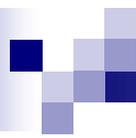


Sistemas de Información para el Análisis de GVDatos

Instituto de Computación - Facultad de Ingeniería
Abril 2024



Diseño Conceptual

Modelo CMDM

InCo – Grupo CSI

[Car00]

Modelo CMDM

- **Objetivos principales:**
 - Representación directa, expresiva y precisa de esquemas multidimensionales.
- **Características:**
 - Basado en conceptos aceptados de MMD.
 - Independiente de las implementaciones.
- **Desarrollado por el grupo CSI (InCo).**
 - Tesis de maestría [Car00].

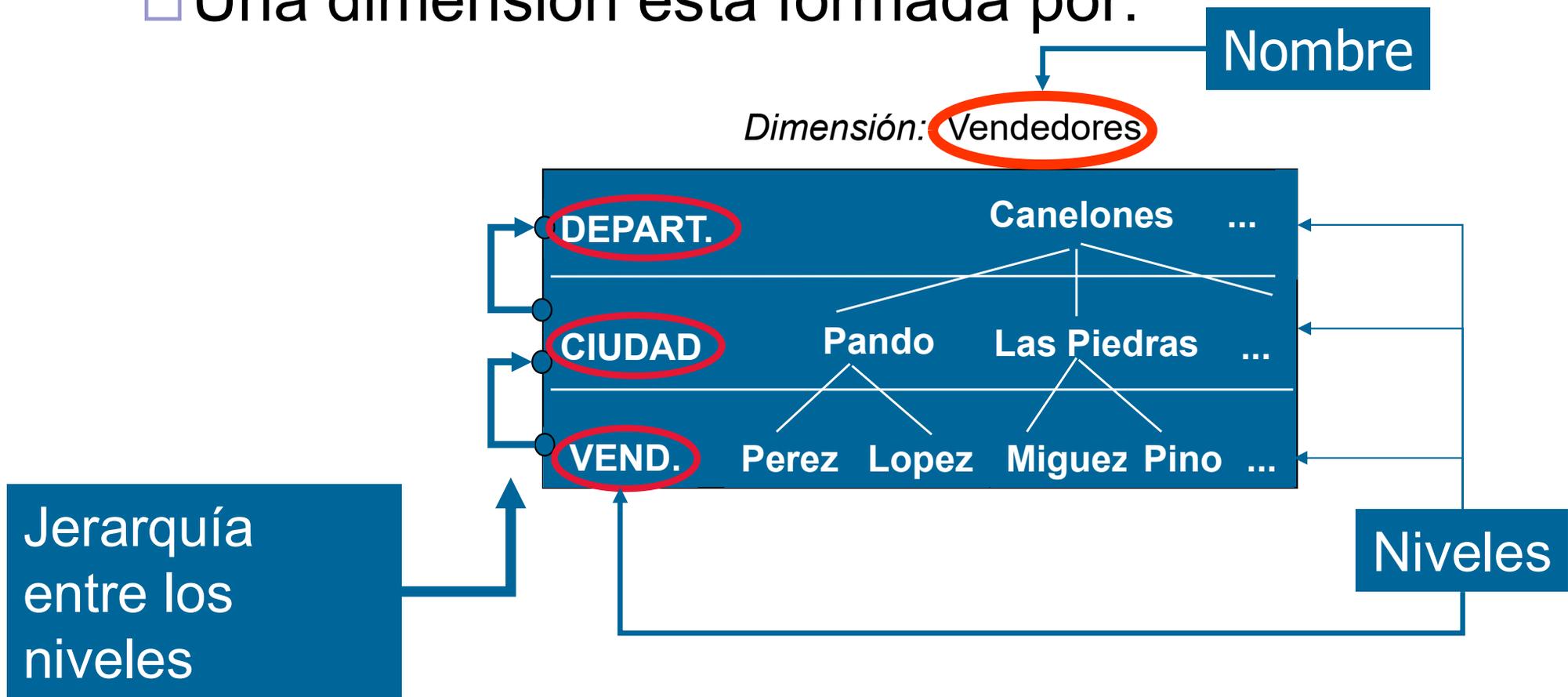
Modelo CMDM

- Estructuras básicas.
 - Niveles.
 - Dimensiones.
 - Con Jerarquías, formadas por Niveles.
 - Incluye Medidas (Dimensionalidad Genérica).
 - Relaciones dimensionales.
 - Cubos.
 - Cruzamientos específicos.

Modelo CMDM

■ Dimensiones:

- Una dimensión esta formada por:

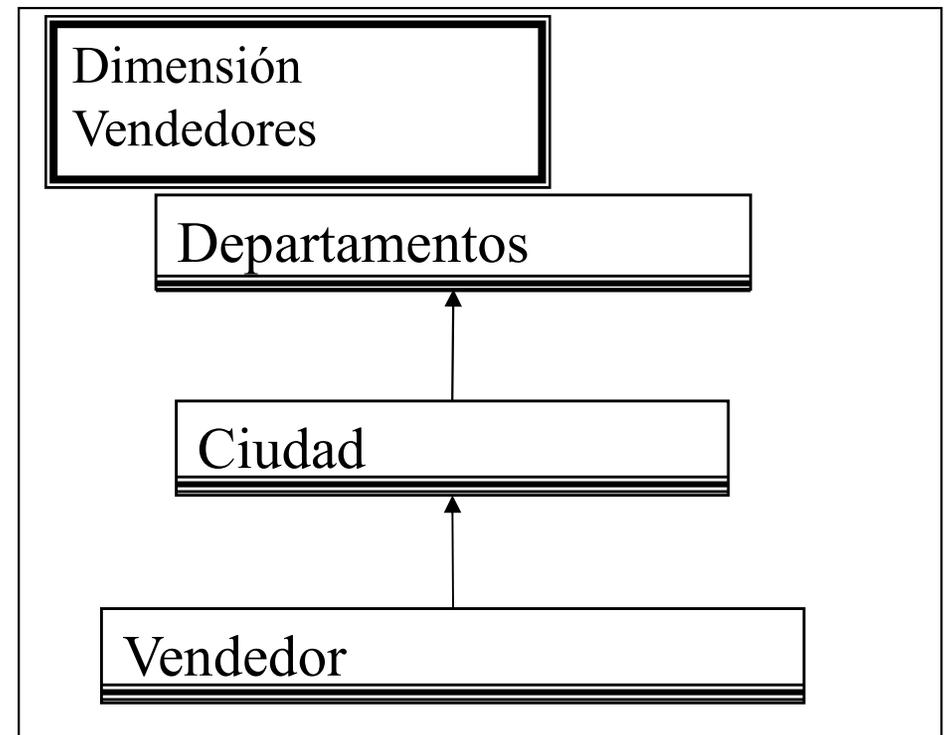
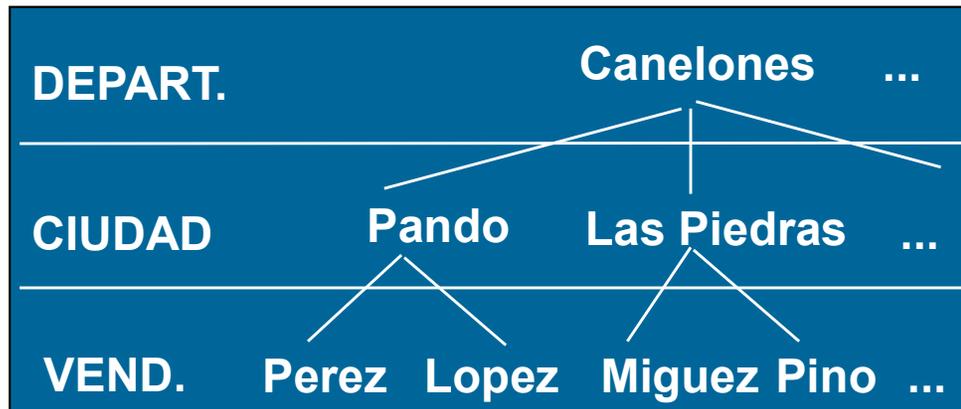


Modelo CMDM

■ Dimensiones:

□ Ejemplo:

Dimensión: Vendedores

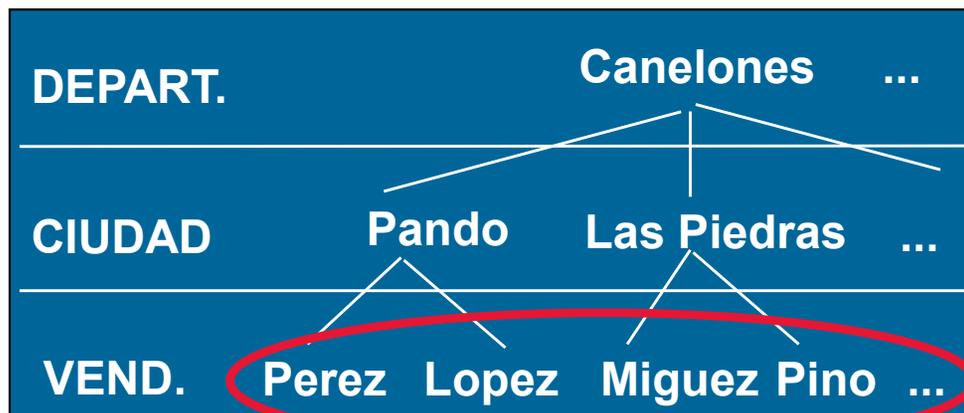


Modelo CMDM

■ Niveles:

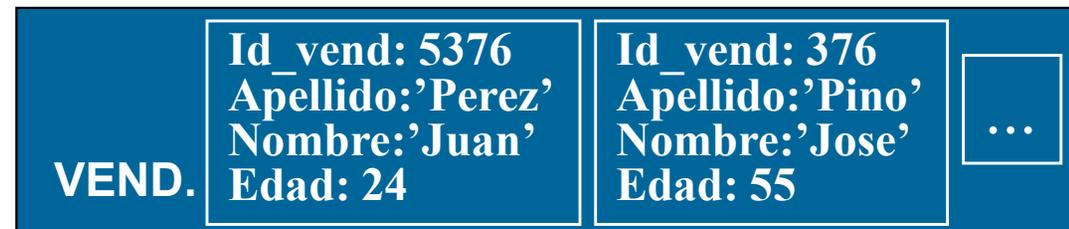
- Un nivel representa un conjunto de datos.

Dimensión: Vendedores



Los datos pueden ser no atómicos.

