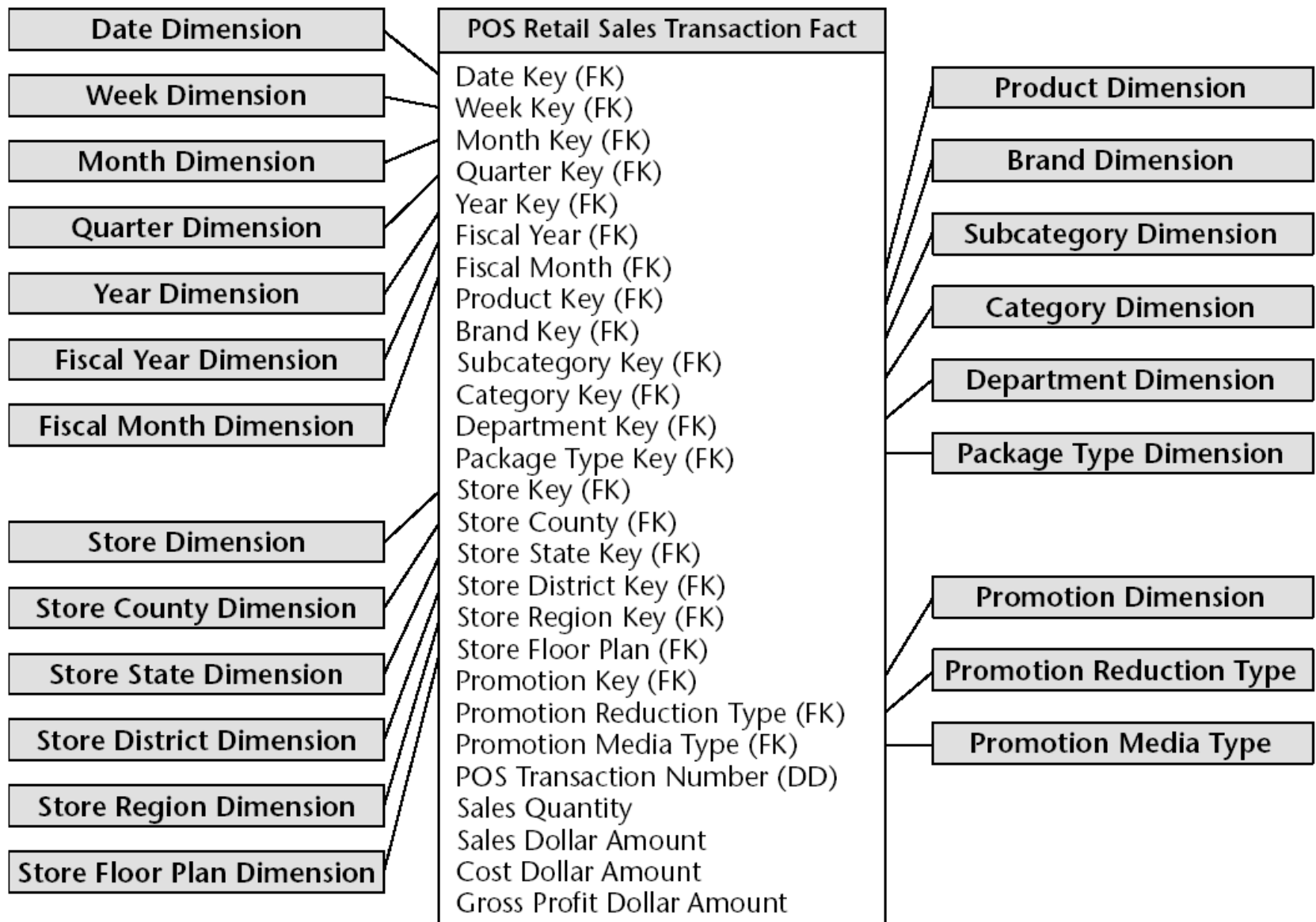
Pérdida de tiempo tratar de normalizar la dimensión!

