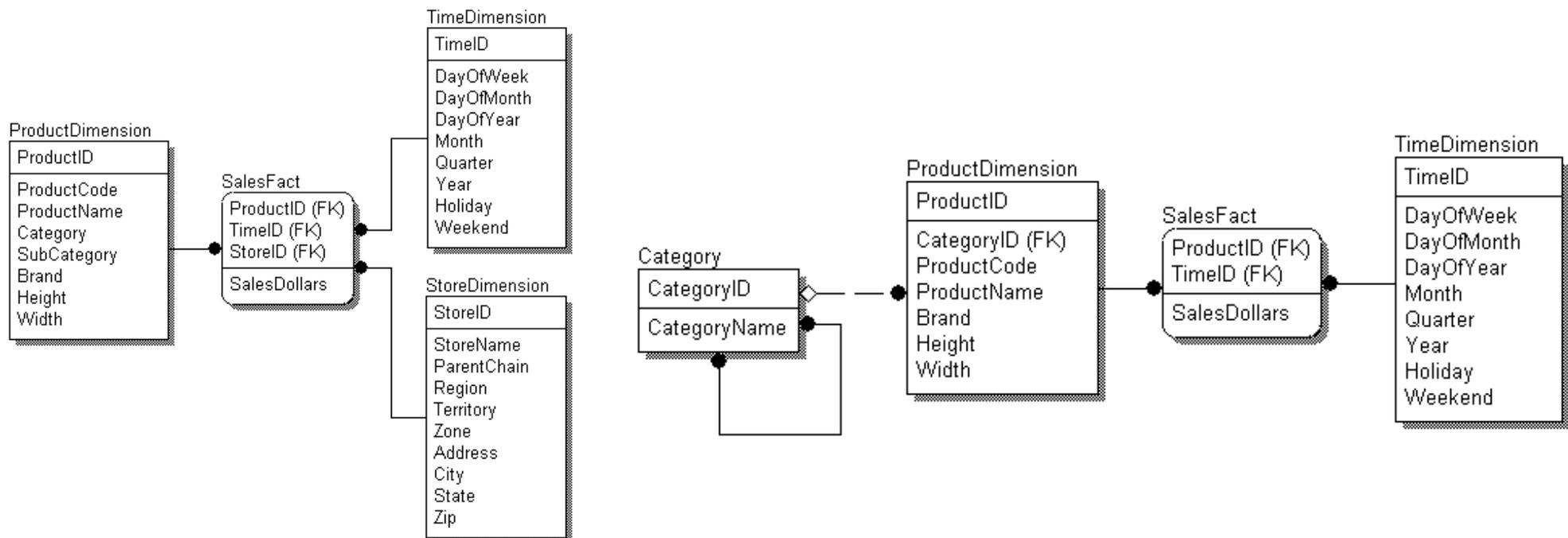


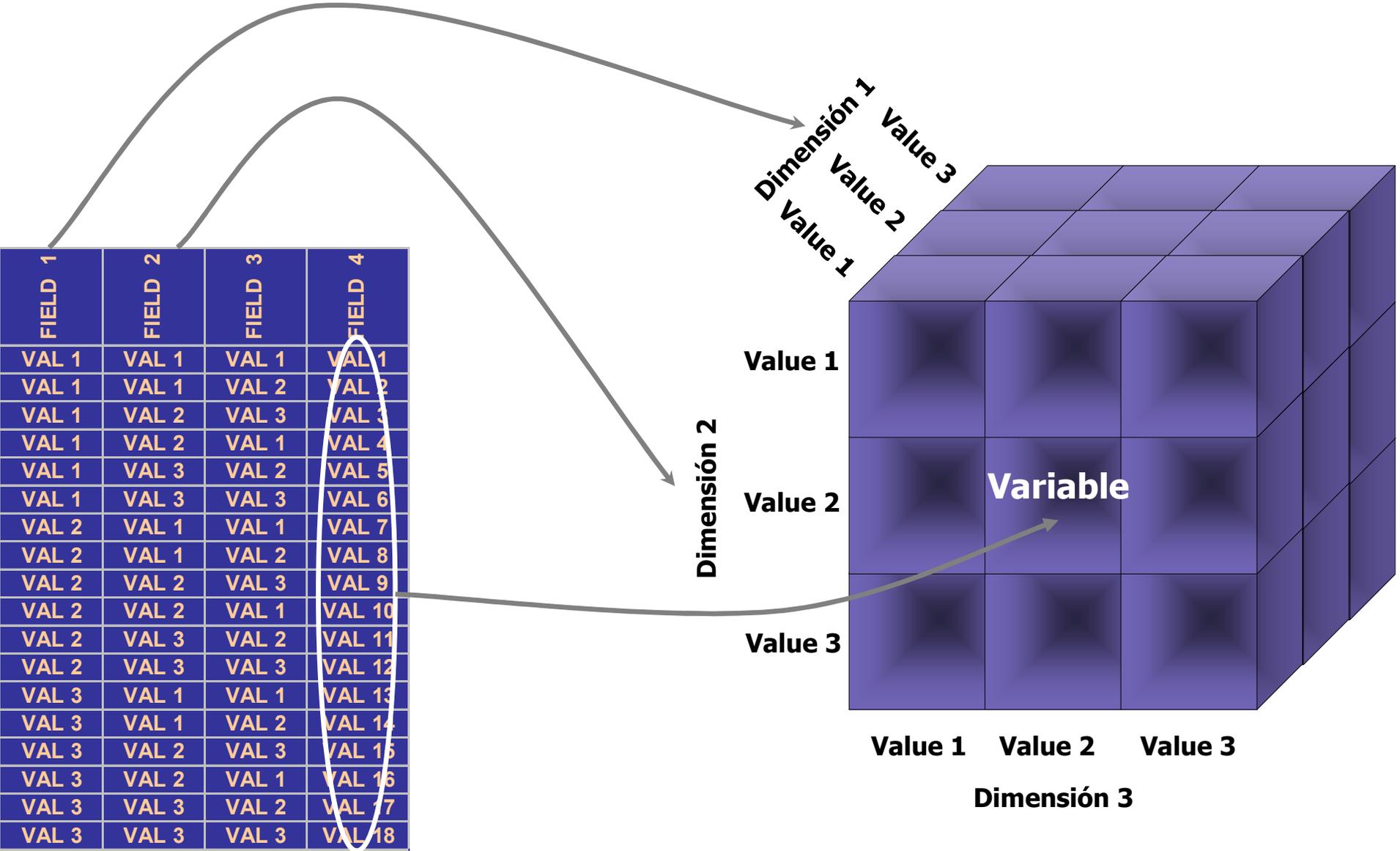
# Modelado del Data Warehouse: Estructuras de datos

Las estructuras de datos que son comunmente encontradas en un Data Warehouse son:

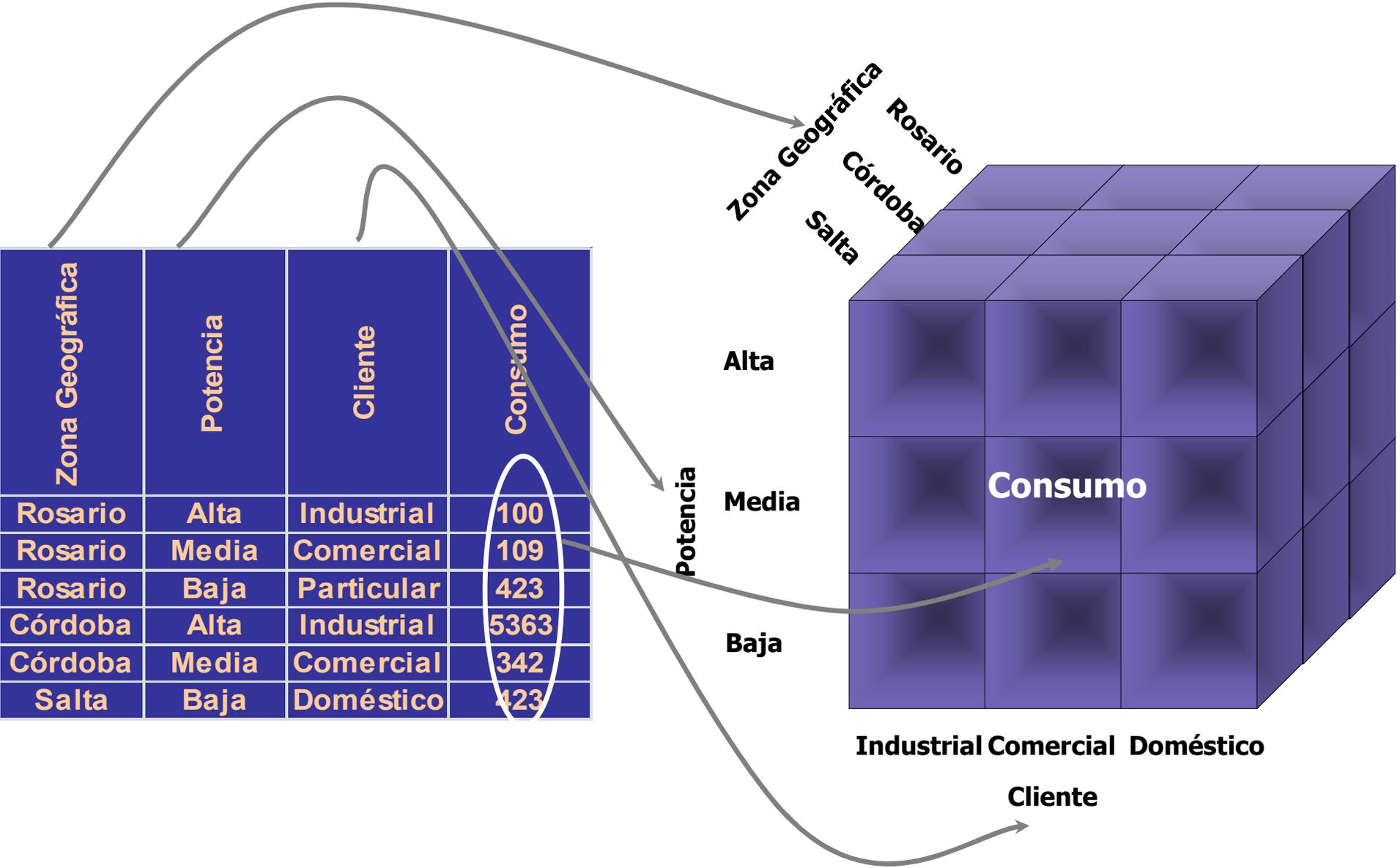
- ✓ Multidimensional:
  - Esquema estrella
  - Esquema copo de nieve



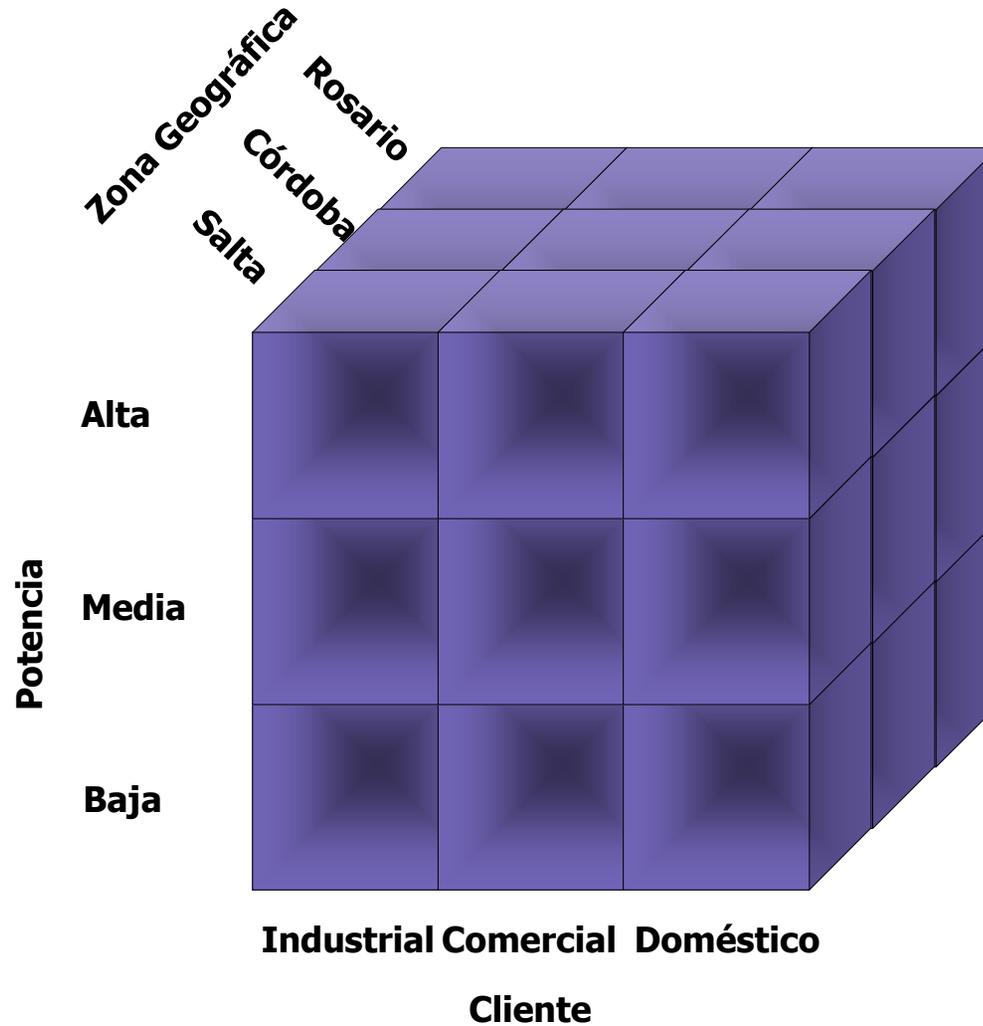
# Modelado del Data Warehouse: Modelo Multidimensional



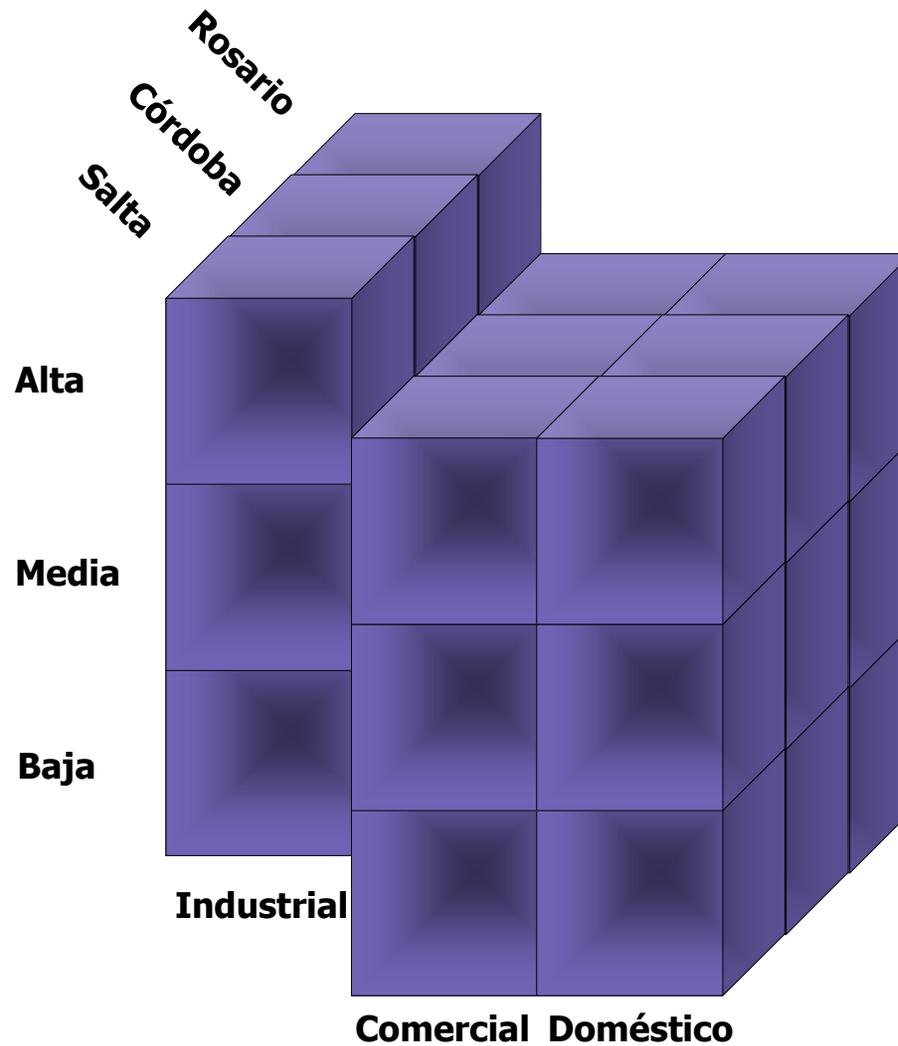
# Modelado del Data Warehouse: Consulta Multidimensional