Nivel: Vendedor



Modelo CMDM

- Niveles:
 - Ejemplo:

Nivel: Vendedor

VEND.	Id_vend: 5376 Apellido: 'Perez' Nombre: 'Juan' Edad: 24	Id_vend: 376 Apellido: 'Pino' Nombre: 'Jose' Edad: 55	...

Vendedor
Id_vendedor#: Integer
Apellido: String
Nombre: String
Edad: [18-60]

Modelo CMDM

■ Jerarquías:

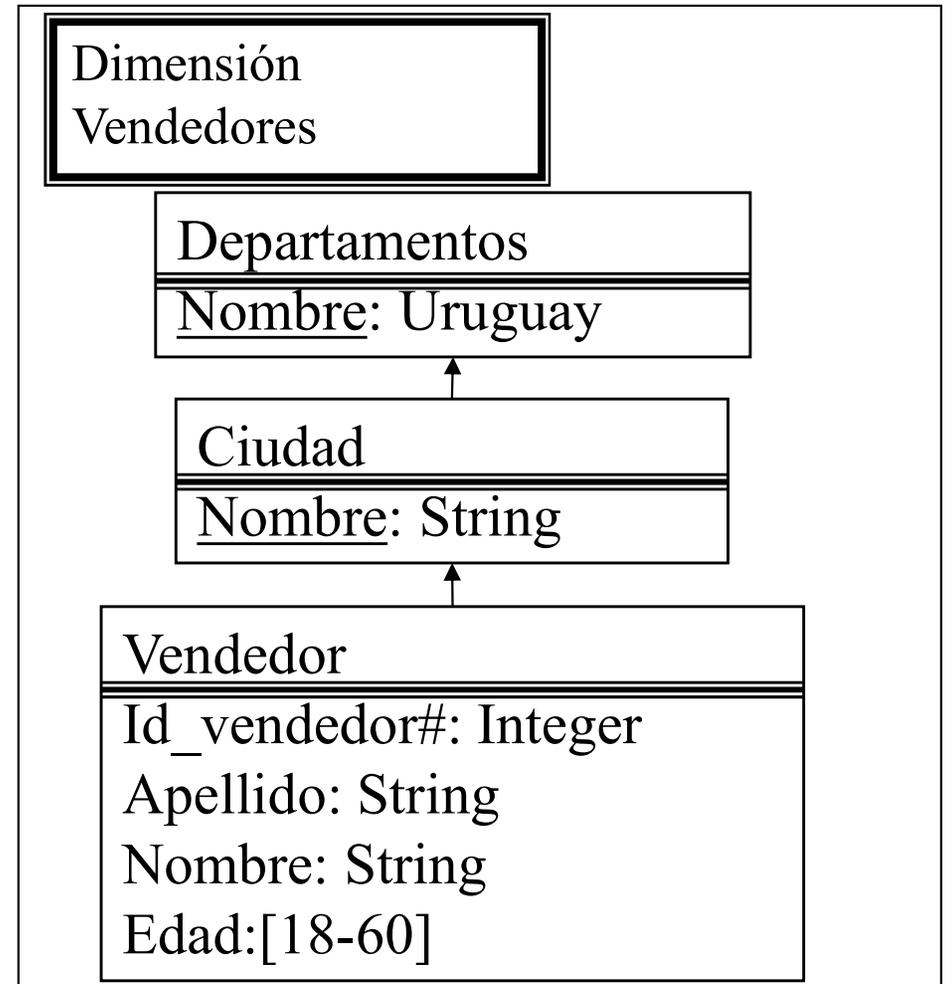
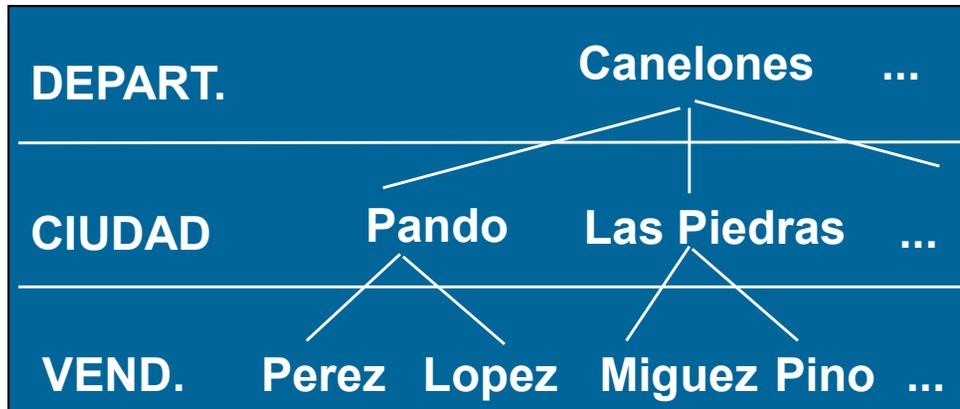
- Los niveles se organizan en jerarquías.
- Cada jerarquía está compuesta por uno o varios niveles.
- En cada jerarquía:
 - Se tiene una relación <1-n> entre objetos de nivel superior e inferior.

Modelo CMDM

■ Jerarquías:

□ Ejemplo:

Dimensión: Vendedores



Modelo CMDM

■ Relaciones Dimensionales:

- Representan cruzamientos entre Dimensiones.

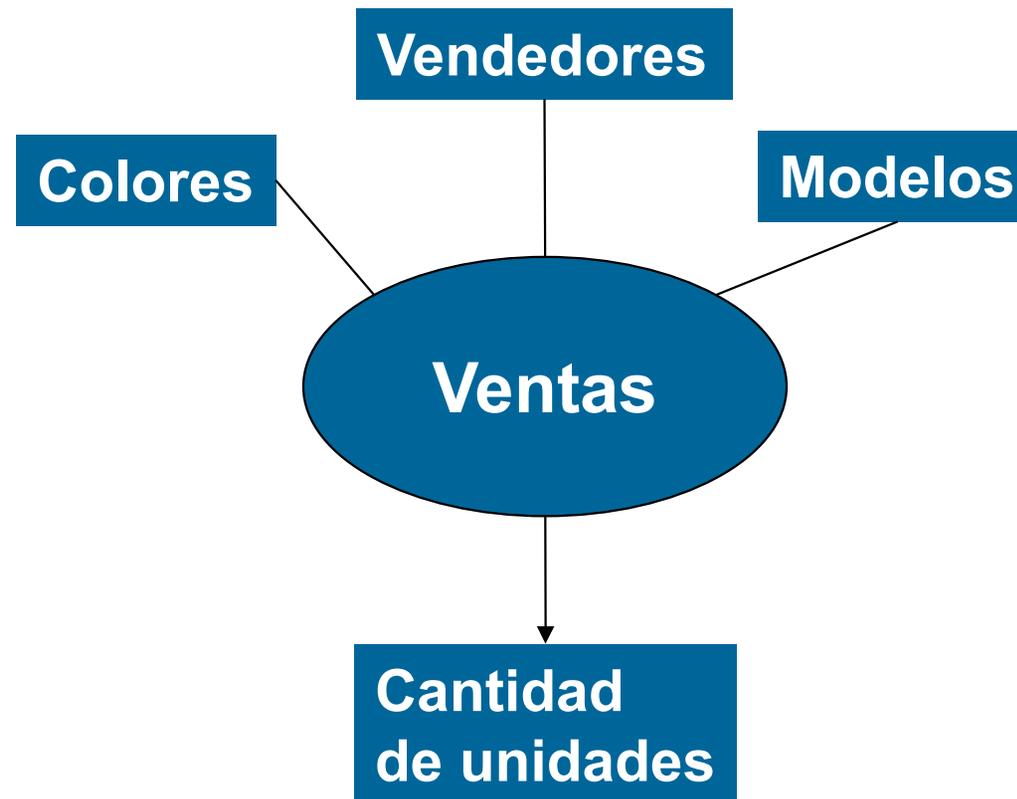
- Vista como una relación:

- Se tiene un elemento en el conjunto relación si y solo si hay un cruzamiento.
- Esto obliga a que las Dimensiones participantes realmente sean cruzables.

Modelo CMDM

■ Relaciones Dimensionales.

□ Ejemplo.



Modelo CMDM

■ Esquema Conceptual MD:

- Está formado por un conjunto de Relaciones Dimensionales.
- Las Relaciones Dimensionales pueden compartir Dimensiones.
 - Permite el Drill-Across.

