



Sistemas de Inteligencia de Negocios

Prof. Lorena Zúñiga S.

Febrero, 2020

¿Qué es un Data Warehouse?

Es una base de datos

“Orientada a un tema, integrada,
variante en el tiempo, no- volátil”

William Inmon

Características

- Orientada a un área
- Integrada
- Variante en el tiempo
- No volátil

Componentes

- Sistemas operacionales
- Área de staging
- Área de Presentación de datos
- Herramientas de acceso

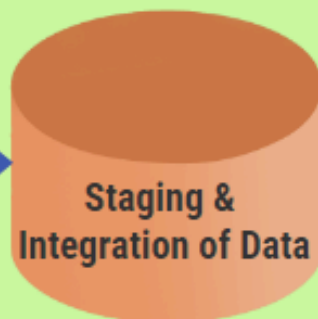
Different Layers of Data Warehouse Architecture

Data Source

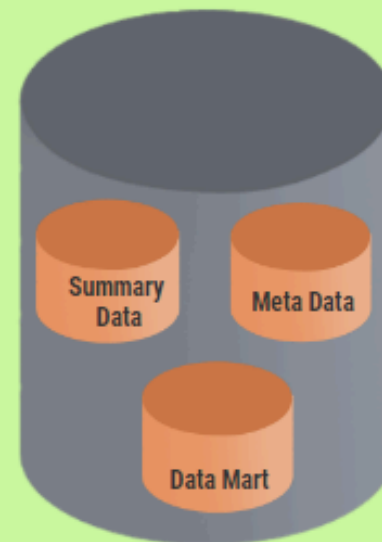


ETL

Data Staging



Data Storage



Data Presentation



Analysis of Data



Reporting Tool



Data Mining

Top Tier



Query/Report



Analytics



Data Mining

Front-end tools



Output

Middle Tier



OLAP Server



OLAP Server

OLAP Server

Bottom Tier



Monitoring



Administration



Metadata
Repository



Data Warehouse



Data Marts

Data Warehouse Server



Extract | Clean | Transform
Load | Refresh



Operational Databases



External Sources

Data

Datamart

- Versión especializada de un *DW*
- Es sobre un área de negocios específica (servicio al cliente, mercadeo, finanzas, entre otros)
- Puede servir como herramienta analítica para las necesidades de un departamento

Construcción

1. Identificar necesidades y requerimientos.
2. Reconocer fuentes de datos originales y sus estructuras.
3. De acuerdo a los requerimientos, definir las tablas auxiliares y los procesos de selección, transformación e importación de datos.
4. Construir el esquema multidimensional.

DISEÑO DEL ALMACÉN DE DATOS

Diseño del DW

Modelado Dimensional

Dimensión: colección de atributos relacionados

– Ejemplos:

- Productos
- Ubicaciones
- Fechas
- Clientes

Diseño del DW

Modelado Dimensional

Hechos (facts):

- datos operativos almacenados.
- Son medidas.
 - Ejemplo:
 - Cantidad de unidades vendidas
 - Monto de ventas
 - Precio unitario
 - % Descuento promedio

Tablas de hechos

- Contienen
 - Gran cantidad de registros
 - Pocos atributos
 - Reflejan el nivel de granularidad de los datos

Tablas de hechos

- Contienen
 - Llaves de tablas de dimensión
 - Medidas totalmente aditivas
 - Medidas semi – aditivas (atributos derivados)

Medidas

- **Aditivas**
 - Se pueden sumarizar por todas las dimensiones
 - Ejemplo: ventas, costos, ganancias
- **Semi – aditivas**
 - Se pueden sumarizar por algunas dimensiones, pero no todas (tiempo)
 - Ejemplo: niveles de inventario, saldos de una cuenta

Medidas

- **No aditivas**
- No es posible sumarizarlas por ninguna dimensión
- Ejemplo: precio unitario (histórico)

