

Modelo de Datos Conceptual Canónico

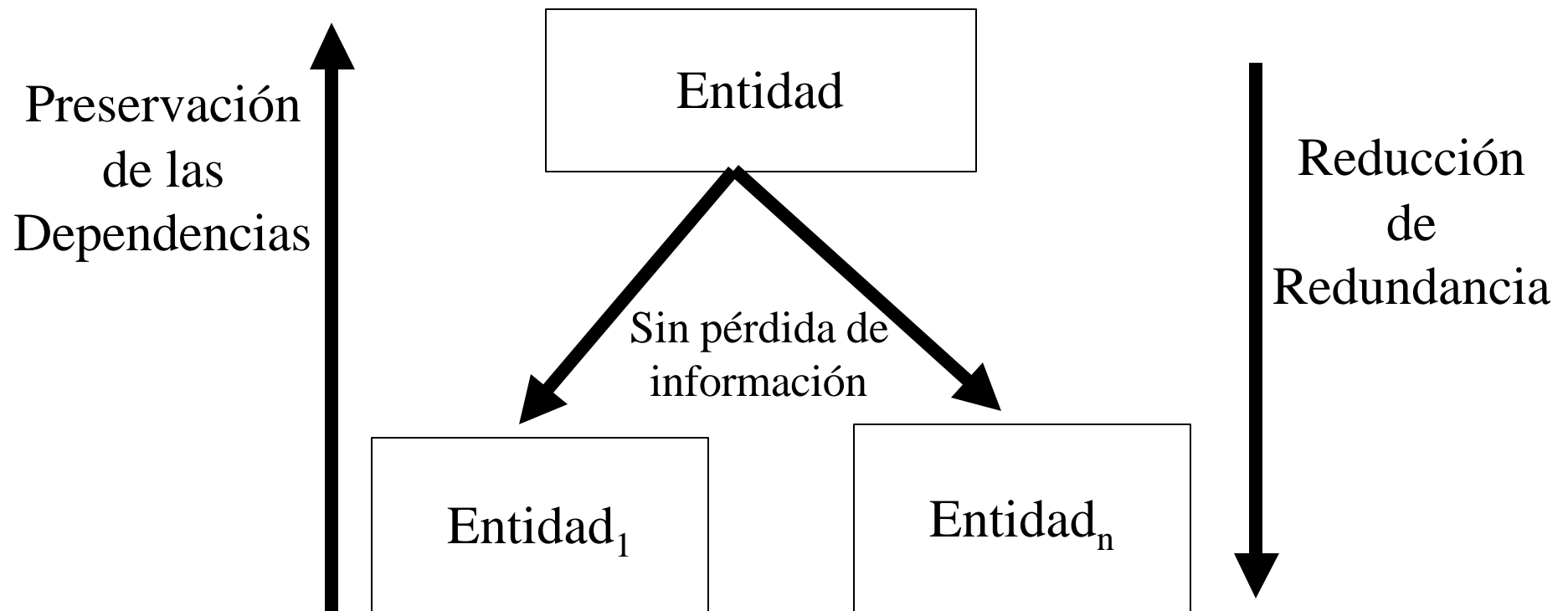
Sistemas de Información Administrativos
Departamento de Ingeniería Industrial
Universidad de Chile



Diseño de Datos

- Objetivo: generar un modelo de datos que permita almacenar los datos sin innecesaria redundancia y fácil acceso a ellos.
- Problemas
 - Información Repetida
 - Inhabilidad de Representación
 - Pérdida de Información
- La mayoría de estos problemas pueden ser resueltos en el modelo lógico de datos (Entidad-Relación).

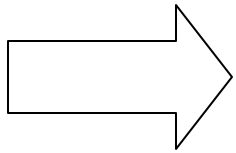
Criterios en el Diseño





Ejemplo: Información Repetida

- Suponga una entidad Banco = (nombre_sucursal, activos, ciudad, #créditos, cliente, cantidad).
- Se desea agregar un crédito a una sucursal y cliente existente



Se debe repetir nombre_sucursal, activos y ciudad

- Repetir información NO es deseable, ya que requiere espacio, presenta problemas al actualizar (hay que repetirla).

Ejemplo: Inhabilidad de Representación

- La representación basada en Banco NO permite representar una SUCURSAL directamente.
- Una alternativa es introducir valores NULL, pero son difíciles de manejar
 - ¿Qué significan?
 - ¿Cómo los uso para extraer información?