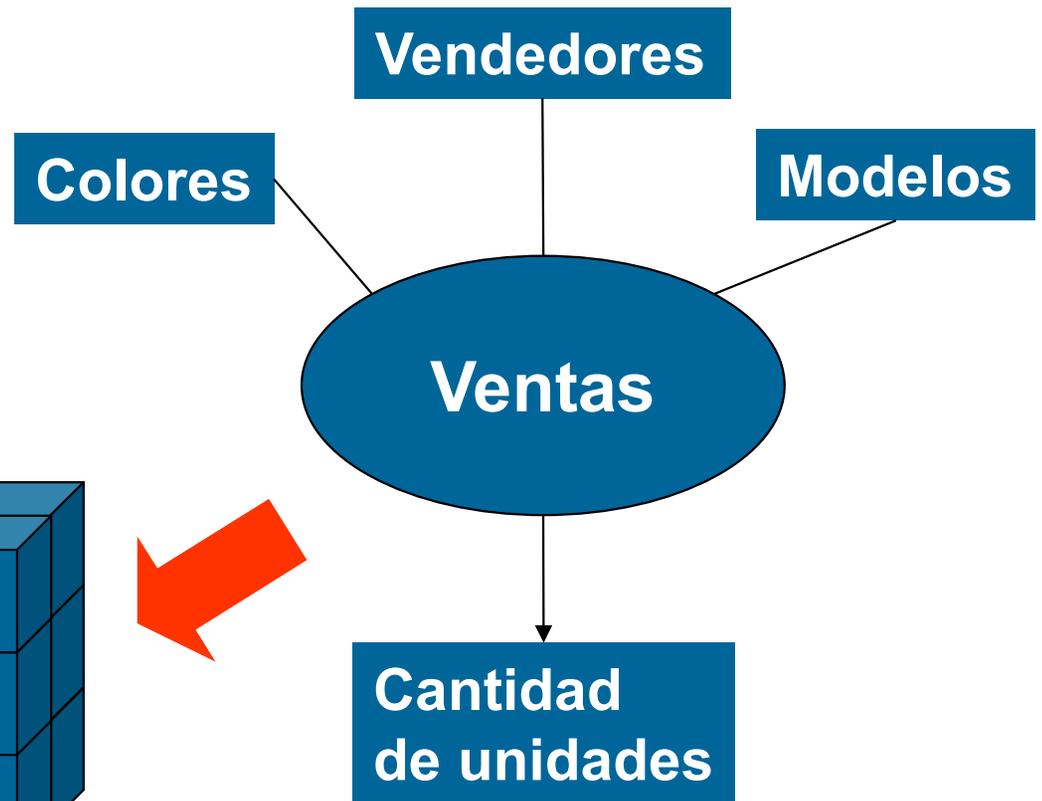
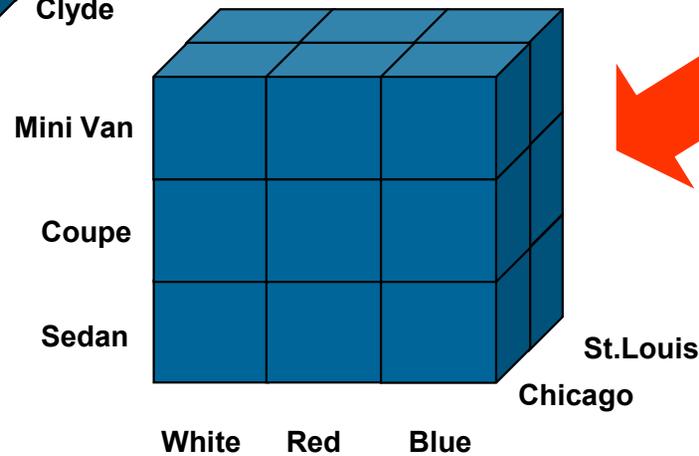
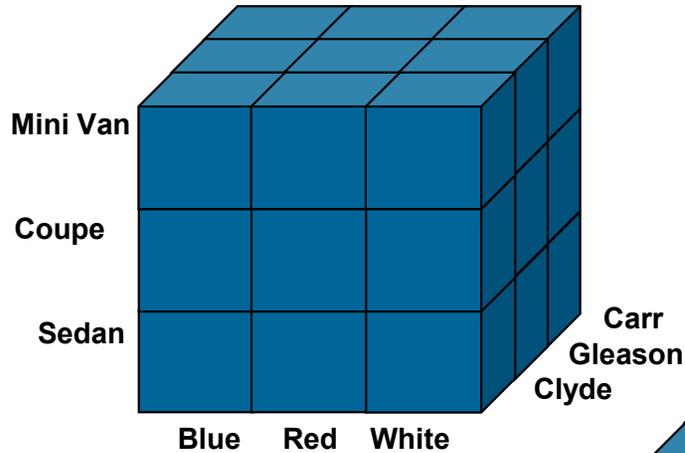
■ Cubos:

- Dada una Relación Dimensional, un cubo representa un cruzamiento concreto entre niveles determinados de las Jerarquías de la RD.

Modelo CMDM

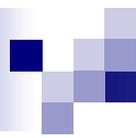
■ Cubos:

□ Ejemplo.



Modelo CMDM

- Conclusiones sobre CMDM.
 - Apunta a dar un modelo de especificación de esquemas conceptuales Multidimensionales.
 - Cumpliendo un rol equivalente al del Modelo ER.
 - Los esquemas conceptuales MD luego serán traducidos a esquemas lógicos MD.
 - Prioriza simplicidad y practicidad sobre expresividad.



Diseño Conceptual

Pautas y problemas en diseño conceptual

Pautas de diseño

- Preguntas básicas en diseño:
 - ¿Qué es una Dimensión y qué es una Medida?
 - ¿Tiene sentido construir cualquier cruzamiento de niveles?
 - ¿Cuándo definir una Dimensión con varias jerarquías o varias Dimensiones?
 - ¿Con qué criterios definir una o varias Relaciones Dimensionales?
 - Dada una relación dimensional, ¿cualquier operación de roll-up es correcta?
 - ¿Cómo clasificar las medidas calculadas?

Def. de Dimensiones y Medidas

- ¿Qué items pueden ser medidas y cuáles dimensiones?
 - No hay ningún criterio ni método efectivo para tomar la decisión.
 - Pueden haber “indicios” basados en:
 - Objetos del problema y/o criterios de análisis Vs. Indicadores y/o medidas.
 - Interés de totalizar según Drill-Ups.
 - Identificadores Vs. No Identificadores.
 - Estudio de distribución de valores.
 - Estudio de independencia de variables.

Def. de Dimensiones y Medidas

- Separar objetos vs. indicadores
 - Se analizan los requerimientos (lenguaje natural).
 - Ejemplos:
 - Los vendedores son un objeto del problema → tienen posibilidades de funcionar bien como una dimensión.
 - El tiempo es típicamente un criterio de análisis → candidato a dimensión.
 - Varios reportes necesitan totales por departamento → la geografía es candidata a dimensión
 - El promedio de autos vendidos es un indicador → es casi seguro una medida.
 - Contra-ejemplo:
 - A veces se habla de autos vendidos (objetos del problema) pero se quiere referir a cantidades de autos vendidos (indicadores).
 - Para facilitar la tarea, tratar de re-expresar los requerimientos como totalizaciones y cruzamientos.
 - Ej.: promedio de autos vendidos por vendedor y por fecha

Def. de Dimensiones y Medidas

- Estudio de Independencia de las variables.
 - ***Variables Independientes:*** Las dos pueden ser dimensiones simultáneamente.
 - ***Variables Dependientes:*** Dos niveles en la misma jerarquía.

No deberían existir niveles distintos de la misma jerarquía que sean independientes entre sí.

Jerarquías en dimensiones

■ Ejemplos:

- Los supervisores de ventas estudian zonas, que son conjuntos de departamentos.
- Los jefes de sección deben presentar informes trimestrales y anuales de ventas.
- Cada cliente puede realizar compras en diferentes ciudades del país.
- El pago de comisiones a vendedores depende de su antigüedad (porcentajes mayores cada 5 años de antigüedad).

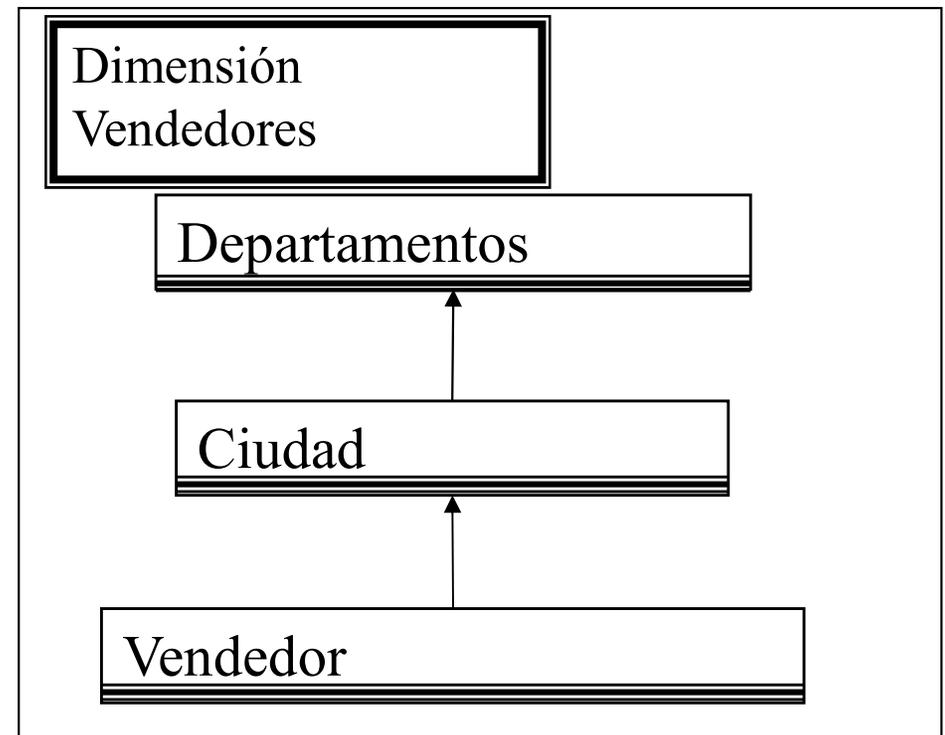
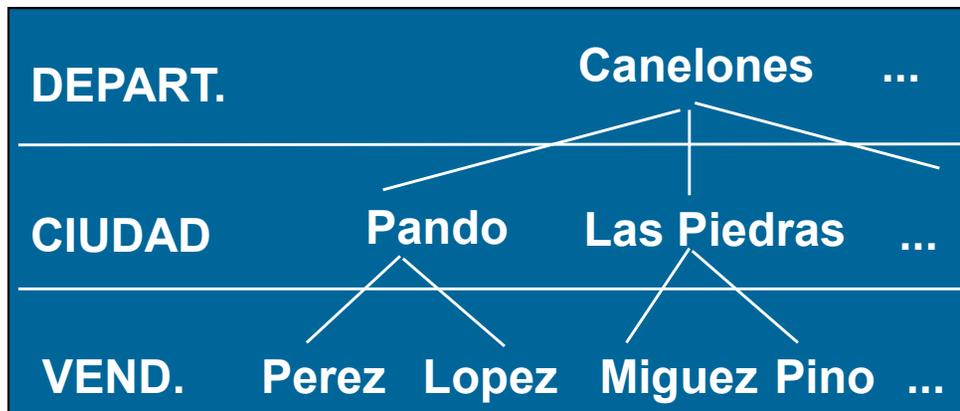
Jerarquías en dimensiones

- Criterios para construir jerarquías:
 - Definir caminos de navegación (drill-down/up).
 - Definir agrupamientos naturales de los datos.
 - Asociar diferentes niveles de cálculo de indicadores.
- Atención!
 - Verificar que las jerarquías estén bien formadas:
 - Cada objeto debería tener un y un sólo predecesor en la jerarquía

Jerarquías Estrictas

- Todas las relaciones entre niveles son N-1
 - Ejemplo:

Dimensión: Vendedores

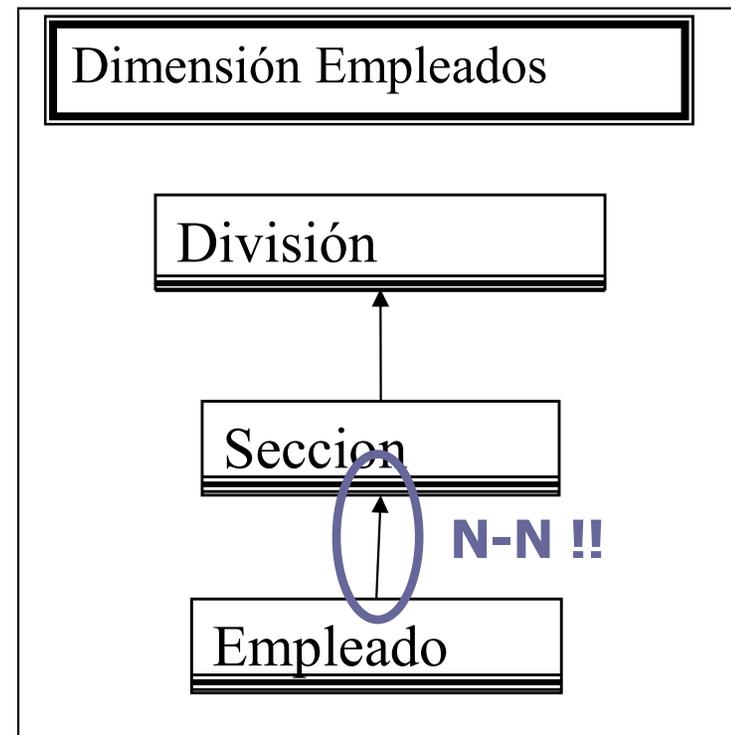
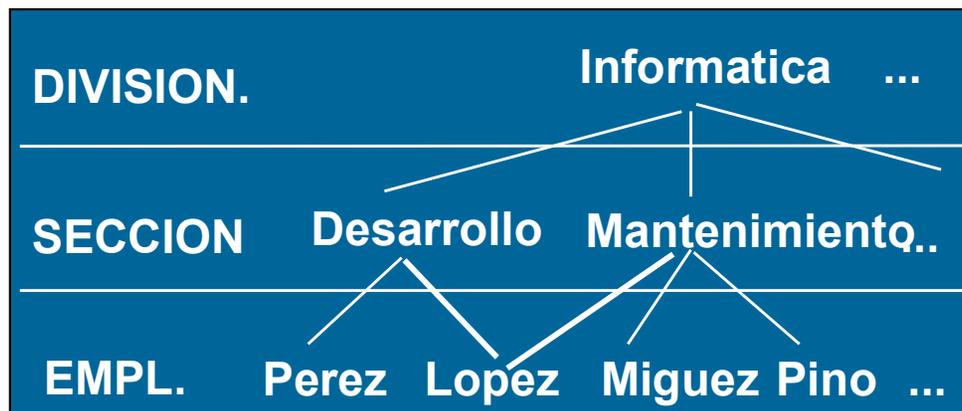


Jerarquías No-Estrictas

- Qué pasa si existe una relación N-N entre 2 niveles de una jerarquía?

□ Ejemplo:

Dimensión: Empleados



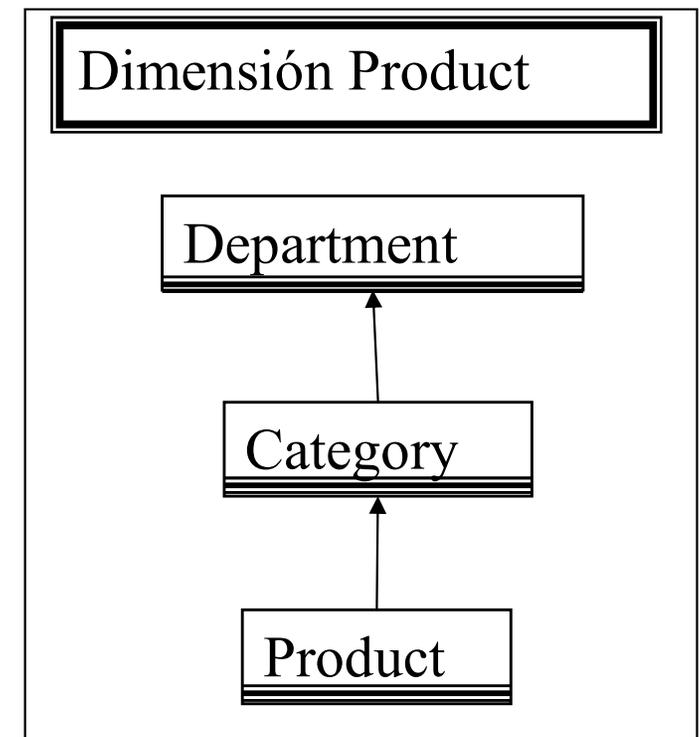
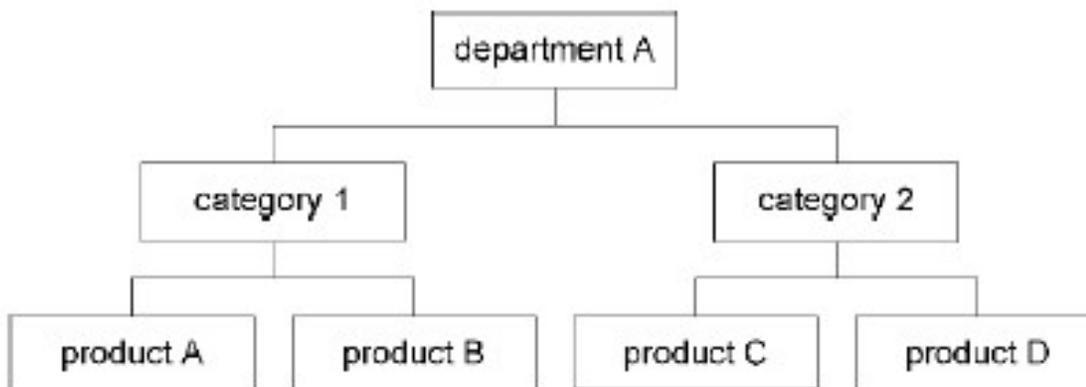
Jerarquías No-Estrictas

- Problema: doble-conteo
 - Cuando se hace roll-up a través de la jerarquía
 - Ej: salario de los empleados, cant. empleados
- Soluciones posibles
 - Transformar la jerarquía en estricta
 - Creando nuevos miembros que son la unión de otros
 - Ej: Sección Desarrollo-Mantenimiento
 - Eligiendo uno de los miembros padre para cada hijo
 - Decir como se distribuyen las medidas al subir de nivel
 - Crear una nueva dimensión

Tipos de jerarquías [Ma108,Vai14]

■ Balanceadas

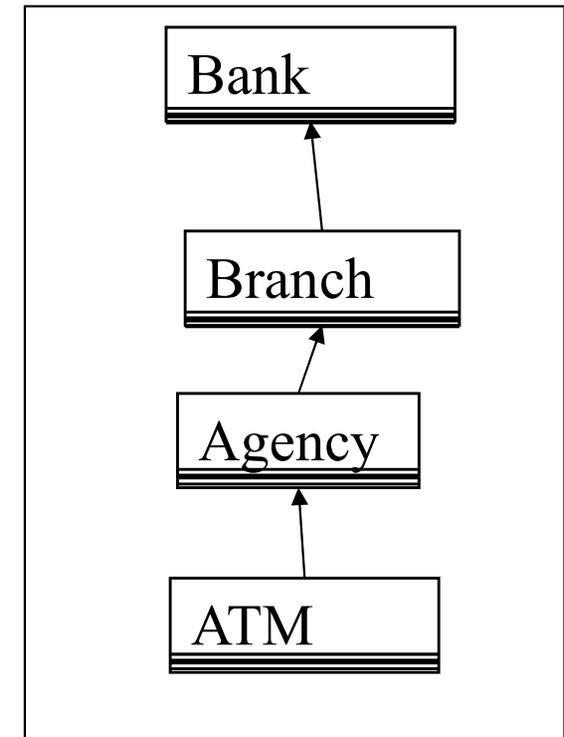
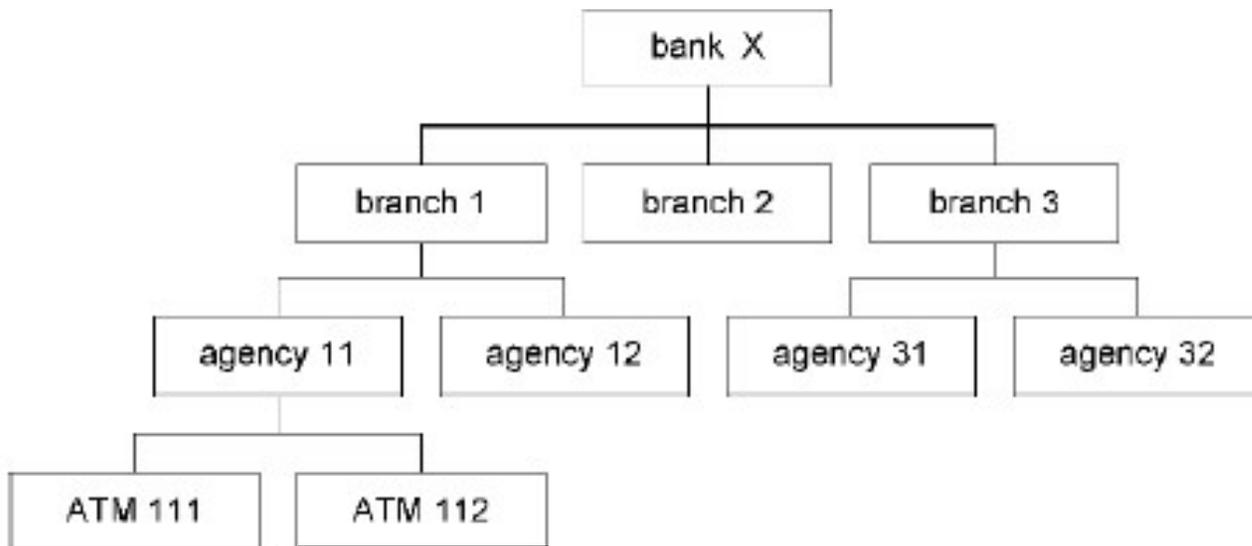
- Todos los miembros padres tienen por lo menos un miembro hijo y todos los miembros hijos tienen solamente un padre



