

Sistemas de Información para el Análisis de GVDatos

Instituto de Computación - Facultad de Ingeniería
Abril 2024



Diseño Conceptual



Temario: Diseño Conceptual

- Introducción.
- Modelos Multidimensionales.
- Modelo CMDM
- Pautas y problemas en Diseño Conceptual
- Modelo MultiDim

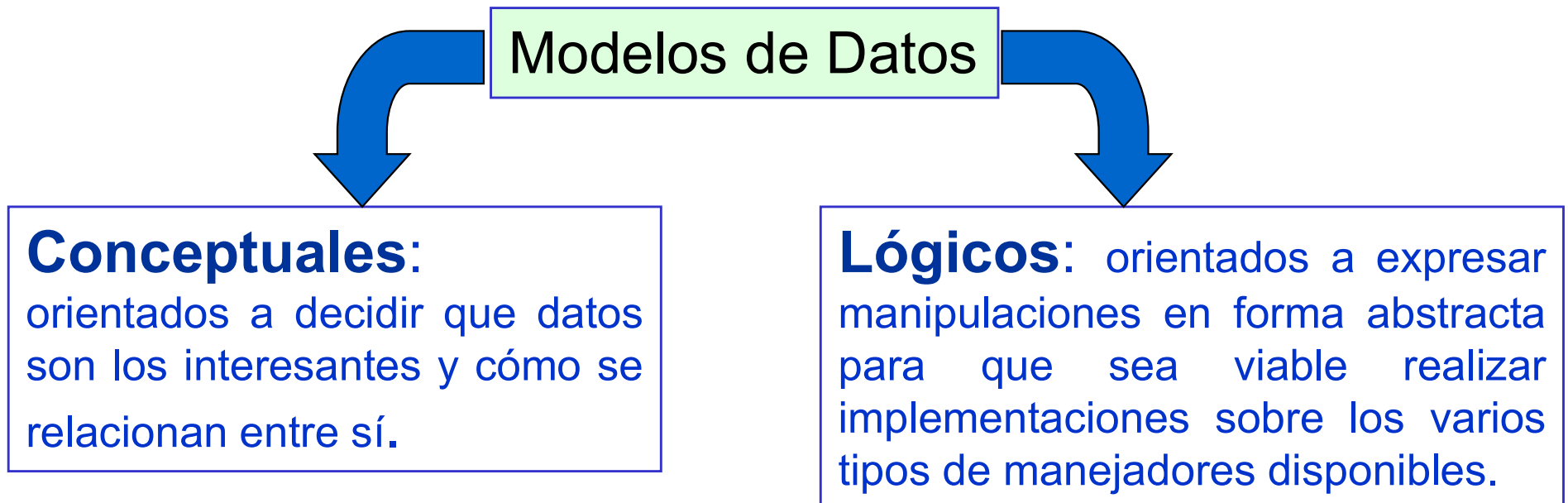


Diseño Conceptual

Introducción

Motivación

- ¿ Cuáles son las herramientas que necesita el diseñador para poder razonar sobre los datos y presentárselos al usuario ?

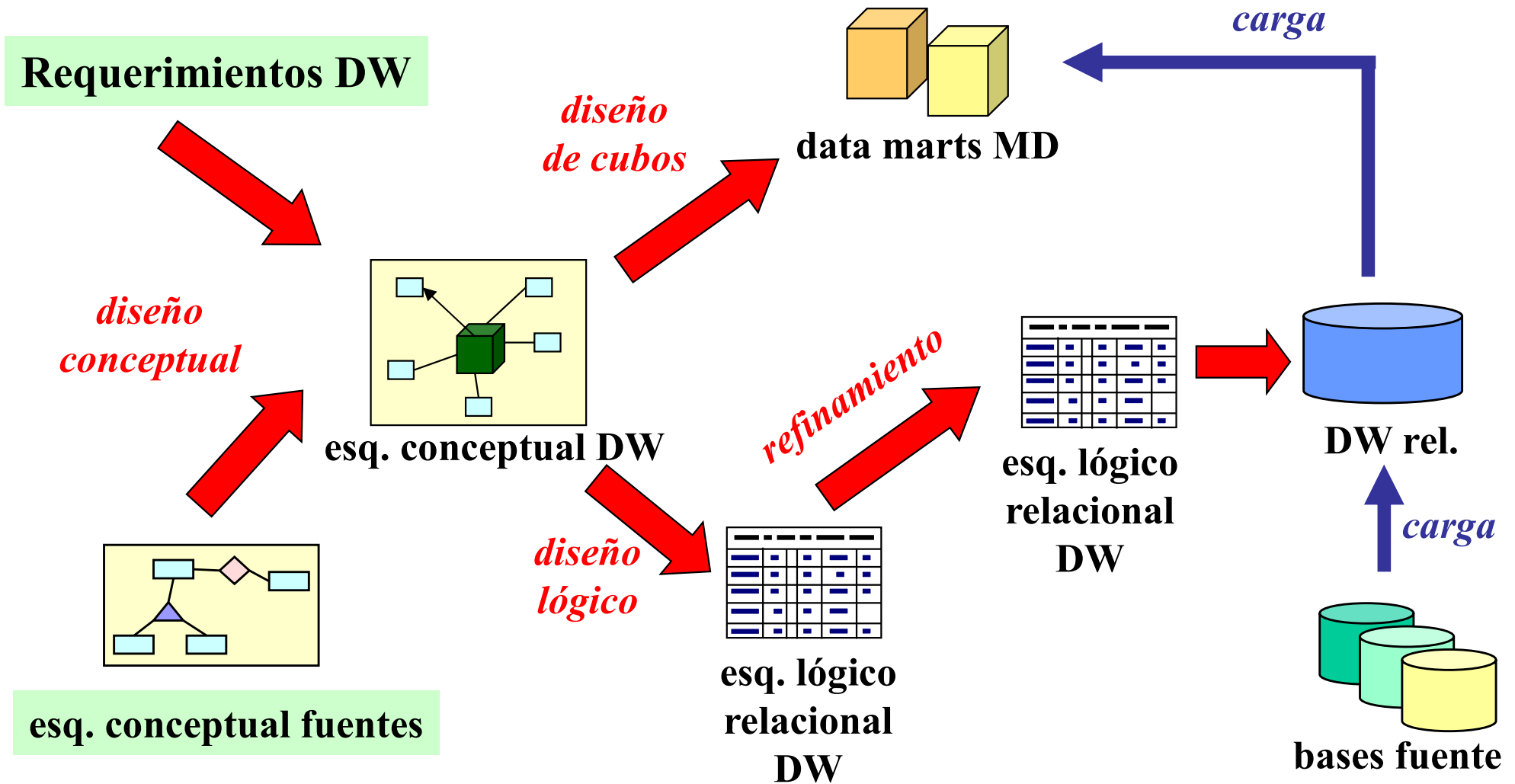


Modelos de Datos

■ Los niveles en diseño de BDs:

| | Enfasis | Modelos de Registros | Modelo Multidimensional |
|-------------------|----------------------|----------------------|---|
| CONCEPTUAL | Objetos y relaciones | Entidad-Relación | - No hay estándares - CMDM (InCo) |
| LOGICO | Operaciones | Relacional | - MD específicos - Modelo Estrella (Rel) |
| FISICO | Almacenamiento | Relacional (DBMS) | - Parámetros en servidores MD - Relacional (DBMS), otros |

Proceso de Diseño del DW



Enfoques de Diseño Conceptual

- **Análisis desde requerimientos:**
 - Los requerimientos son el universo de información.
 - Las bases fuente se relacionarán luego.
 - Aplicable cuando se tienen bases fuentes complejas. (Se analizan con los requerimientos en mente).