Ejemplo: Pérdida de Información

- El diseño basado en BANCO es malo, y nos gustaría descomponerlo para mejorar su calidad.
- Sin embargo un diseño descuidado puede generar problemas.
- Suponga que BANCO es descompuesto en:
 - Sucursal = (nombre_sucursal, activos, ciudad)
 - Crédito = (nombre_sucursal, #crédito, cliente, cantidad)
 - Cantidad = (cantidad, cliente)
 - Préstamo = (nombre_sucursal, #crédito, cantidad)

No se puede reconstruir Crédito!!

Modelo de datos conceptual canónico (MDCC)

La forma canónica del modelo de datos conceptual es aquella en que:

- El modelo está normalizado (en 5ª forma normal),
- No hay datos redundantes, y
- Se preservan las dependencias de los datos

Normalización (según Codd)

- Es un proceso reversible mediante el cual los datos de un modelo de datos se ponen en la forma más simple e independiente
- Permite que, si la estructura física para los datos que se implemente esta normalizada, siempre será posible satisfacer cualquier requerimiento sobre los datos

MDC ↔ 1FN ↔ 2FN ↔ 3FN ↔ 4FN ↔ 5FN ↔ MDCC

- Veremos sólo las tres primera formas normales que son las más frecuentes

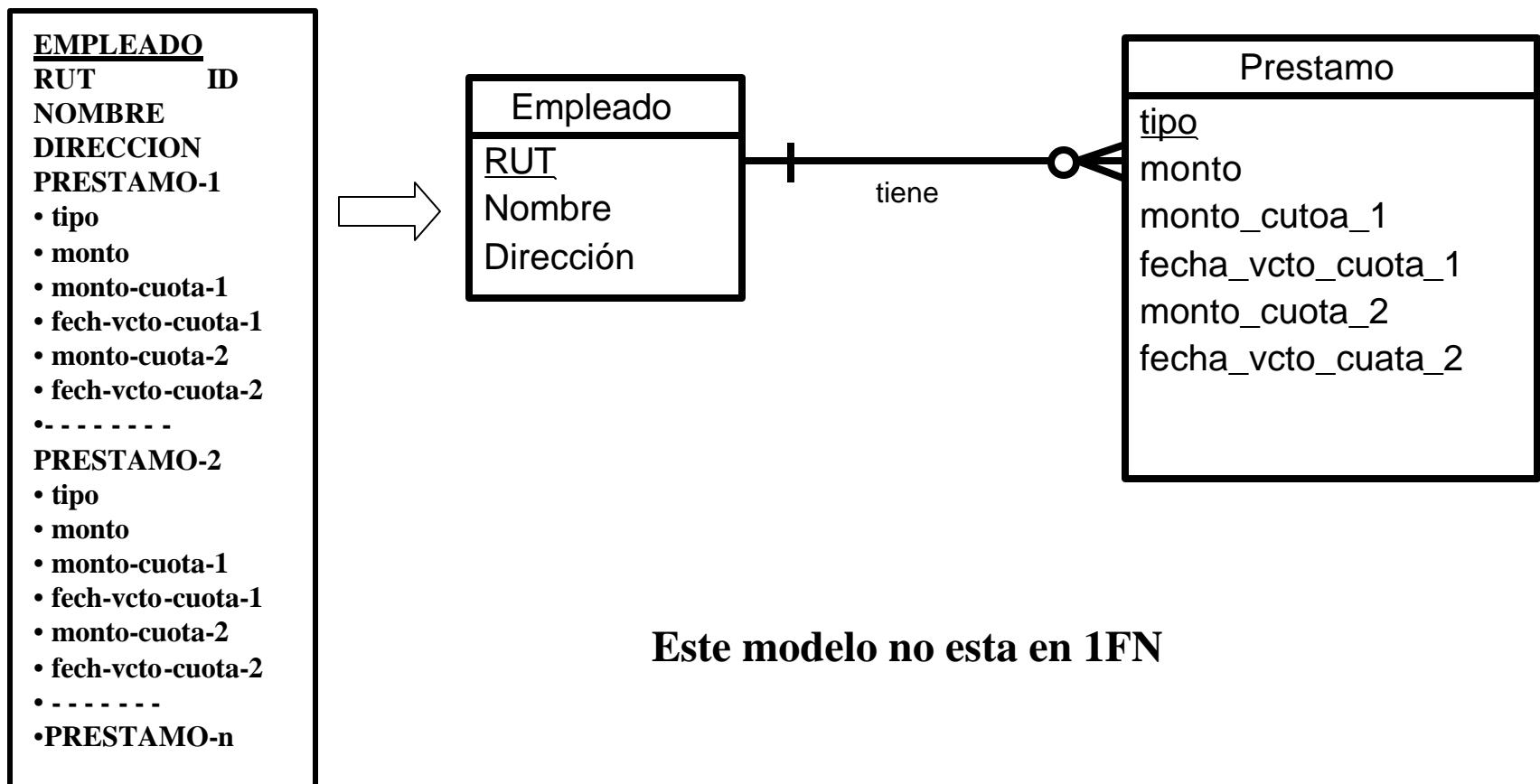
1FN - Primera forma normal (1)