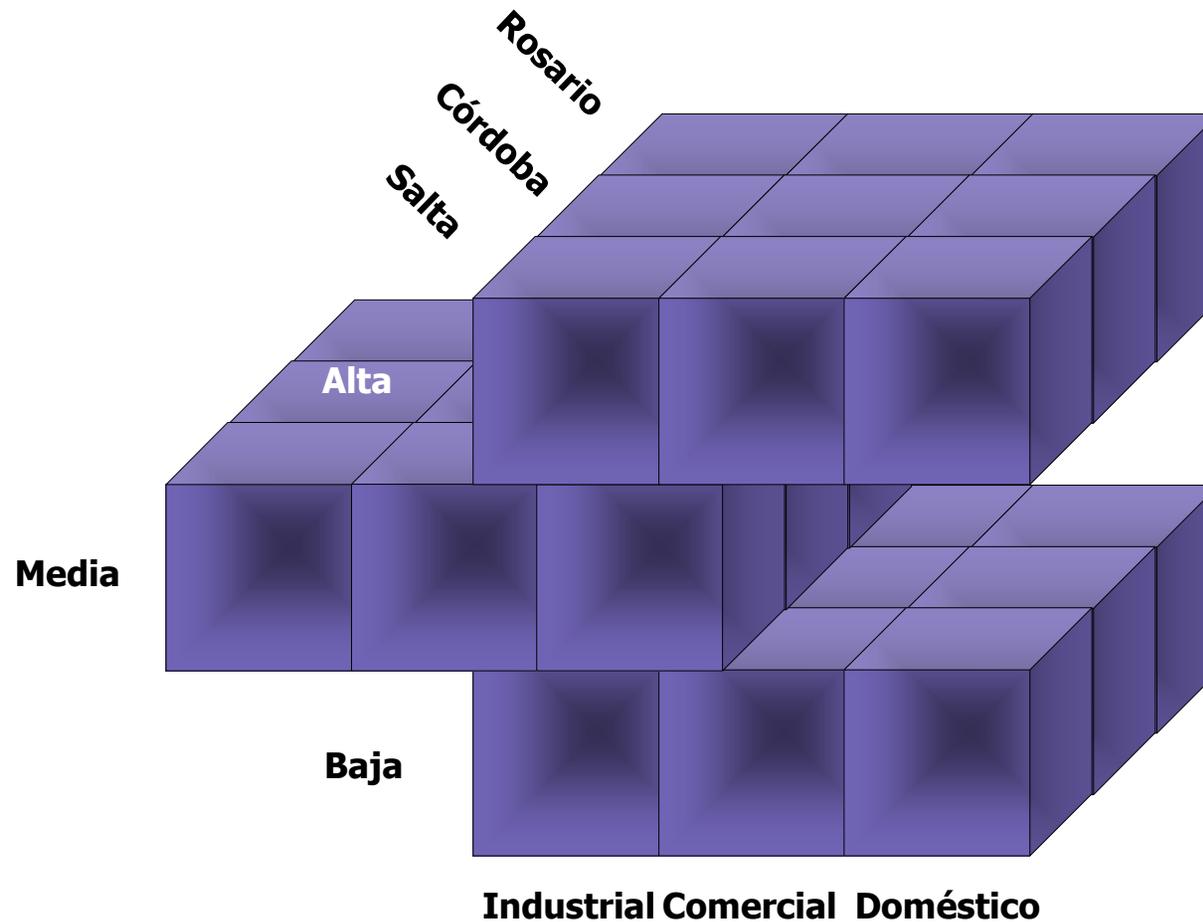
# Modelado del Data Warehouse: Consulta Multidimensional



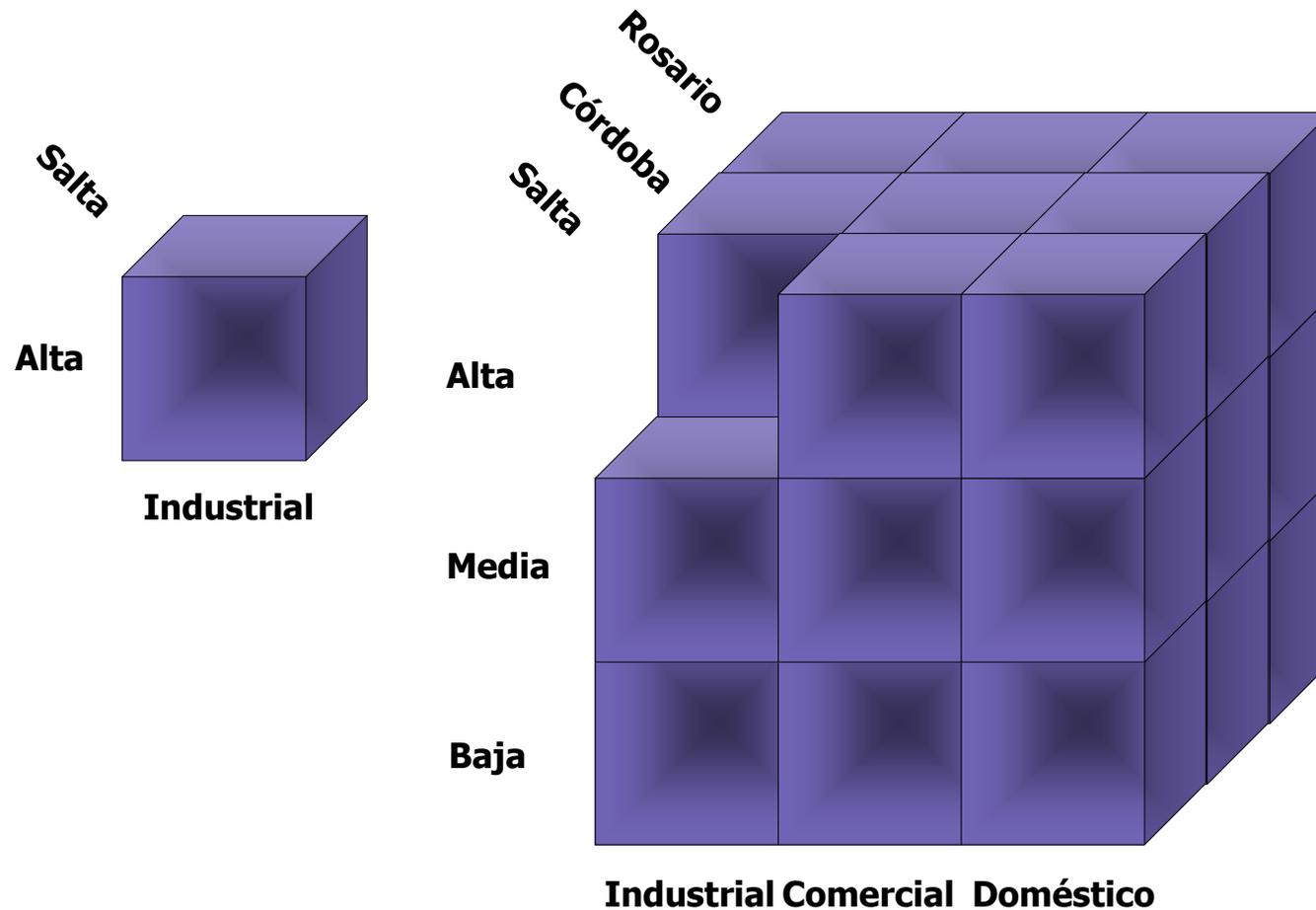
# Modelado del Data Warehouse: Consulta Multidimensional