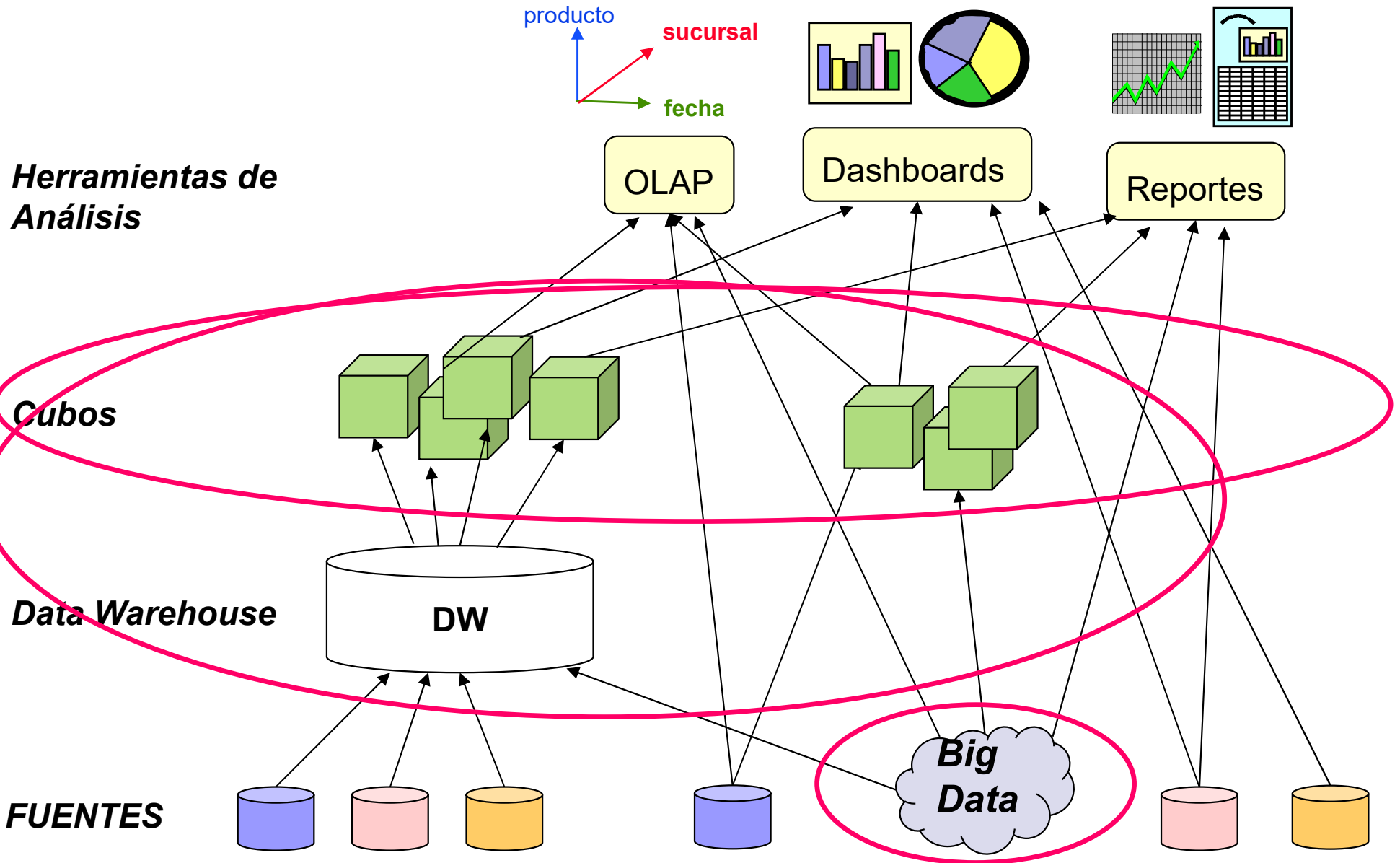
- **Análisis desde datos:**
 - Datos fuentes son el universo de información.
 - El modelo multidimensional se obtiene transformando las fuentes.
 - Aplicable cuando los requerimientos están poco claros.



Diseño Conceptual

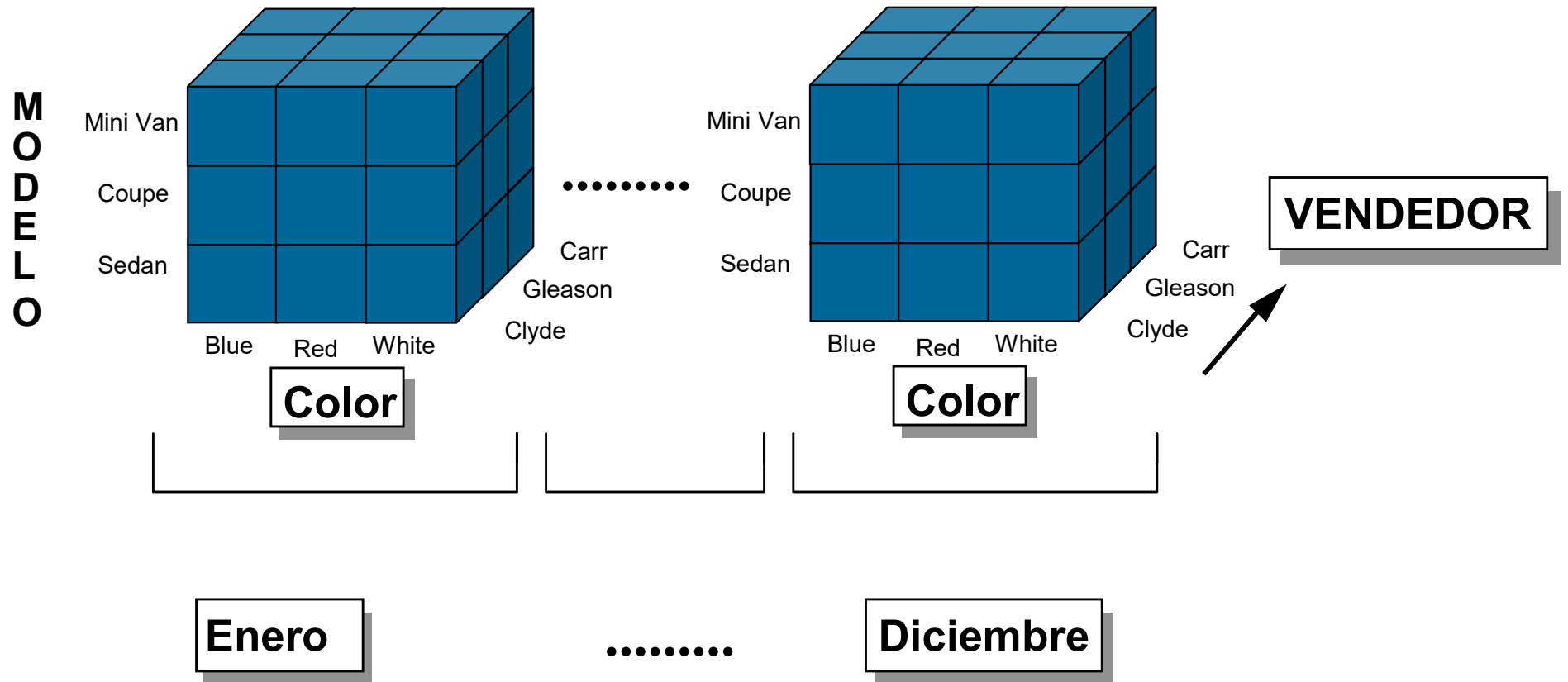
Modelos Multidimensionales

En la arquitectura



Cubos

■ Ej. de 4 dimensiones



Estructuras básicas

- Los Cubos o Hipercubos constan de:
 - Dimensiones:
 - Criterios de análisis de los datos.
 - Macro-objetos del problema.
 - Variables independientes.
 - Ejes en el hipercubo.