<u>EMPLEADO</u>	
RUT	ID
NOMBRE	
DIRECCION	
PRESTAMO-1	
• tipo	
• monto	
• monto-cuota-1	
• fech-vcto-cuota-1	
• monto-cuota-2	
• fech-vcto-cuota-2	
• - - - - -	
PRESTAMO-2	
• tipo	
• monto	
• monto-cuota-1	
• fech-vcto-cuota-1	
• monto-cuota-2	
• fech-vcto-cuota-2	
• - - - - -	
PRESTAMO-n	
• tipo	
• - - - - -	

- Un modelo de datos está en primera forma normal si no tiene atributos repetitivos

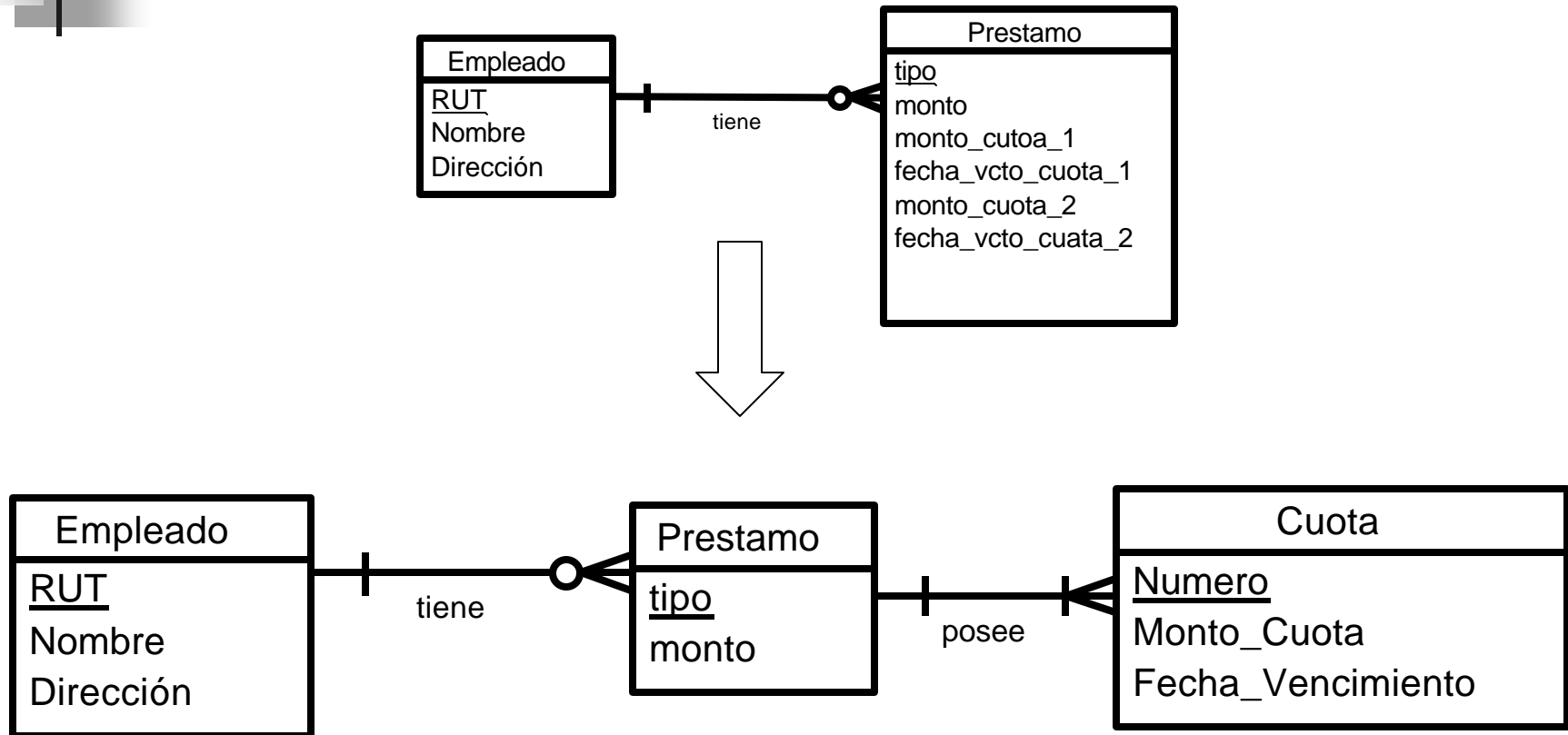
Este modelo no esta en 1FN

1FN - Primera forma normal (2)



Este modelo no esta en 1FN

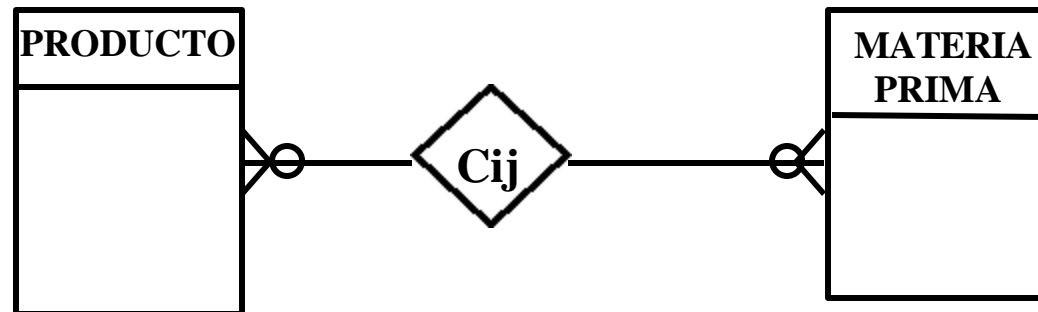
1FN - Primera forma normal (3)



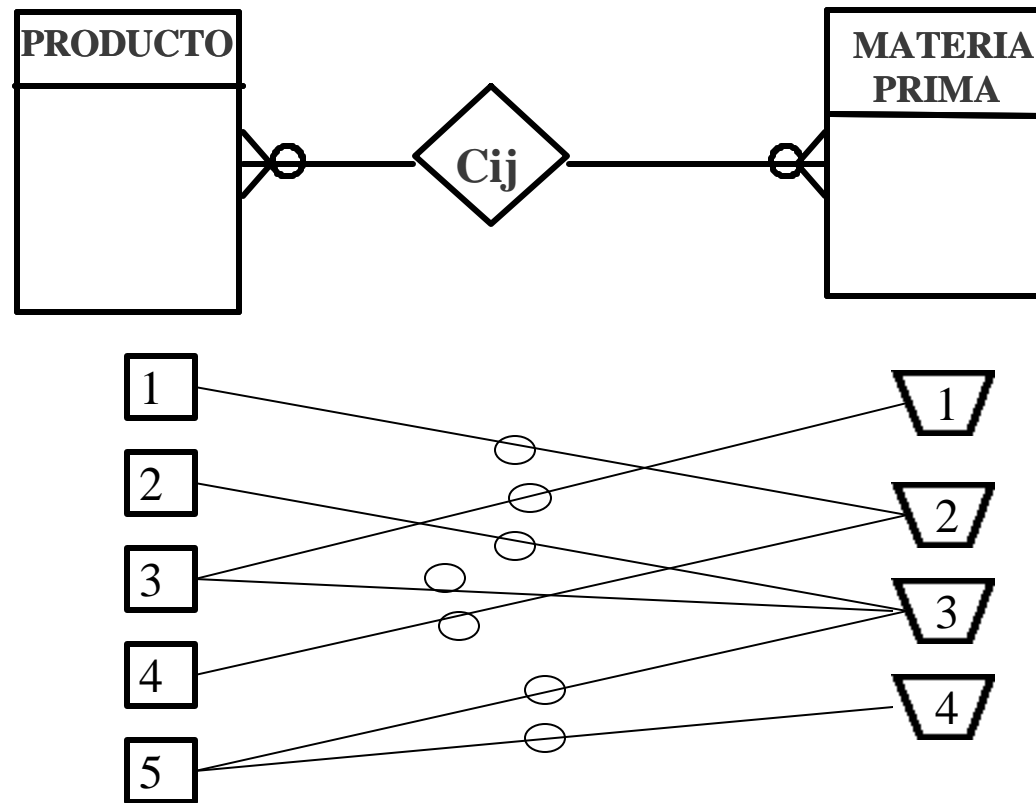
Este modelo si está en 1FN

2FN - Segunda forma normal (1)