Tipos de jerarquías

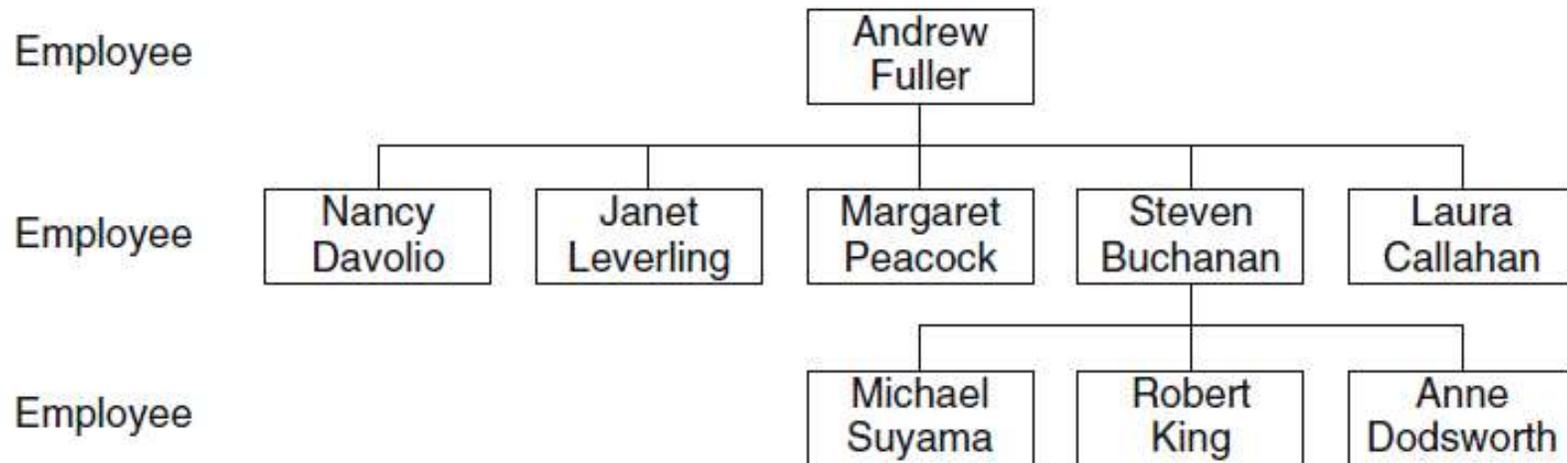
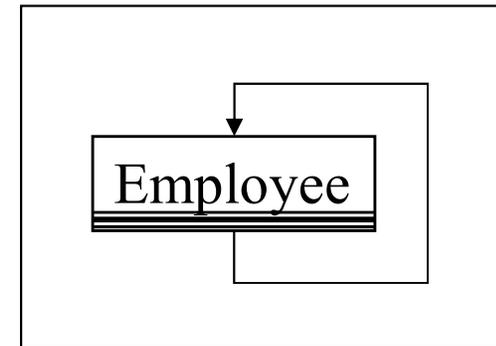
■ Desbalanceadas

- Miembros padre pueden no tener hijos asociados
- Incluyen el caso de jerarquías recursivas



Tipos de jerarquías

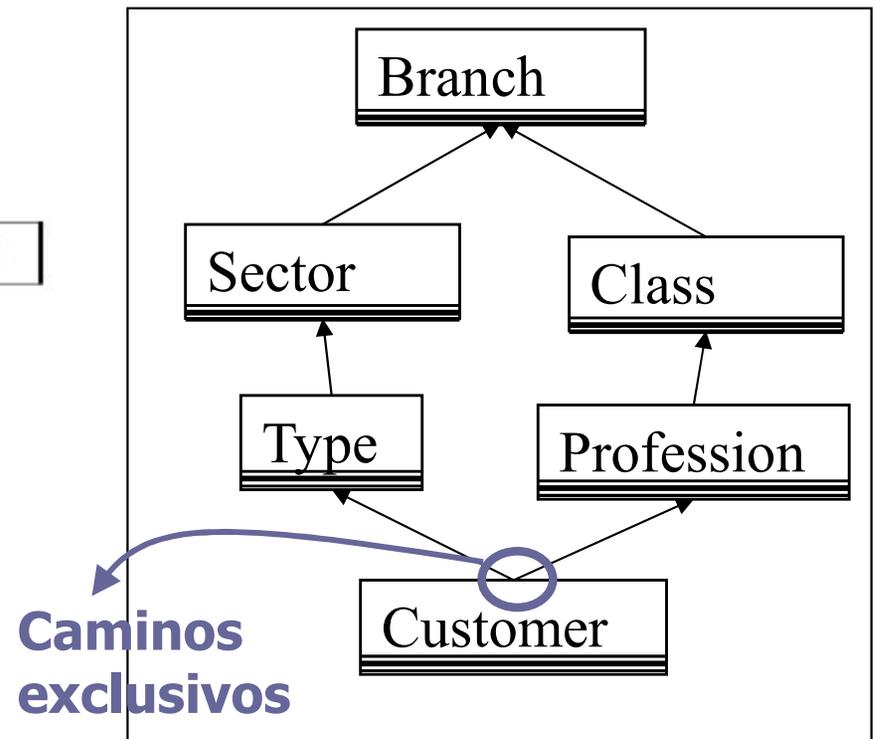
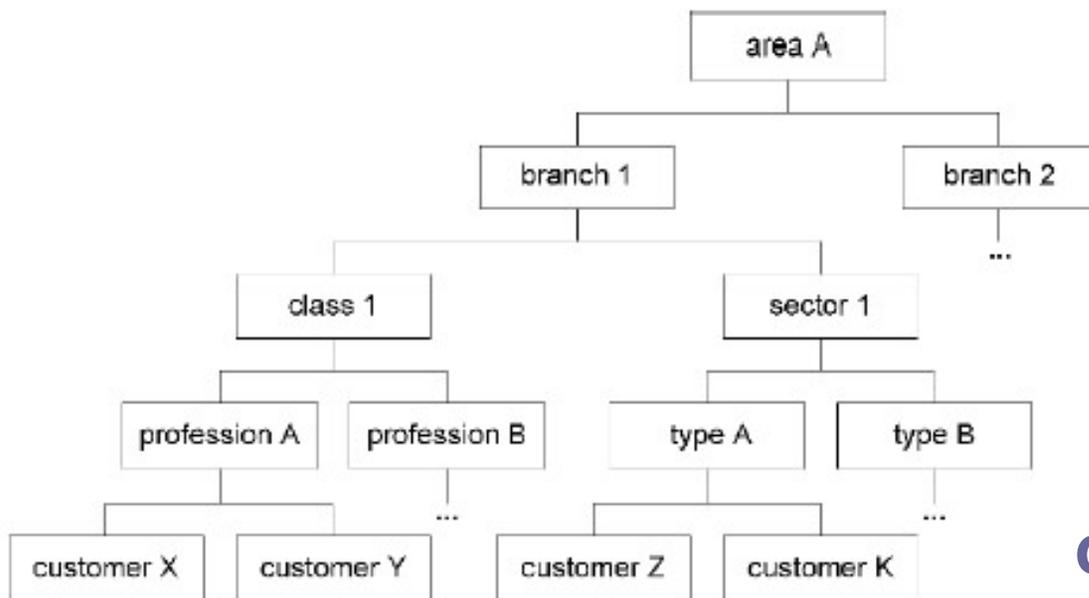
- Jerarquías recursivas



Tipos de jerarquías

■ Generalizadas

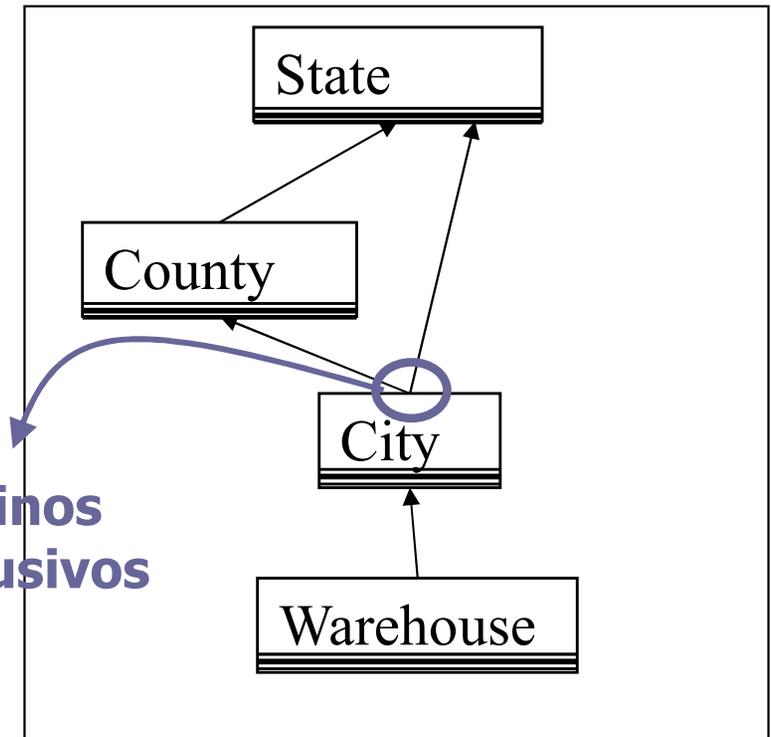
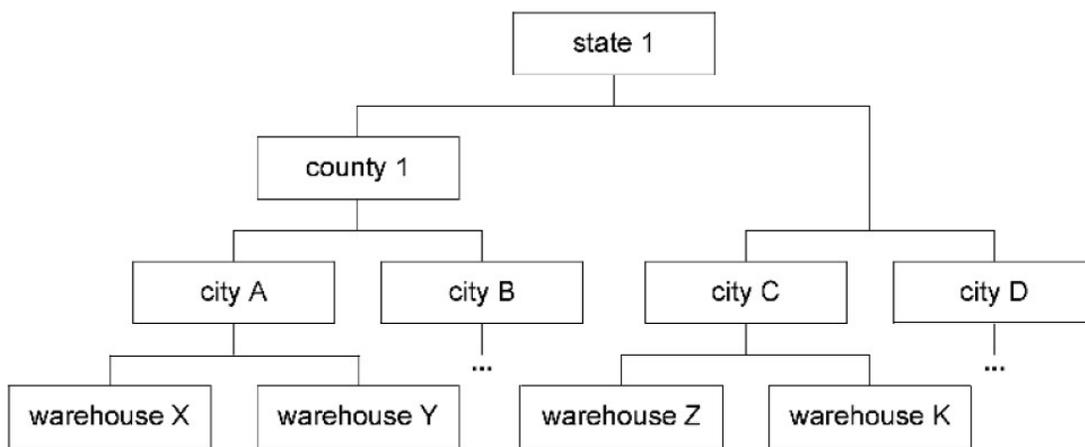
- Cuando una dimensión incluye especializaciones (sub-tipos), y estos sub-tipos tienen su propia jerarquía.