 - Medidas
 - Valores o indicadores a analizar.
 - Datos asociados a relaciones entre los objetos del problema.
 - Variables dependientes.
 - Variables en la intersección de las dimensiones.

Estructuras básicas

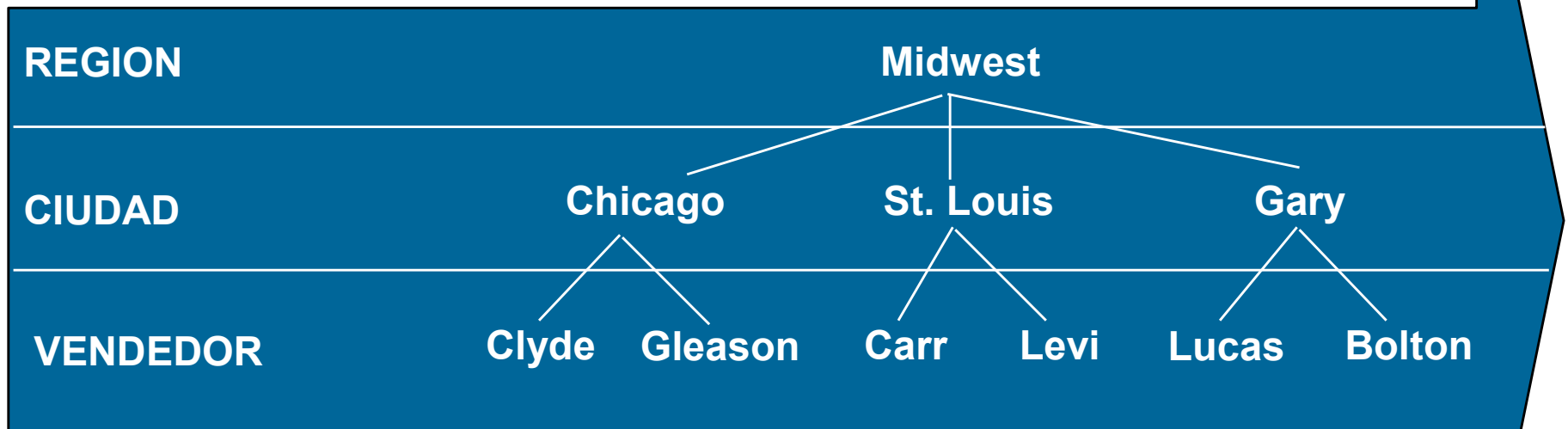
- En el ejemplo anterior:
 - Dimensiones:
 - Modelo
 - Color
 - Vendedor
 - Fecha
 - Medida:
 - Cantidad Vendida

Dimensiones

■ Jerarquías:

- Los valores se organizan en jerarquías (categorías).
- Por ejemplo:

Dimensión: Vendedores



Dimensiones

- Jerarquías alternativas:

- Pueden haber varias jerarquías para una misma dimensión.

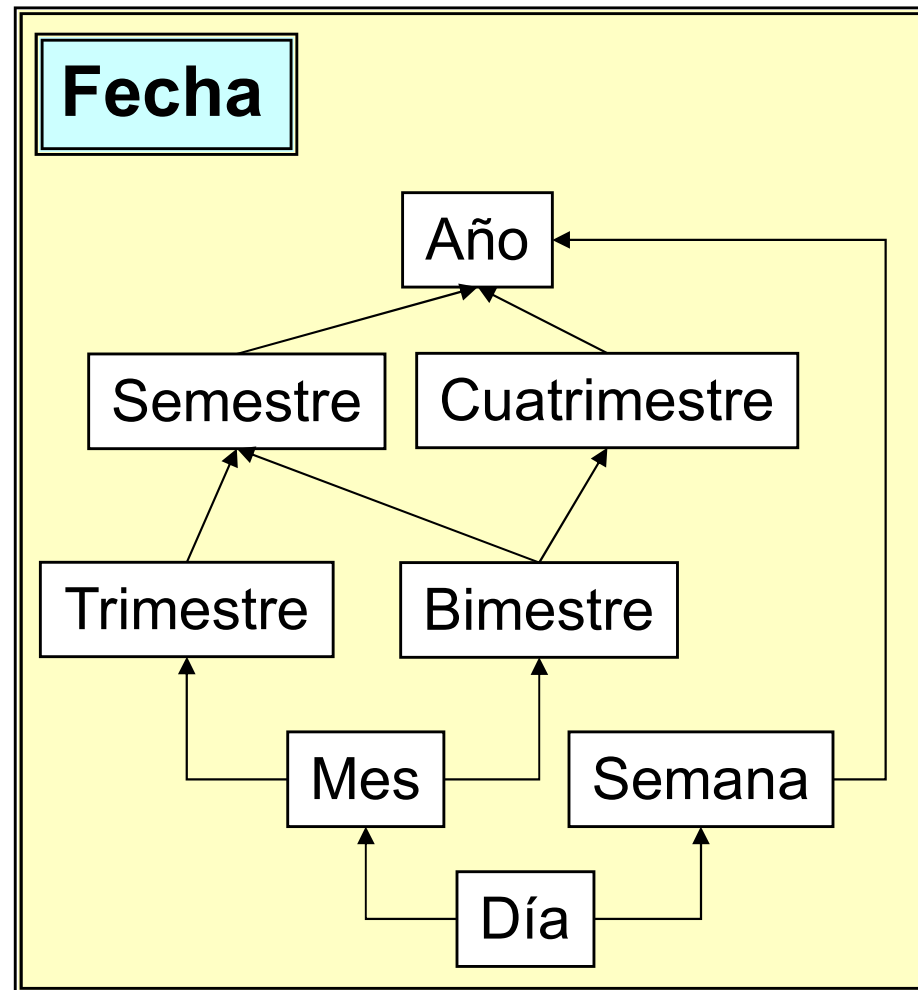
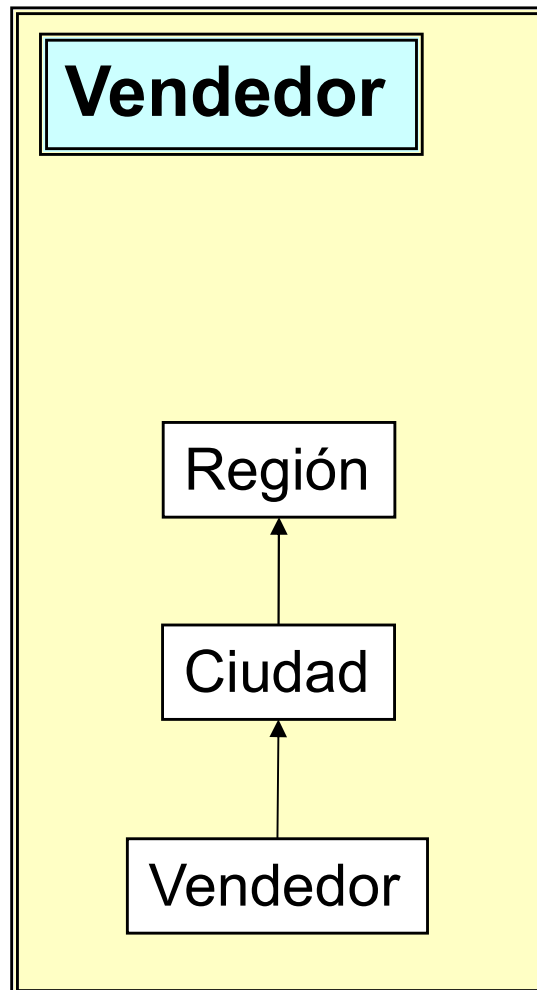
- Por ejemplo:**

- Dimensión Vendedores:

- Vendedor / Ciudad / Región
 - Vendedor / Sexo
 - Vendedor / Rango_Edad

Dimensiones

■ Jerarquías Arbitrariamente Complejas



Medidas

■ Propiedades:

- 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.