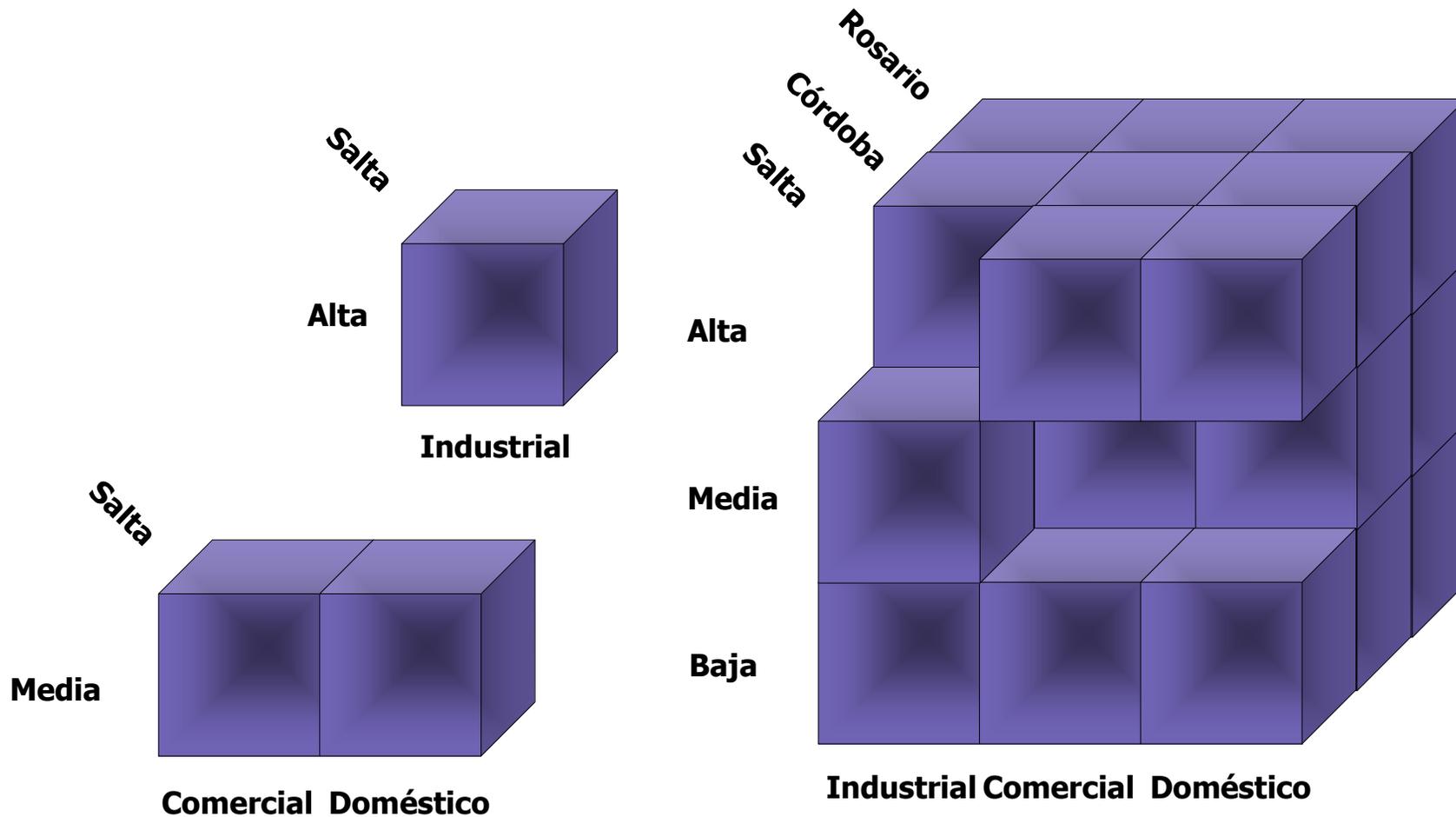
# Modelado del Data Warehouse: Consulta Multidimensional



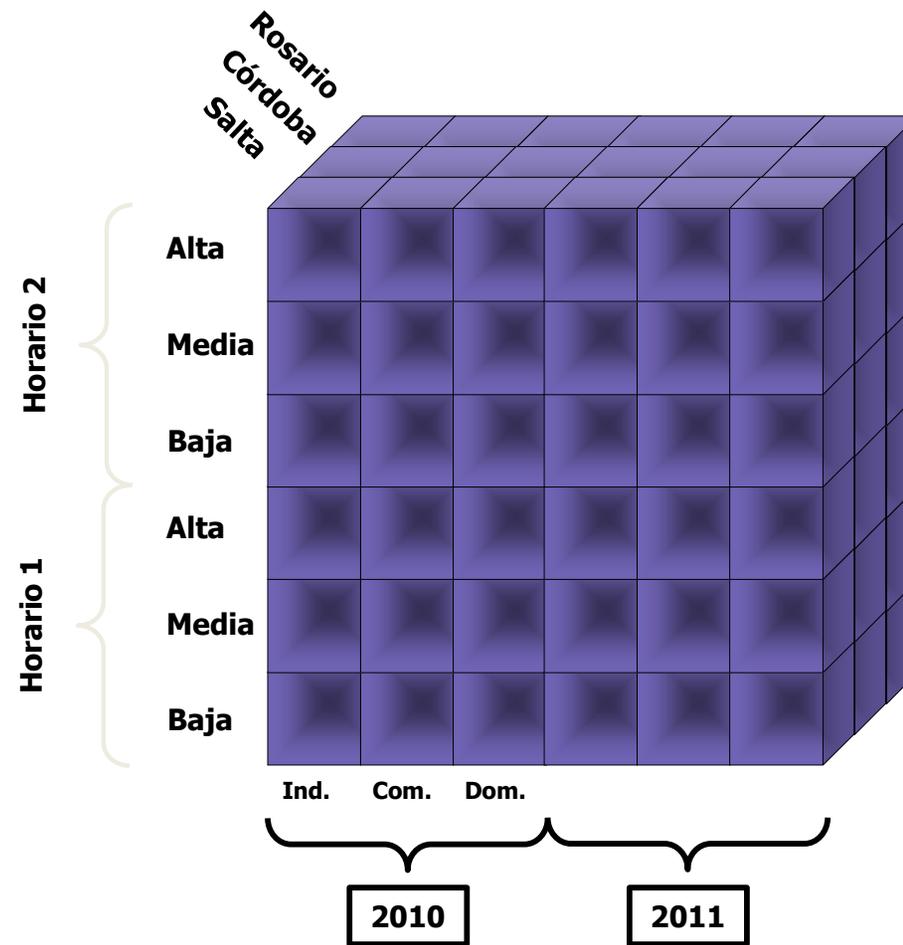
# Modelado del Data Warehouse: Consulta Multidimensional



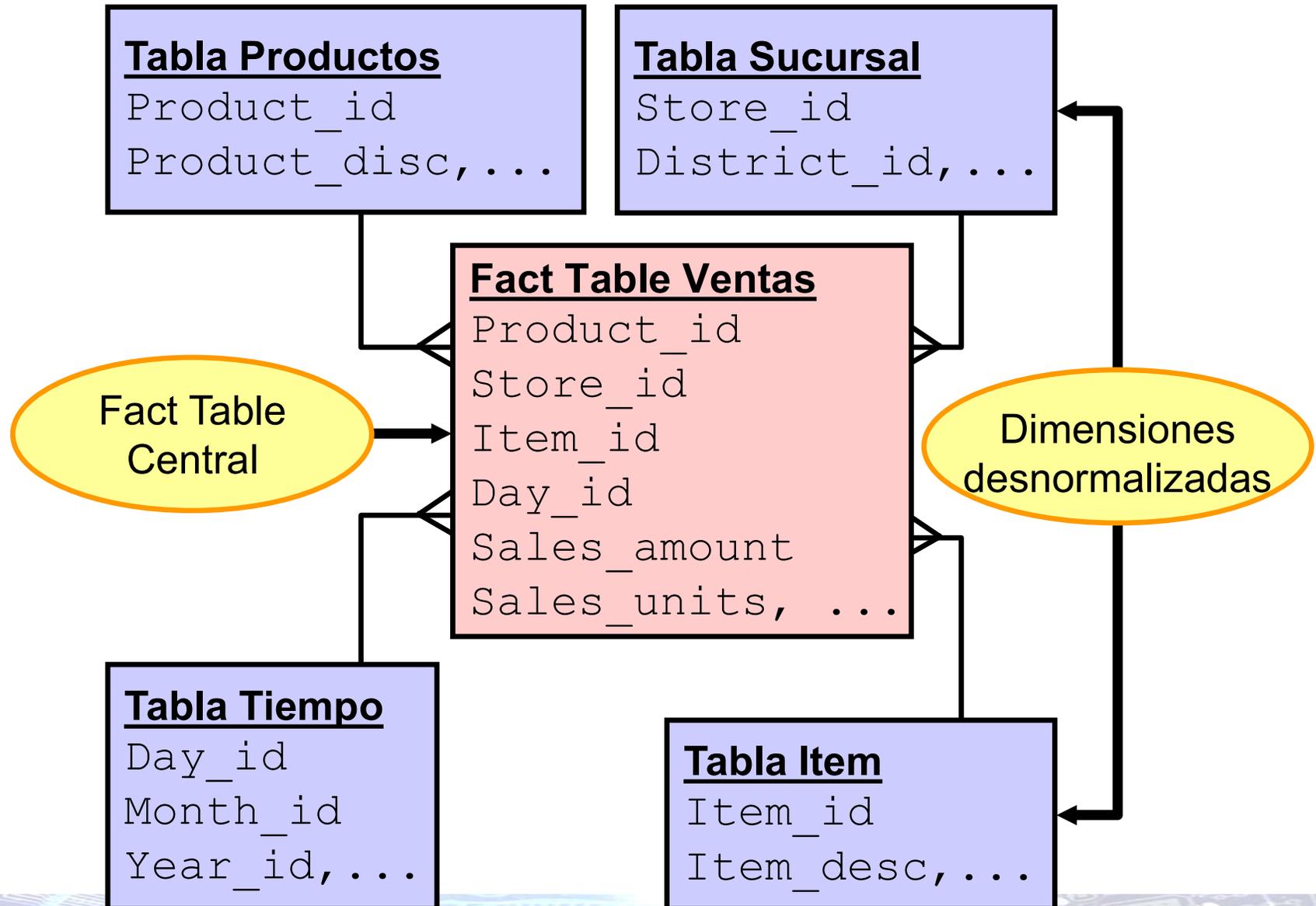
# Modelado del Data Warehouse: Consulta Multidimensional