- Un modelo está en segunda forma normal si no tiene relaciones M:N (muchos a muchos)

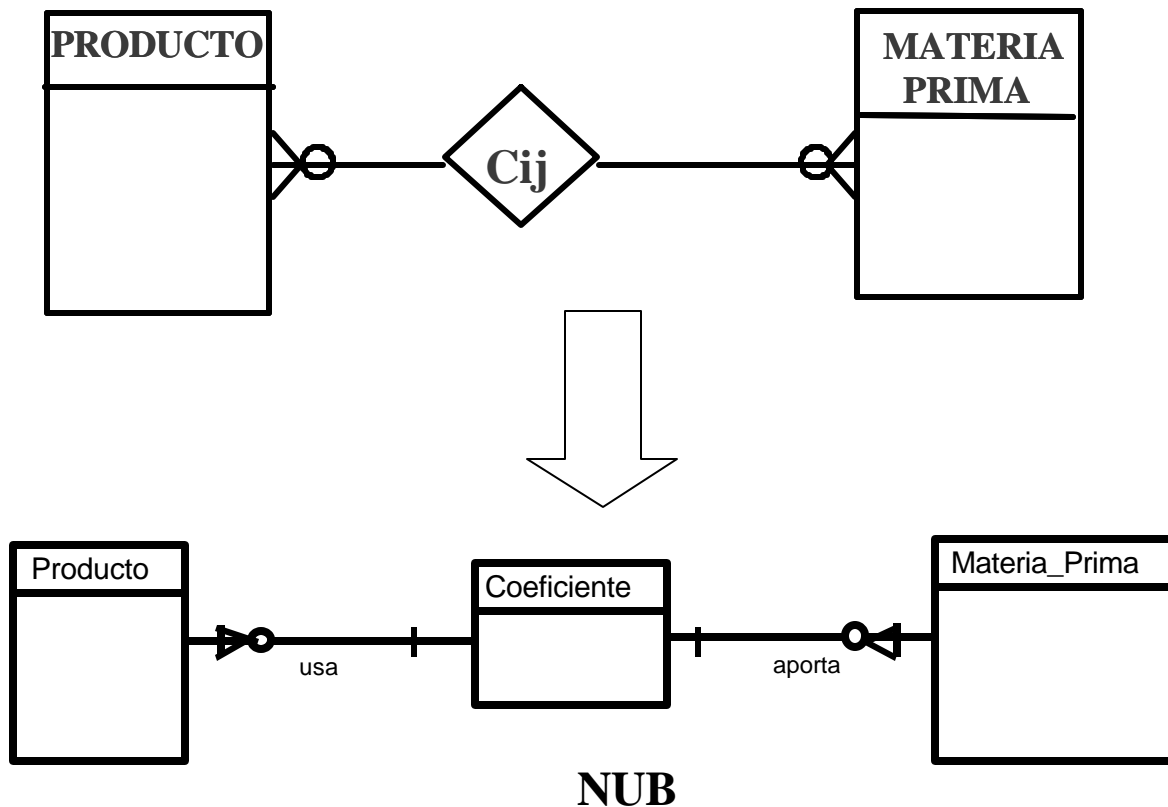


2FN - Segunda forma normal (2)



Agreguemos una nueva entidad (NUB) de la cual habrá una ocurrencia por cada ocurrencia de la relación o sea una ocurrencia por cada par (i,j) que esté efectivamente relacionado

2FN - Segunda forma normal (3)



- La representación requiere de una nueva entidad (NUB)
- Se debe utilizar como entidad débil

2FN - Segunda Forma Normal (4)

- El NUB puede asociarse al elemento de una matriz de orden n , en que n corresponde al número de entidades que participan en la relación.
- De esta forma, el NUB tendrá tantos padres (e índices en los atributos) como entidades participen en la relación.
- Ej: C_{ijk} : cantidad del producto \underline{i} programado para ser producido en la máquina \underline{j} en el período \underline{k} .