Definición: Se llama *coordenada* a una tupla formada por un valor de cada dimensión.

Medidas

| | | Blue | Red | White | COLOR |
|--------|----------|------|-----|-------|-------|
| MODELO | Mini Van | 6 | 5 | 4 | |
| | Coupe | 3 | | 5 | |
| | Sedan | 4 | 3 | 2 | |

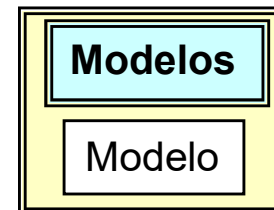
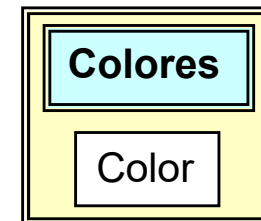
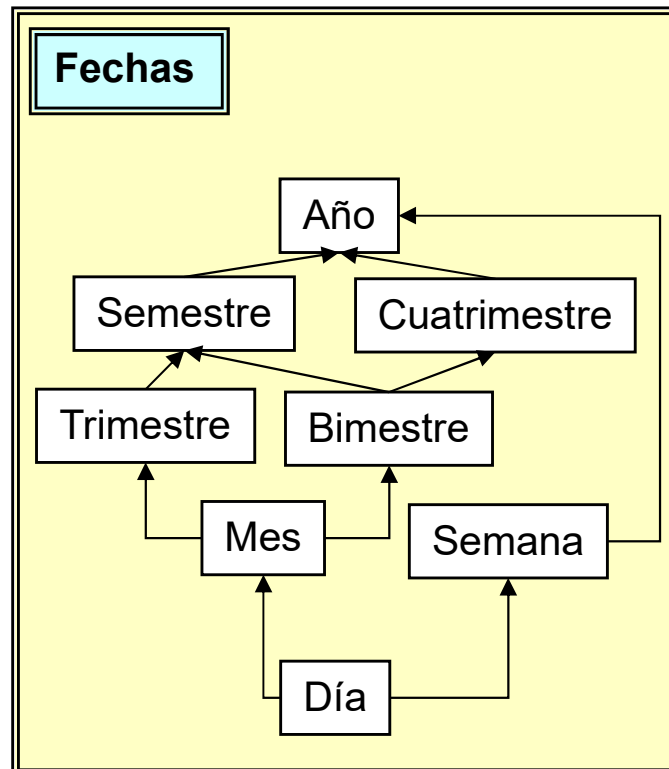
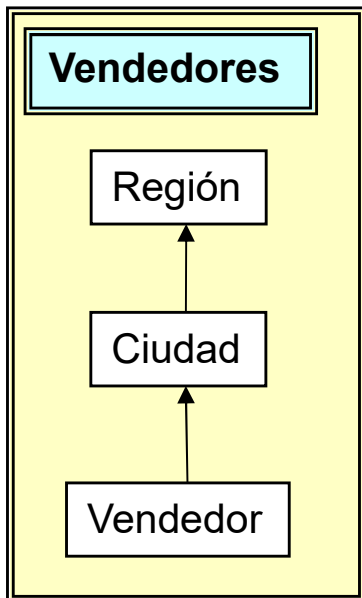
VENTAS("Mini Van", "Blue")=6

Cubos

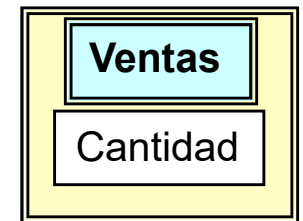
- La realidad se modela como un conjunto de cubos.
 - Cada cubo, esta formado por:
 - Un conjunto de Dimensiones organizadas en jerarquías.
 - Un conjunto de Medidas asociadas a cada Coordenada.
 - Es posible moverse en las jerarquías de las dimensiones y observar de esa forma, diferentes visiones de las medidas.