Muchas dimensiones

- Una tabla de hechos es normalizada y compacta. Tentación de desnormalizar
- Una tabla de hechos con muchas dimensiones lleva a un crecimiento significativo de requerimientos de espacio
- La mayoría de procesos del negocio pueden ser representados con menos de 15 dimensiones


Muchas dimensiones (2)

- Un número largo de dimensiones es signo de que varias dimensiones no son completamente independientes y pueden ser combinadas en una sola dimensión
- Si el negocio tiene mas de 25 dimensiones, se debe combinar dimensiones correlacionadas en una sola dimensión
- Es un error representar elementos de una jerarquía como dimensiones separadas



Llaves subordinadas(1)

- Llaves artificiales, llaves sintéticas, llaves enteras
- Se recomienda el uso de estas llaves
- Son enteros que son asignados secuencialmente al poblar una dimensión



empleados		
1		
2		
3		

Llaves subordinadas(2)

- Cada *join* entre dimensiones y tabla de hechos debe estar basado en llaves subordinadas.
- Evitar el uso de códigos naturales de los sistemas que soportan la operación
- Beneficios:
 - Mantener flexibilidad al cambio.
 - Permite integración
 - Ventajas de rendimiento computacional (pequeñas)

Análisis de la canasta de mercado

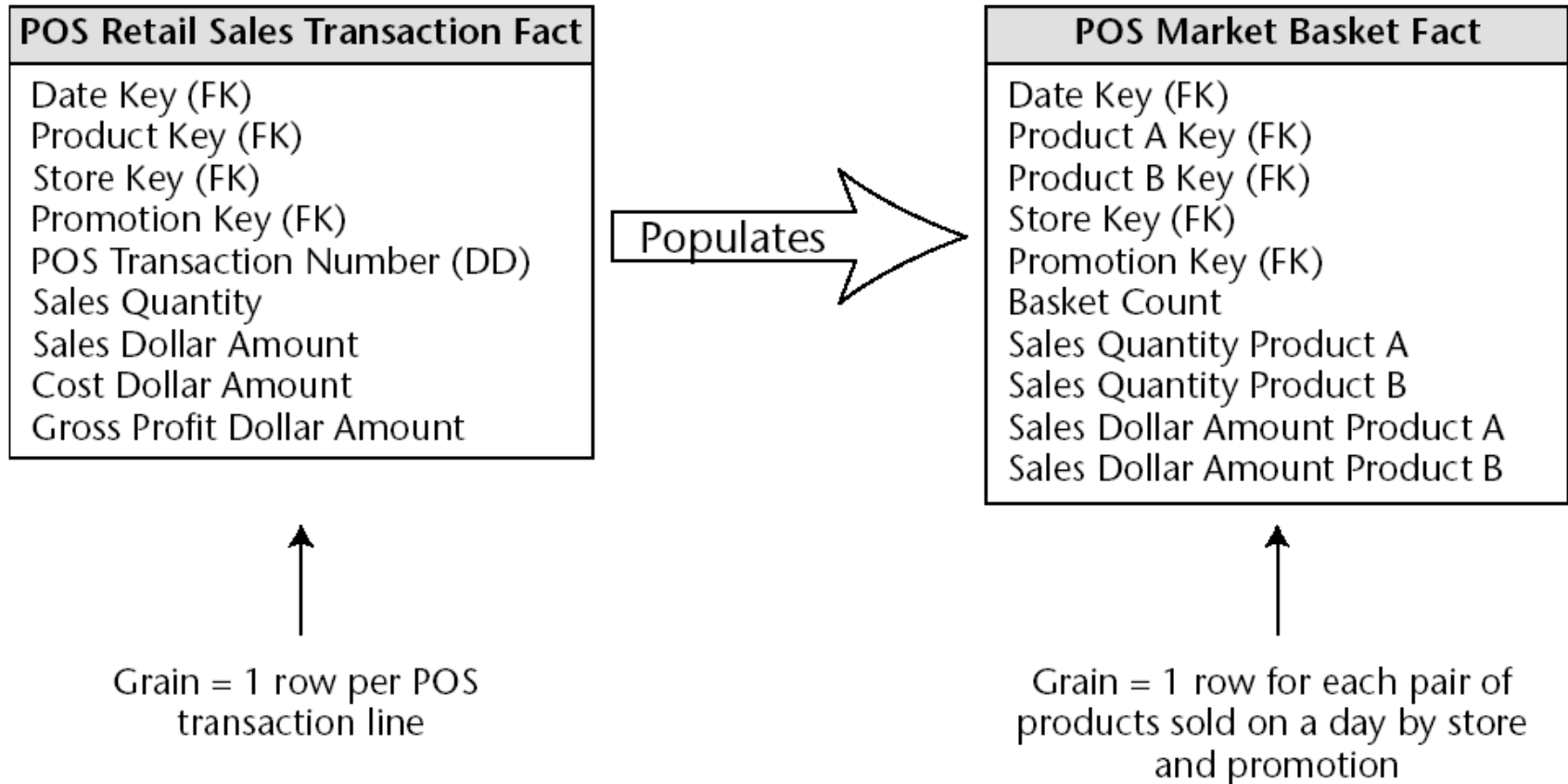
- Analizar que productos son vendidos en la misma canasta de mercado!

