

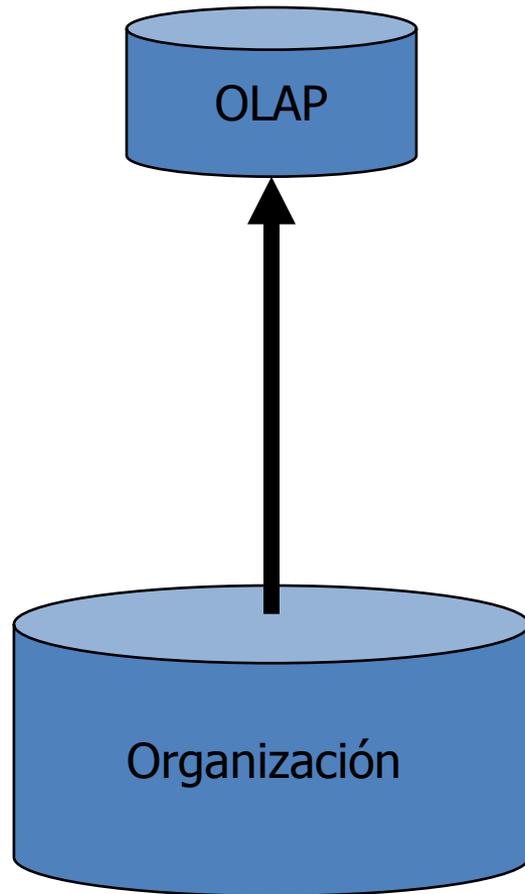


# Datawarehousing: Modelado de Datos y OLP

Jose Aguilar  
CEMISID, Escuela de Sistemas  
Facultad de Ingeniería  
Universidad de Los Andes  
Mérida, Venezuela

# Modelado de Datos y de Información

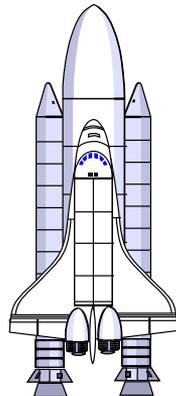
## Usuarios tienen diferentes vistas de los datos



**Turistas:** Navegar por la información recolectada



**Agricultores:** información de los caminos a los datos, etc.



**Exploradores:** Busca cosas desconocidos que se esconden en los datos detallados

# Consideraciones para el Diseño Data warehouse

- Al construir un Data Warehouse se necesitan herramientas para ayudar a **la migración y a la transformación de los datos** hacia el almacén.
- Ya construido, se requieren **medios para manejar grandes volúmenes** de información.
- Se **diseña su arquitectura** dependiendo de la estructura interna de los datos del almacén, y especialmente del **tipo de consultas a realizar**.

# Consideraciones para el Diseño Data warehouse

Para abordar un proyecto de data warehouse es necesario hacer un estudio de algunos temas generales de la organización:

- **Situación actual.**- Cualquier solución propuesta de data warehouse debe estar muy orientada por las necesidades del negocio, debe ser compatible con la arquitectura técnica existente y planeada de la compañía.
- **Tipo y características del negocio.**- Tener el conocimiento exacto sobre el tipo de negocios de la organización y el soporte que representa la información dentro de todo su proceso de toma de decisiones.

# Consideraciones para el Diseño Data warehouse

Para abordar un proyecto de data warehouse es necesario hacer un estudio de algunos temas generales de la organización:

- **Entorno técnico.**- hardware (servidores, redes,...) así como aplicaciones y herramientas. Se dará énfasis a los Sistemas de Soporte a Decisiones (DSS).
- **Expectativas de los usuarios.**- Es una forma de vida de las organizaciones y como tal, tiene que contar con el apoyo de todos los usuarios y su convencimiento sobre su bondad.

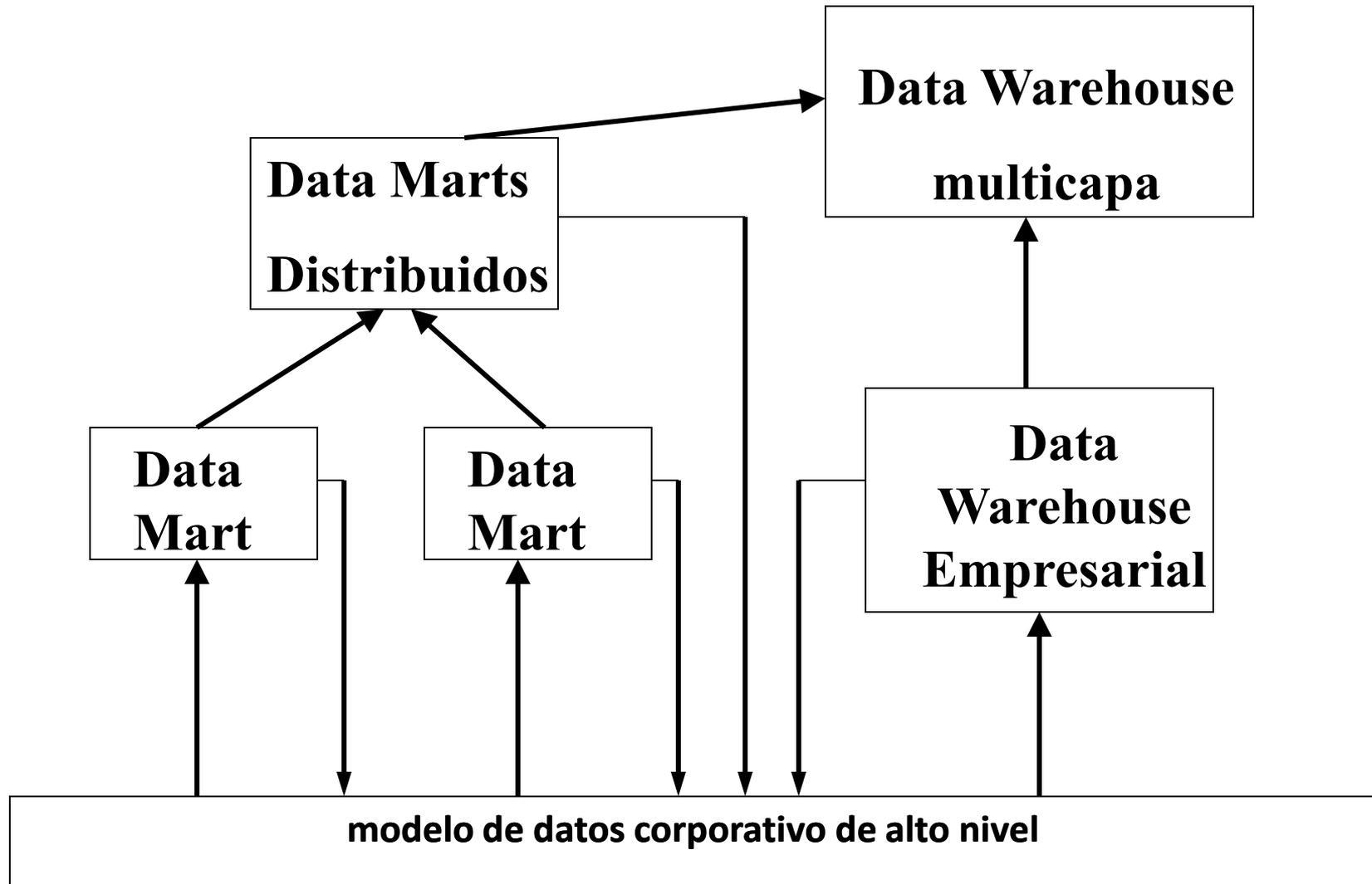
# Proceso de diseño de Data Warehouse

- **De arriba hacia abajo, de abajo hacia arriba o combinación de los dos**
  - De arriba hacia abajo: inicia con el diseño general y la planificación (va madurando)
  - De abajo hacia arriba: Empieza por experimentos y prototipos (rápido)
- **Desde el punto de vista de la ingeniería de software**
  - Cascada: estructurada y análisis sistemático de cada paso antes de pasar al siguiente
  - Espiral: rápida generación de sistemas cada vez más funcionales, corto tiempo de vuelta, vuelta rápida
- **Típico proceso de diseño de data warehouse**
  - Elija un proceso de negocio para modelar, por ejemplo, pedidos, facturas, etc.
  - Elija el grano (nivel atómico de los datos) de los procesos de negocio
  - Elija las dimensiones que se aplicarán a cada registro de la tabla
  - Elija la medida que se usará para explotar cada registro de la tabla

# Proceso de diseño de Data Warehouse

- **Exportación de Fuentes de datos**
  - Heterogéneas y diversas
  - Archivos, patrimonio BD, Web, ...
  - Definir vistas exportadas
- **Definición de un esquema general**
  - Integrar datos relevantes
  - Se basa en el modelo relacional
- **Necesidad de gestionar los metadatos**
  - Descripción de las fuentes
  - Descripción de puntos de vista exportados
  - Descripción del esquema general

# Un enfoque de desarrollo de Data Warehouse



# Modelos Perfilados de Visualización

Analistas

Usuarios Casuales

Análisis Estratégico

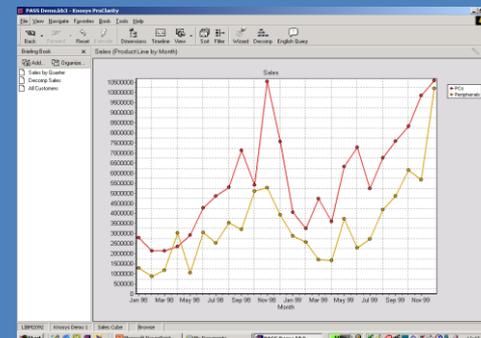
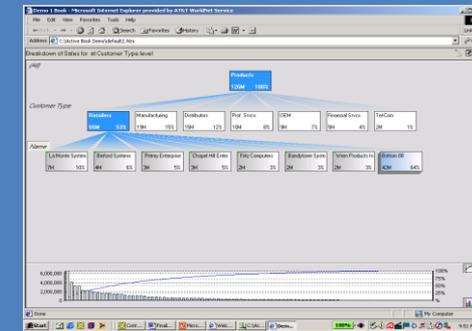
Consultas Analíticas

Reportes Interactivos

Tablero de Control

Herramienta de Consulta

Reportes Dinámicos



# Modelos dimensionales

Es una técnica de **diseño lógico** comúnmente utilizada para Data Warehouses, que busca presentar los datos en una arquitectura estándar y permita una **alta performance de acceso** a los usuarios finales.

Técnica que busca **fortalecer la capacidad de consultas** de que los usuarios disponen, haciendo la interfaz más entendible y el procesamiento más eficiente

El modelo se basa en esquemas estrella, conformados por Tablas de Hechos y Tablas Dimensionales (p.ej. cubos).

# Modelos dimensionales

- Un modelo relacional desnormalizado
  - Compuesto por tablas con atributos
  - Las relaciones definidas por claves nuevas y claves externas
- Organizado por la comprensibilidad y facilidad de presentación de informes en lugar de facilitar la actualización
- Consultado y mantenido por herramientas especiales de gestión analítica

# Entidad-Relación vs Modelos dimensionales

- **Optimizar la actualización**
- Modelo de Procesamiento de Transacciones
- En el ambiente transaccional se diseña alrededor de las **aplicaciones y funciones** tales como préstamos, ahorros, tarjeta bancaria y depósitos para una institución financiera.
- **Maximizar la comprensibilidad**
- **Optimizado para la recuperación**
- El modelo data warehousing
- En el ambiente data warehousing se organiza alrededor de **sujetos** tales como cliente, vendedor, producto y actividad.

# Entidad-Relación vs Modelos dimensionales

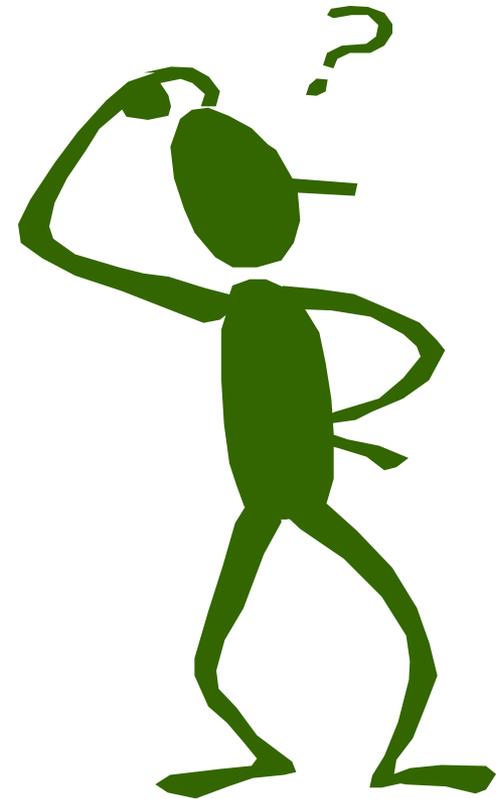
- Una tabla por entidad
  - Minimizar la redundancia de datos
  - Simétrico
  - Divide los datos en muchas entidades
  - Describe entidades y relaciones
- Una tabla de hechos para la organización de datos
  - Asimétrico
  - Divide los datos en dimensiones y hechos
  - Describe las dimensiones y medidas
  - Alienta a la redundancia de datos

# Fortalezas del Modelo Dimensional

- Marco estándar predecible,
- Responde bien a los cambios en las necesidades de información del usuario
- Relativamente fácil de agregar los datos sin necesidad de recargar las tablas
- Se han desarrollado enfoques de diseño estándar
- Existen una serie de productos que apoyan el modelo dimensional

# Proceso de diseño de Data Warehouse

- Datawarehouse incluye, historias, resúmenes de los datos de la empresa , etc.
- El diseñador define:
  - esquema exportado e integrado y
  - Esquema como quedara en DW



# Diseño de Esquemas

- **Los datos se organizan por temas importantes:**  
Los clientes, los productos, las ventas, ...
- **Tema = datos + dimensiones**
  - Recopilación de datos útiles sobre un tema  
Ejemplo: ventas
  - Sintetizar una visión única de los temas a analizar  
Ejemplo: Ventas (producto, período, tienda, número)
  - Detallar la vista según dimensiones  
Ejemplo:  
Productos (IDprod, descripción, color, tamaño ...)  
Tiendas (IDmag, nombre, ciudad, departamento, país)  
Periodo (IDper, año, trimestre, mes, día)

# Diseño de Esquemas

## Los tipos de Esquema

- En estrella
- Constelación
- Copo de nieve

### 1. Aislar Datos a tener en cuenta

- Esquemas de las Tablas de hechos

### 2. Definir las dimensiones

- Ejes de análisis

### 3. Estandarizar dimensiones

- Dividir en varias tablas unidas por las restricciones de referencia

### 4. Integrar todo

- Varias tablas de hechos comparten algunas tablas de dimensiones (constelación de la estrella)