Medidas

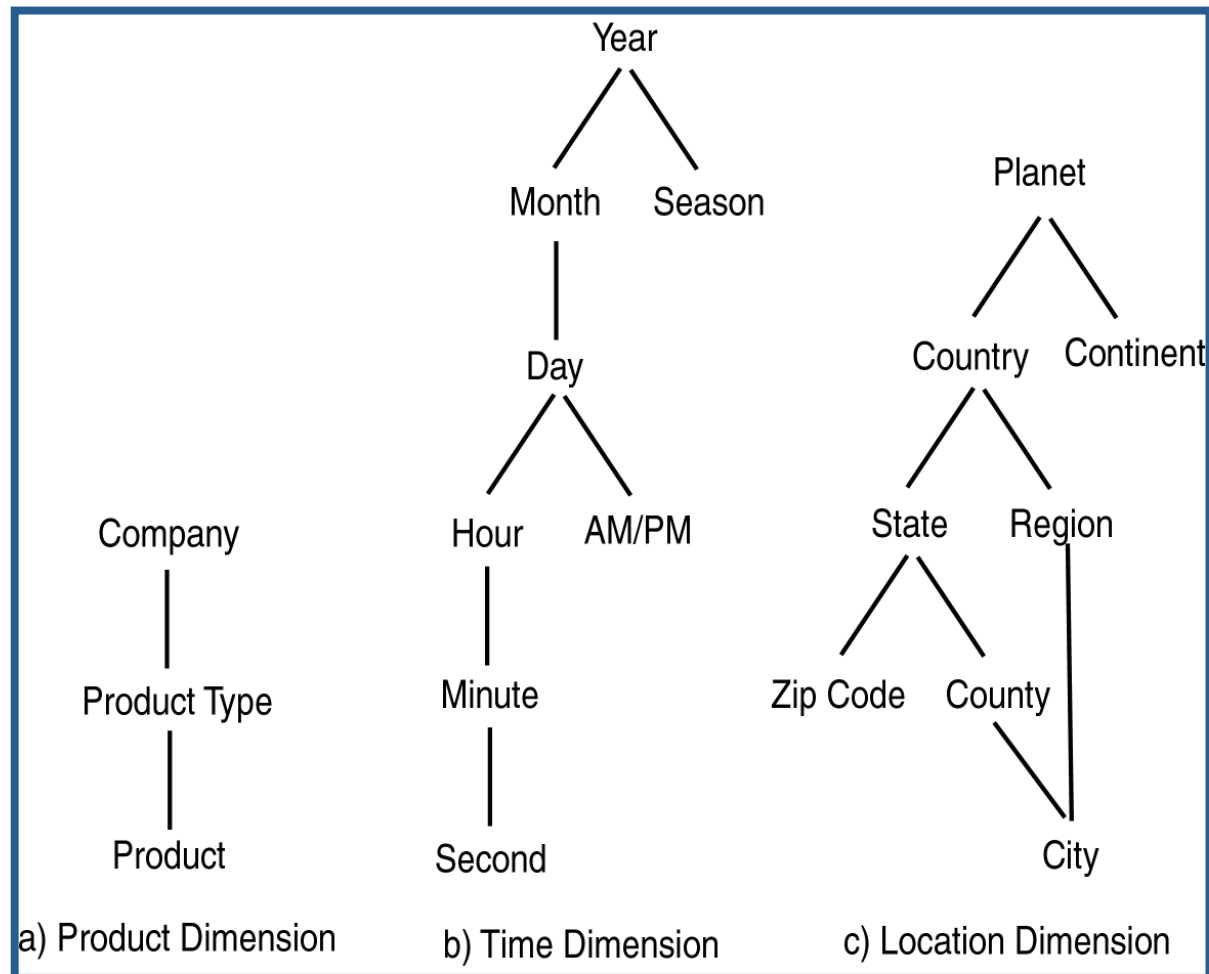
- Una medida es una columna cuantitativa en la tabla de hechos.
- Representan los valores que son analizados.



Agregaciones

- Son resúmenes de datos pre-calculados
- Mejoran el tiempo de respuesta, ya que tienen preparadas las respuestas antes de que se planteen las preguntas.