Grado de afinidad de los productos: “pañales son vendidos con cerveza”

- OLAP y minería de datos pueden ser usados
- Se puede tener una solución a nivel de bodega!

Caso de estudio

Análisis de la canasta de mercado



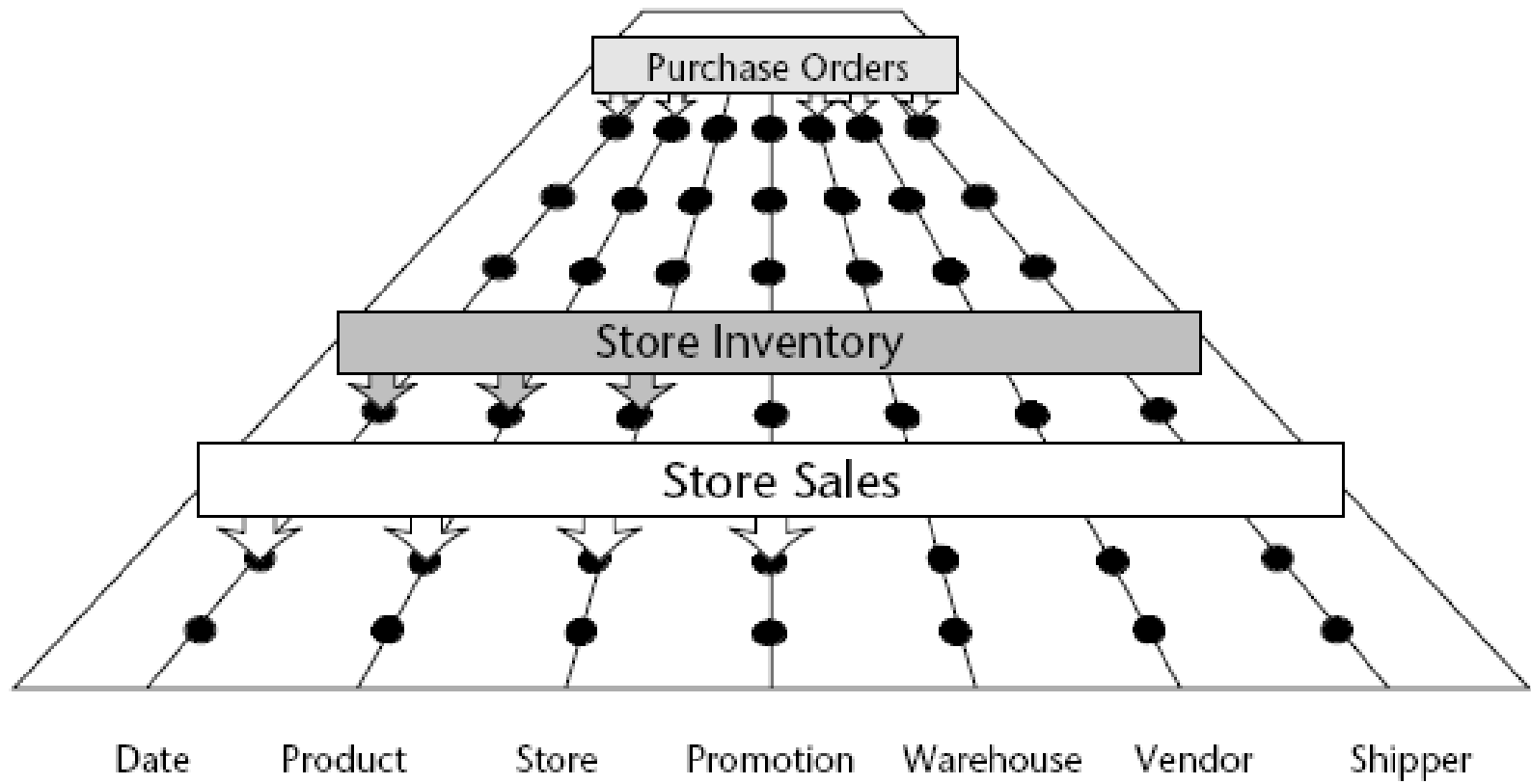
Arquitectura en Bus

- Construir la bodega en un solo paso?
 - Usar una arquitectura incremental
 - Arquitectura de bus!

Los Data Marts son contruidos por tema de negocio en tiempos diferentes o con grupos diferentes.

Los Data Marts pueden ser “conectados” y ser usados consistentemente si se adhieren al estándar

Arquitectura en Bus



Arquitectura en Bus

- Permite una mejora en la administración de los componentes de la bodega.
- Basado en el principio reduccionista de control sobre cada uno de los elementos.
- Los elementos de desarrollo de los Data Mart siguen las pautas de la arquitectura
- Fácil incorporación de nuevos Data Marts a la arquitectura (**plug and play**)

Bus Matrix

- Herramienta para documentar y comunicar la arquitectura de bus.

		COMMON DIMENSIONS							
BUSINESS PROCESSES		Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Data marts	Retail Sales	X	X	X	X				
	Retail Inventory	X	X	X					
	Retail Deliveries	X	X	X					
	Warehouse Inventory	X	X			X	X		
	Warehouse Deliveries	X	X			X	X		
	Purchase Orders	X	X			X	X	X	X

Bus Matrix

1. Lista de **data marts de primer nivel**

- ❑ Derivados de una fuente
- ❑ Minimizar el riesgo en implementaciones ambiciosas (ETL, diseño e implementación)