3FN - Tercera forma normal (1)

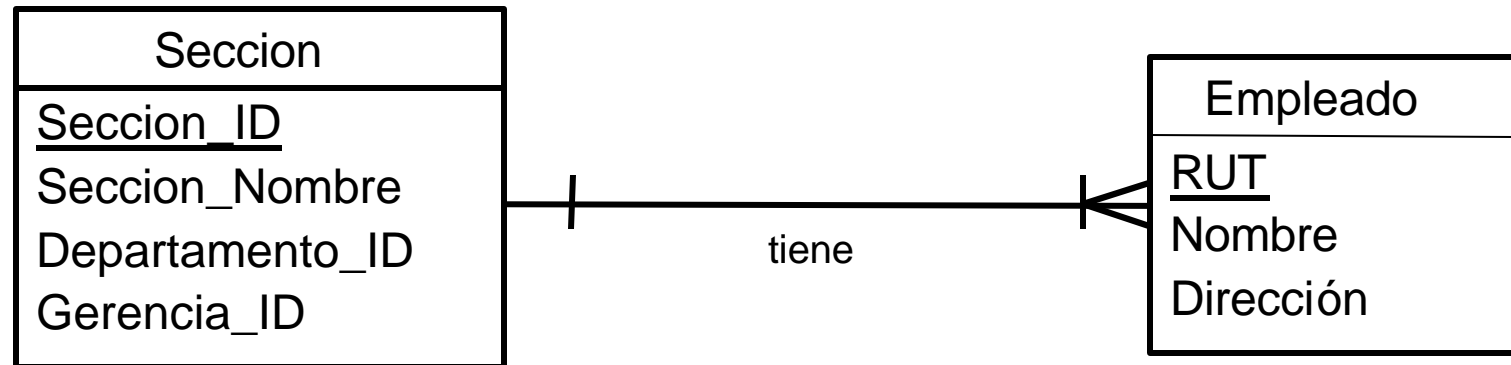
- Un modelo está en tercera forma normal si no tiene transitividades (dependencias funcionales), o sea, no tienen atributos que estén determinados linealmente por otros.

EMPLEADO

. RUT
. Nombre
. Dirección
. #Sección
. #Depto.
. #Gerencia

Este modelo no está en 3FN

3FN - Tercera forma normal (2)