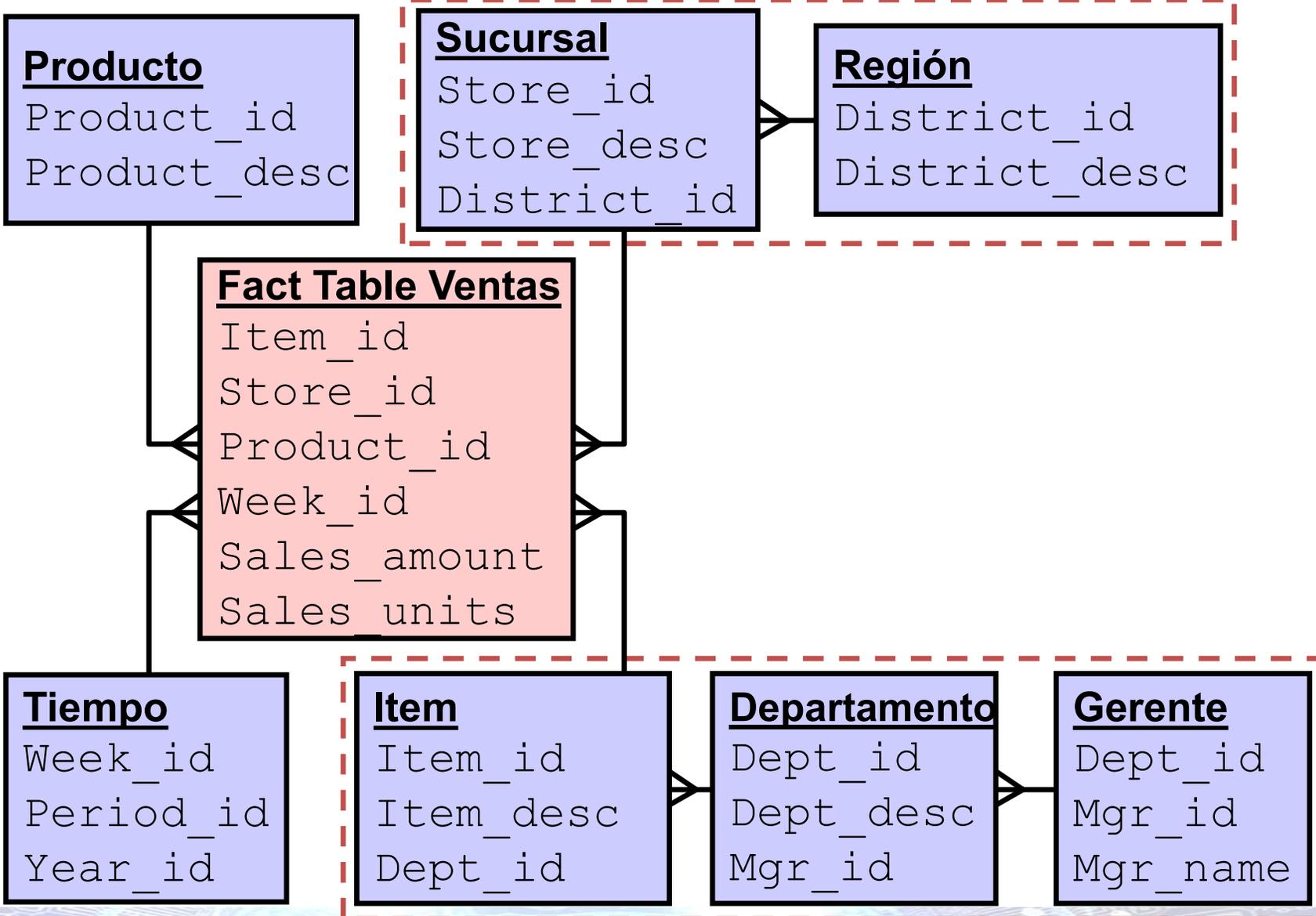
# Modelado del Data Warehouse: Consulta Multidimensional



# Modelado del Data Warehouse: Modelo estrella



# Modelado del Data Warehouse: Modelo copo de nieve



# Modelado del Data Warehouse: Fases de diseño del DW

- ✓ **Fase 1**
  - Definir el modelo de negocio
- ✓ **Fase 2**
  - Definir el modelo dimensional
- ✓ **Fase 3**
  - Definir el modelo físico



# Modelado del Data Warehouse: Fase 1: Definir el modelo de negocio

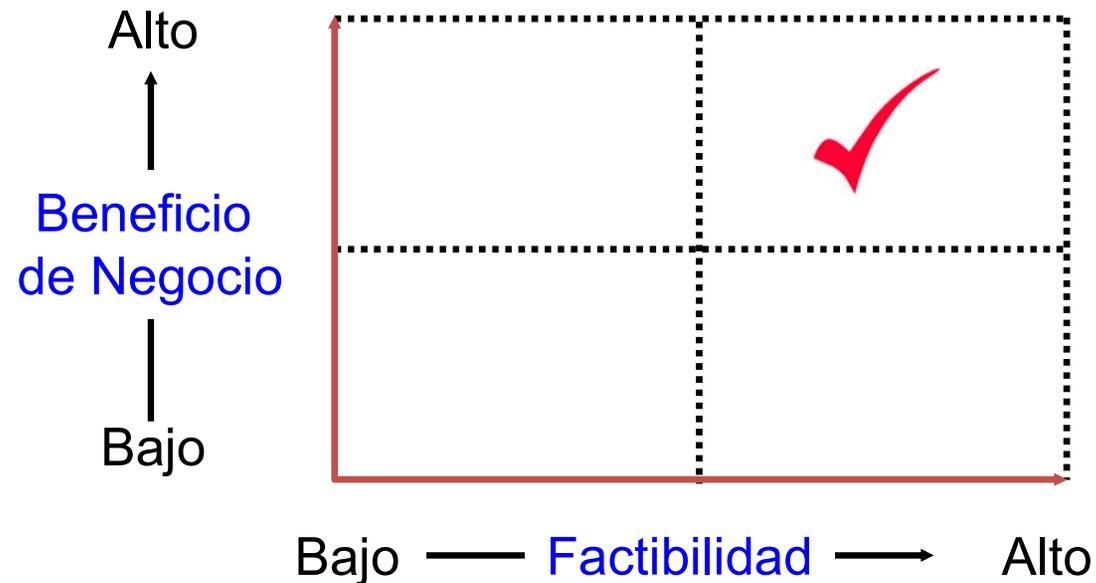
- 1. Ejecutar el análisis estratégico**
- 2. Crear el modelo de negocio**
- 3. Documentar la metadata**