- ❑ Suficientes datos interesantes para los usuarios, mientras se continua trabajando en implementaciones más difíciles

1. Identificar **data marts consolidados**

- ❑ Con múltiples fuentes
- ❑ Más difíciles de implementar

Dimensiones que cambian lentamente

- La dimensión puede cambiar con el tiempo (no estática): cambian atributos
- Inaceptable colocar todo en la tabla de hechos o crear dimensiones dependientes

Si el valor de un atributo cambia en el sistema operacional

¿Cómo responder a ese cambio en el modelo dimensional?

Dimensiones que cambian lentamente

1. Sobrescribir el valor

12345

NiñoInte1.0

Software

ABC922-Z

Se decide que NiñoInte1.0 debe ser retirado del depto de software y adicionado al depto de estrategia en enero 5 del 2006

12345

NiñoInte1.0

Estrategia

ABC922-Z

- El atributo siempre refleja el nuevo valor
- La historia mostrará que NiñoInte1.0 siempre ha pertenecido al depto de estrategia

“Es fácil de implementar, pero no mantiene la historia de los cambios los atributos”

Dimensiones que cambian lentamente

1. Adicionar una fila a la dimensión

12345	NiñoInte1.0	Educación	ABC922-Z
35896	NiñoInte1.0	Estrategía	ABC922-Z

No puede
ser llave
primaria

Llave
subordinada

- Representa la historia correctamente
- En la tabla de hechos, antes de enero 5 del 2006 se referencia el articulo con llave primaria 1234, después de esta fecha el articulo tendrá la llave 35896