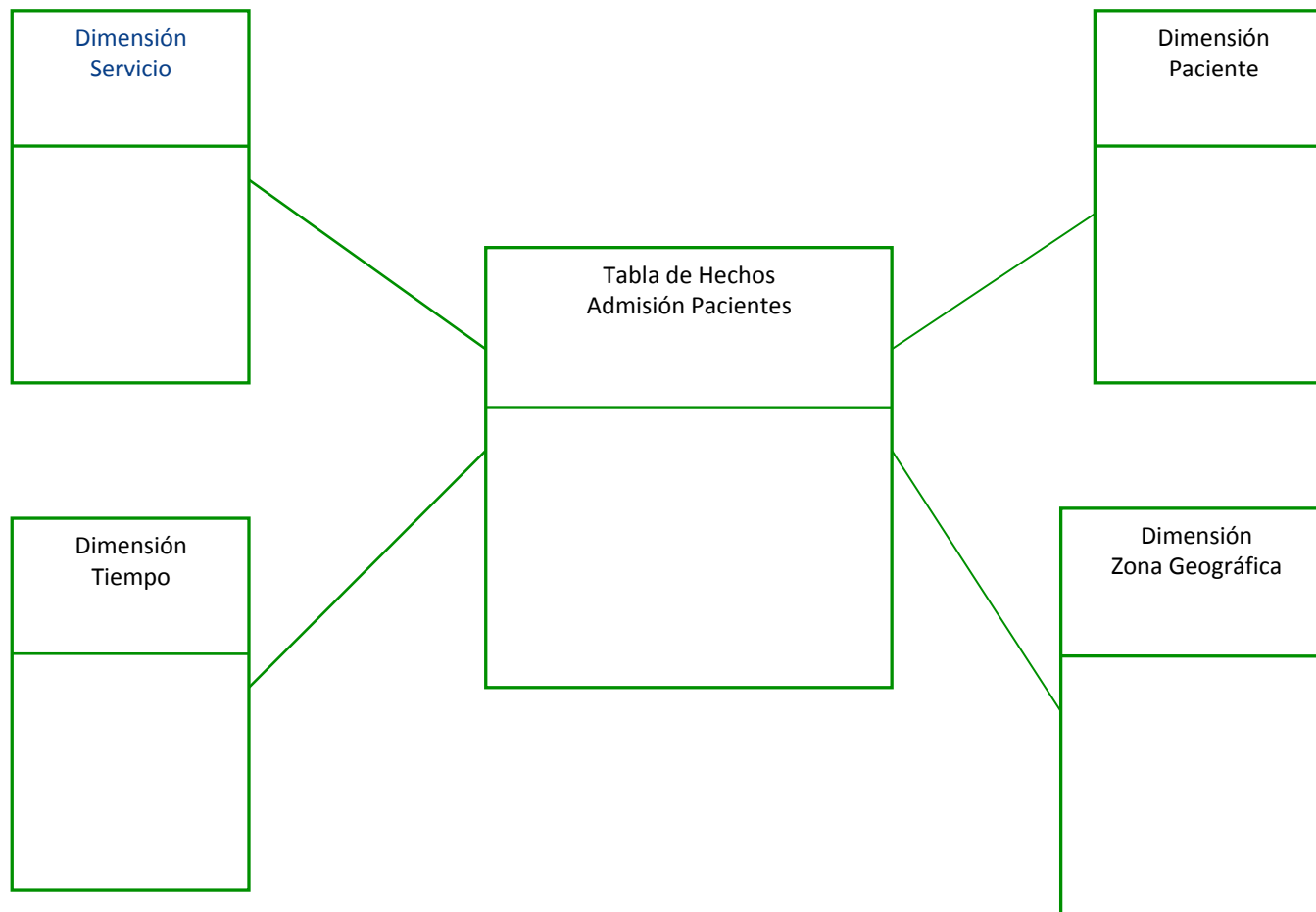
Jerarquías de agregación



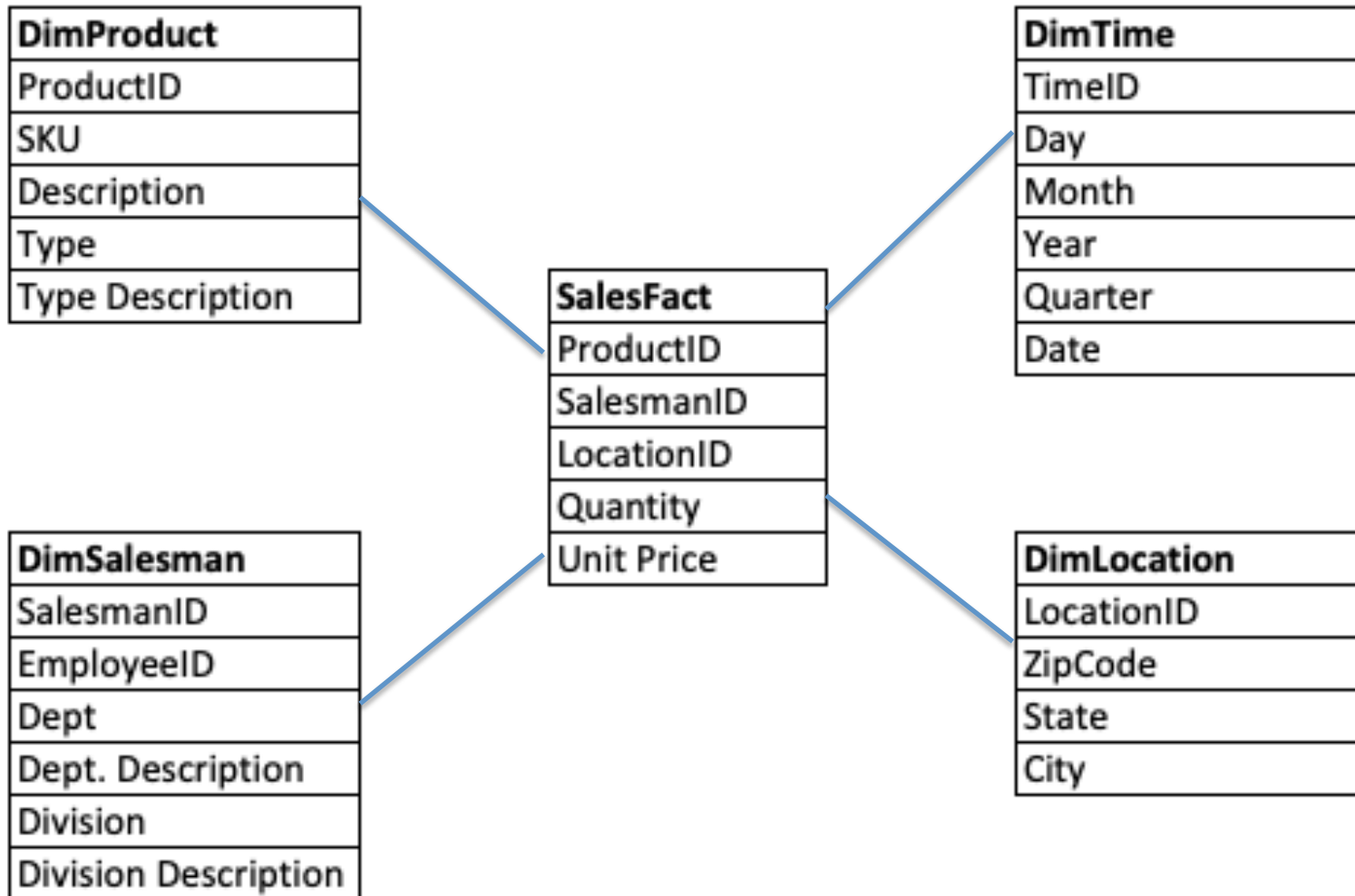
Modelos dimensionales

- El **esquema estrella** muestra hechos y dimensiones:
 - Centro de la estrella: hechos en tablas de hechos
 - Cada dimensión se muestra por separado en tablas de dimensión.