# Esquema en estrella

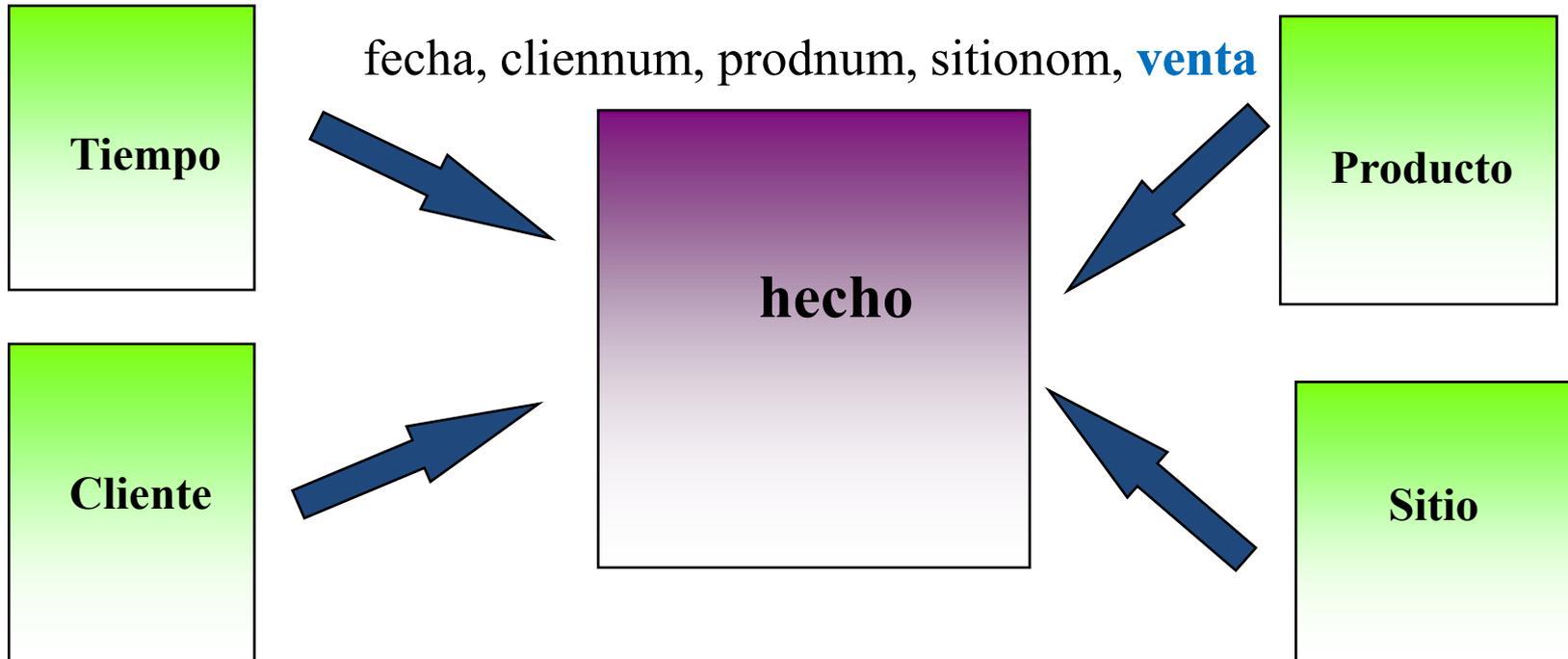
- Tecnologías de modelado relacional actual no satisfacen las necesidades actuales
- Tecnología de diseño especial para las representaciones de datos multidimensionales
- Técnica para mapear los sistemas de apoyo a decisiones multidimensionales en una base de datos relacional
- Optimizar las operaciones de consulta de datos en lugar de las operaciones de actualización de datos
- Los datos no representan una transacción del negocio en particular.
- Los datos pueden obtenerse mediante cálculos o agregaciones.

# Esquema en estrella

- El modelo estrella es una representación de una vista de la organización.
  - Ventas
  - Mercadeo
- El modelo estrella consolida hechos en relación a unas dimensiones o filtros.
- Esquema en estrella
  - Hecho rodeado de 4-15 dimensiones
  - Las dimensiones se de-normalizan
- Una tabla de hechos en el medio conectado a un conjunto de tablas de dimensiones

# Esquema en estrella

- Una sola tabla de hechos y para cada dimensión una tabla de dimensiones
- No captura jerarquías directamente



# Esquema en estrella: Componentes

- Datos (hechos)
- Dimensiones
- Atributos
- Jerarquías de atributos

# Esquema en estrella: Hechos

- Mediciones numéricas (valores) que representan un aspecto del negocio o actividad específica
- Almacenado en una tabla de hechos en el centro del esquema de estrella
- Contiene hechos que se vinculan a través de sus dimensiones
- Se pueden calcular o derivar en tiempo de ejecución
- Actualizado periódicamente con los datos de las bases de datos operacionales

# Esquema en estrella: Tabla de Hechos

## Tabla central

- Representar un proceso o reporta el entorno que es de valor para la organización
- Especifica exactamente lo que representa.
- Por lo general corresponden a una entidad asociativa en el modelo ER
- Guarda Medidas de interés del negocio
- Varía constantemente datos

# Esquema en estrella: Tabla de Hechos

## Tabla central

- Ejemplo típico: los registros de ventas individuales
- elementos numéricos principalmente en bruto
- filas estrechas, algunas columnas como máximo
- gran número de filas (millones a un mil millones)
- Acceso por dimensiones: Enlaces directos a las dimensiones
- Clave principal de varias partes
- Contiene dos o más claves foráneas

# Esquema en estrella: Dimensiones

- Características cualitativas que proveen perspectivas adicionales a un hecho
- Normalmente dimensiones se almacenan en tablas de dimensiones
- Definir negocio en términos ya familiares para los usuarios
- Dimensiones típicas:  
períodos de tiempo, áreas geográficas (mercados, ciudades), productos, clientes, vendedores, etc.
- Típicamente contienen atributos para consultas desde SQL.

# Esquema en estrella: Tabla de Dimensiones

- Se enlaza a la tabla de hechos (clave primaria única)
- Guarda los Atributos del negocio
- Más o menos constante datos  
Por ejemplo, Tiempo, Producto, Cliente, tienda, etc.
- Contiene información textual descriptiva
- Hileras anchas con mucho texto descriptivo
- Tablas pequeñas (alrededor de un millón de filas)
- Ingresó a la tabla de hechos mediante una clave externa
- fuertemente indexados

# Esquema en estrella: Atributos

- Tablas de dimensiones contienen atributos
- Los atributos se utilizan para buscar, filtrar o clasificar los hechos
- Dimensiones proporcionan características descriptivas acerca de los hechos a través de sus atributos
- Debe definir los atributos comunes de negocios que se utilizará para reducir la búsqueda, agrupar información, o describir las dimensiones. (por ejemplo, tiempo/lugar/producto)
- Sin límite matemático para el número de dimensiones (3-D Hace que sea fácil modelar)

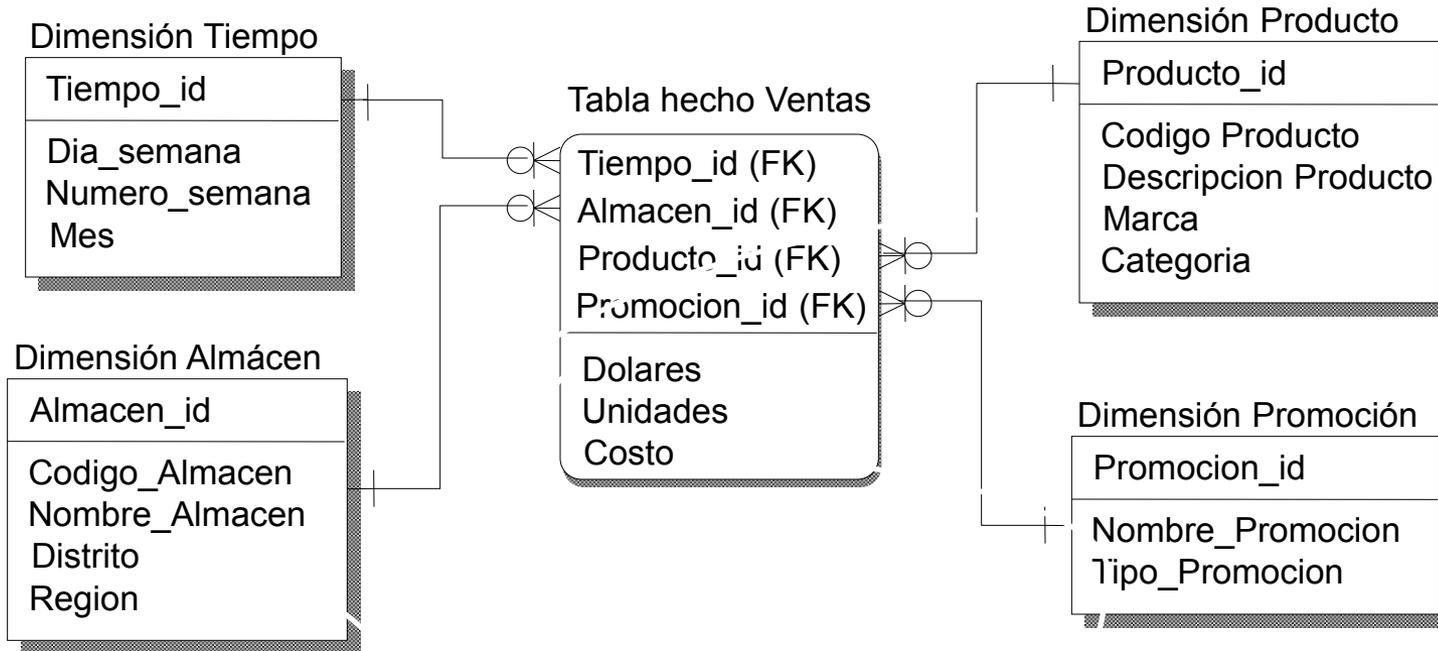
# Jerarquía entre atributos

- Proporciona una organización de los datos de arriba abajo
- agregación de atributos
- Análisis de datos: Drill-down / Roll-Up
- Atributos de diferentes dimensiones se pueden agrupar para formar una jerarquía

# Ejemplo de esquema en estrella para ventas



# Ejemplo de esquema en estrella para ventas Reporte

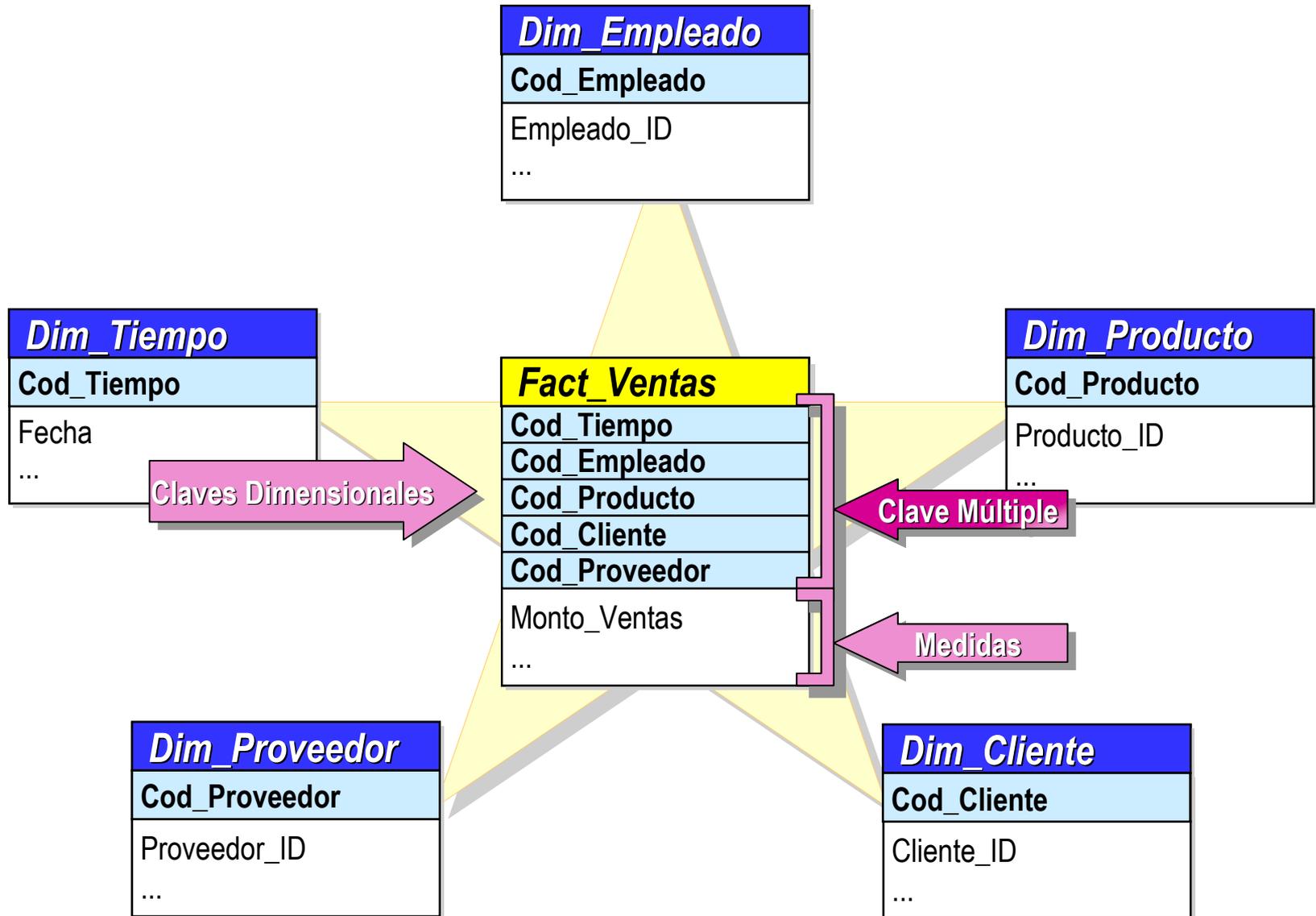


<b>Distrito</b>	<b>Marca</b>	<b>Total Dolares</b>	<b>Total Costo</b>	<b>Utilidad</b>
Atherton	Clean Fast	\$ 1,233	\$ 1,058	\$ 175
Atherton	More Power	\$ 2,239	\$ 2,200	\$ 39
Atherton	Zippy	\$ 848	\$ 650	\$ 198
Belmont	Clean Fast	\$ 2,097	\$ 1,848	\$ 249
Belmont	More Power	\$ 2,428	\$ 2,350	\$ 78
Belmont	Zippy	\$ 633	\$ 580	\$ 53

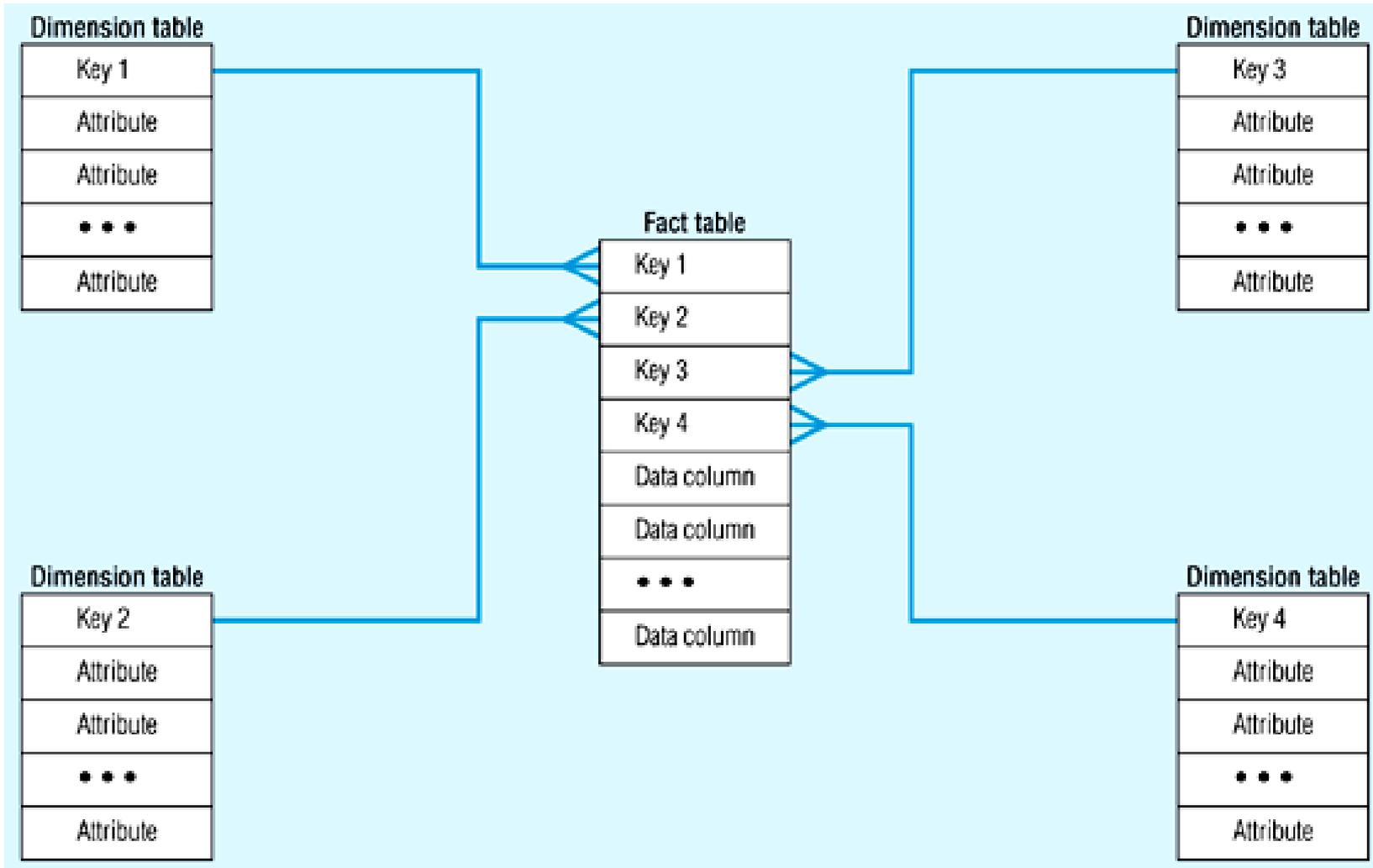
# Esquema en estrella

- Hechos y dimensiones son representadas por tablas físicas en la base de datos de data warehouse
- Las tablas de hechos están relacionados a cada tabla de dimensión en una relación Muchos a Uno (clave externa)
- Tabla de hechos está relacionado con muchas tablas de dimensiones
- La clave principal de la tabla de hechos es compuesta de las claves principales de las tablas de dimensiones
- Cada tabla de hecho está diseñada para responder a una pregunta específica de IN