# Modelado del Data Warehouse: Fase 1: Definir el modelo de negocio

## 1. Ejecutar el análisis estratégico:

- Identificar procesos de negocio cruciales
- Entender los procesos de negocio
- Priorizar y seleccionar los procesos de negocio a implementar

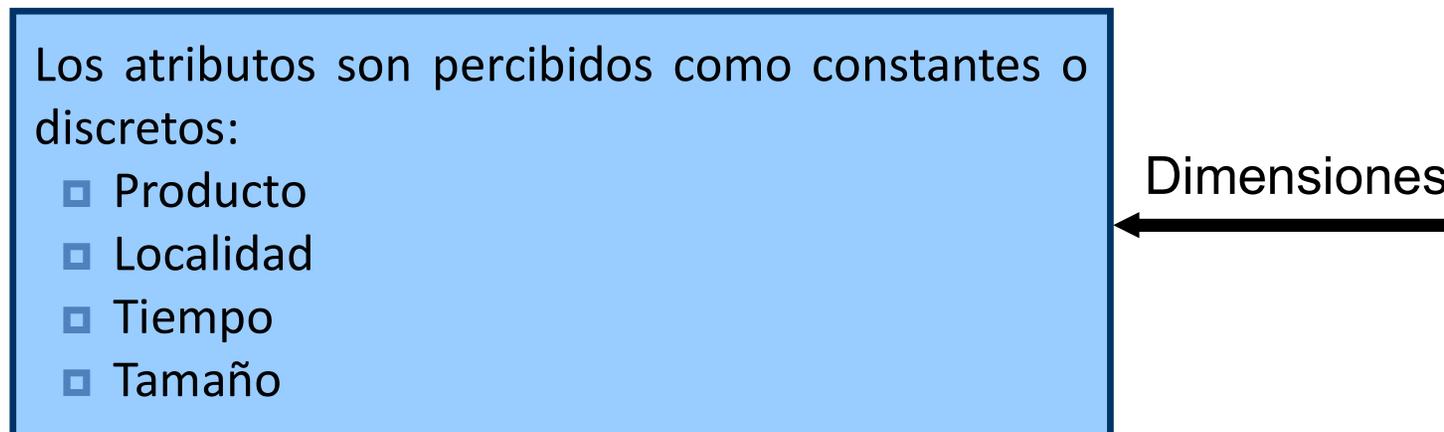
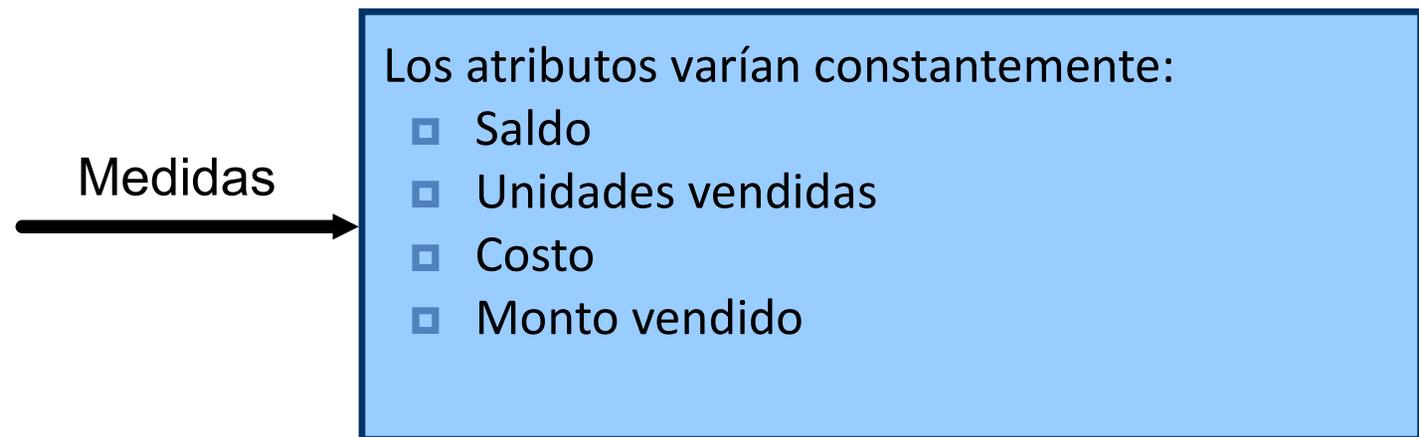


## 2. Crear el modelo de negocio:

- ✓ Definir los requerimientos de negocio
  - Identificar las medidas del negocio
  - Identificar las dimensiones
  - Identificar la granularidad
  - Identificar las definiciones de negocio y reglas
- ✓ Verificar las fuentes de datos



## Identificar medidas y dimensiones



# Modelado del Data Warehouse:

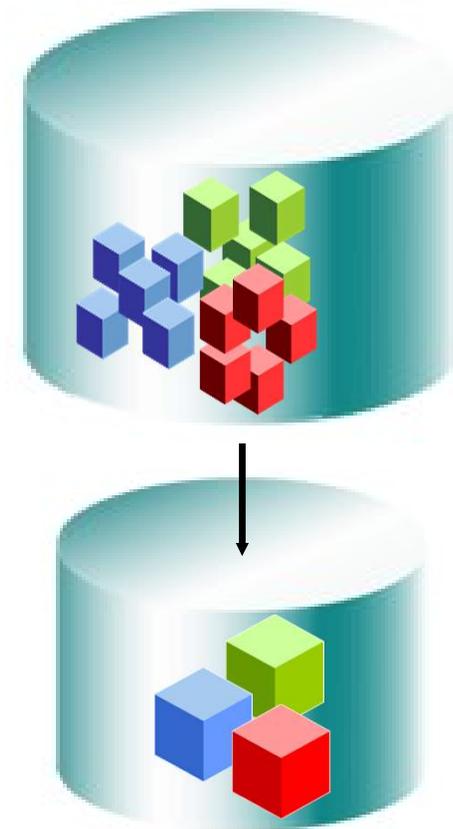
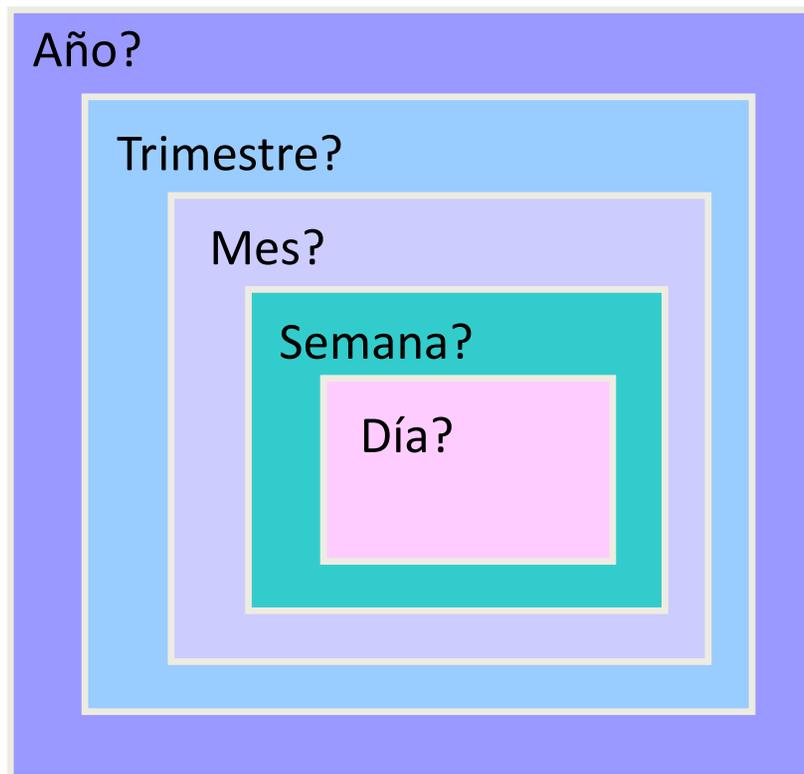
## Fase 1: Definir el modelo de negocio

### Matriz de proceso de negocio

Dimensiones de Negocio	Processes de Negocio		
	Ventas	Devolución	Inventario
Cliente	✓	✓	
Fecha	✓	✓	✓
Producto	✓	✓	✓
Canal	✓		
Promoción	✓	✓	

Ejemplo de matriz de proceso de negocio

## Determinar la granularidad

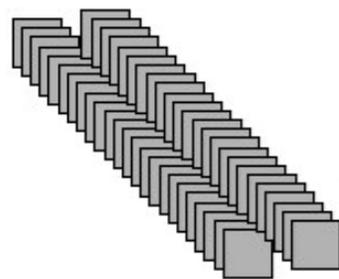


# Modelado del Data Warehouse: Fase 1: Definir el modelo de negocio

## Determinar la granularidad

### Alto nivel de detalle

El detalle de cada llamada realizado por un cliente en un mes.