Caso de Estudio

□ Dimensiones



□ Medidas

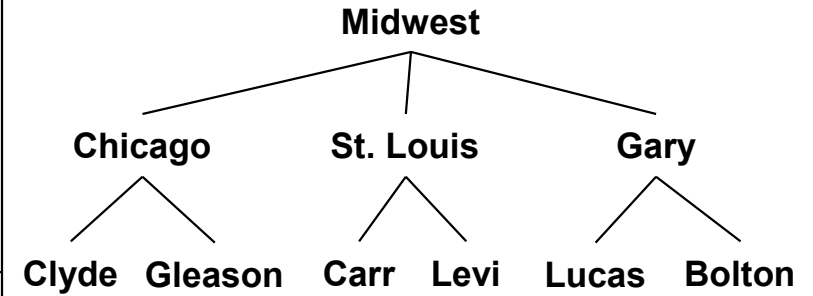


Caso de Estudio

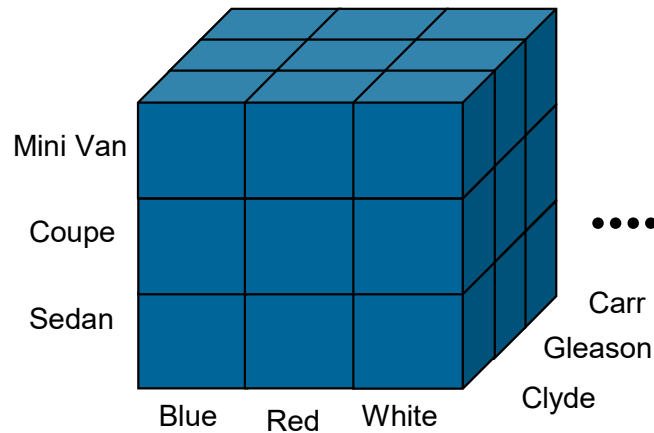
□ Instancias

FECHAS

VENDEDORES

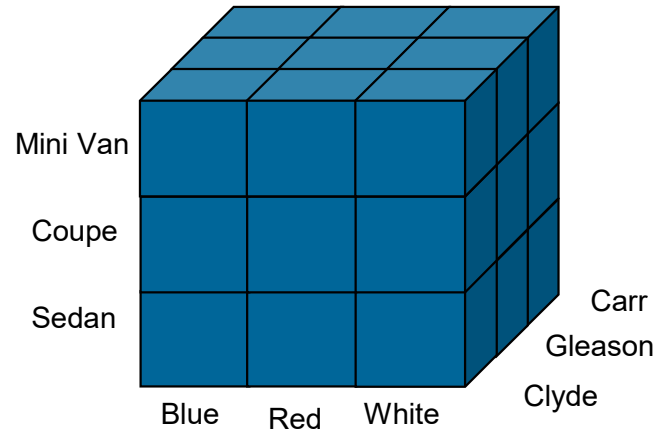


**M
O
D
E
L
O**



COLOR

1/1/1997



COLOR

31/12/2006

VENDEDOR

FECHAS

Operaciones Multidimensionales

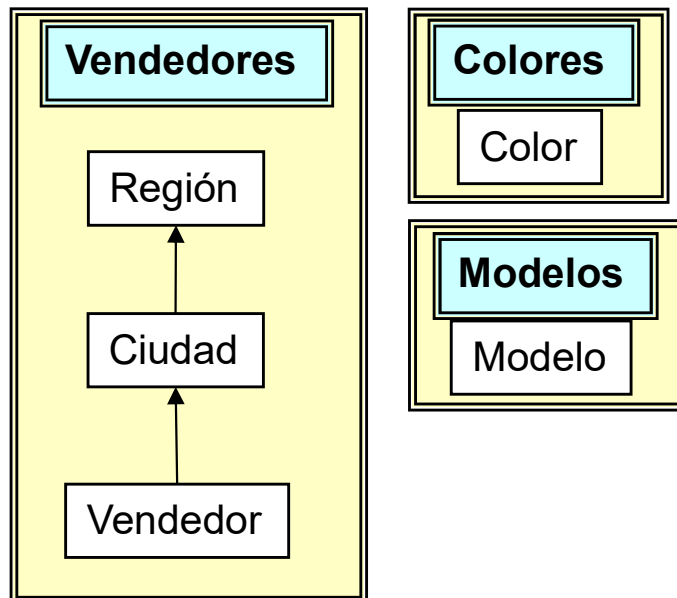
- Principales operaciones en modelos MD:
 - Slice
 - Dice
 - Pivot
 - Drill-down
 - Drill-up
 - Roll-up
 - Drill-across
 - Drill-through

Operaciones: Slice

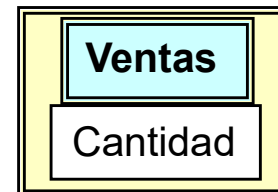
■ Seleccionar Dimensiones (Slice)

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

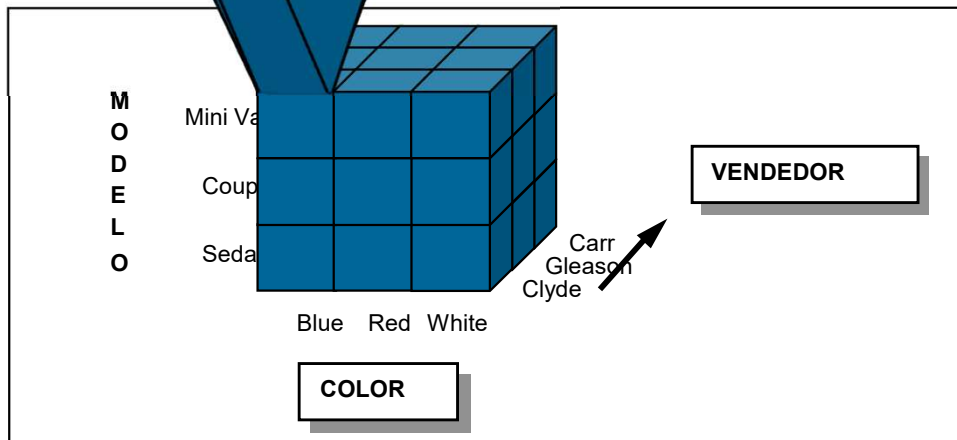
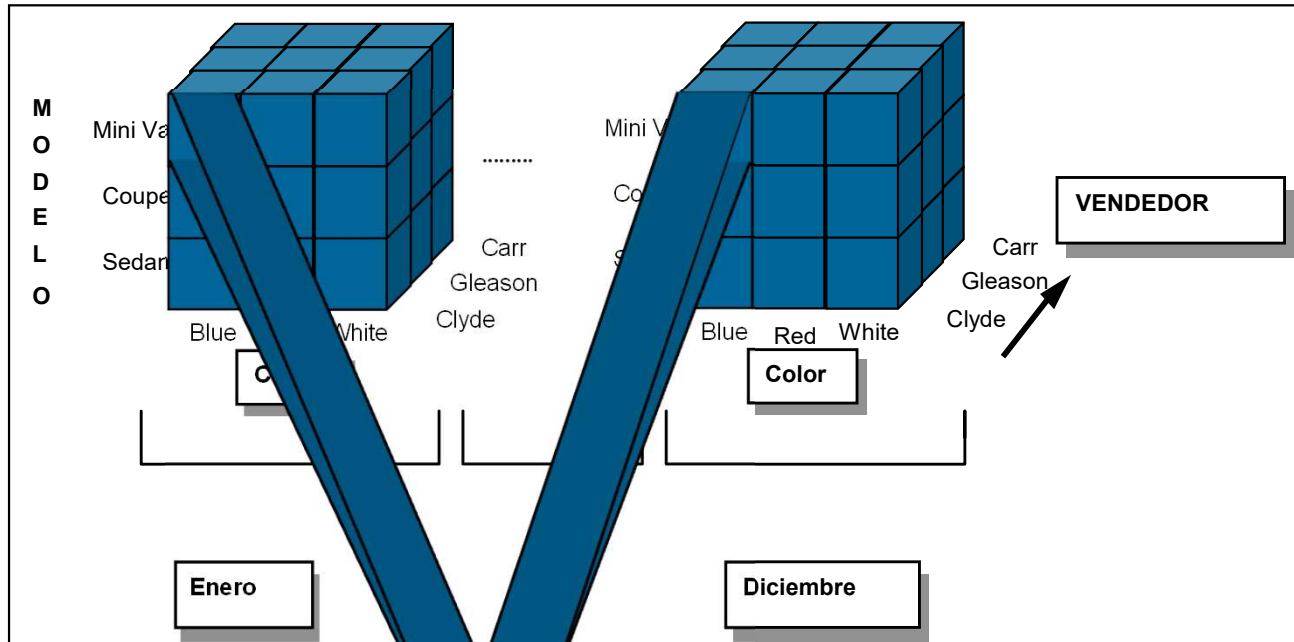
□ Dimensiones



□ Medidas



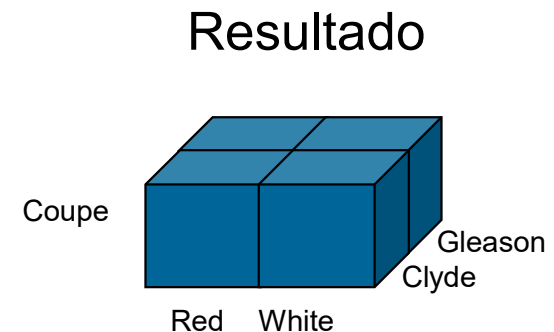
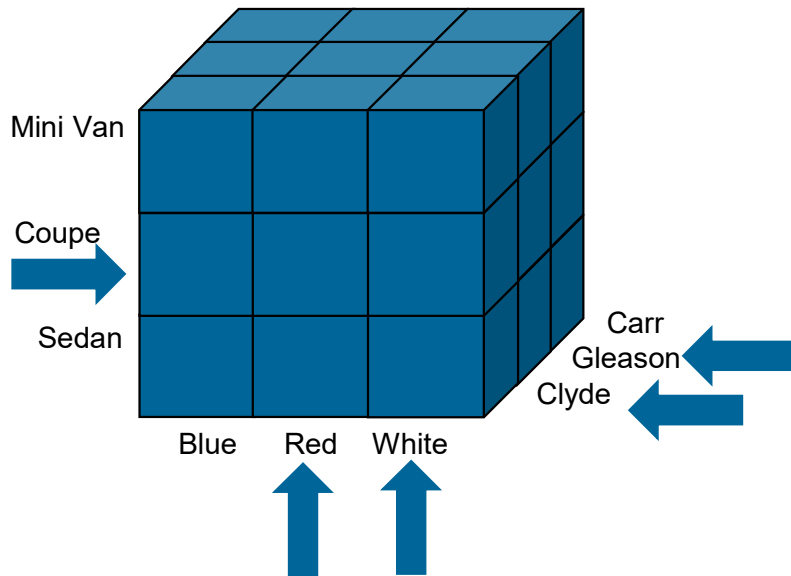
Operaciones: Slice