# Ejemplo de esquema en estrella para ventas



# Star schema

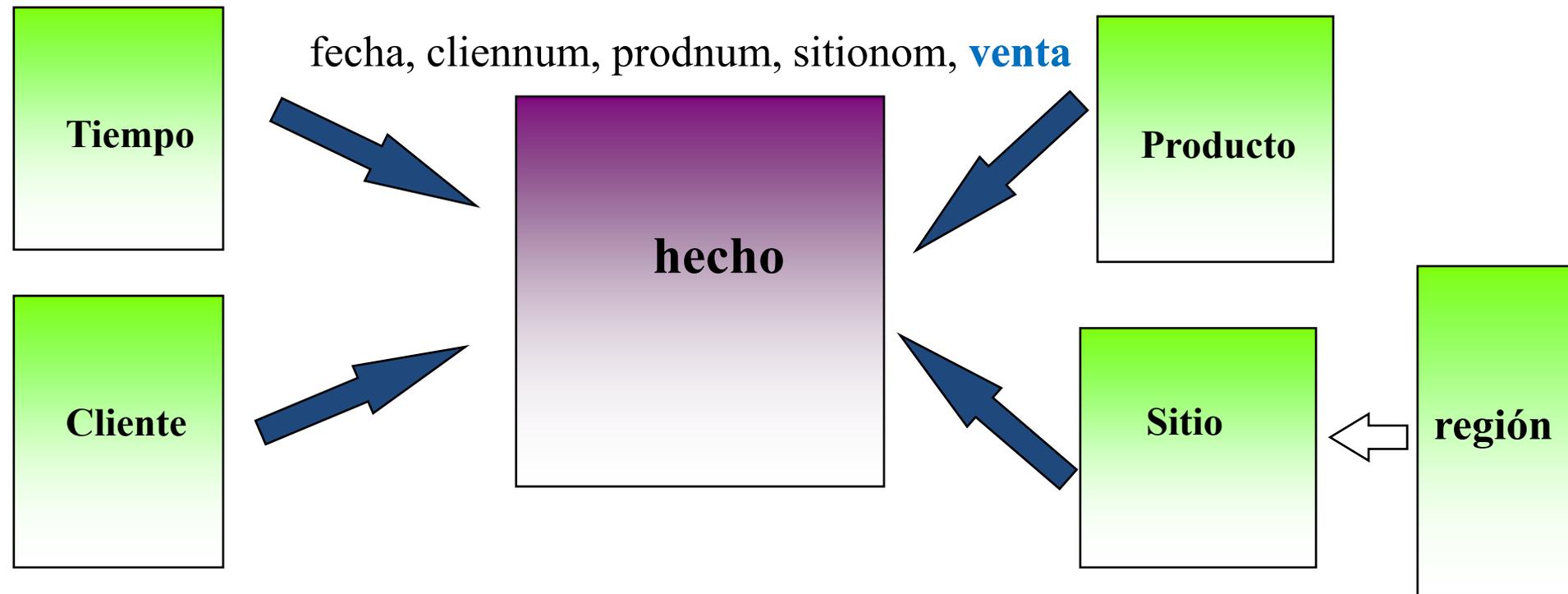


# Esquema Copo de nieve

- Un refinamiento del esquema en estrella donde alguna jerarquía dimensional se normaliza en un conjunto de tablas de dimensiones más pequeñas, formando una forma similar a la del copo de nieve
  - Esquema en estrella con dimensiones secundarias
  - No ahorra de espacio
- Copo de nieve, si las dimensiones secundarias tienen muchos atributos

# Esquema Copo de nieve

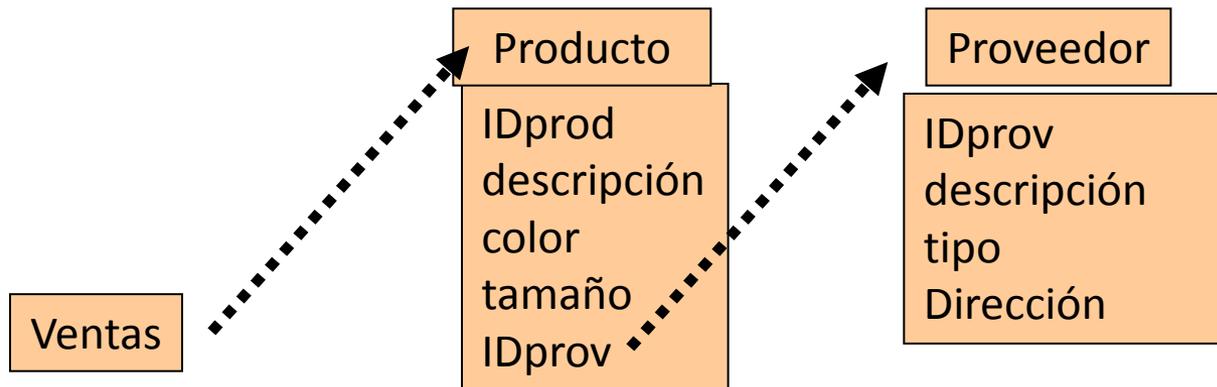
- Representa jerarquía dimensional directamente definida por la normalización de las tablas.
- Fácil de mantener y ahorra almacenamiento



# Esquema Copo de nieve

## Beneficios

- Evita la duplicación
- Conduce a las constelaciones (varias tablas de hechos con dimensiones compartidas)



# Esquema Copo de nieve

Dimensión del almacén

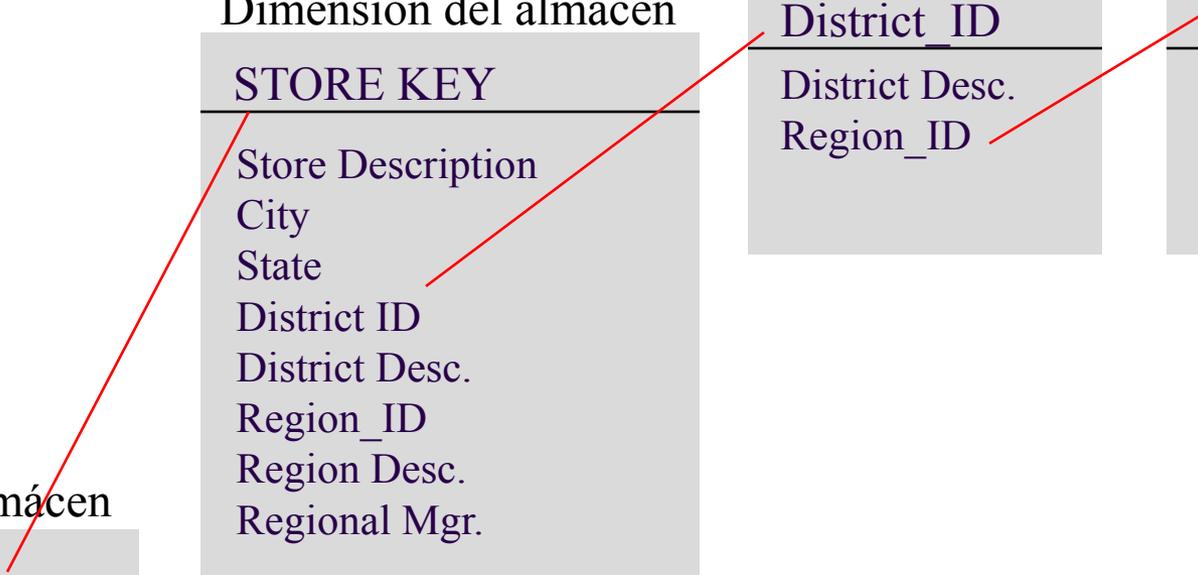
<u>STORE KEY</u>
Store Description
City
State
District ID
District Desc.
Region_ID
Region Desc.
Regional Mgr.

<u>District_ID</u>
District Desc.
Region_ID

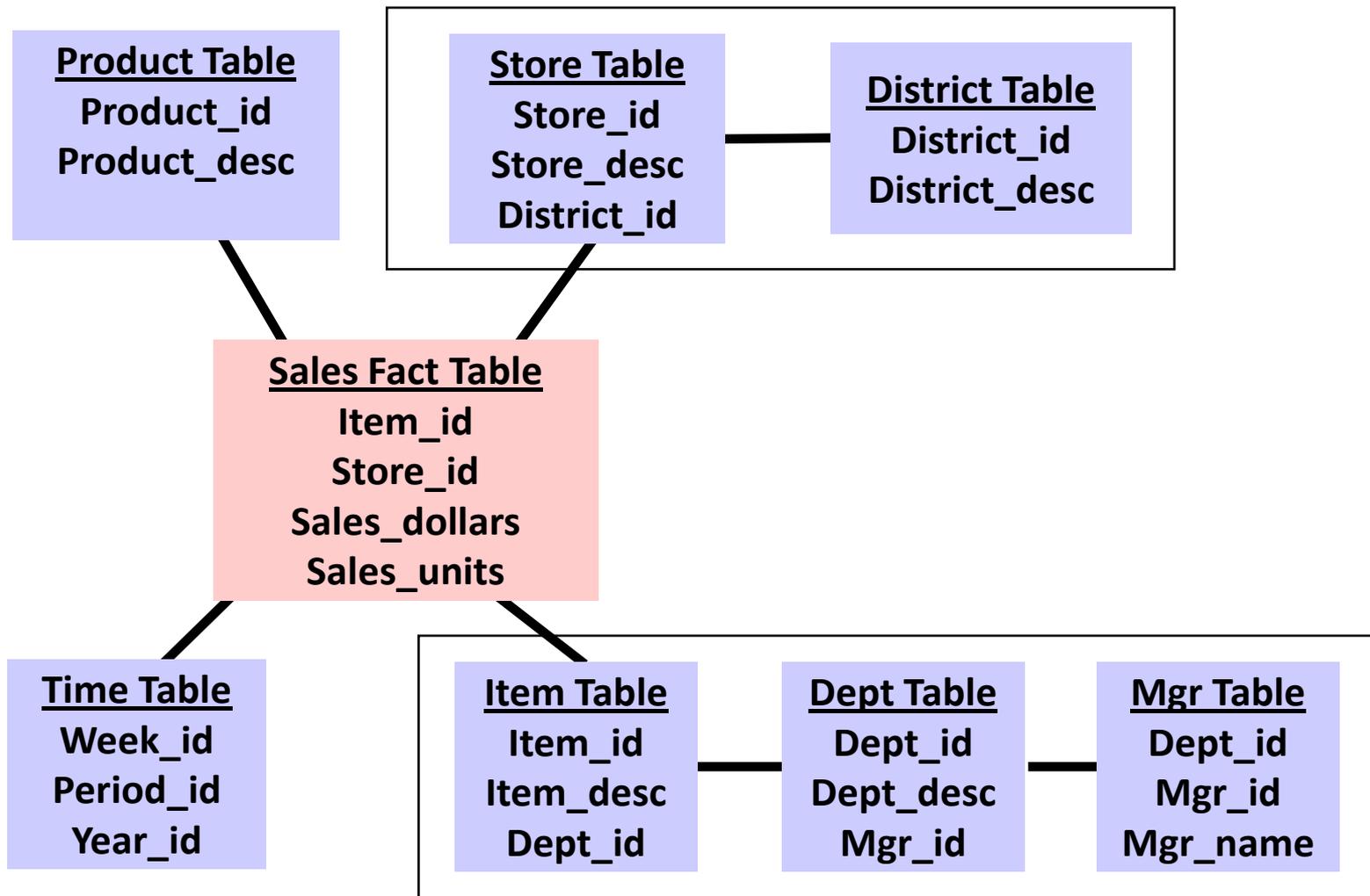
<u>Region_ID</u>
Region Desc.
Regional Mgr.

Tabla hecho almacén

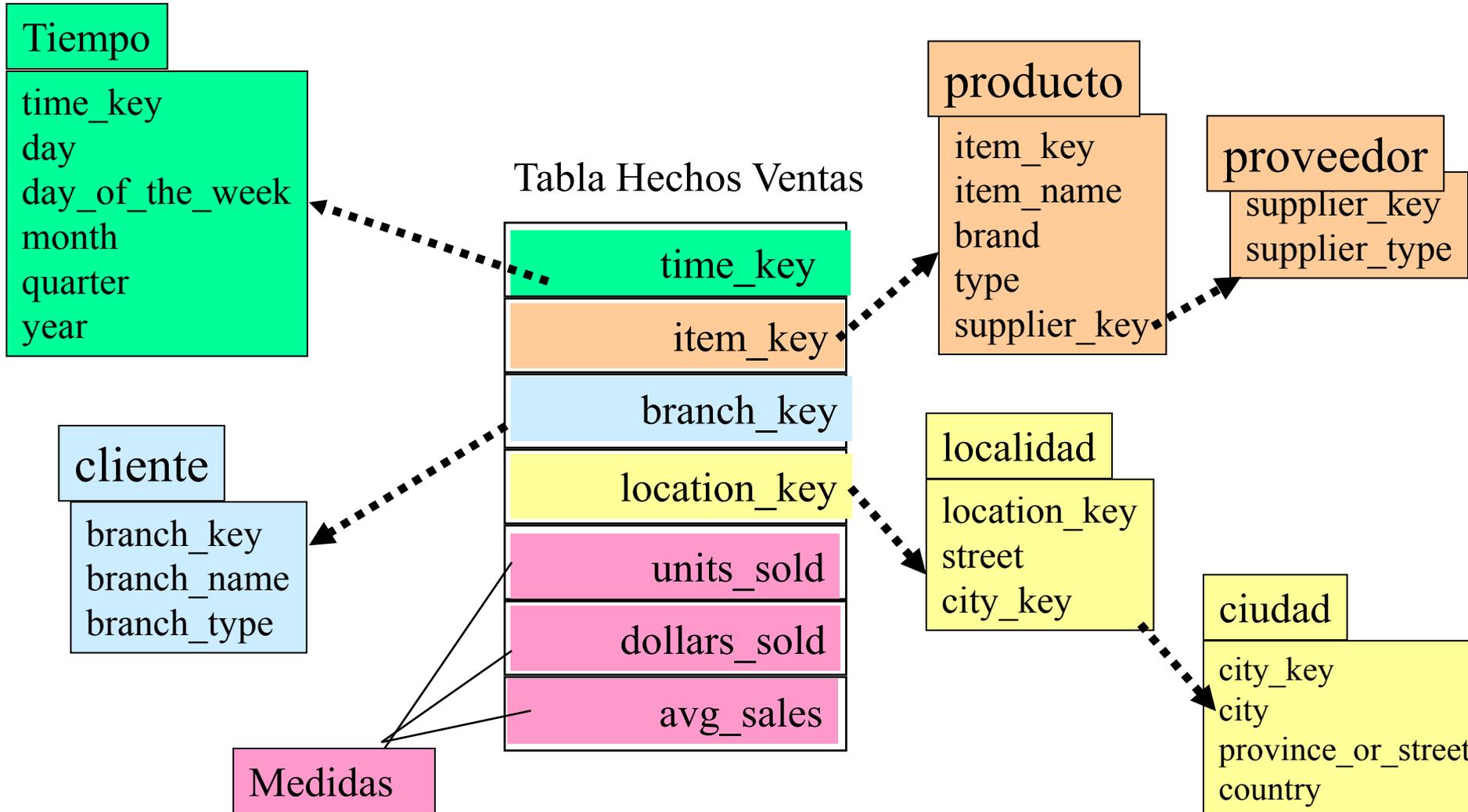
<u>STORE KEY</u>
<u>PRODUCT KEY</u>
<u>PERIOD KEY</u>
Dollars
Units
Price