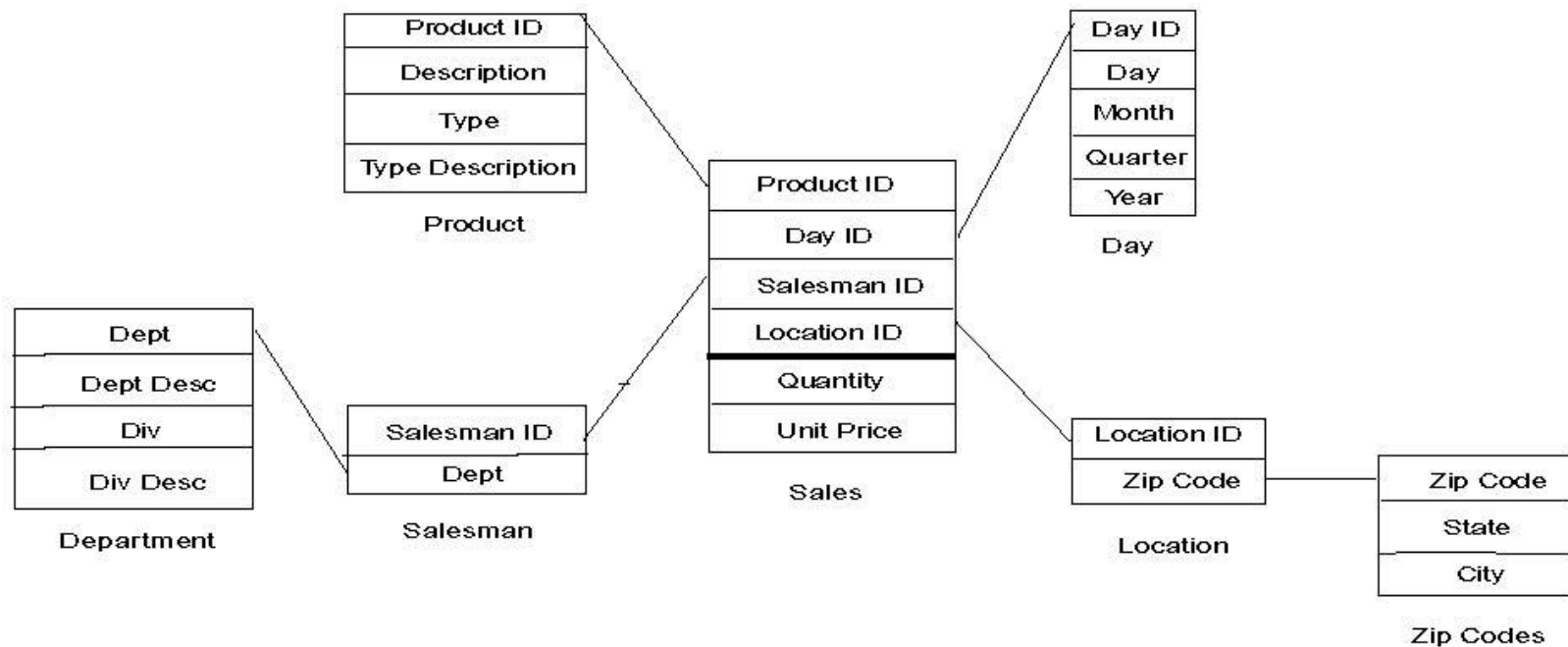
Esquema estrella



Esquema estrella



Esquema copo de nieve



Llaves en el esquema estrella

- **Llaves primarias**
 - en una dimensión: cada fila está identificada de manera única
 - ¿Se deben utilizar las mismas llaves de la fuente operacional ?

Llaves en el esquema estrella

- **Llaves subrogadas**
 - implica mantener la llave 'original' como un atributo adicional en la dimensión
- **Llaves foráneas**
 - tablas de hechos

OnLine Analytical Processing (OLAP)

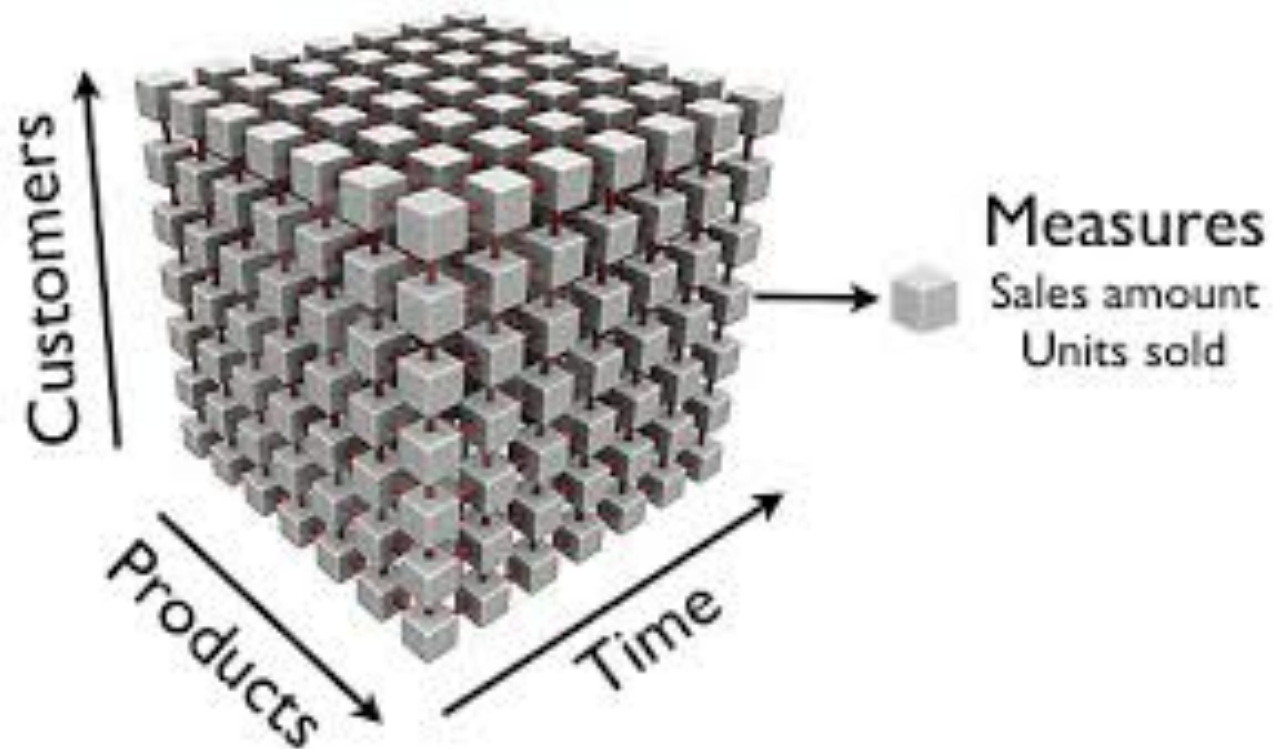
“Es una categoría de software que posibilita a analistas, gerentes y ejecutivos profundizar en los datos, **mediante acceso interactivo, rápido, consistente, en una amplia variedad de posibles vistas de información**, que ha sido transformada a partir de datos crudos para reflejar la dimensionalidad real de una empresa tal y como **la entiende el usuario** ”

OLAP Council

OLAP

- **Objetivos**
 - Representar los datos en forma más cercana a la intuición del usuario.
 - Resolver problemas planteados en sistemas relacionales.
- La información se representa como:
 - cuadros de doble o triple entrada.
 - cubos de "n" dimensiones.