Operaciones: Dice

■ Filtrado (DICE)

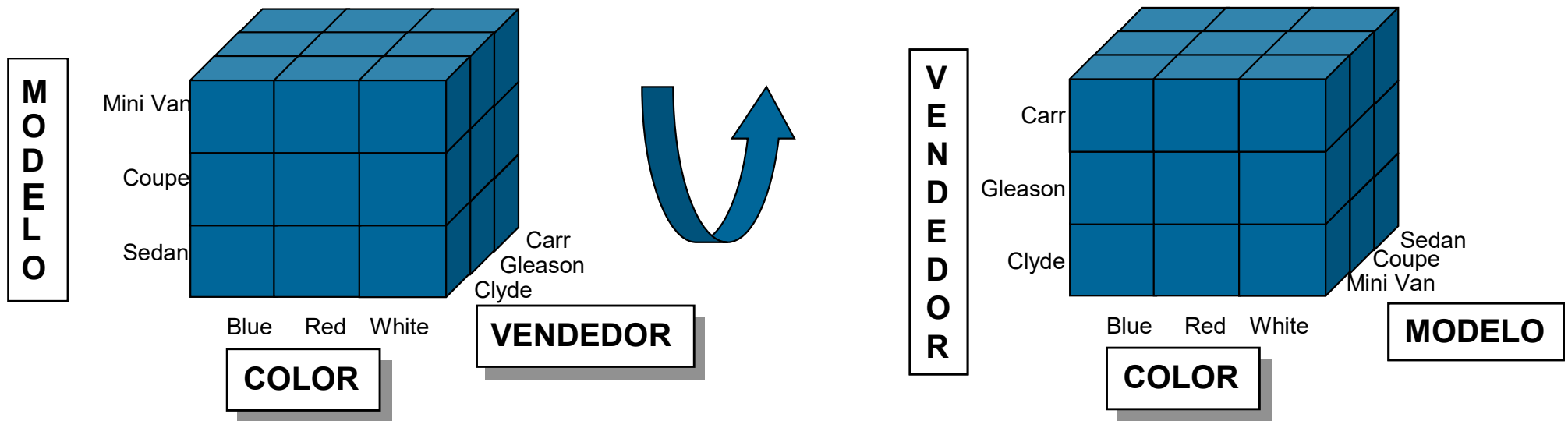
- Se fijan valores para algunas dimensiones.



Operaciones: Pivot

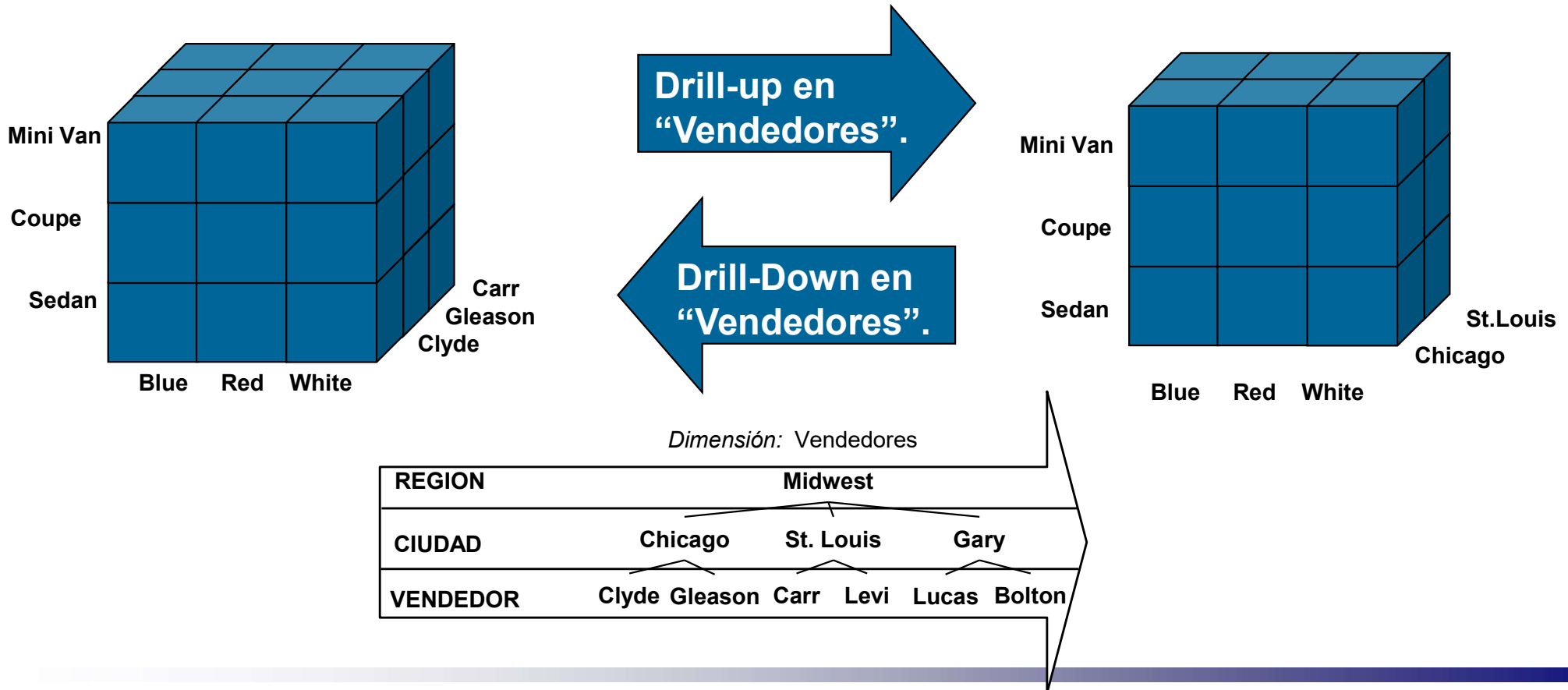
■ Pivot (Rotación)

- Selecciona el orden de visualización de las dimensiones.