# Esquema Copo de nieve



# Esquema Copo de nieve

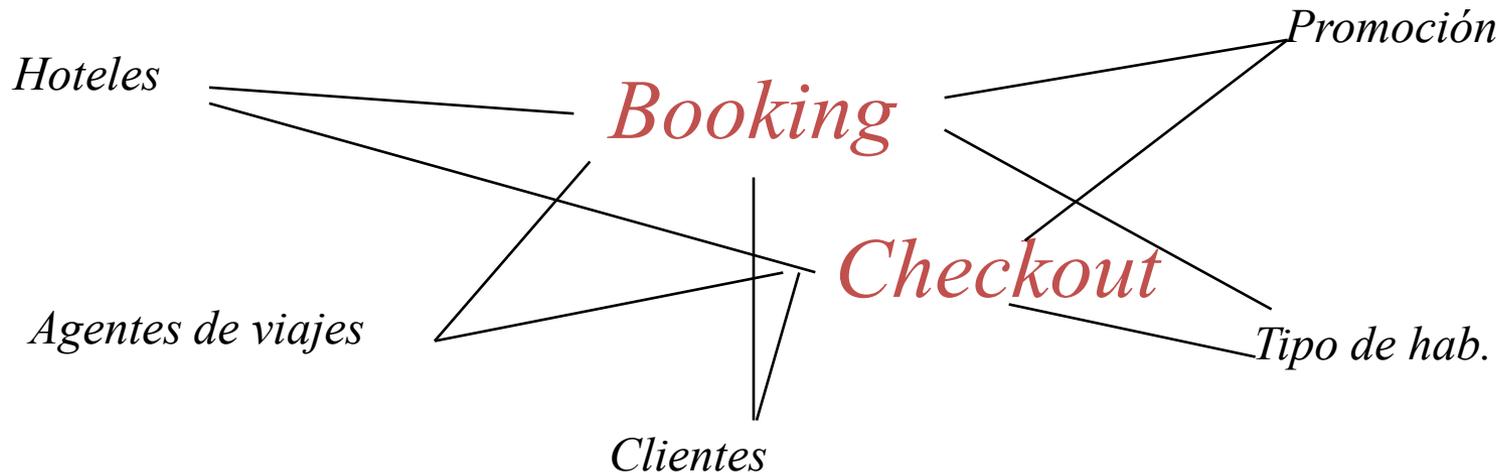


# Esquema de Constelación de hechos

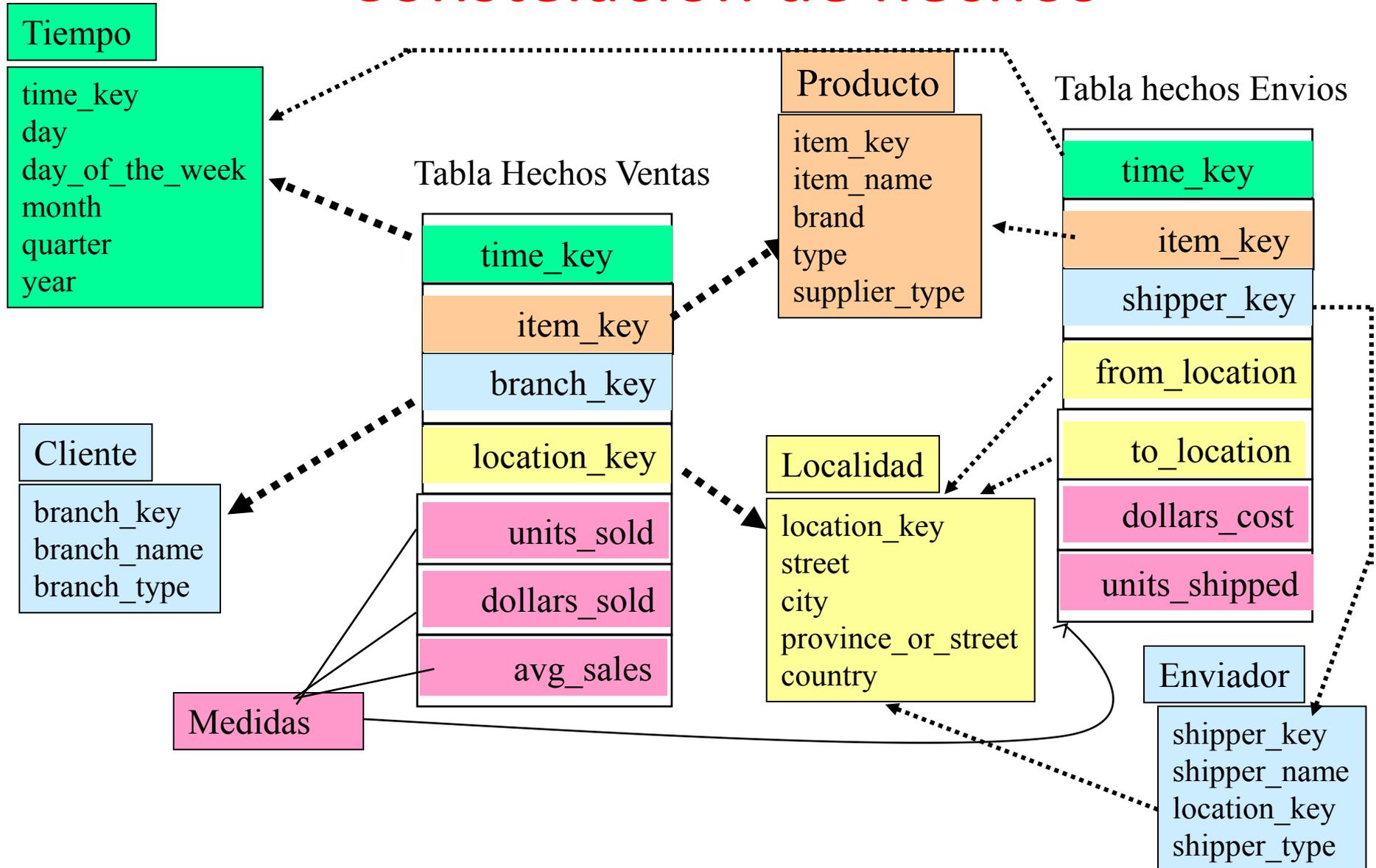
Varias tablas de hechos comparten tablas de dimensiones, vistos como una colección de estrellas, por lo tanto, llamados esquema de galaxia o constelación de hecho

# Esquema de Constelación de hechos

Reservas (Booking ) y Checkout pueden compartir tablas de dimensiones en la industria hotelera



# Esquema de Constelación de hechos



# De Tablas y Hojas de cálculo a los cubos de datos

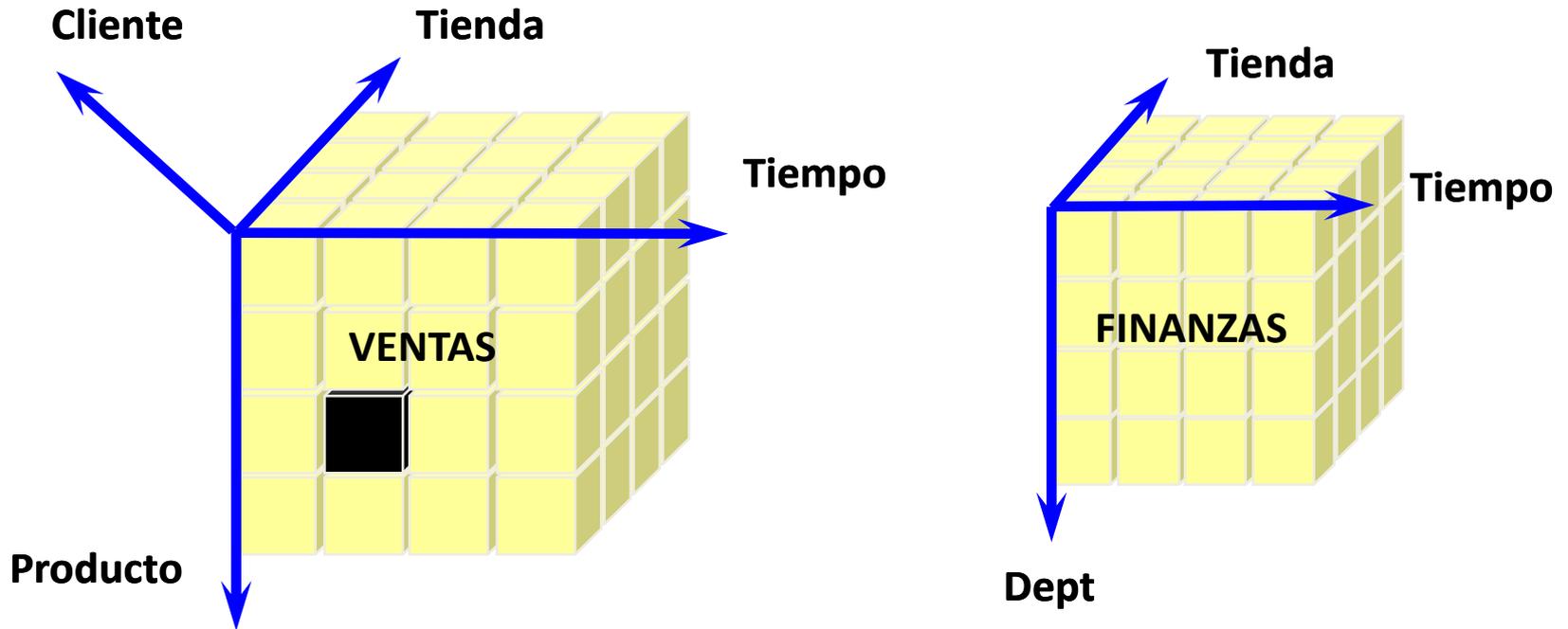
- Un data warehouse se basa en un modelo de datos multidimensional
- Todo que ve los datos en la forma de un cubo de datos
- Un cubo de datos: por ejemplo, ventas, permite realizar el modelado y ver múltiples dimensiones
- Las tablas de dimensiones: son items (inombre, marca, tipo), o tiempo (día, semana, mes, año)
- Tabla de hechos contiene medidas (tales como: dolares\_vendidos) y claves para cada una de las tablas de dimensiones
- Un cubo n-D se llama un paralelepípedo.

# Base de datos relacional

	Atributo 1 Nombre	Atributo 2 edad	Atributo 3 sexo	Atributo 4 No. Emp
Fila 1	Anderson	31	F	1001
Fila 2	Green	42	M	1007
Fila 3	Lee	22	M	1010
Fila 4	Ramos	32	F	1020

Tabla de empleados

# Modelo BD multidimensional



Los datos se encuentra en la intersección de las dimensiones

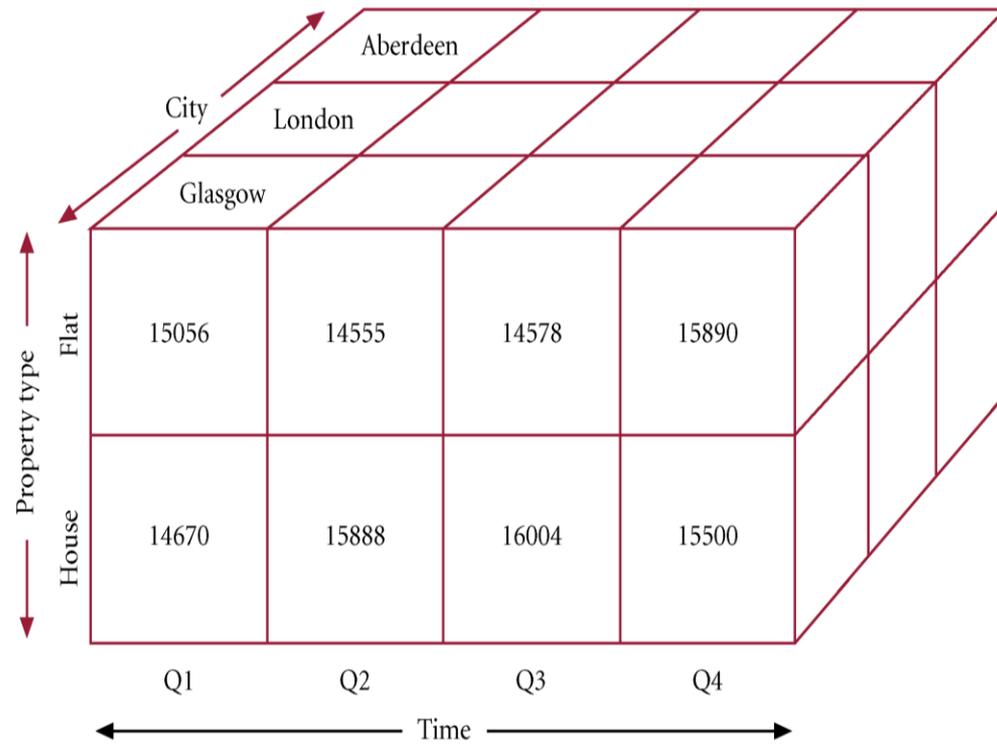
# Dos dimensiones

City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
.....	.....	.....
.....	.....	.....