Cubos de datos



Características

- **Se representan los datos como una matriz.**
 - En los ejes están los criterios de análisis.
 - En los cruces están los valores a analizar.
 - A esta estructura se le llama Cubo

| | | | | |
|----------------------------|----------|-------|-----|-------|
| M O D E L O | Mini Van | 6 | 5 | 4 |
| | Coupe | 3 | 5 | 5 |
| | Sedan | 4 | 3 | 2 |
| | | Blue | Red | White |
| | | COLOR | | |

Características

- **Se representan los datos como una matriz.**
 - En los ejes están los criterios de análisis.
 - En los cruces están los valores a analizar.
 - A esta estructura se le llama Cubo

| | | | | |
|--------------------------------------|-----------|---------|---------|---------|
| P R O D U C T O | | | | |
| | Farmacia | 100 | 200 | 75 |
| | Abarrotes | 87 | 62 | 30 |
| | Lácteos | 59 | 45 | 38 |
| | | Tienda1 | Tienda2 | Tienda3 |

Estructura de un cubo

–Dimensiones:

- Criterios de análisis de los datos.
- Objetos del problema.
- Variables independientes.
- Ejes

Estructura de un cubo

–Medidas

- Valores o indicadores a analizar.
- Datos asociados a relaciones entre los objetos del problema.
- Variables dependientes.

Estructura de un cubo

- Se ubican en la intersección de algunos valores de las dimensiones.
- Dado un valor para cada dimensión se puede determinar un valor para la medida.

Estructura de un cubo

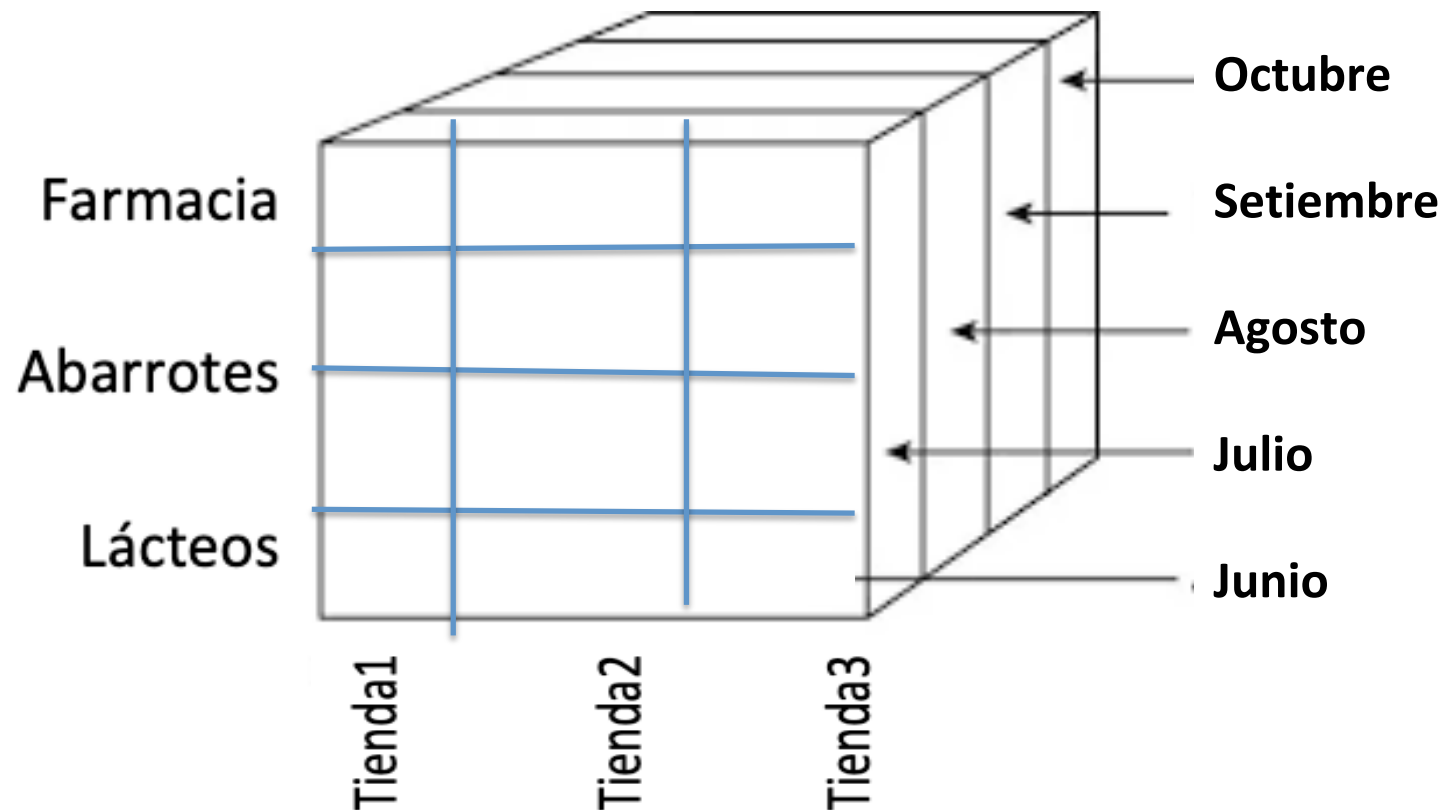
P
R
O
D
U
C
T
O

| | | | |
|-----------|------------|------------|-----------|
| Farmacia | 100 | 200 | 75 |
| Abarrotes | 87 | 62 | 30 |
| Lácteos | 59 | 45 | 38 |
| | Tienda1 | Tienda2 | Tienda3 |