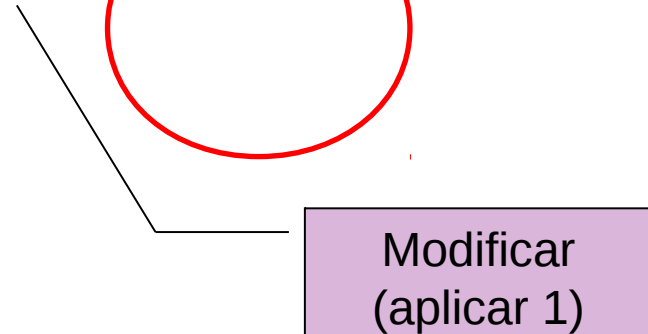
“Es poderosa ya que la nueva fila divide automáticamente la historia en la tabla de hechos”

“Sin embargo no permite asociar el nuevo valor del atributo con la historia de los hechos y viceversa”

Dimensiones que cambian lentamente

1. Adicionar una columna a la dimensión

12345 NiñoInte1.0 Estrategia Educación ABC922-Z



- Es apropiado cuando se necesita soportar dos vistas del mundo al mismo tiempo
- Es inapropiado si se desea hacer seguimiento de valores intermedios de los atributos
- Menos usada

Dimensiones que cambian lentamente

□ Técnicas Híbridas

- Más flexibles pero más complejas
- Combinan las técnicas simples
- Usarlas si son necesarias para cumplir con los requerimientos

- 1. Cambios predecibles con versiones múltiples superpuestas**
- 2. Cambios no predecibles con única versión superpuesta**