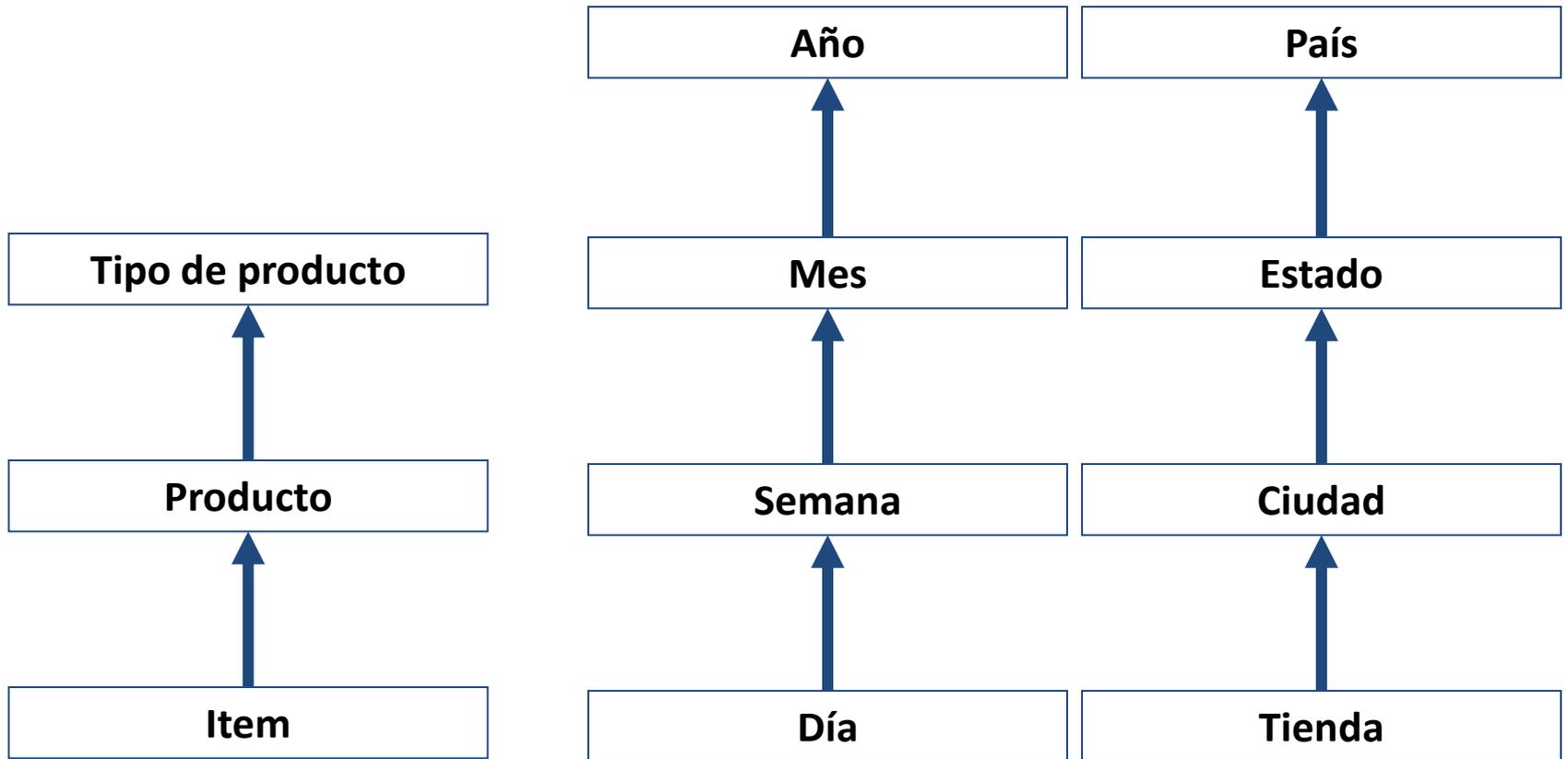
City	Glasgow	London	Aberdeen	.....
Quarter	29726	43555	53210	.....
Q1	29726	43555	53210	.....
Q2	30443	48244	34567	.....
Q3	30582	56222	45677	.....
Q4	31390	45632	50056	.....

# Tres dimensiones

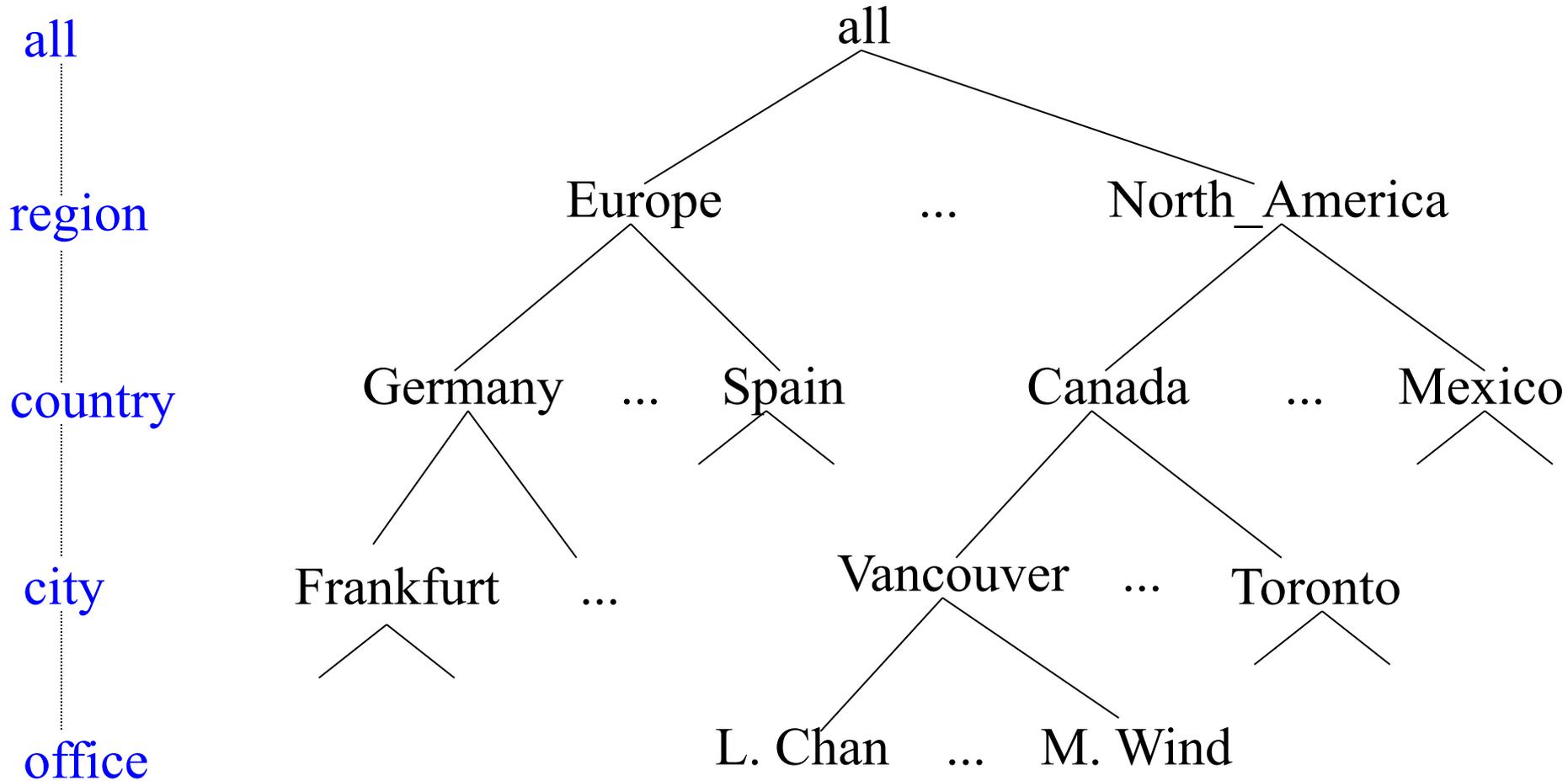
Property Type	City	Time	Total Revenue
Flat	Glasgow	Q1	15056
House	Glasgow	Q1	14670
Flat	Glasgow	Q2	14555
House	Glasgow	Q2	15888
Flat	Glasgow	Q3	14578
House	Glasgow	Q3	16004
Flat	Glasgow	Q4	15890
House	Glasgow	Q4	15500
Flat	London	Q1	19678
House	London	Q1	23877
Flat	London	Q2	19567
House	London	Q2	28677
.....	.....	.....	.....
.....	.....	.....	.....



# Jerarquía Dimensional



# Jerarquía Dimensional (localidad)



# Jerarquía Dimensional

- jerarquía de esquema

day < {month < quarter; week} < year

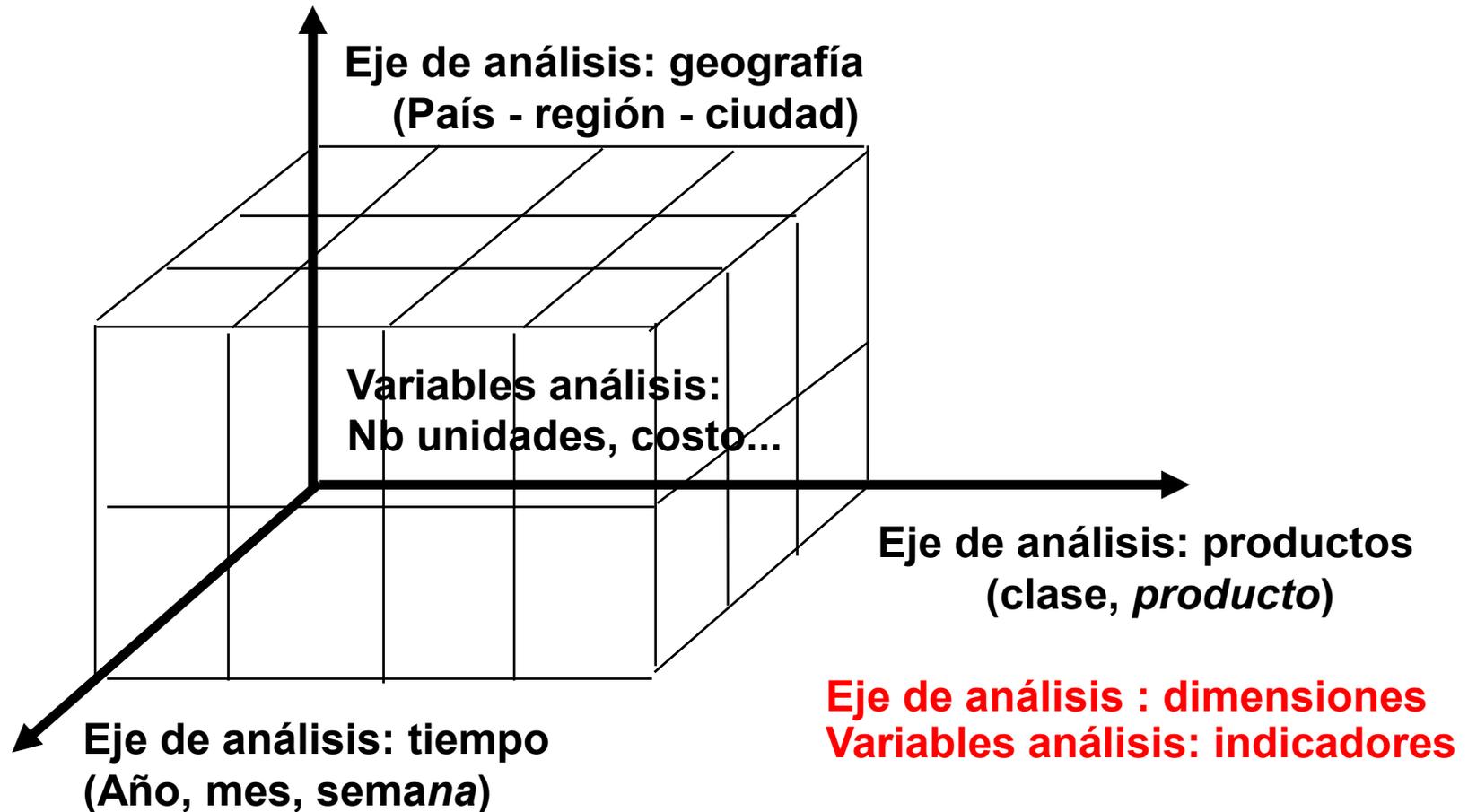
- Agrupando jerarquía

{1..10} < inexpensive

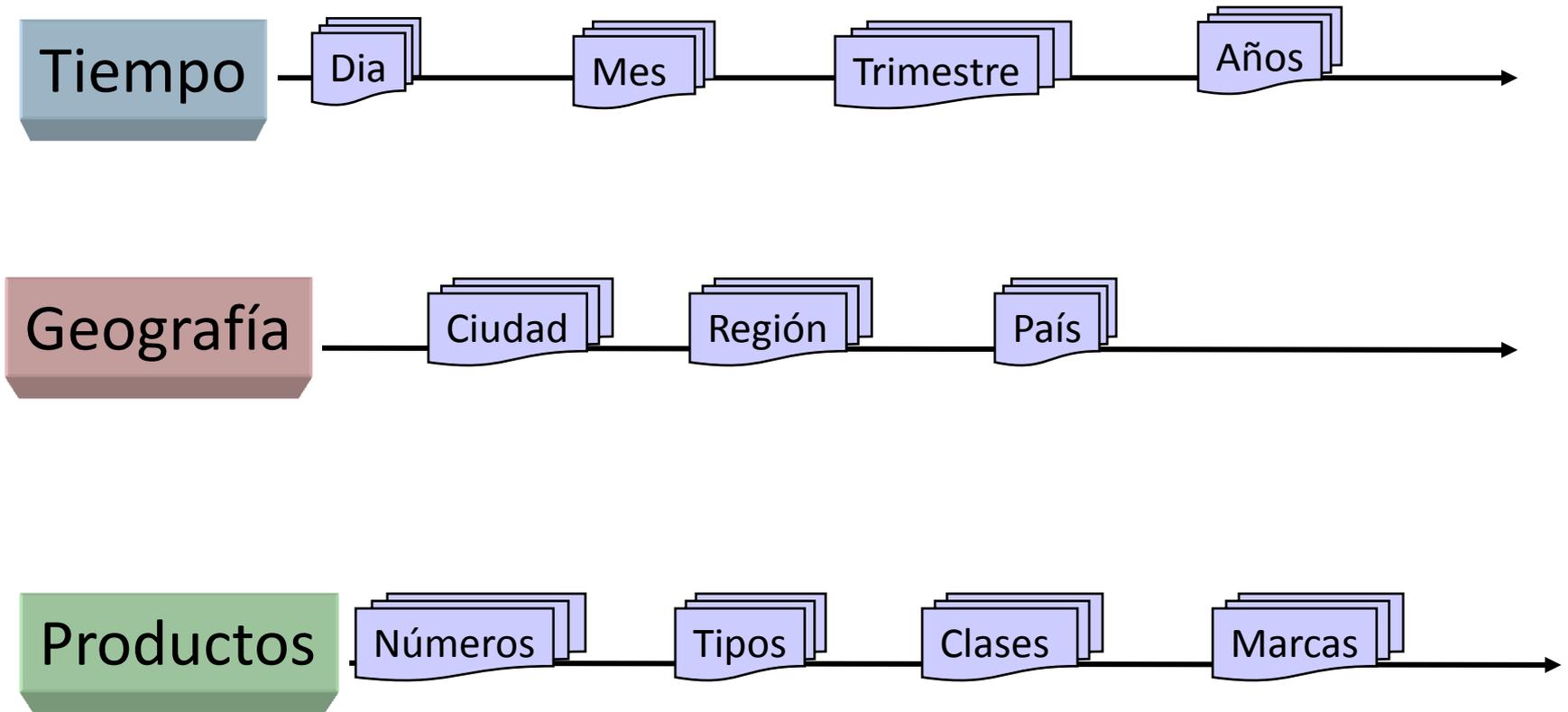
# Las multidimensiones

- Dimensiones:
  - Tiempo
  - Geografía
  - Productos
  - Clientes
  - Canales de ventas.....
- Indicadores:
  - Número de unidades vendidas
  - Costo