Ventas(Farmacia,Tienda1) = 100

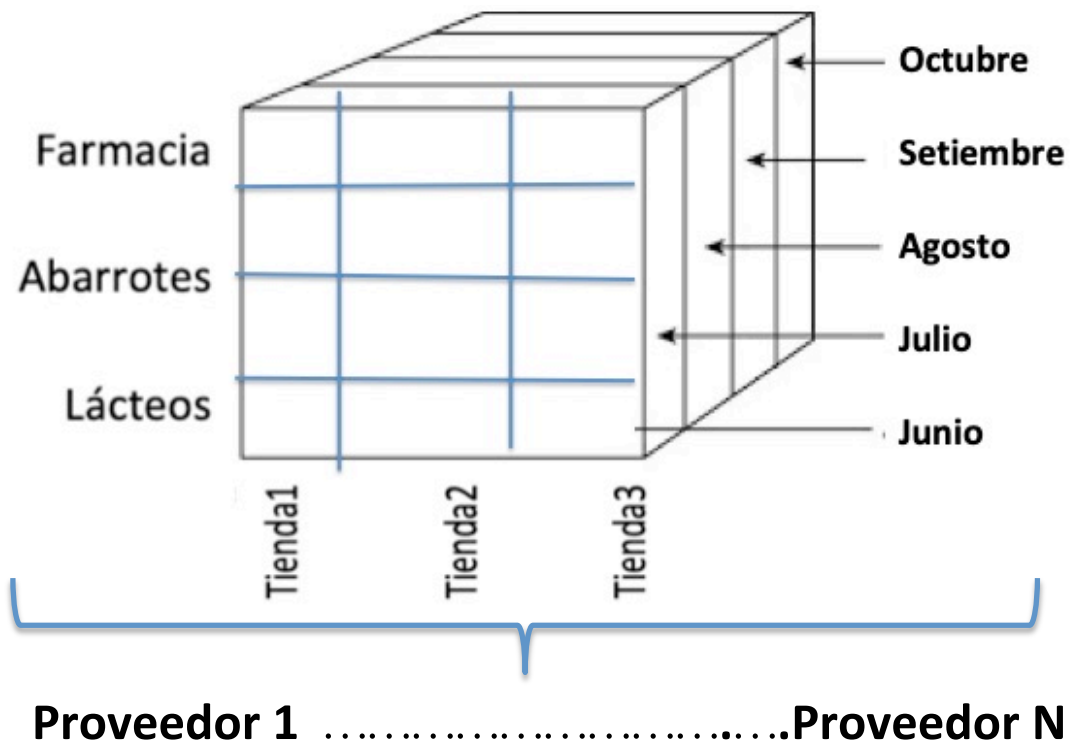
Ejemplo

- Si se agrega una 3a. dimensión:



Ejemplo

- Agregando una 4a. dimensión:



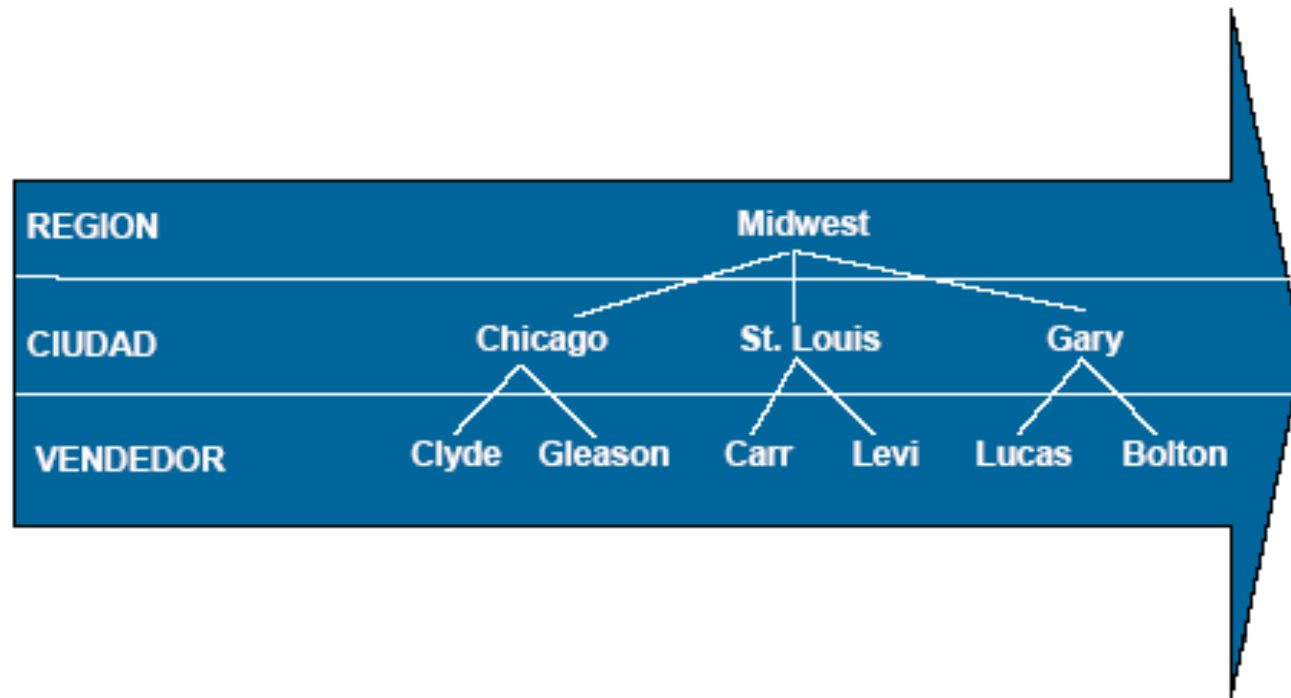
Ejemplo

- En el ejemplo anterior las **dimensiones** son:
 - Producto
 - Tiempo
 - Tienda
 - Proveedor
- Las **medidas** son:
 - Cantidad Vendida