Tipos de jerarquías

■ Incompletas

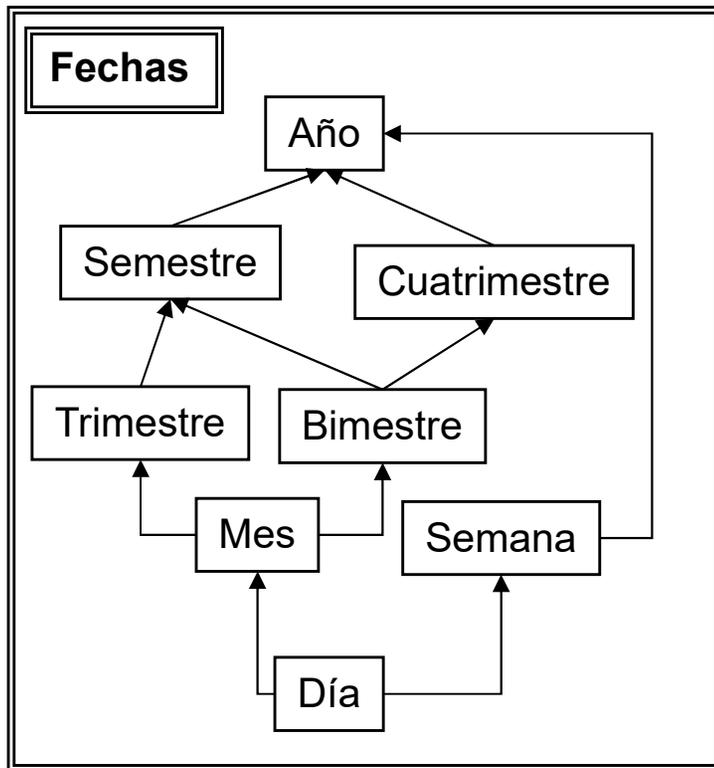
- Caso particular de Generalizada. Algunas instancias saltean niveles intermedios.



Tipos de jerarquías

■ Alternativas

- Varias jerarquías no exclusivas que comparten por lo menos el nivel más bajo.



Disyuntiva:

Cuándo es mejor crear otra dimensión en vez de agregar una jerarquía nueva en la misma dimensión.

Cambios en Jerarquías

- Problema:
 - Los agrupamientos de objetos pueden cambiar.
- Ejemplos:
 - Antigüedad de vendedores
 - Edad de clientes
 - Rubro de empresas
 - Rutas de venta
- Factores a tener en cuenta:
 - Frecuencia de los cambios
 - Interés de visualizar la historia (vs. conservar sólo el valor actual)

Cambios en Jerarquías

■ Soluciones:

□ Modificar valores (no guardar la historia).

- Puede ser necesario informar a los usuarios, ya que el cambio puede afectar análisis precedentes.
- Ej. Reestructuración de rutas de ventas.

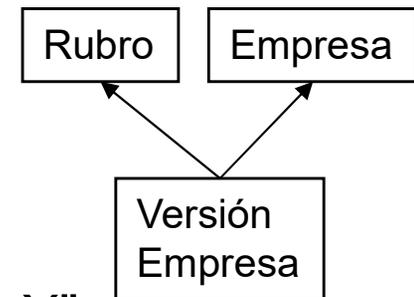
□ Mantener versiones

- Ej. “Empresa X v1” está en rubro R1,
“Empresa X v2” está en rubro R2.

Si es necesario, ambas totalizan como “Empresa X”.

□ Considerar dimensiones separadas

- Ej: Vendedores y Antigüedades
- Atención: Si se abusa de esta estrategia se puede llegar a tener todas dimensiones planas (de un solo nivel).

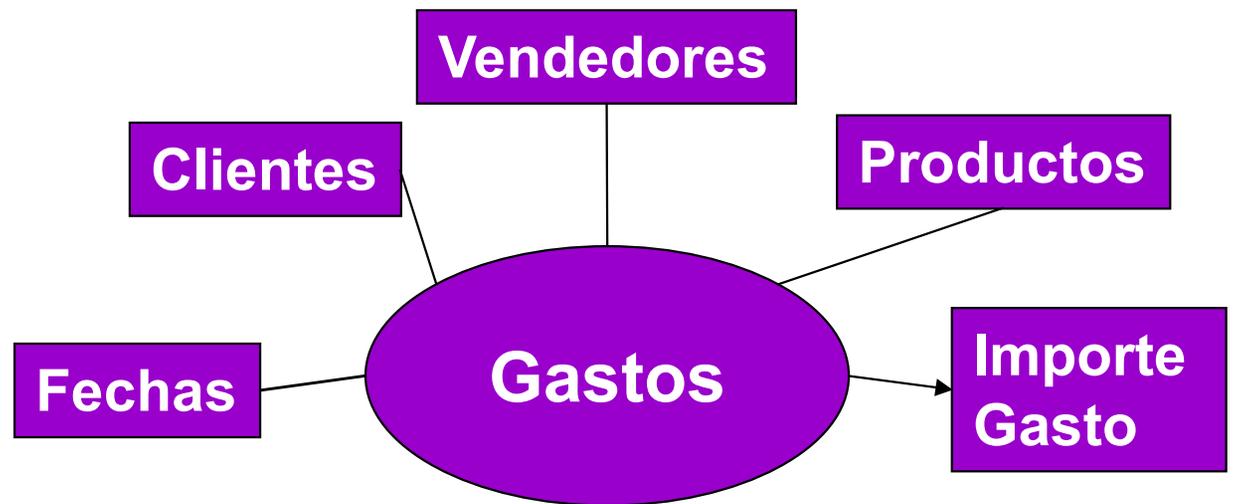
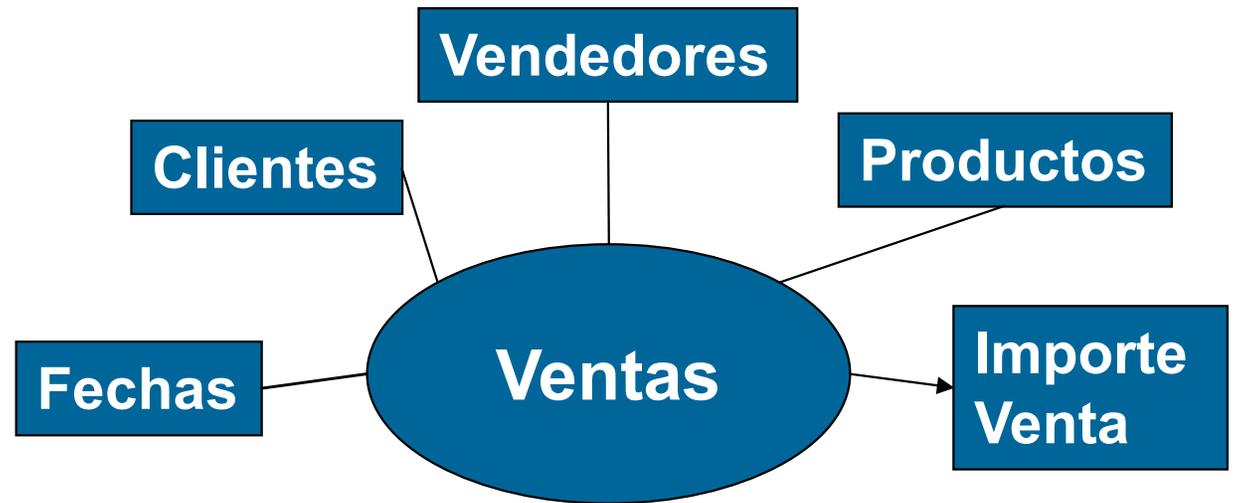
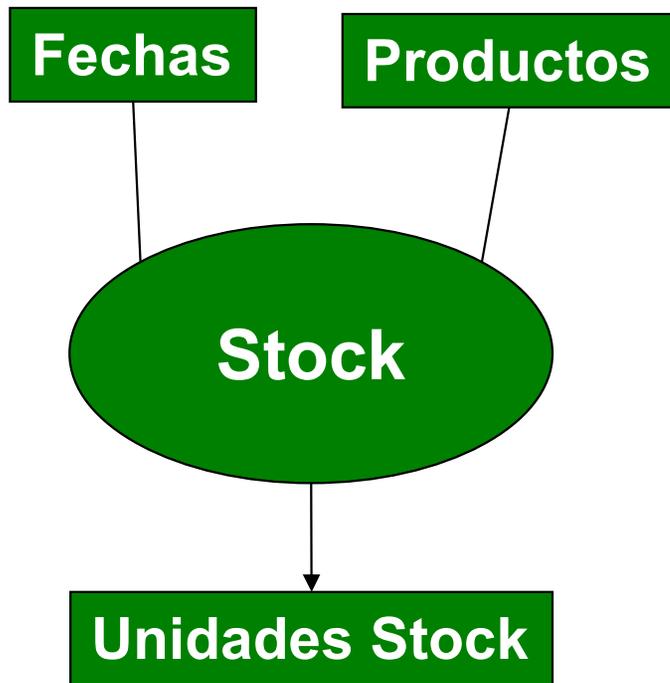


Relaciones dimensionales

- Disyuntiva:
 - Definir una o varias Relaciones Dimensionales
- Criterios para construir relaciones dimensionales:
 - Cada relación representa un concepto del problema.
 - Las dimensiones involucradas se cruzan entre sí.
 - Las medidas están relacionadas (cálculos).