# Cubo de dato y las dimensiones

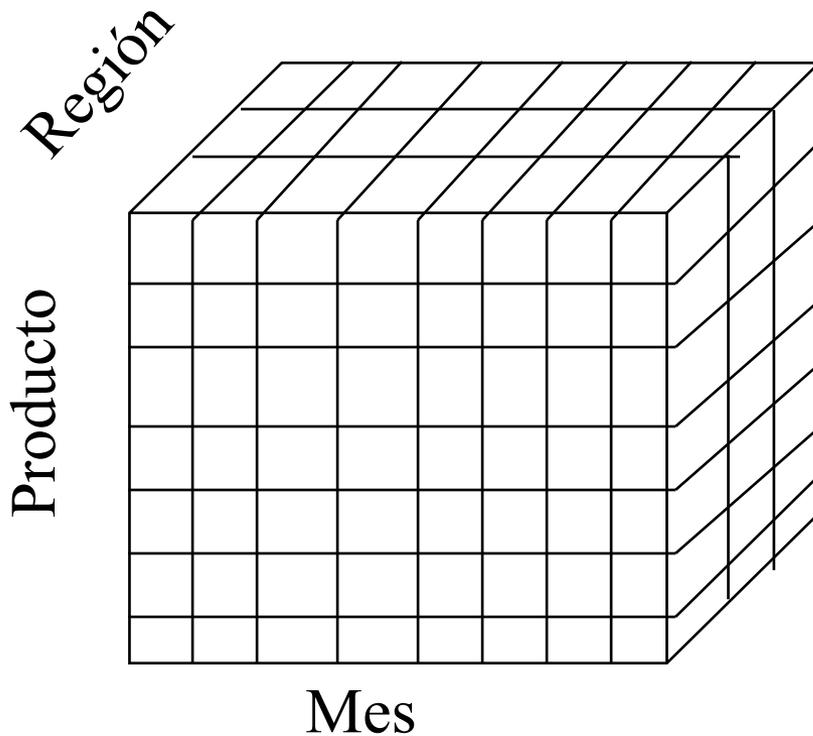


# La granularidad de las dimensiones

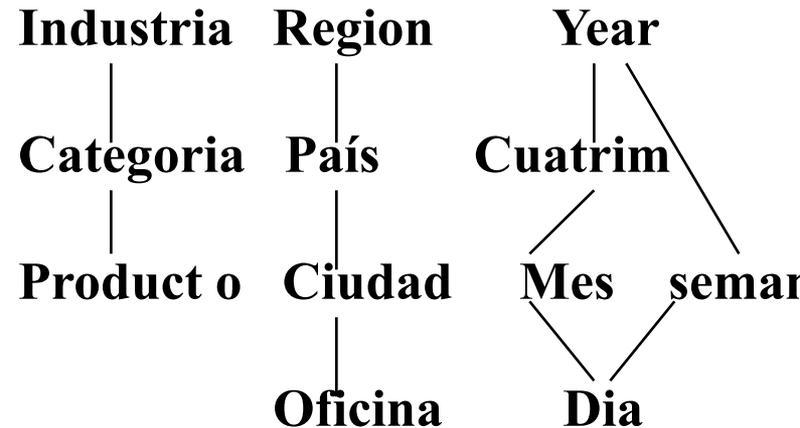


# Datos Multidimensionales

El volumen de ventas en función del producto, el mes, y el área

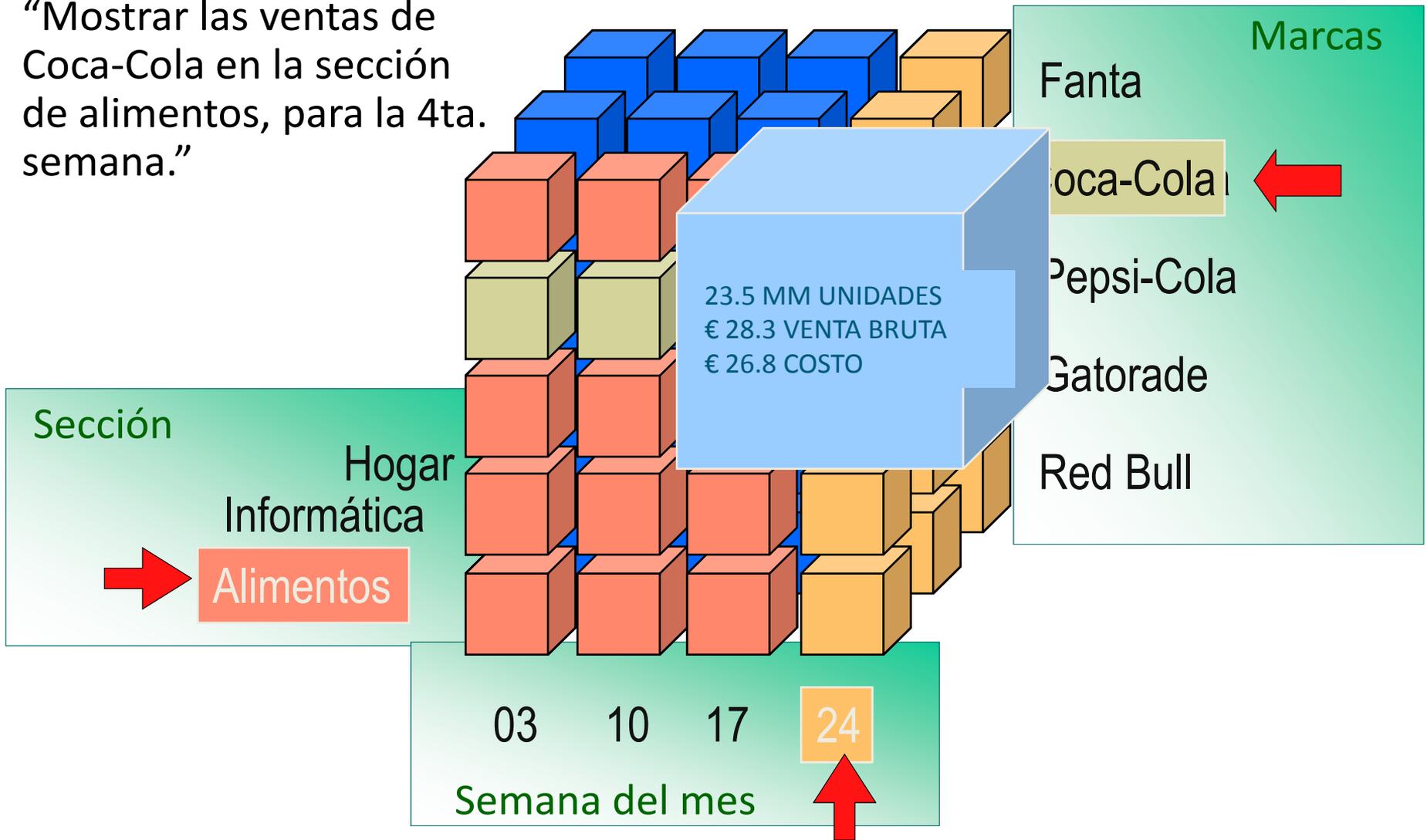


Dimensiones: Producto, Localidad, Tiempo  
Caminos jerarquicos

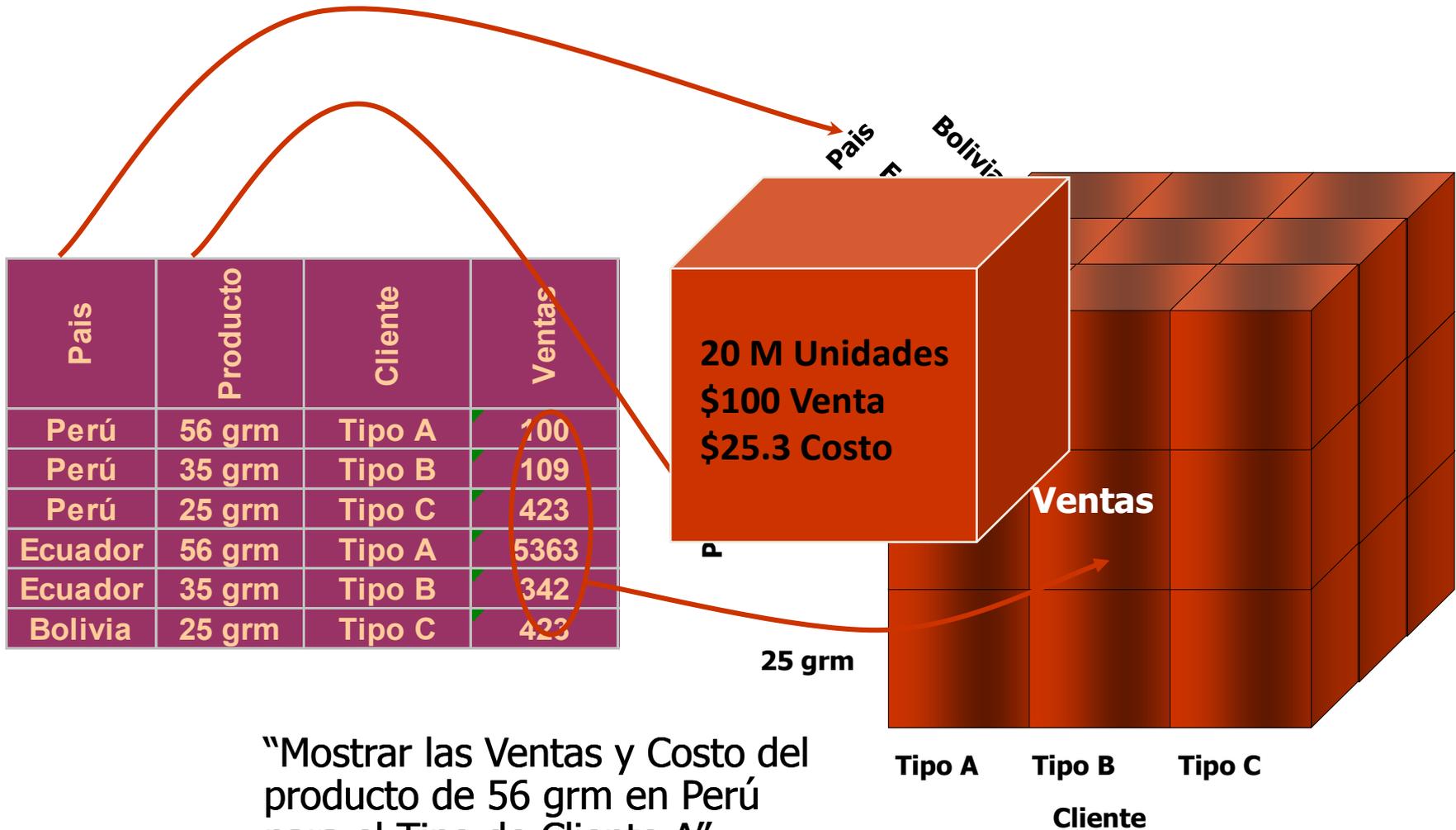


# Cubo Multidimensional

“Mostrar las ventas de Coca-Cola en la sección de alimentos, para la 4ta. semana.”

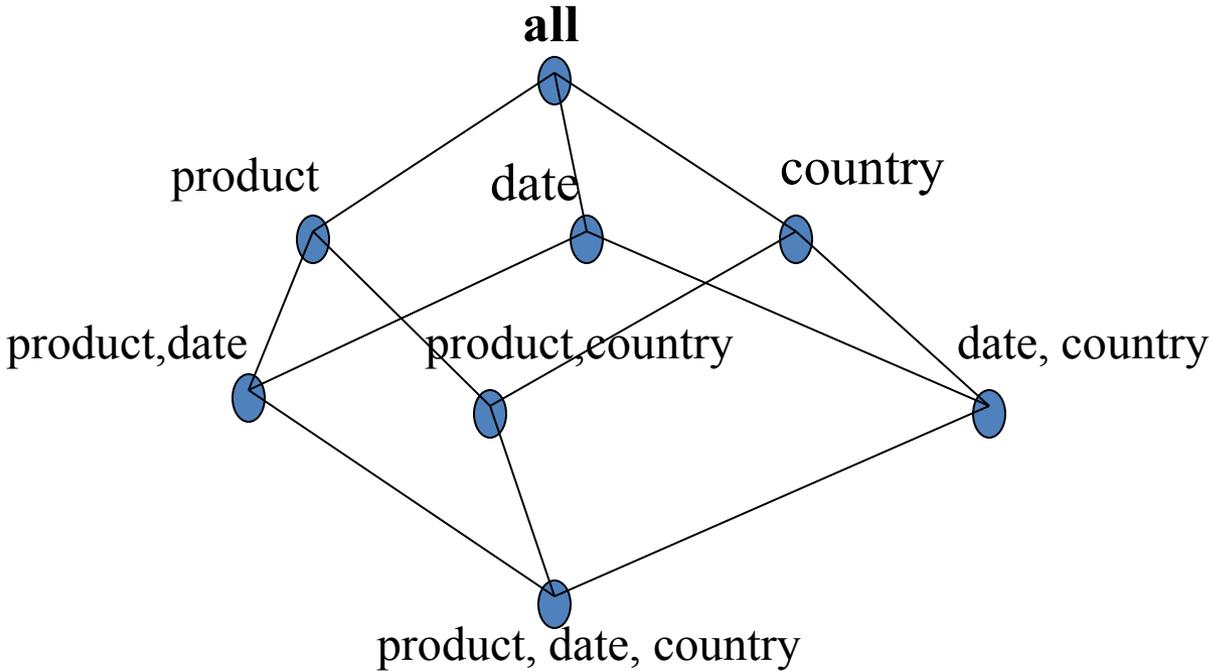


# Cubo Multidimensional



“Mostrar las Ventas y Costo del producto de 56 grm en Perú para el Tipo de Cliente A”