40.000 bytes  
200 registros por mes

### Bajo nivel de detalle

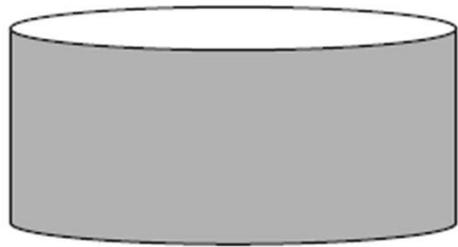
El resumen de llamadas realizadas por un cliente en un mes.



200 bytes  
1 registro por mes

## Determinar la granularidad

Debemos analizar el trade-off para encontrar la mejor solución



- Nivel de detalle para responder cualquier pregunta
- Gran volumen de datos



- Flexibilidad lo suficientemente pequeña para ser capaz de manipular
- Poco volumen de datos

### Identificar las reglas de negocio

Locación
<u>Proximidad geográfica</u>
0 - 1 km
1 - 5 km
> 5 km

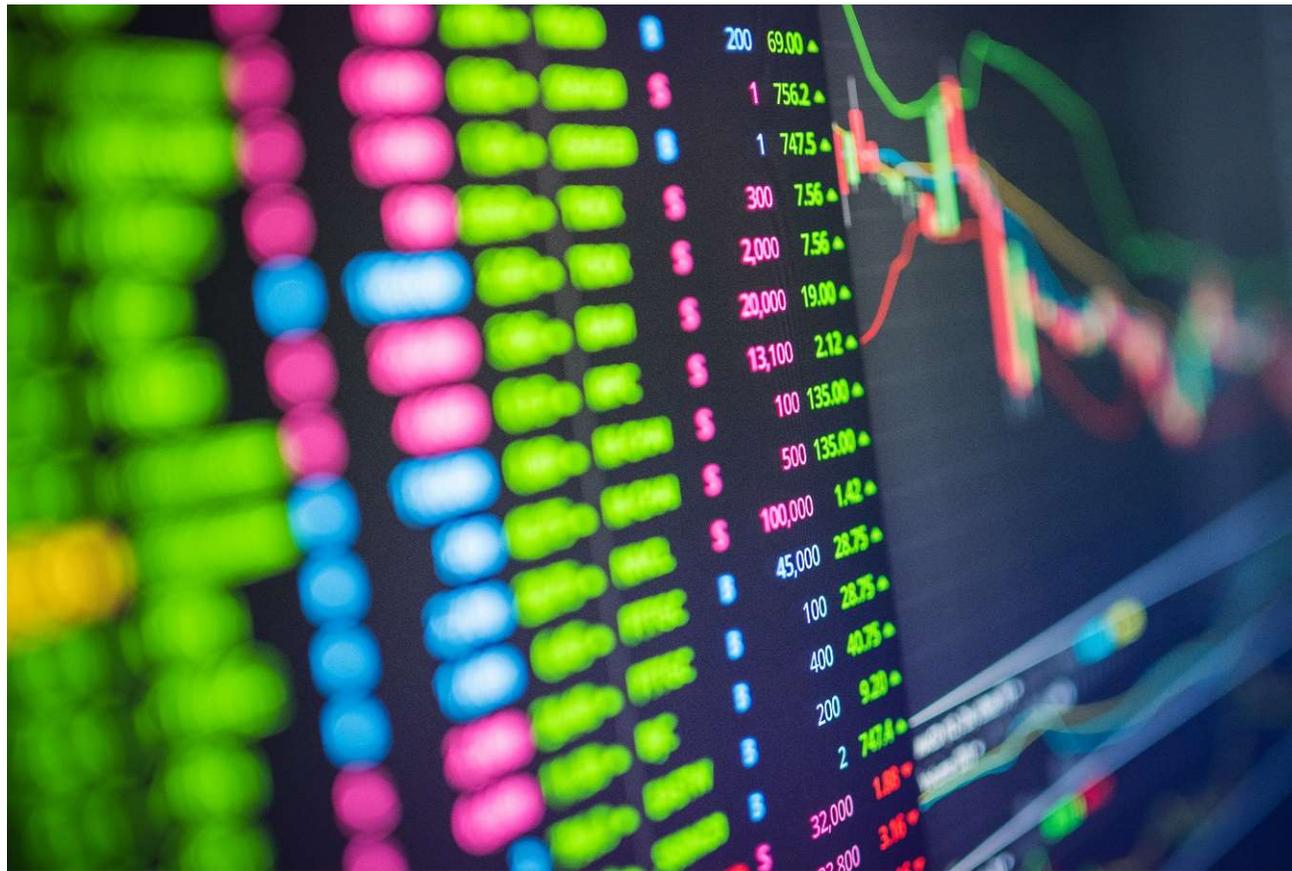
Producto		
<u>Tipo</u>	<u>Monitor</u>	<u>Estado</u>
PC	15 “	Nuevo
Server	17 “	Usado
	19 “	

Tiempo
Mes > Trimestre > Año

Sucursal
Sucursal > Distrito > Región

# Modelado del Data Warehouse: Fase 1: Definir el modelo de negocio

Verificar las fuentes de datos



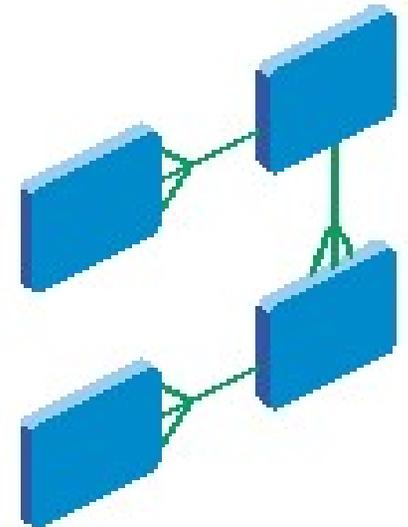
## Documentar la metadata

- ✓ Automatizada
  - Herramienta de modelado de datos
  - Herramienta de ETL
  - Herramienta de usuario final
- ✓ Manual