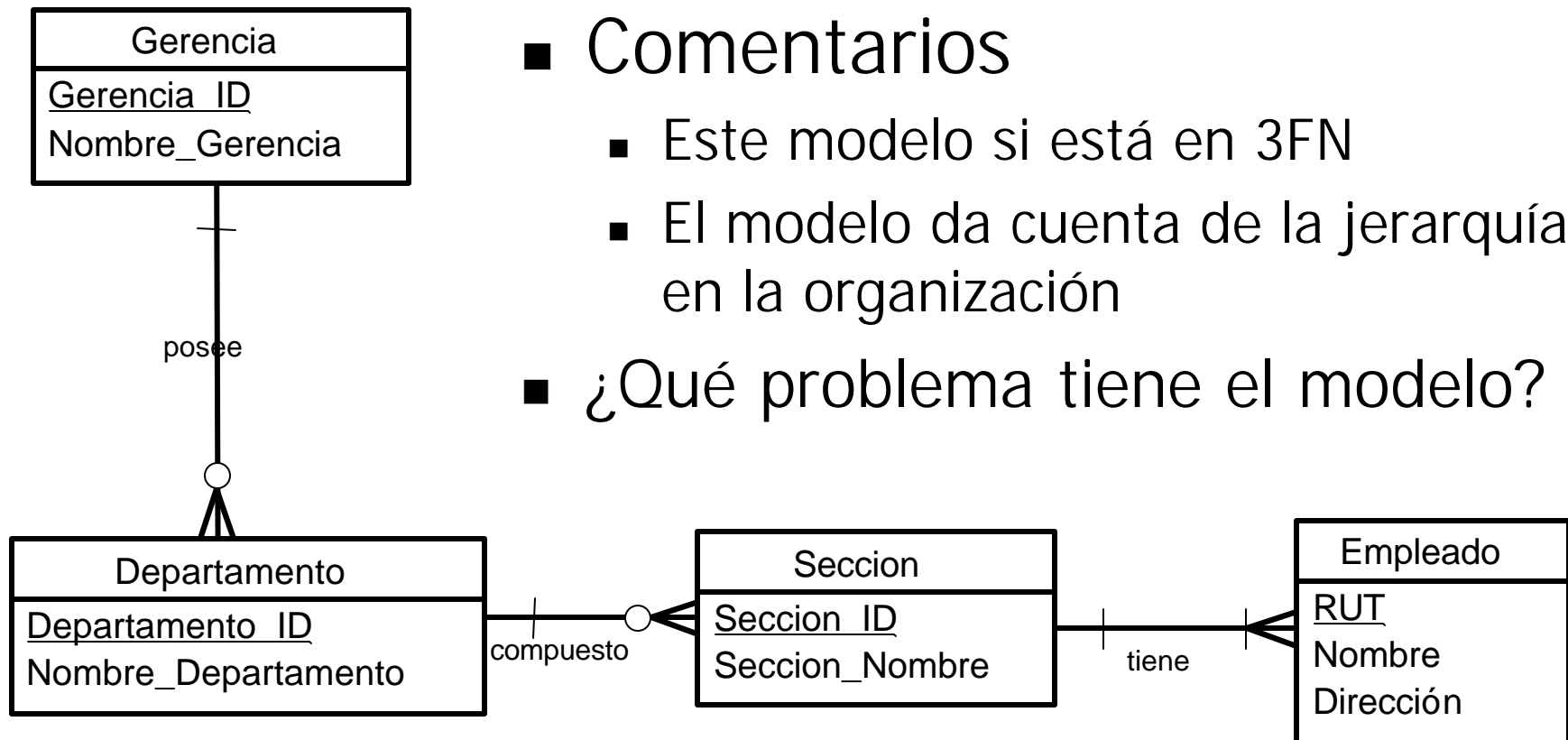


Este modelo no está en 3FN

■ Razones

- En un departamento hay una o más secciones
- En una gerencia hay uno o más departamento
- ¿Cómo se puede modelar esto?

3FN - Tercera forma normal (4)



■ Comentarios

- Este modelo si está en 3FN
- El modelo da cuenta de la jerarquía en la organización
- ¿Qué problema tiene el modelo?

Este modelo si está en 3FN



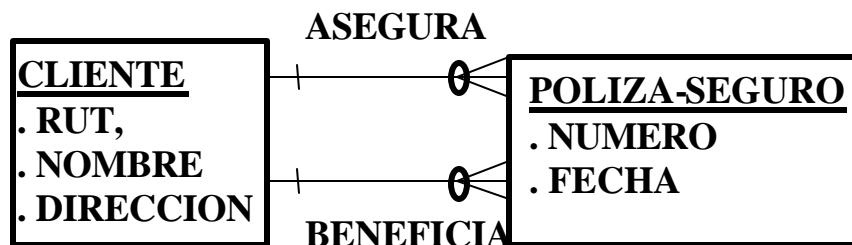
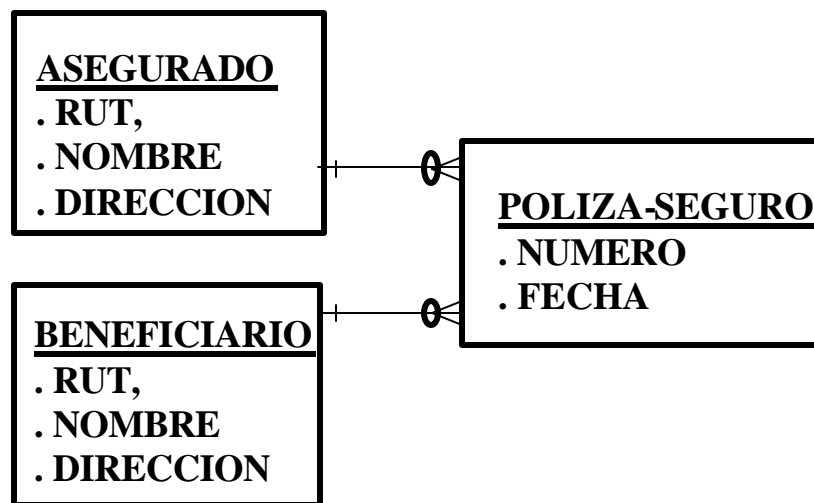
Eliminación de redundancia

- Entidades:

- Dos entidades tienen redundancia de estructura si sus atributos son iguales (o un % dado), pero no hay una ocurrencia del mismo sujeto en cada entidad.
- Dos entidades tienen redundancia de datos si una misma ocurrencia puede tener los mismos datos en ocurrencias distintas de las entidades.