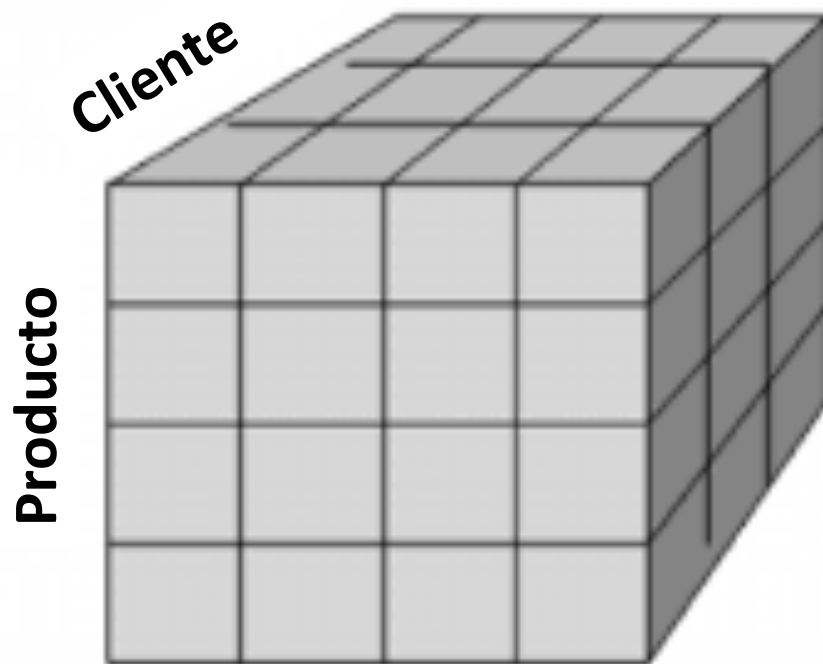
Dimensiones y Jerarquías

Ejemplo

– Dimensión: Vendedores



Dimensiones y Jerarquías



Producto
Categoría
|
Producto

Tiempo
Año
|
Mes
|
Fecha

Cliente
País
|
Estado
|
Ciudad
|
Cliente

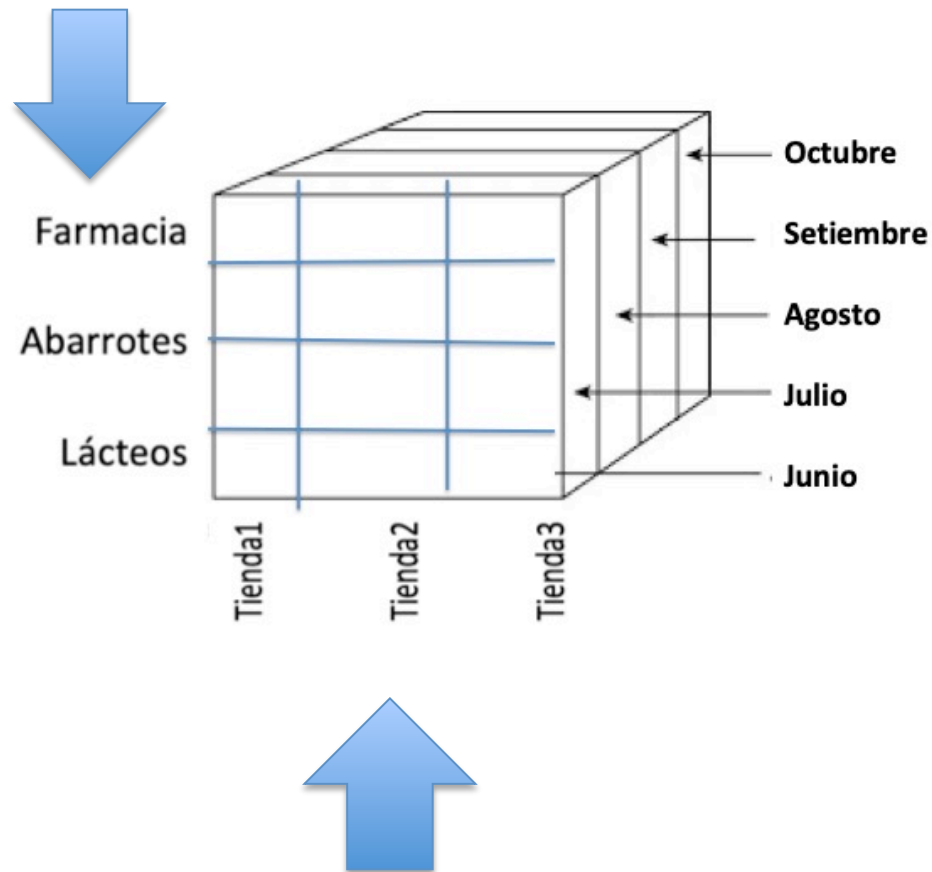
Operaciones

- Principales operaciones
 - Slice.
 - Dice.
 - Rotación.
 - Drill-down.
 - Drill-up.

Slice

- Seleccionar dimensiones
- Se define un subconjunto del cubo especificando sobre qué dimensiones interesa analizar qué medida.