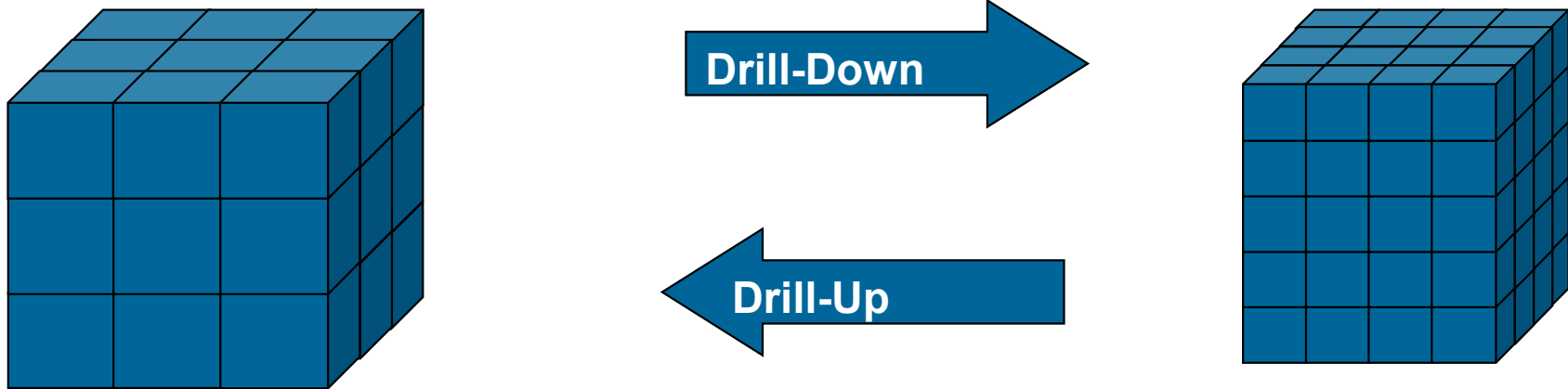
Opns: Drill-up, drill-down

- Movimientos en la Jerarquía de una Dimensión (Drill-up, Drill-down)



Opns: Drill-up, drill-down

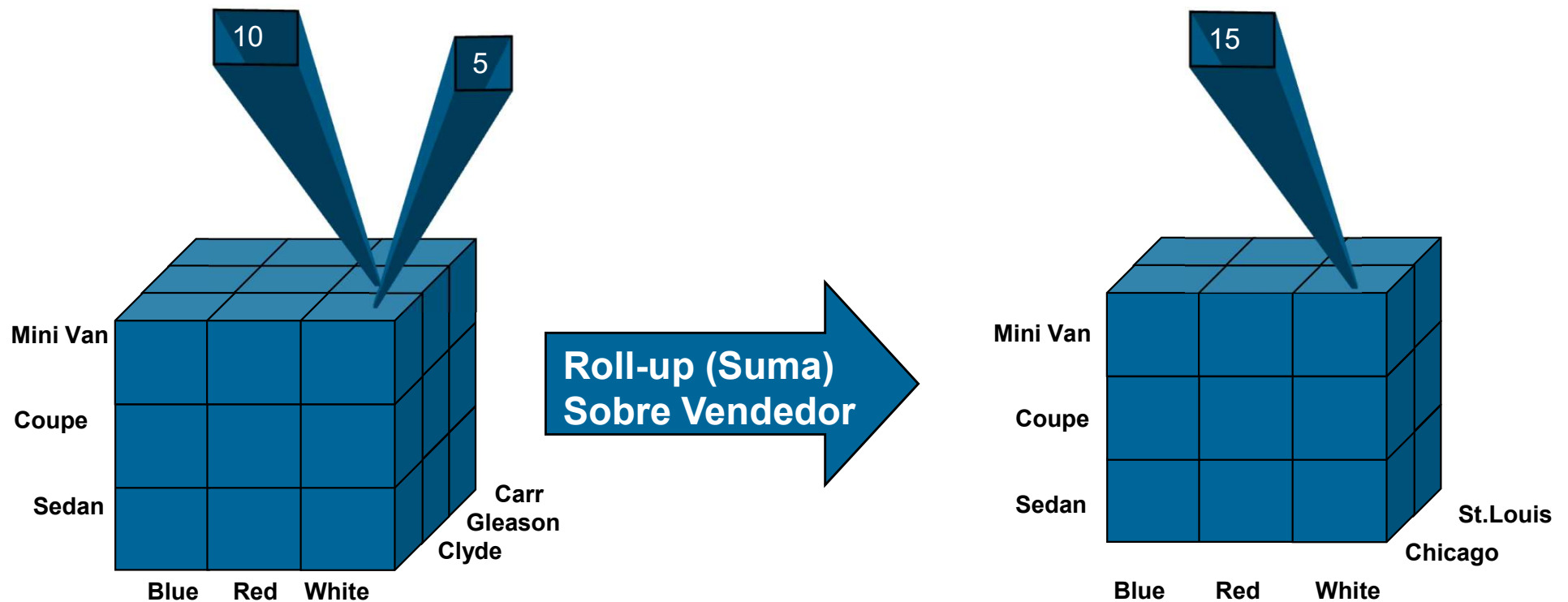
- Drill-Up o Drill-Down pueden verse como ajuste en las escalas de los ejes.
- Son agrupamientos y des-agrupamientos.



Operaciones: Roll-up

- Roll-Up (Consolidación)

- Calcula las medidas en función de agrupamientos.



Operaciones: Roll-up

■ Propiedades:

- Se debe especificar cuál es la operación que calcula el nuevo valor de la medida.
 - Esta operación puede ser: suma, promedio, etc.
 - Pueden haber medidas con comportamientos diferentes.
 - Por ejemplo:
 - Cantidades de productos vendidos se acumulan.
 - Notas en exámenes se promedian.

Ejemplo

| Time (Quarter) | Store (City) | | | | books | CDs | games | DVDs |
|----------------|--------------|------|------|-------|-------|-----|-------|------|
| | Milan | Rome | Nice | Paris | | | | |
| Q1 | 21 | 10 | 18 | 35 | 14 | 23 | 19 | |
| Q2 | 27 | 14 | 11 | 30 | 12 | 20 | 17 | |
| Q3 | 26 | 12 | 35 | 32 | 10 | 33 | 18 | |
| Q4 | 14 | 20 | 47 | 31 | | | | |

| Time (Quarter) | Store (Country) | | | | books | CDs | games | DVDs |
|----------------|-----------------|--------|----|----|-------|-----|-------|------|
| | Italy | France | | | | | | |
| Q1 | 33 | 30 | 42 | 68 | | | 41 | |
| Q2 | 27 | 14 | 11 | 30 | | | 57 | |
| Q3 | 26 | 12 | 35 | 32 | | | 51 | |
| Q4 | 14 | 20 | 47 | 31 | | | | |

ROLL-UP a nivel Country

Ejemplo

| Time (Quarter) | Store (City) | | | |
|----------------|--------------|------|------|-------|
| | Milan | Rome | Nice | Paris |
| Q1 | 21 | 10 | 18 | 35 |
| Q2 | 27 | 14 | 11 | 30 |
| Q3 | 26 | 12 | 35 | 32 |
| Q4 | 14 | 20 | 47 | 31 |

Product (Category): games, DVDs, books, CDs

| Time (Month) | Store (City) | | | |
|--------------|--------------|------|------|-------|
| | Milan | Rome | Nice | Paris |
| Jan | 7 | 2 | 6 | 13 |
| Feb | 8 | 4 | 8 | 12 |
| Mar | 6 | 4 | 4 | 10 |
| ... | ... | ... | ... | ... |
| Dec | 4 | 4 | 16 | 7 |