Relaciones dimensionales

■ Ejemplos:

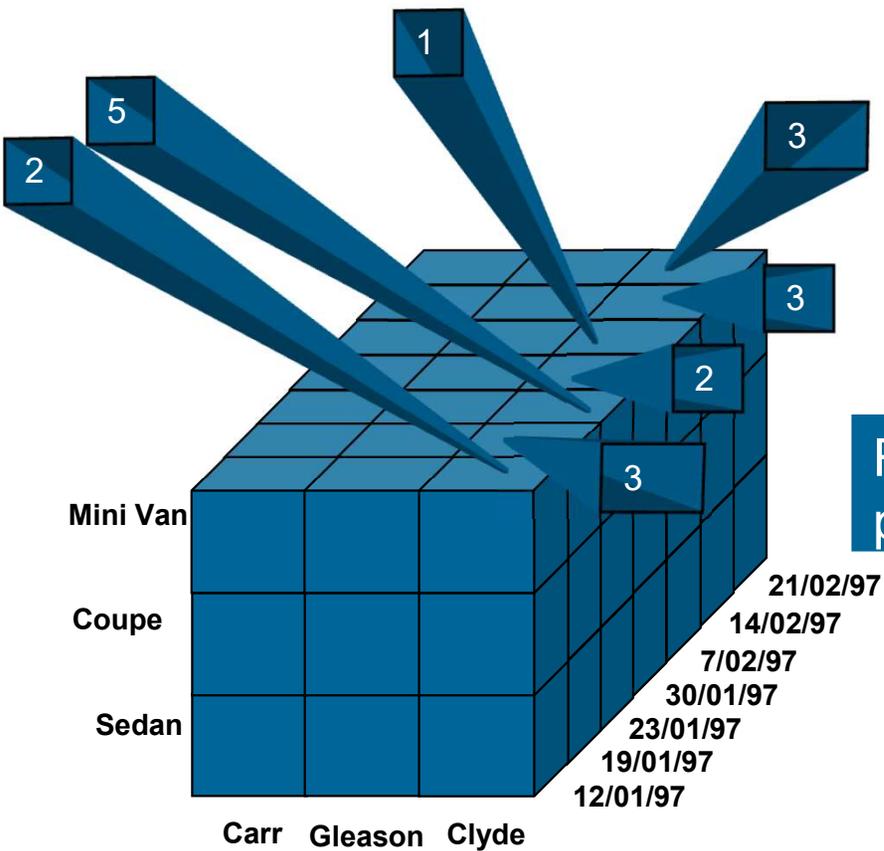


Funciones Roll-Up

- Dada una Relación Dimensional y una medida:
 - ¿Cualquier función de roll-up es válida?
- NO !
 - En particular no siempre se puede sumar valores.

Aditividad

Ventas

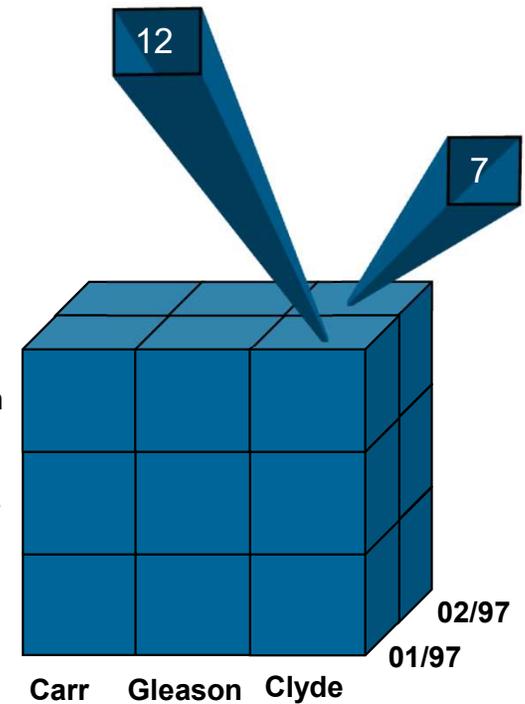


Resumen (suma)
por Fecha.

Mini Van

Coupe

Sedan



Problema: Medidas no disjuntas

Medida: Cantidad de vendedores diferentes que vendieron determinado modelo en determinada fecha.

$\sum_{\text{mes}=1 \wedge \text{año}=97} (\text{CantVendedores}(\text{"mini-van"}, \text{fecha}))$

Mini Van	2	0	2	2	1	1	1
Coupe	0	2	3	0	1	1	1
Sedan	1	1	0	1	2	1	1

12/01/97 19/01/97 23/01/97 30/01/97 7/02/97 14/02/97 21/02/97

RollUp (suma)
por Fecha ????

¿Cómo eliminar
repetidos?

6	3
5	3
3	4

01/97 02/97

Problema: Medidas no disjuntas

- Puede no haber disjuntez por ninguna dimensión:

- Ejemplo:

- Dimensiones: materias, fechas, institutos.
- Medida: cantidad de alumnos que rindieron examen.
 - Un alumno puede dar exámenes en varias fechas, de varias materias, de varios institutos.

Problema: Medidas no disjuntas

- Puede haber disjuntez sólo para algunas dimensiones:
 - Dimensiones: fechas, geografía, rubro, producto.
 - Medida: cantidad de clientes que compraron.
 - Un cliente sólo puede estar en un rubro y un lugar geográfico, pero puede comprar varios productos en varias fechas.
 - Es disjunta para las dimensiones que clasifican a los clientes.
 - Hay intersección para las otras dimensiones.

Problema: Cálculos

Medida: **Indice de venta por vendedor**

Definición: Cantidad vendida / Cantidad de vendedores

$\sum_{\text{mes}=1 \wedge \text{año}=97} (\text{IndiceVenta}(\text{"mini-van"}, \text{fecha}))$

Mini Van	1	1	4	1	2	1	1
Coupe	0	2	1	0	4	1	1
Sedan	3	4	0	3	3	1	1

12/01/97 19/01/97 23/01/97 30/01/97 7/02/97 14/02/97 21/02/97

RollUp (suma)
por Fecha ????

*total ventas /
total de vendedores*

7	4
3	6
10	5

01/97 02/97

Problema: Cálculos

■ Cuándo hacer el roll-up

□ Qué debo hacer?

- Aplicar el roll-up a cada medida y luego el cálculo.
- Hacer el cálculo y luego aplicar algún roll-up.

□ Los dos casos pueden ser válidos en distintos contextos.

Problema: Cálculos

■ Ejemplo:

□ Nota Promedio= suma notas / cantidad materias

- 1o (matematicas 3, fisica 5). Prom=4.
 - 2o (programacion 10). Prom=10.
 - Cantidad=3, Suma=18, Prom=6.
- } Prom=7

□ Caudal promedio

- Me interesa promediar los promedios de cada día, no importando la cantidad de muestras.
- Lunes (10,11,12,11,11), martes (15,15), miercoles(14,14). Prom=13,3.
- Cantidad=9, Suma=113, Prom=12,6.

Tipos de Medidas

■ Aditivas (Flow)

- Conserva la semántica al aplicar la suma como RollUp, realizando DrillUp por cualquier dimensión.
- Refieren a un evento o período, y son registradas al final del mismo.
- Ejemplos:
 - \$ en una factura, cantidad de nacimientos por mes, expedientes procesados en una semana, cantidad de ventas en el día.

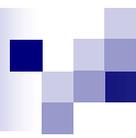
Tipos de Medidas

■ Semi-aditivas (Stock)

- Conserva la semántica al aplicar la suma como RollUp, al realizar DrillUp, en algunas de las dimensiones.
- Son registradas en un punto específico del tiempo y refieren a ese instante.
- Ejemplos:
 - Inventarios (de cualquier tipo).
 - Saldos de cuentas.
 - Censos de ciudadanos.

Tipos de medidas

- No aditivas (Value-per-unit)
 - No conserva la semántica al aplicar la suma como RollUp, al realizar DrillUp, en ninguna dimensión.
 - Son determinados en un momento dado pero su unidad de medida las diferencia de stock.
 - Ejemplos:
 - Precio por ítem, tasa de cambio, edades, notas.
 - Mediciones: temperaturas, caudales.
 - Anomalías de disjuntez.



Diseño Conceptual

Estrategia basada en requerimientos

Requerimientos

- Dimensiones y medidas involucradas en cada requerimiento.

Referencia: [Bal98]

<i>Table 1. Dimensions, Measures, and Related Questions</i>									
Dimensions and Measures	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
<i>Dimensions</i>									
Sales		X		X	X			X	X
Manufacturing	X		X						
Product	X	X	X		X	X	X	X	X
<i>Measures</i>									
Average quantity on hand	X								
Total cost		X	X				X		
Total revenue		X	X				X		
Quantity sold							X		
Percentage of models eligible for discount				X					
Percentage of models eligible for discount that are actually discounted				X					
Percentage of a model sold through a retail outlet					X				
Percentage of a model sold through a corporate sales office order desk					X				
Percentage of a model sold through a sales person					X				

1a versión del esq. conceptual

■ Definir:

- Jerarquías para las dimensiones.
- Medidas.
- Relaciones dimensionales.

■ Se utiliza un modelo conceptual:

- Por ejemplo CMDM.

Correspondencias

- Correspondencias entre:
 - Esquema conceptual.
 - Bases fuentes.
- Objetivo:
 - Ubicar en las fuentes los objetos conceptuales.
 - Verificar que los datos existen.
- Documentación:
 - Cuadros de correspondencias.
 - Parejas: item de un nivel, atributo de una tabla o cálculo.

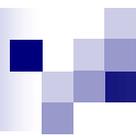
Refinamiento del Esquema

- Aplicar pautas de diseño.
 - Refinar jerarquías de dimensiones.
 - Medidas calculadas.
 - Funciones de roll-up.
 - Estudio de aditividad (roll-up suma):
 - Para cada dimensión.
 - Para los niveles de las dimensiones.
- Documentación:
 - Esquema conceptual.
 - Cuadros de roll-up.

Estudio de Aditividad

■ Cuadros de Roll-Up.

Dimension Medida		Cant. Vendida	Importe
Fecha	Dia -> DiaSemana	Prom	Prom, +
	Dia -> Mes	+	+
	Mes -> Año	+	+
Producto	Producto -> Tipo-Prod	+	NA
	Producto -> Tamaño	+	NA
Vendedor	Todos los niveles	+	+



Diseño Conceptual

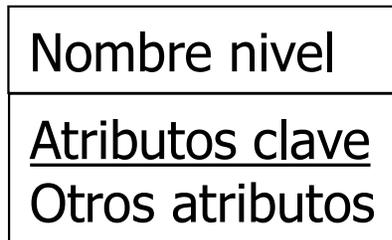
Modelo MultiDim

A. Vaisman – E. Zimányi

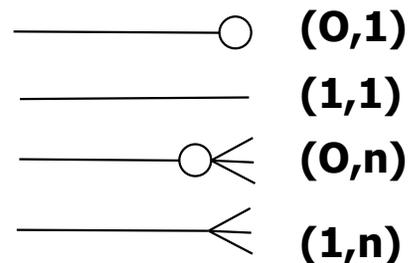
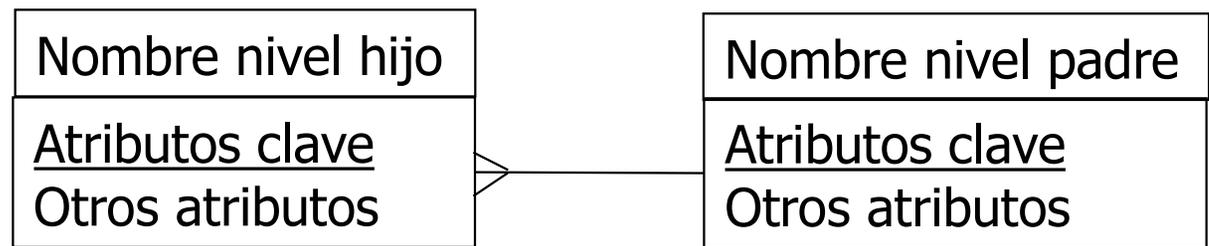
[Vai14]