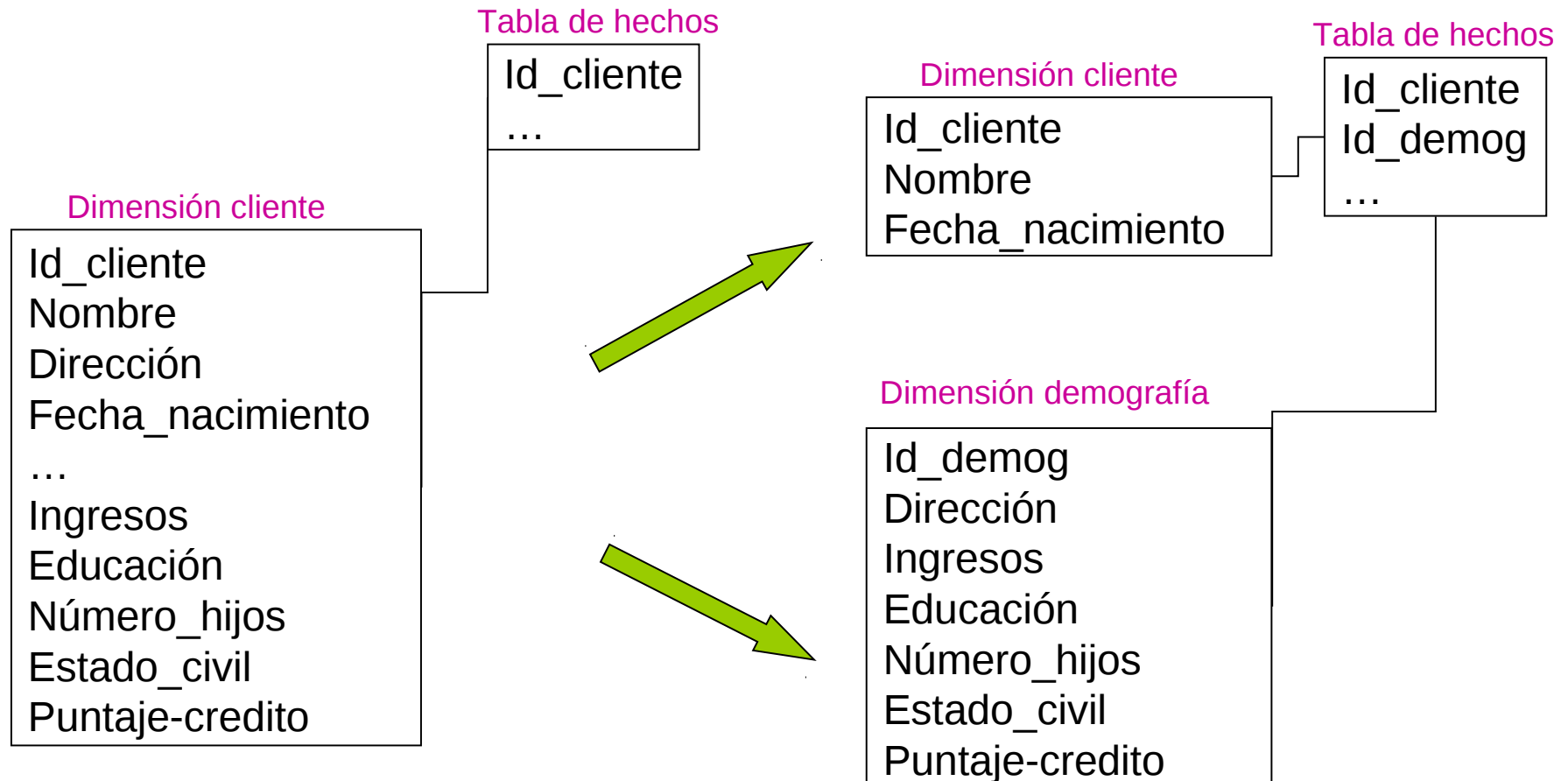
Dimensiones que cambian rápidamente

- Mensualmente
- Separar los atributos que cambian en una o más dimensiones separadas
 - La tabla de hechos tendrá dos llaves foráneas

Dimensiones que cambian rápidamente - Ejemplo

Compañías de seguros

Actualización de los clientes periódicamente!



Dimensiones “Junk”

- Banderas, indicadores, atributos que no son usados frecuentemente (parecen no organizados coherentemente!)

No se debe:

- ▣ Dejarlos en la tabla de hechos
 - ▣ Colocar cada uno en su propia dimensión
- Estudiarlos cuidadosamente y colocarlos en una o más “dimensiones junk”

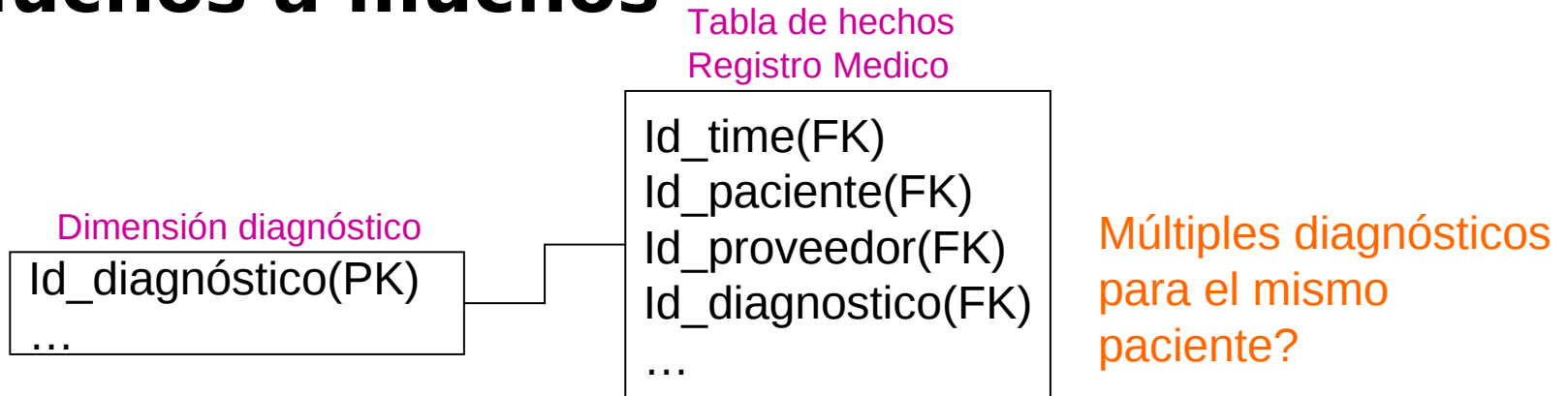
Dimensiones “Junk”

- Ejemplo: 10 indicadores Si/No
 - 2^{10} registros
- Encuestas con preguntas de si y no.