# Cuboides correspondientes al Cubo



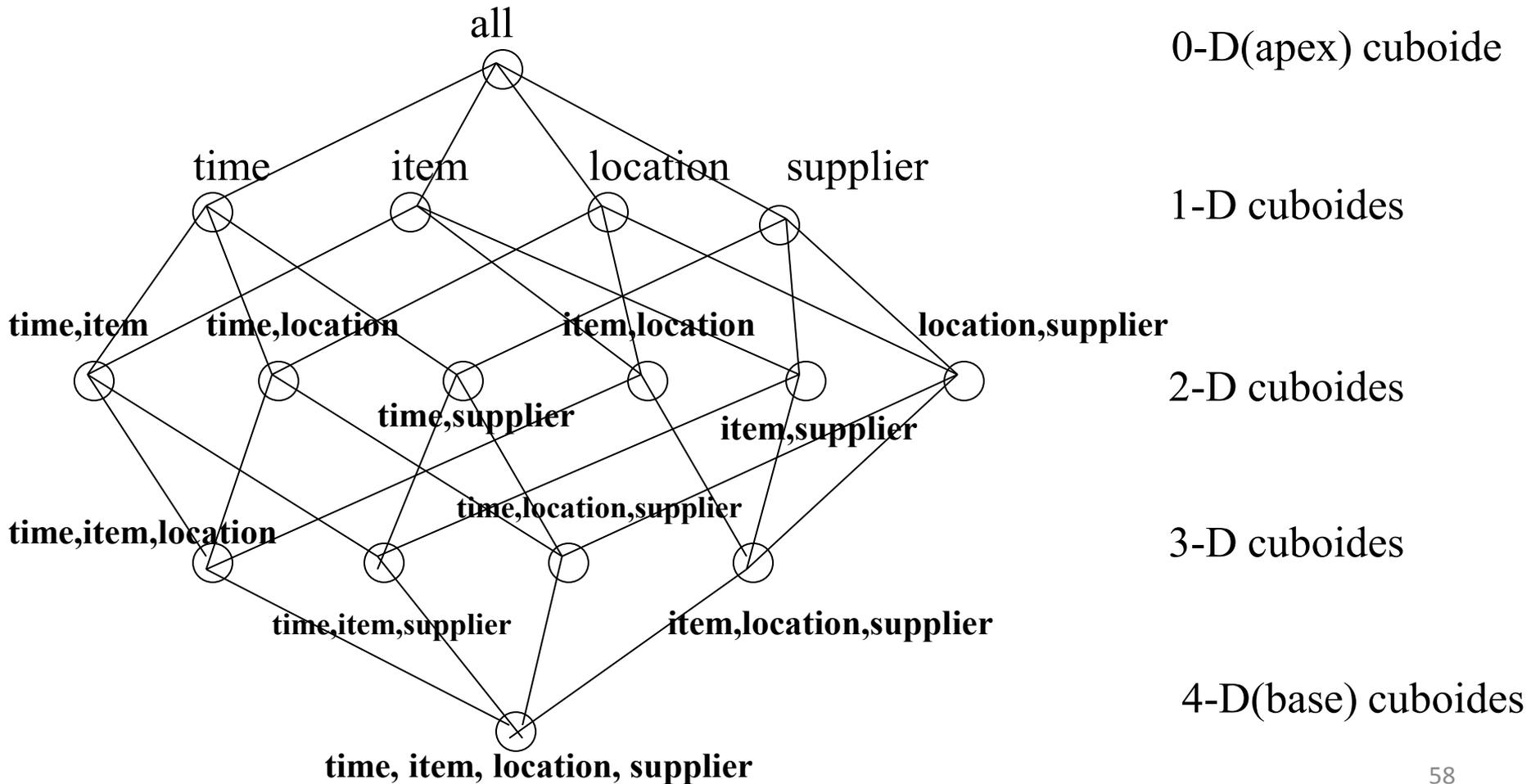
0-D(apex) cuboide

1-D cuboides

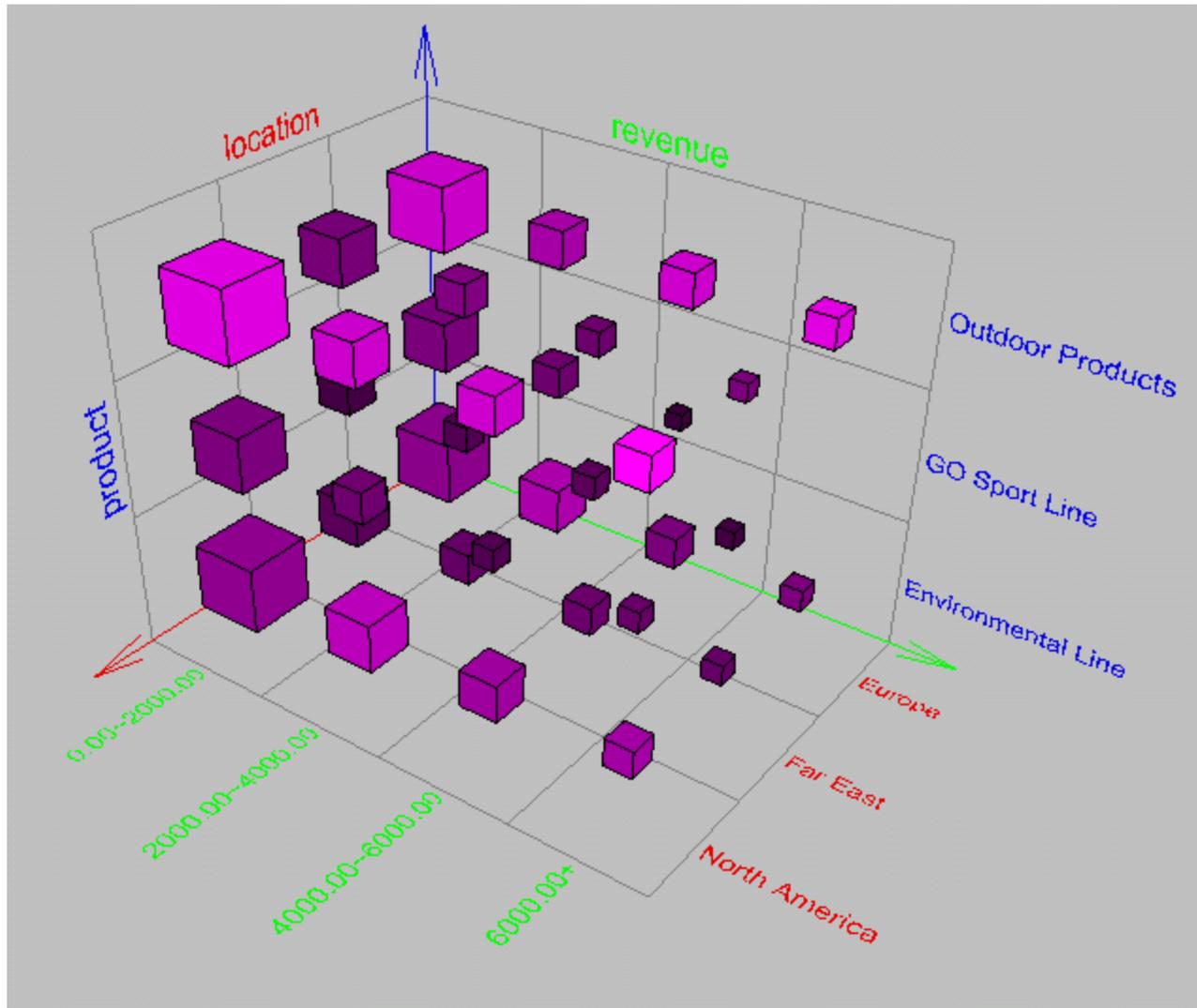
2-D cuboides

3-D(base) cuboides

# Cuboides correspondientes al Cubo

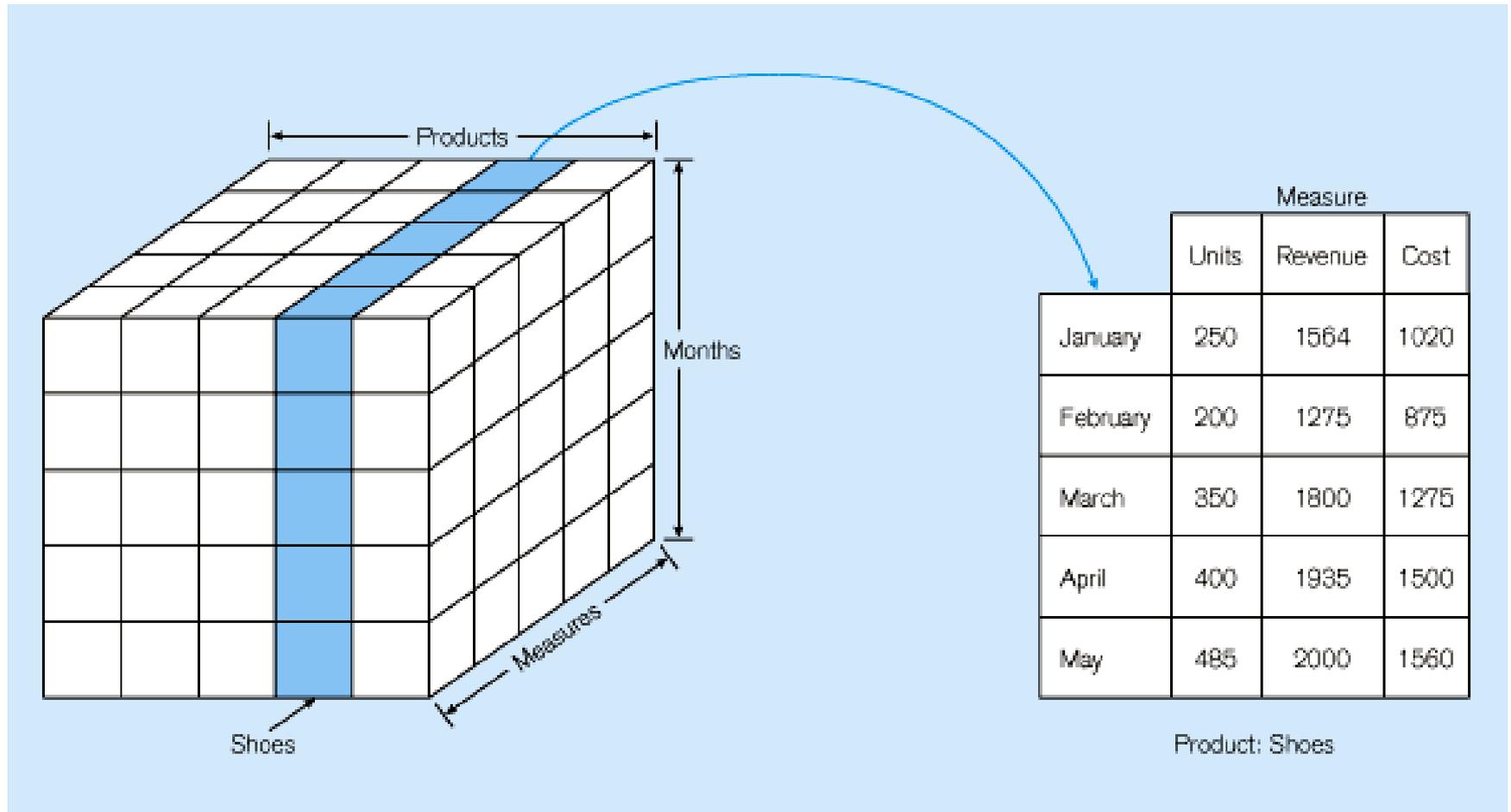


# Navegar por un cubo de datos



- Visualización
- OLAP

# Cortando/rebanando un cubo de datos

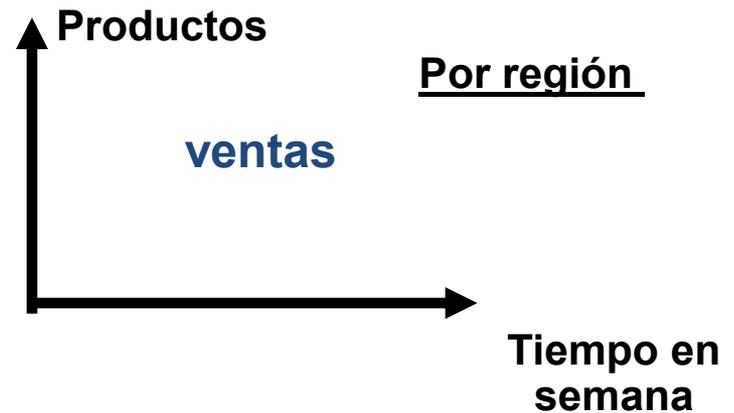


# Navegar por un cubo de datos

En 2 dimensiones



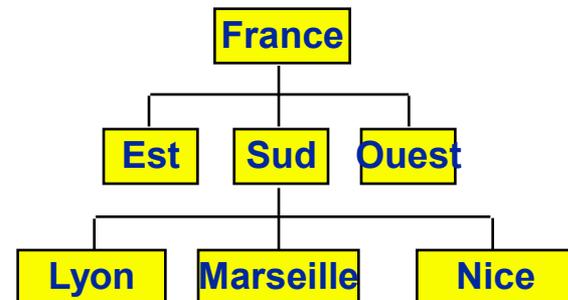
Copa de nieve



1 dimensión

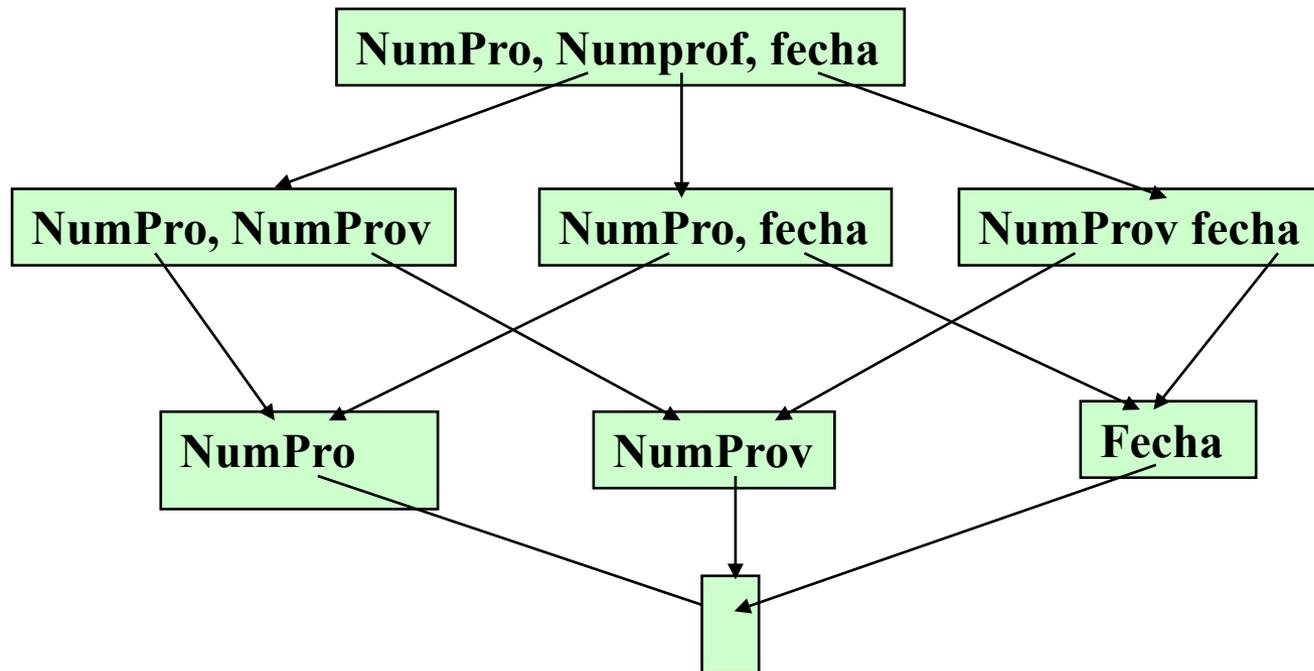


Zoom a una dimensión

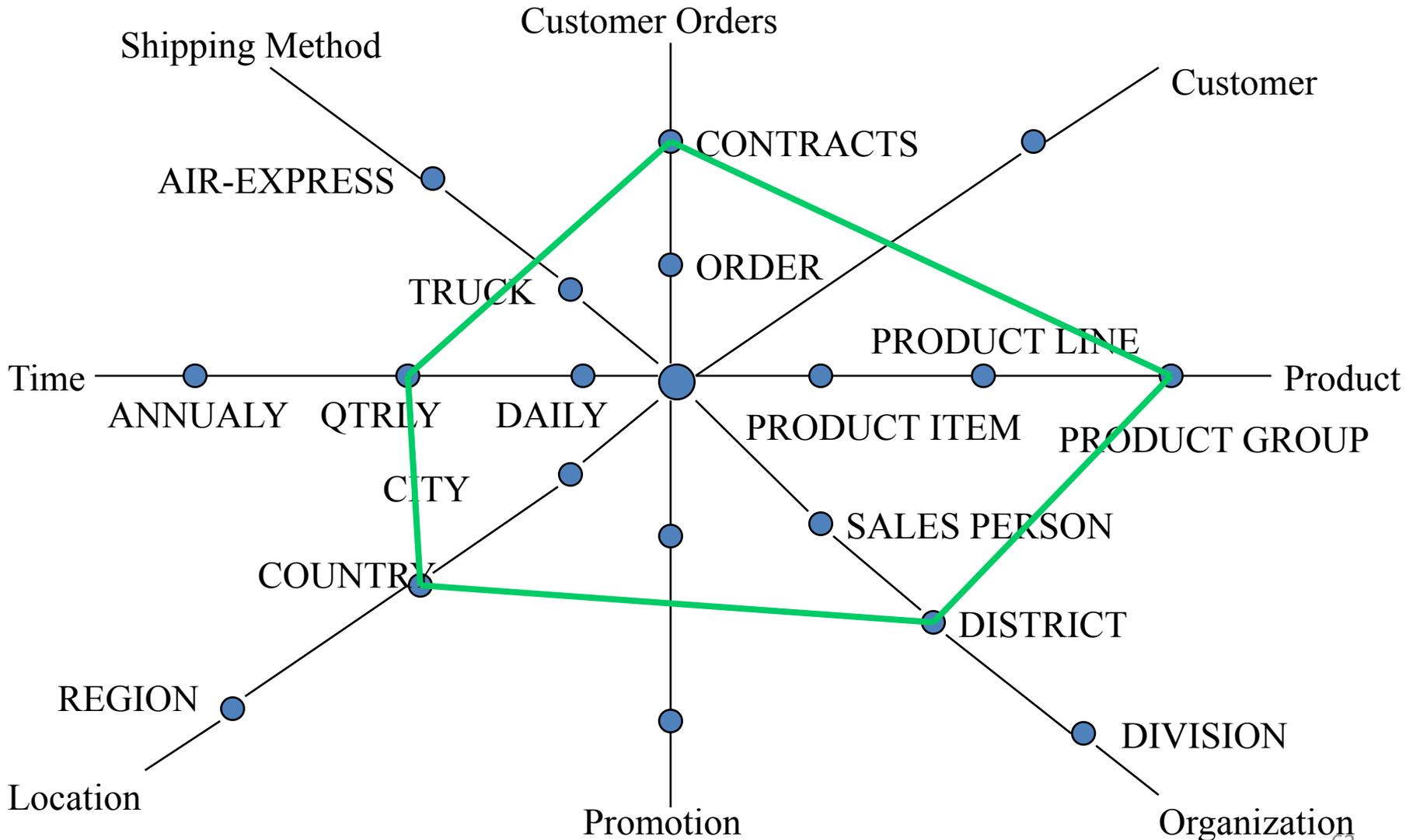


# Navegar por un cubo de datos

- A partir de un cubo en 3D, es posible agregar dimensión en una rotación  
Se obtiene una retícula de puntos de vista



# Modelo de consulta



# Medidas: Tres Categorías

- **distributiva:** si el resultado derivado mediante la aplicación de la función a  $n$  valores es el mismo que el derivado por la aplicación de la función a todos los datos sin partición
  - E.g., `count()`, `sum()`, `min()`, `max()`.
- **algebraica:** si se puede calcular una función algebraica con  $M$  argumentos (donde  $M$  es un entero) , cada uno de los cuales se obtiene mediante la aplicación de una función distribuida.
  - E.g., `avg()`, `min_N()`, `standard_deviation()`.
- **Holística:** describe un subagregado.
  - E.g., `median()`, `mode()`, `rank()`.