Niveles - Jerarquías

Nivel



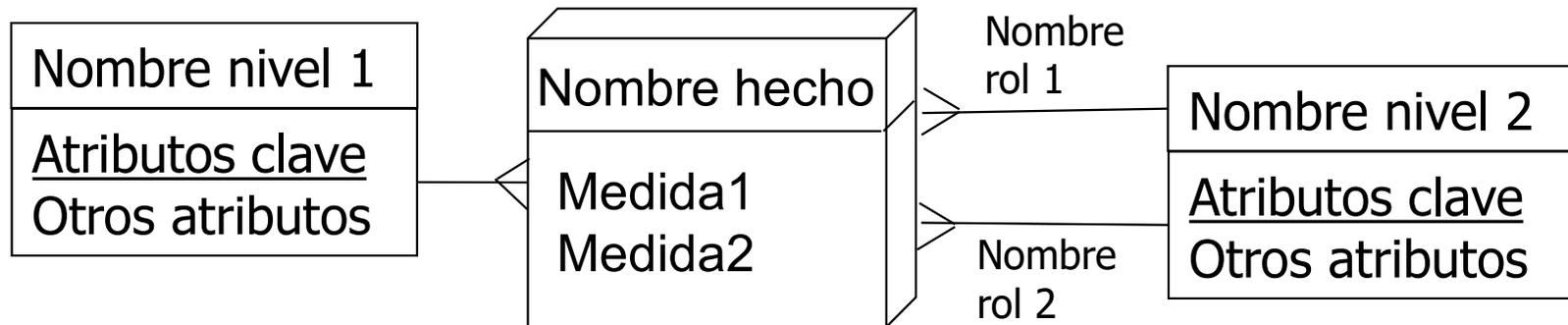
Jerarquía



Cardinalidades

Hechos - medidas - jerarquías

Hecho con medidas



Tipos de medidas

Aditiva
Semiaditiva +!
No aditiva ≠
/Derivada

Jerarquía

Nombre Jerarq.

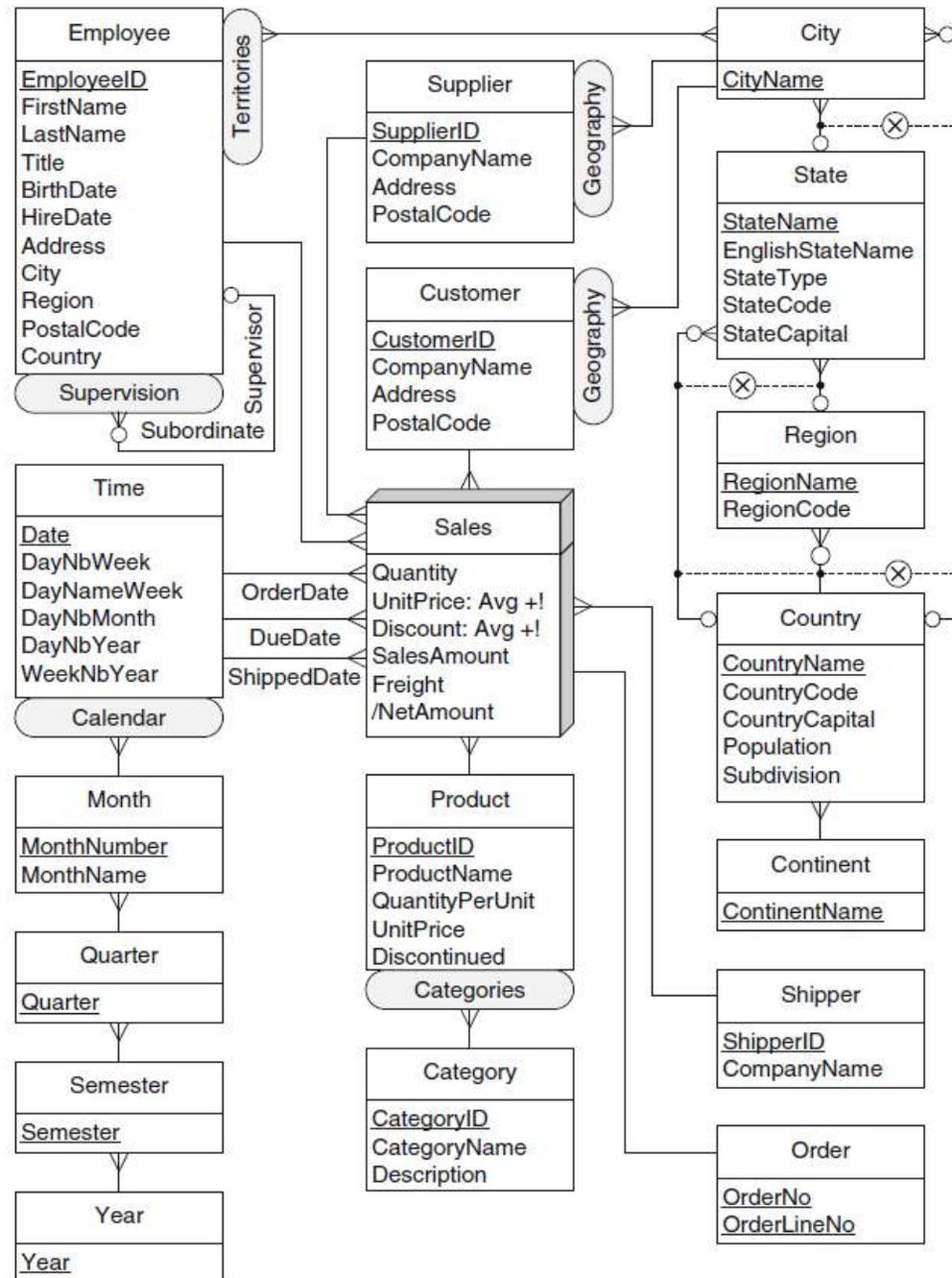
Factor de distribución



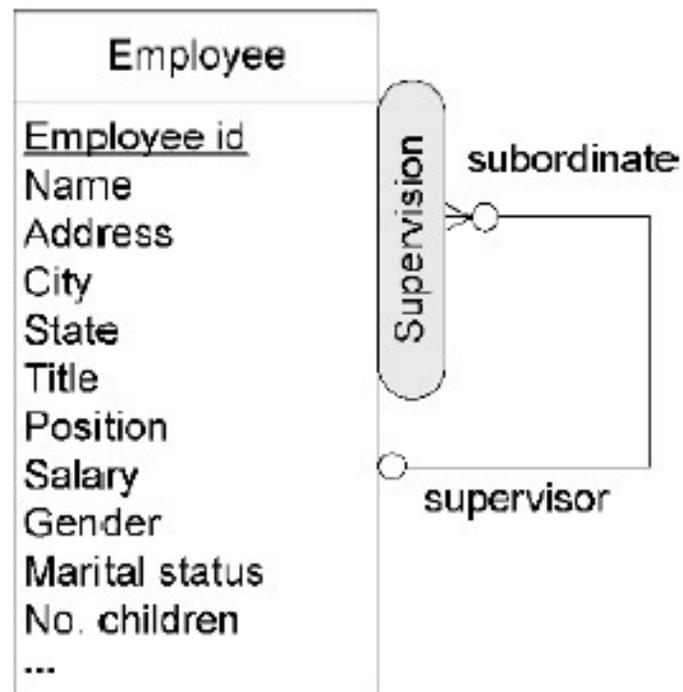
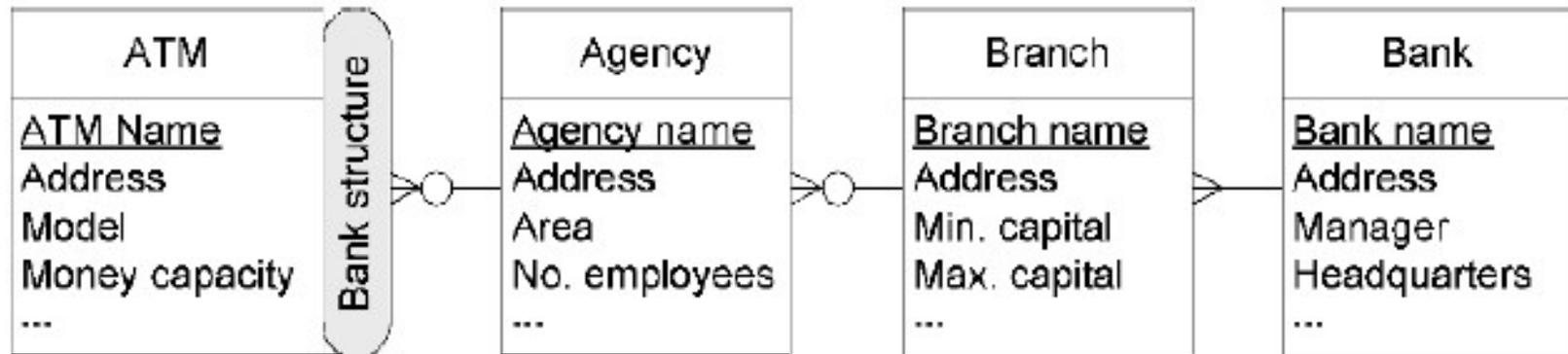
Relación exclusiva



Ejemplo



Ejemplos - jerarquías



Referencias principales

- [Car00] "CMDM: A conceptual multidimensional model for Data Warehouse". F. Carpani. Tesis de Maestría. InCo - Pedeciba, Udelar, Uruguay, 2000.
- [Gol10] "Data Warehouse Design. Modern Principles and Methodologies". M. Golfarelli, S. Rizzi. McGraw Hill, 2009.
- [Mal08] "Advanced Data Warehouse Design", Elzbieta Malinowski, Esteban Zimanyi. Springer, 2008. ISBN: 978-3-540-74404-7
- [Vai14] "Data Warehouse Systems Design and Implementation". Alejandro Vaisman, Esteban Zimányi. Springer Series: Data-Centric Systems and Applications 2014, XXVI, 603 p. ISBN 978-3-642-54654-9