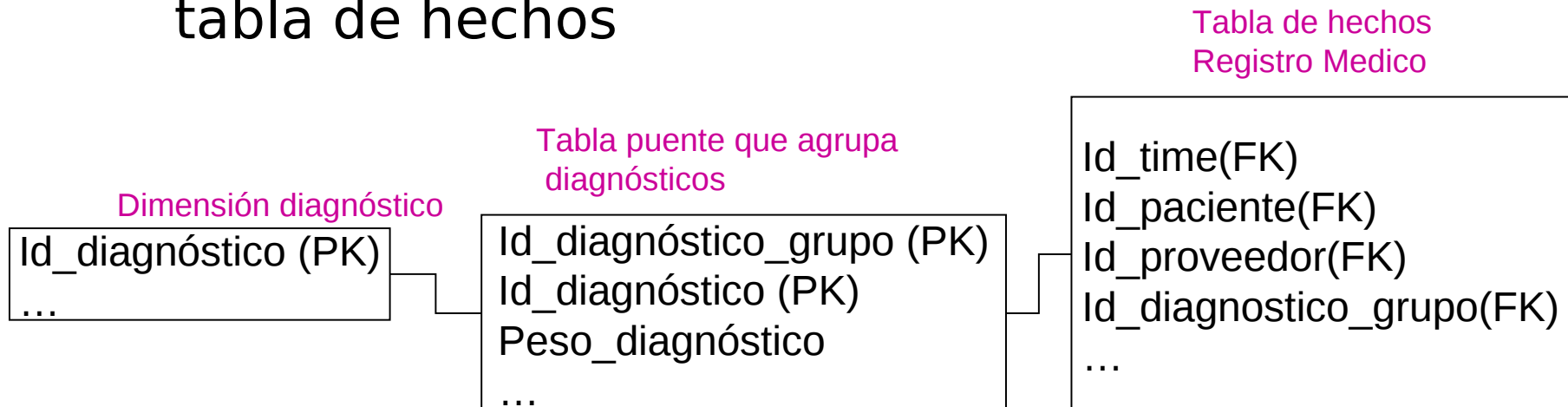
Una dimensión junk compone un grupo de banderas y atributos de texto removidos de la tabla de hechos

Tablas de dimensiones extendidas

- Muchos a muchos



- Crear una tabla puente entre la dimensión y la tabla de hechos



Tablas de dimensiones extendidas

- Con el tiempo el paciente puede tener varios grupos de diagnóstico



Ejercicio

- Cuentas de balance. Una cuenta puede pertenecer a más de una persona.
 - Tabla de hechos Balance mensual de la cuenta

Dimensiones Role-playing

- Una sola dimensión aparece muchas veces en la tabla de hechos

Tabla de hechos

fecha_orden (FK)
fecha_de_empaque (FK)
fecha_de_compra (FK)
fecha_de_envio (FK)
fecha_de_pago (FK)
fecha_de_devolución (FK)

Todos los atributos
son FK a la dimensión
fecha

-
- Construir la matrix
 - Lista de data marts
 - Lista de dimensiones
 - Usar el método de 4 pasos para cada tabla de hechos (data mart)
 - Descripción detallada de los atributos de las dimensiones
 - Hoja de datos derivados o calculados
 - Definiciones de los datos fuente
 - Mapa de fuente-destino (entrada al ETL)

Referencias

- [1] Wiley - Mastering Data Warehouse Design - Relational And Dimensional Techniques - 2003.
- [2] Wiley - Data Analysis -The Data Warehouse Toolkit - Second Edition.
- [3] Wiley - Building The Data Warehouse - Third Edition
- [4] Wiley - The Data Warehouse ETL Toolkit -2005.
- [5] Wiley - The Data Warehouse Lifecycle Toolkit 1998
- [6] MicroStrategy - Business Intelligence - 2006