Product (Category): games, DVDs, books, CDs

DRILL-DOWN a nivel Month

Ejemplo

| Store (City) | Time (Quarter) | | | | Product (Category) | | | |
|--------------|----------------|----|----|----|--------------------|-------|-----|------|
| | Q1 | Q2 | Q3 | Q4 | books | games | CDs | DVDs |
| Milan | 21 | 10 | 18 | 35 | 14 | 20 | 47 | 31 |
| Rome | 27 | 14 | 11 | 30 | 12 | 23 | 25 | 18 |
| Nice | 26 | 12 | 35 | 32 | 10 | 14 | 12 | 20 |
| Paris | 33 | 25 | 23 | 25 | 19 | 33 | 19 | 17 |

| Store (City) | Time (Quarter) | | | | Product (Category) | | | |
|--------------|----------------|----|----|----|--------------------|-------|-----|------|
| | Q1 | Q2 | Q3 | Q4 | books | games | CDs | DVDs |
| Paris | 21 | 27 | 26 | 14 | 17 | 21 | 12 | 10 |
| Nice | 12 | 14 | 11 | 13 | 26 | 20 | 18 | 11 |
| Rome | 33 | 28 | 35 | 32 | 19 | 40 | 19 | 17 |
| Milan | 24 | 23 | 25 | 18 | 14 | 12 | 20 | 10 |



Ejemplo

| | | | | | |
|-----------------------|-------|---------------------------|-------|-----|------|
| Store (City) | Milan | 24 | 19 | 29 | 14 |
| | Rome | 33 | 25 | 23 | 25 |
| | Nice | 12 | 20 | 24 | 33 |
| | Paris | | | | |
| | | | | | |
| Time (Quarter) | Q1 | 21 | 10 | 18 | 35 |
| | Q2 | 27 | 14 | 11 | 30 |
| | Q3 | 26 | 12 | 35 | 32 |
| | Q4 | 14 | 20 | 47 | 31 |
| | | books | games | CDs | DVDs |
| | | Product (Category) | | | |

| | | | | | |
|-----------------------|----|---------------------------|-------|-----|------|
| Time (Quarter) | Q1 | 21 | 10 | 18 | 35 |
| | Q2 | 27 | 14 | 11 | 30 |
| | Q3 | 26 | 12 | 35 | 32 |
| | Q4 | 14 | 20 | 47 | 31 |
| | | books | games | CDs | DVDs |
| | | Product (Category) | | | |

SLICE por:
Store.City = 'Paris'

Ejemplo

| | | | | | |
|-------------------|-------|---------------------------|-----|-------|------|
| Store (City) | Milan | 24 | 19 | 29 | 14 |
| | Rome | 33 | 25 | 23 | 25 |
| | Nice | 12 | 20 | 24 | 33 |
| | Paris | | | | |
| Time (Quarter) | Q1 | 21 | 10 | 18 | 35 |
| | Q2 | 27 | 14 | 11 | 30 |
| | Q3 | 26 | 12 | 35 | 32 |
| | Q4 | 14 | 20 | 47 | 31 |
| | | books | CDs | games | DVDs |
| | | Product (Category) | | | |

| | | | | | |
|-------------------|-------|---------------------------|-----|-------|------|
| Store (City) | Nice | 12 | 20 | 24 | 33 |
| | Paris | | | | |
| Time (Quarter) | Q1 | 21 | 10 | 18 | 35 |
| | Q2 | 27 | 14 | 11 | 30 |
| | | books | CDs | games | DVDs |
| | | Product (Category) | | | |

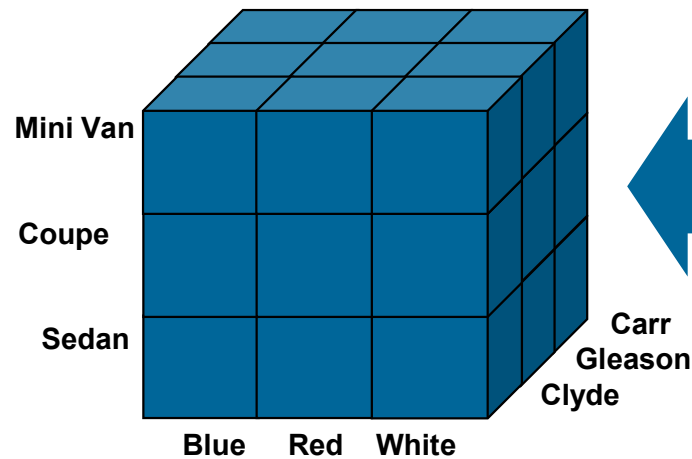
DICE por:
Store.Country = 'France'
Time.Quarter = 'Q1' or 'Q2'

Operaciones: Drill-Across

■ Drill-Across

■ Relaciona dos cubos.

Ventas



| | | | | |
|---------|---|---|---|---|
| Marzo | 6 | 5 | 4 | 7 |
| Febrero | 3 | | 5 | 9 |
| Enero | 4 | 3 | 2 | 3 |

Clyde Gleason Carr Carey

| | | | |
|---------|---|---|---|
| Marzo | 6 | 5 | 4 |
| Febrero | 3 | | 5 |
| Enero | 4 | 3 | 2 |

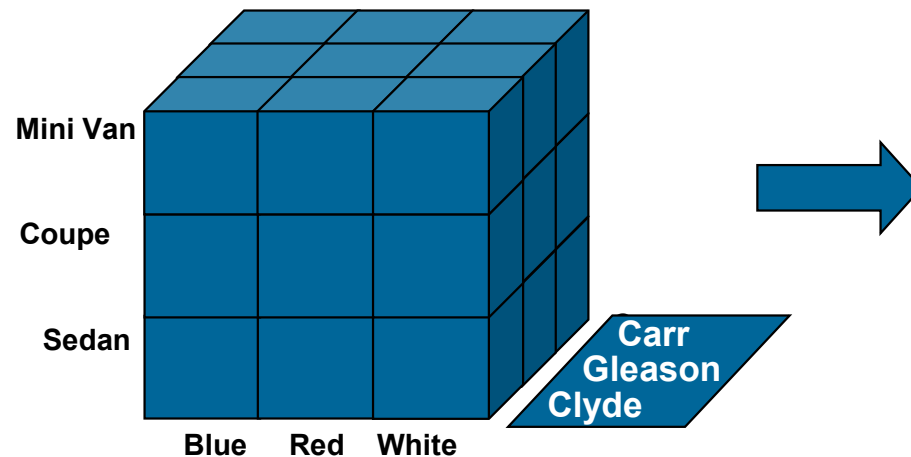
Clyde Gleason Carr

Total Comisión
Por
Vendedor/mes.

Operaciones: Drill-Through

■ Drill-Through

- Accede desde un dato del cubo a datos en el DW de donde se deriva el cubo.
 - Para acceder a datos descriptivos o seguir la traza de un dato.



| Apellidos | Nombres | Edad |
|-----------|---------|------|
| Clyde | John | 23 |
| Gleason | Susan | 38 |
| Carr | Robert | 29 |

Límitaciones de los MMD

- Las implementaciones actuales de MMD no realizan adecuadamente ciertas operaciones:
 - Tienen pocas facilidades para representación y consulta de datos descriptivos o secundarios.
 - Es complicado hacer diferentes RollUp's según la medida y la jerarquía.
- Entonces:
 - Se deben resolver con otras técnicas
 - Se deben considerar en el diseño

Resumen MMD

- Permiten describir una realidad en términos de matrices multidimensionales (Cubos).
- Se utilizan para describir Data Marts , un DW completo o fuentes directamente.
- Las Dimensiones pueden tener una o más Jerarquías.
- Las operaciones permiten construir diferentes visiones de los datos.
- Hay ciertas operaciones que no son totalmente resueltas por los productos actuales.

Pendientes...

- ¿Cómo se describen los modelos multidimensionales?
 - ¿Con qué lenguaje?
 - ¿Con qué criterios se construye esa descripción?
- Respuestas Rápidas...
 - No hay un lenguaje aceptado.
 - Hay criterios muy generales pero no hay criterios absolutos...