# Las operaciones en el modelo de datos multidimensional

- **Agregación (roll-up)**
  - reducción de dimensiones: por ejemplo, las ventas totales de la ciudad
  - resumen sobre la jerarquía global: por ejemplo, las ventas totales de la ciudad y del año -> ventas totales por región y por año
- **Selección (rebanado) define subcubos**
  - por ejemplo, las ventas de ciudad = Palo Alto y fecha = 1/15/96
- **Navegación para detallar datos (drill-down)**
  - por ejemplo, (ventas - gastos) por ciudad, tope 3% de las ciudades por el ingreso promedio
- **Visualización:(pivote)**

# Las operaciones en el modelo de datos multidimensional: algebra de los cubos

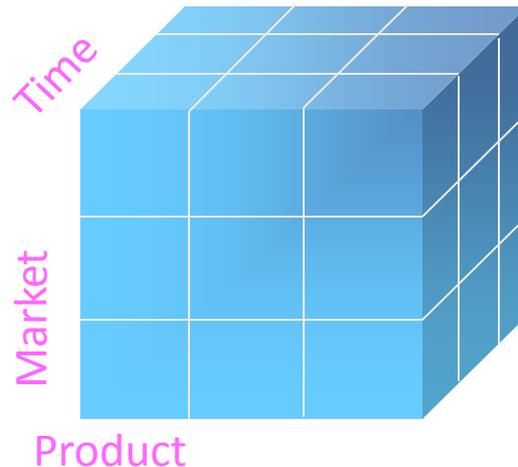
- **Roll up :**
  - Agregar en una dimensión
    - Semana → Mes
- **Drill down :**
  - Détallar en una dimensión
    - Mes → Semana
- **Rebanar:**
  - Selección y proyectar en un eje
    - Mes = 04-2003 ; Proyectar(Región, Producto)
- **Pivote :**
  - Girar cubo para visualizar un lado
    - (Región,Producto)→(Región, Mes)

# Herramienta Multidimensional Especializada

- Beneficios:
  - Acceso rápido a grandes volúmenes de datos
  - Bibliotecas extensas de funciones complejas de análisis
  - Capacidades de modelado y predicción
  - Puede acceder a las estructuras de bases de datos multidimensionales y relacionales
  - Especial para los campos calculados
- Desventajas:
  - Difícil cambiar el modelo
  - Falta de apoyo en caso de grandes volúmenes de datos
  - Puede requerir capacidad de procesamiento

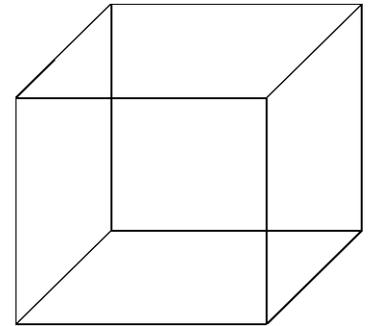
# On-Line Analytical Processing (OLAP)

- Idea básica: los usuarios deben poder manipular los modelos de datos empresariales a través de muchas dimensiones para comprender que se está ocurriendo.
- Los datos utilizados en OLAP deberían estar en la forma de un cubo multidimensional.

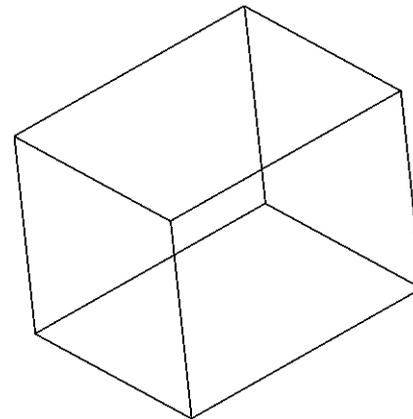


# Operaciones OLAP

- *Rollup*: decrease nivel de detalle
- *Drill-down*: aumenta nivel de detalle
- *Slice-and-dice*: selección y proyección



- *Pivot*: re-orienta vista multidimensional



# Implementando Multi-dimensionada

## Multi-dimensional databases (MDDB)

- Para hacer que las bases de datos relacionales manejen multidimensionalidad, dos tipos de tablas se introducen:
  - Tabla de hechos: contiene datos numéricos. Es larga y delgada.
  - Tablas de dimensiones: contienen punteros a la tabla de hechos. Muestra donde la información se puede encontrar. Una tabla separada se proporciona para cada dimensión. Las tablas de dimensiones son pequeñas, y anchas.

# Extension de SQL

- ROLLUP:
  - SELECT <column list>
  - FROM <table...>
  - GROUP BY  
ROLLUP(column\_list);
- Hace  $n+1$  agregaciones en una columna,  $n$  será el numero de columnas del grupo
- CUBO:
  - SELECT <column list>
  - FROM <table...>
  - GROUP BY  
CUBE(column\_list);
- Crea  $2^n$  combinaciones de agregaciones,  $n$  será el numero de columnas del grupo

# Ejemplo CUBO

Animal	Lieu	Quantite
Chien	Paris	12
Chat	Paris	18
Tortue	Rome	4
Chien	Rome	14
Chat	Naples	9
Chien	Naples	5
Tortue	Naples	1

- SELECT Animal, Lieu,  
SUM(Quantite) as Quantite  
FROM Animaux  
GROUP BY Animal, Magasin  
WITH CUBE

Animal	Lieu	Quantite
Chat	Paris	18
Chat	Naples	9
Chat	-	27
Chien	Paris	12
Chien	Naples	5
Chien	Rome	14
Chien	-	31
Tortue	Naples	1
Tortue	Rome	4
Tortue	-	5
-	-	63
-	Paris	30
-	Naples	15
-	Rome	18

# Ejemplo CUBO

```
SELECT item, city, year, SUM (amount)
FROM SALES
CUBE BY item, city, year
```

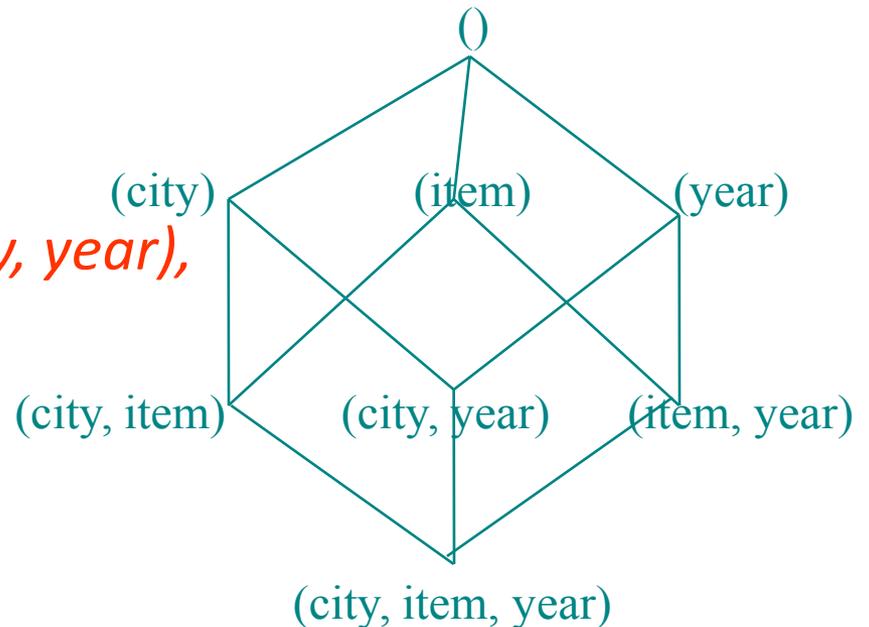
- Se debe calcular

*(item, city, year),*

*(item, city), (item, year), (city, year),*

*(item), (city), (year)*

*()*



# Extension de SQL

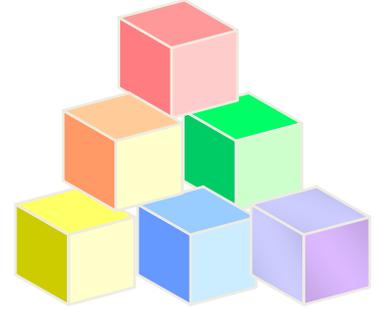
- Agrupar todos los subconjuntos de {item, área, mes}, encontrar el precio máximo en 1997 de cada grupo, y el total de ventas entre todas las tuplas de precios máximos

```
select item, region, mes, max(precio), sum(R.ventas)
from compras
where año = 1997
cube by item, region, mes: R
such that R.precio = max(precio)
```

# Exploración cubos de datos

- Exploración por el usuario, gran espacio de búsqueda
- Muchas medidas posibles guían al usuario en el análisis de los datos, en todos los niveles de agregación
- Excepción: Significativamente diferente del valor anticipado, basado en un modelo estadístico
- Las indicaciones visuales: como color de fondo se utilizan para reflejar el grado de excepción de cada celda

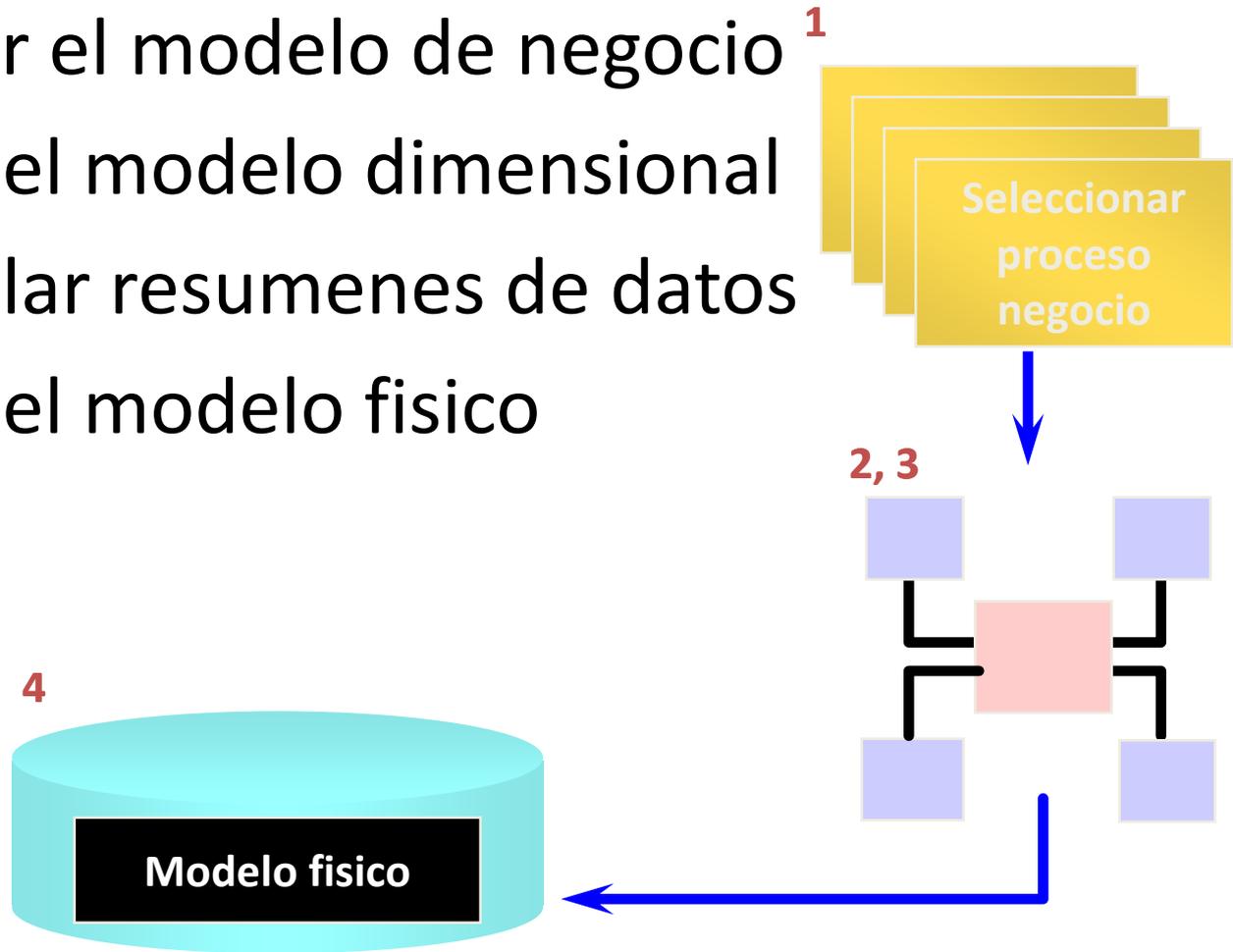
# Modelado en warehouse



- Warehouses difiere de las estructuras operativas:
  - requisitos analíticos
  - orientado a temas
- Los datos se deben orientar a temas:
  - Identificar temas de negocios
  - Definir las relaciones entre los temas
  - Definir los atributos de cada tema
- El modelado es iterativo

# Modelado en Data Warehouse

1. Definir el modelo de negocio
2. Crear el modelo dimensional
3. Modelar resúmenes de datos
4. Crear el modelo físico



# Identificar reglas del negocio

## Localidad

### Proximidad geográfica

0 - 1 kms

1 - 5 kms

> 5 kms

## Producto

### Tipo

### Monitor

### Estado

PC

15 pulg

Nuevo

Servidor

17 pulg

Reparac

19 pulg

operati

Nada

## Tiempo

Mes > Cuatrimestre > año

## Almacen

tienda > Distrito > Region

# Crear Modelo Dimensional

- Seleccionar una entidad para comenzar a armar tabla de hechos
- Determinar granularidad
- Identificar claves operacionales para tabla de hechos
- Buscar jerarquías para la tabla de hechos
- Añadir datos de las dimensiones
- Caracterizar los atributos de las dimensiones

# Granularidad (unidad de análisis)

Determina lo que representa cada registro de la tabla de hechos: el nivel de detalles.

- Ejemplos
  - Transacciones
  - Puntos en el tiempo
  - Lineas en un documento
- Depende del proyecto de IN

# Crear Modelo Dimensional

- Identificar tablas de hechos
  - Traducir medidas empresariales en tablas de hechos
  - Analizar las fuentes de datos para las medidas adicionales
  - Identificar tablas de dimensiones
- Enlazar tabla de hechos con las tablas de dimensiones
- Crear vistas para los usuarios

# Modelar resúmenes de datos

- Proporciona un acceso rápido a los datos precalculados
- Reduce el uso de E/S, la CPU y la memoria
- Se calcula desde las fuentes de datos y otros resúmenes precalculados
- Por lo general, se guardan en las tablas de hechos

# Modelar resúmenes de datos

- Promedio
- Máximo
- Total
- Porcentaje

	Units	Sales(€)	Store
Producto A			
Total			
Producto B			
Total			
Producto C			
Total			

# La construcción de un Data Warehouse desde una base de datos

- Desarrollar un modelo de negocio entidad-relación de la data warehouse.
- Traducir esto en un modelo tridimensional. Este paso refleja las características de la información y de análisis a hacer.
- Traducir esto en el modelo físico.