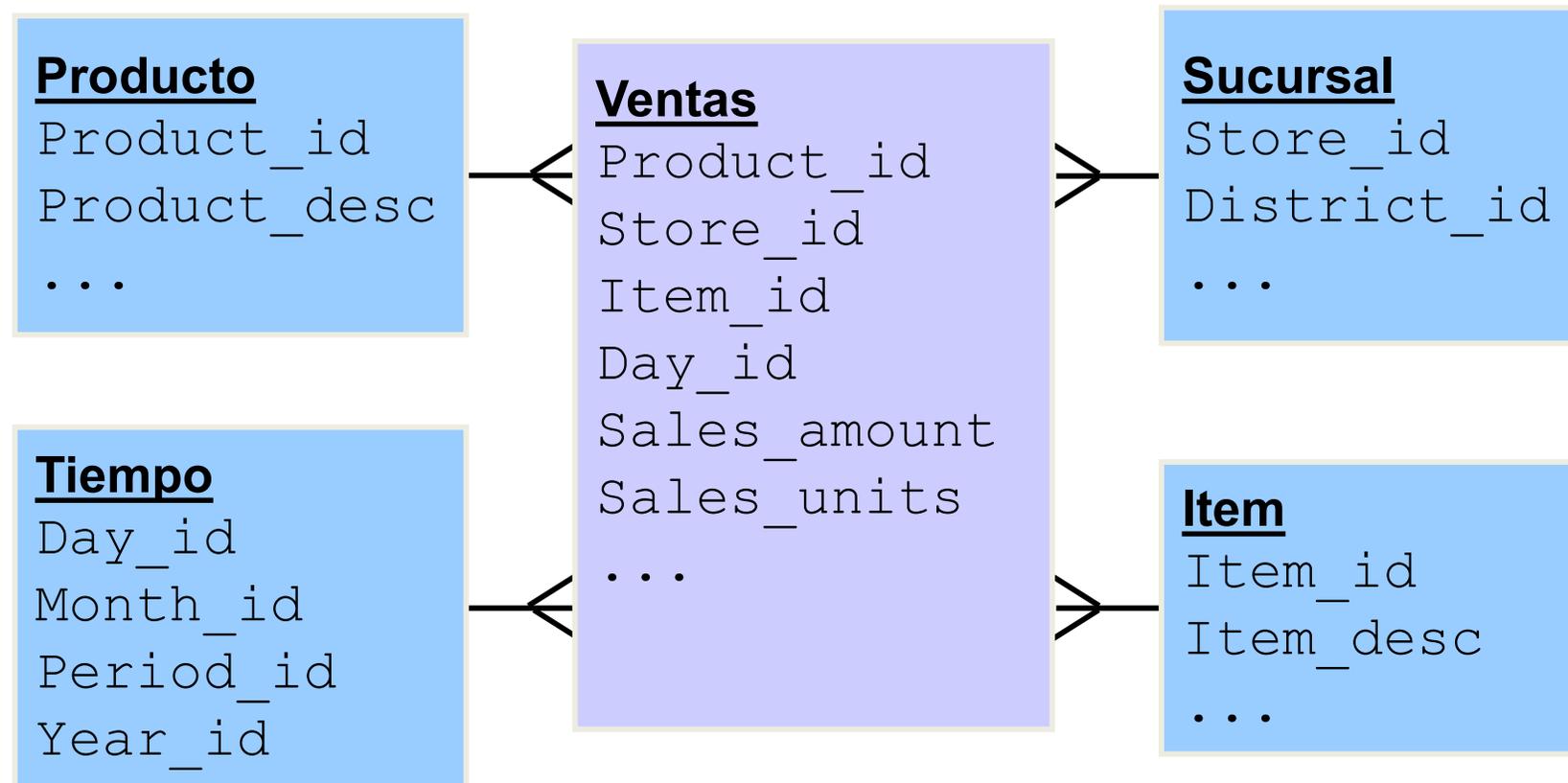
# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

- ✓ Identificar fact tables
  - Trasladar medidas de negocio en fact tables
  - Analizar los sistemas fuentes para encontrar medidas adicionales
- ✓ Identificar tablas de dimensiones
- ✓ Unir las fact tables a las tablas de dimensiones
- ✓ Modelar la dimensión tiempo



# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

## Modelado dimensional estrella



# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

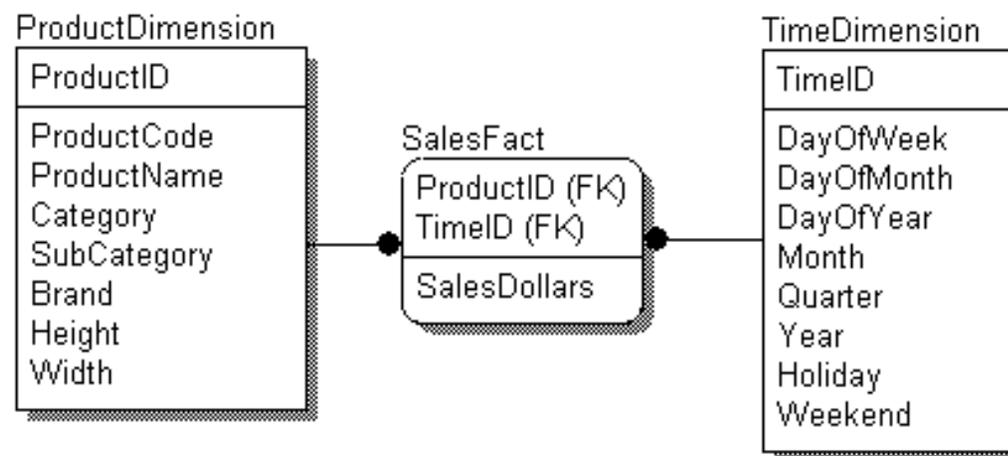
- ✓ Características de la fact table
  - Contienen medidas numéricas del negocio
  - Pueden contener grandes volúmenes de datos
  - Pueden crecer rápidamente
  - Pueden contener datos bases, derivados y resumizados
  - Son típicamente aditivas
  - Se relacionan a las tablas de dimensiones a través de foreign keys

## Ventas

```
Product_id  
Store_id  
Item_id  
Day_id  
Sales_amount  
Sales_units  
...
```

# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

- ✓ Características de las tablas de dimensiones
  - Contienen información textual que representan los atributos del negocio
  - Contienen datos relativamente estáticos
  - Son relacionadas a las fact tables a través de referencias de foreign key



# Modelado del Data Warehouse:

## Fase 2: Definir el modelo dimensional

- ✓ Reglas clave de las relaciones en MD
  - **Claves foráneas en la tabla de hechos** apuntan a las **claves primarias de las dimensiones**.
  - **Cardinalidad:** cada dimensión tiene una relación **1:N** con la tabla de hechos.
  - No debe haber relaciones directas entre dimensiones (no se recomienda en star schema).
  - En Power BI, todas las relaciones deben estar activas y ser unidireccionales (por defecto).
  - NO incluir valores nulos o duplicados en las claves primarias de las dimensiones.

# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

- ✓ Características del modelo estrella
  - El modelo es fácil de entender para los usuarios
  - Las claves primarias representan una dimensión
  - Los campos no claves son valores
  - Los tiempos de respuesta a las consultas son rápidos
  - La performance es mejorada reduciendo las relaciones
  - El usuario final puede realizar consultas complejas
  - El soporte es provisto por varias herramientas de front-end

# Modelado del Data Warehouse: Fase 2: Definir el modelo dimensional

- ✓ Uso del tiempo en el DW
  - La definición del tiempo en el DW es crítico
  - La agregación basada en el tiempo es bastante compleja



## IMPORTANTE:

En el proceso de pasar de un modelo **Entidad-Relación (ER)** a un **modelo Dimensional**, es común que aparezcan **atributos adicionales en las dimensiones que no estaban explícitamente definidos** en el modelo ER. Estas descripciones adicionales suelen tener su origen en:

1. Normalización previa: atributos estaban en otras tablas relacionadas
2. Atributos derivados o calculados
3. Requerimientos de análisis del negocio
4. Jerarquías predefinidas de negocio
5. Necesidades de trazabilidad histórica

Normalización previa: atributos estaban en otras tablas relacionadas

En el modelo ER, muchos atributos descriptivos están **distribuidos en múltiples tablas normalizadas** para cumplir las formas normales (1FN, 2FN, 3FN). Al construir el modelo dimensional, estas tablas se **desnormalizan** para facilitar el análisis.

Ej:

- En ER, Cliente puede estar relacionado con Segmento, Region, Pais, etc (tablas separadas).
- En el MD, todo se incorpora en una sola tabla DimCliente

## Atributos derivados o calculados

Muchos atributos del modelo dimensional provienen de **derivaciones lógicas o cálculos**.

Ej:

- Trimestre se deriva de la fecha (Mes)
- $\text{TotalVenta} = \text{CantidadVendida} * \text{PrecioUnitario}$

## Requerimientos de análisis del negocio

En BI, los usuarios finales necesitan **contexto adicional** para los análisis. Esto motiva la inclusión de **atributos semánticos** que no son necesarios para la operación, pero sí para el análisis

Ej:

- Categoría, Subcategoría del producto, Zona del vendedor, tipo de canal, etc

Estos pueden provenir de:

- Sistemas auxiliares (ERP, CRM)
- Manuales o catálogos de negocio
- Archivos externos (Excel, sistemas satélite)

## Jerarquías predefinidas de negocio

Las dimensiones frecuentemente se modelan con **jerarquías** que no existen explícitamente en el modelo ER.

Ej:

- En DimProducto: Subcategoría... Categoría... Línea
- En DimFecha: Día... Mes... Trimestre... Año

Estas jerarquías pueden estar implícitas en claves o en atributos externos.

## Necesidades de trazabilidad histórica

El modelo dimensional puede incorporar **atributos para gestión de cambios lentos** (Slowly Changing Dimensions - SCD), como fechas de vigencia, versiones, estados previos.

Ej:

- FechaInicioVigencia, FechaFinVigencia, EsActual