Slice



P
R
O
D
U
C
T
O

| | | | |
|-----------|------------|------------|-----------|
| Farmacia | 100 | 200 | 75 |
| Abarrotes | 87 | 62 | 30 |
| Lácteos | 59 | 45 | 38 |
| | Tienda1 | Tienda2 | Tienda3 |

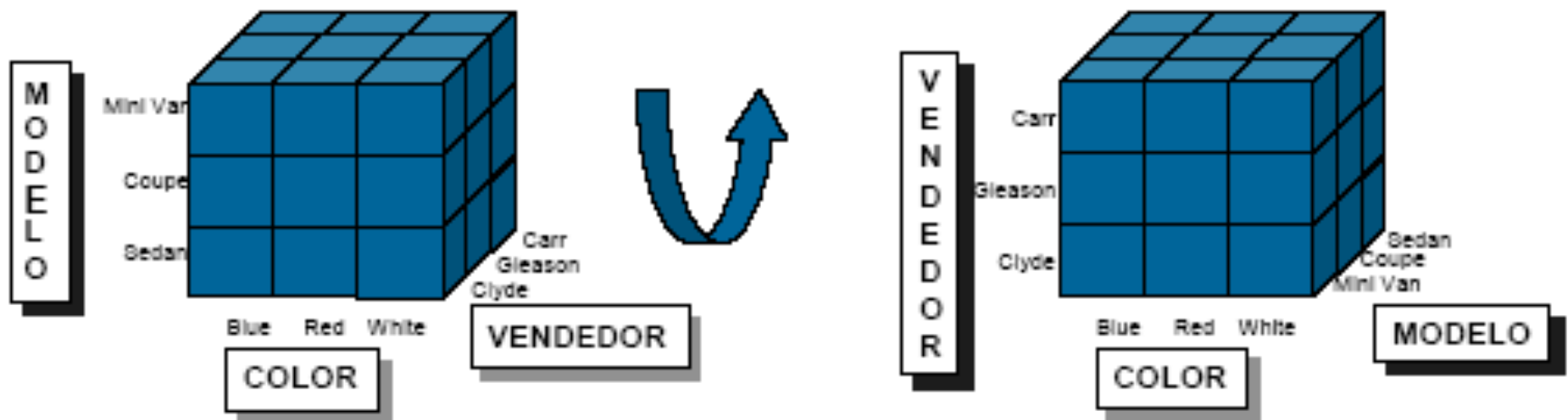
Dice

- Filtrado
- Se fijan valores para algunas dimensiones.

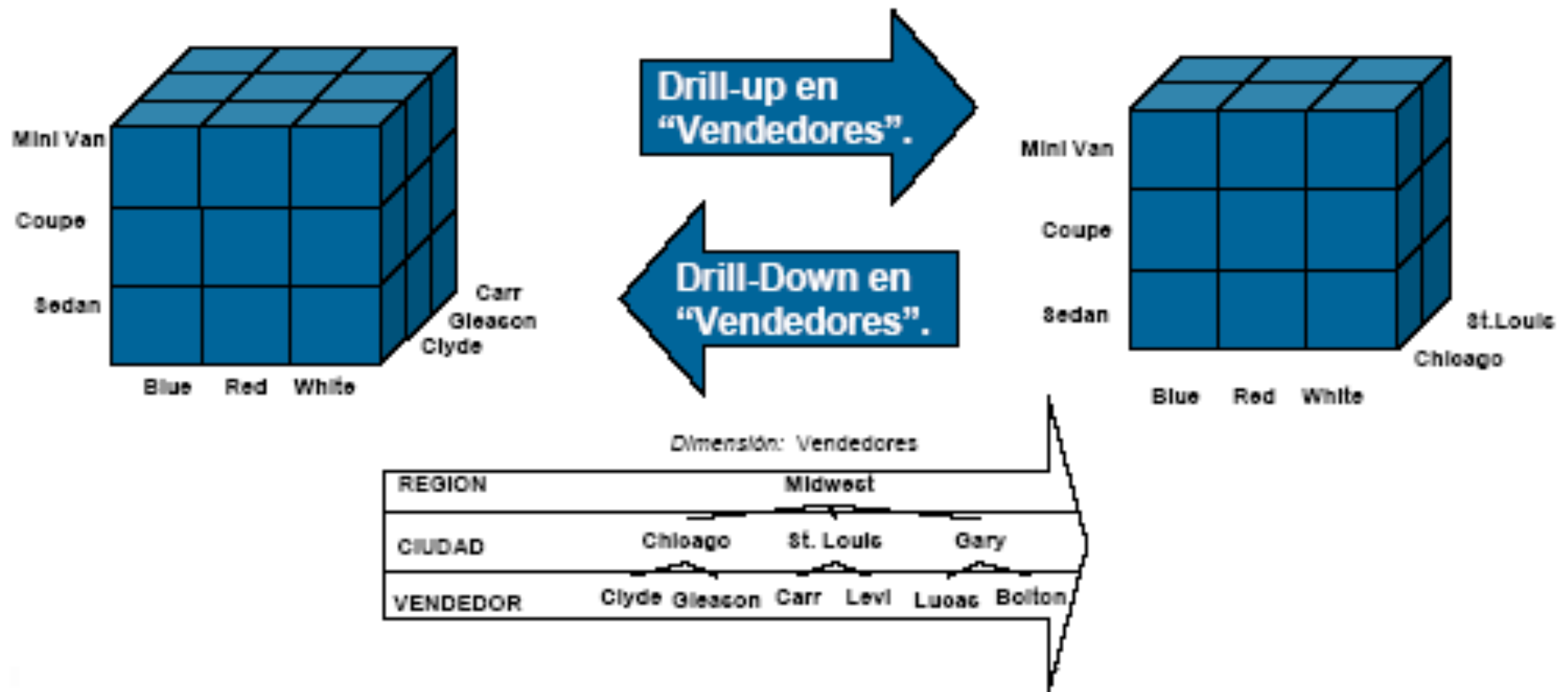


Rotación

- Selecciona el orden de visualización de las dimensiones.



Drill-up, drill-down



¿PREGUNTAS?