Redundancia en las entidades

modelo con redundancia de datos



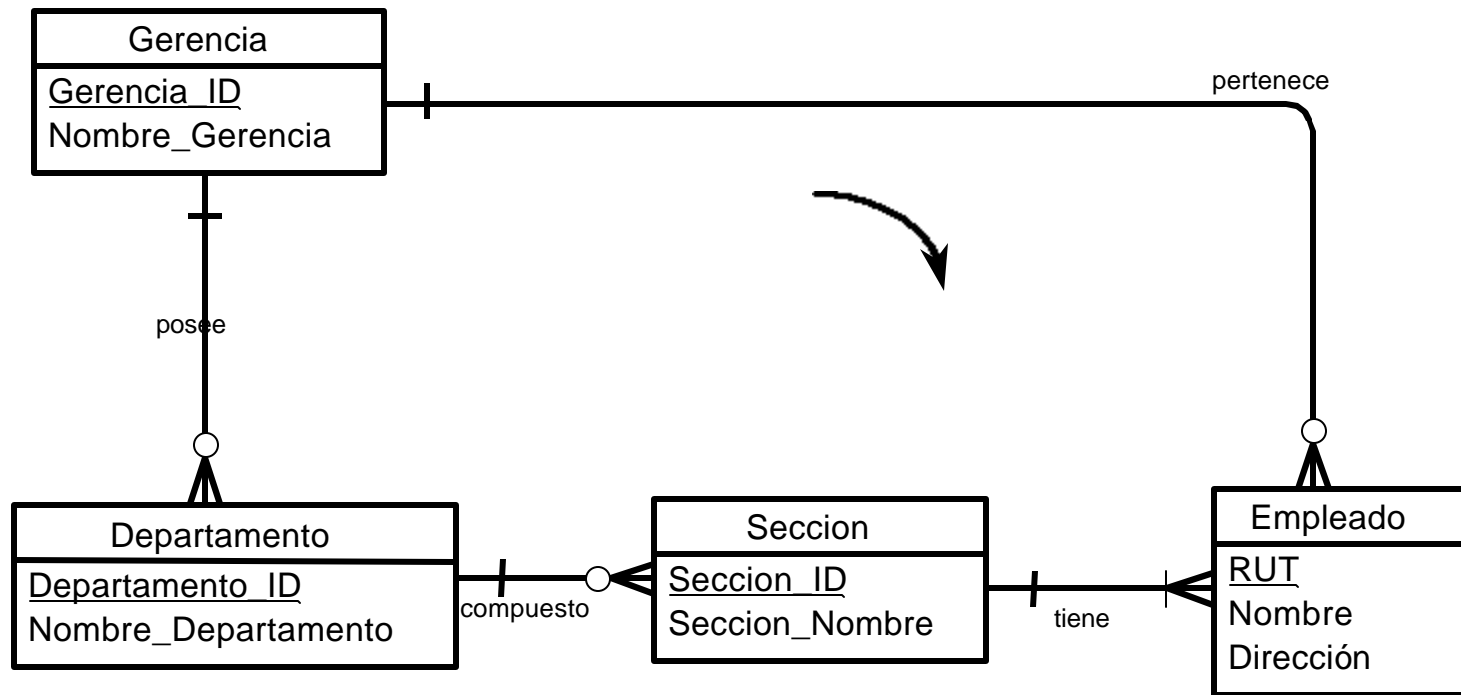
modelo sin redundancia de datos

- Dos entidades son redundantes (o repetidas) si:
 - ambas tienen los mismos atributos o
 - una de ellas tiene todos sus atributos repetidos en la otra
- ¿Qué problema tiene el último modelo?

Redundancia de relaciones 1 de 3

- Solamente en una **trayectoria cerrada**, en que todas las relaciones son de la misma naturaleza, y sólo hay una relación de $M:N$ (normalizada o no), es posible que exista una relación redundante.
- Una relación directa entre dos entidades es redundante si la información que ella entrega es la misma que entrega la relación indirecta entre esas entidades
- Si una relación directa es redundante se **debe** eliminar (en el Modelo de Datos Conceptual), pues la relación indirecta dice lo mismo pero es más completa

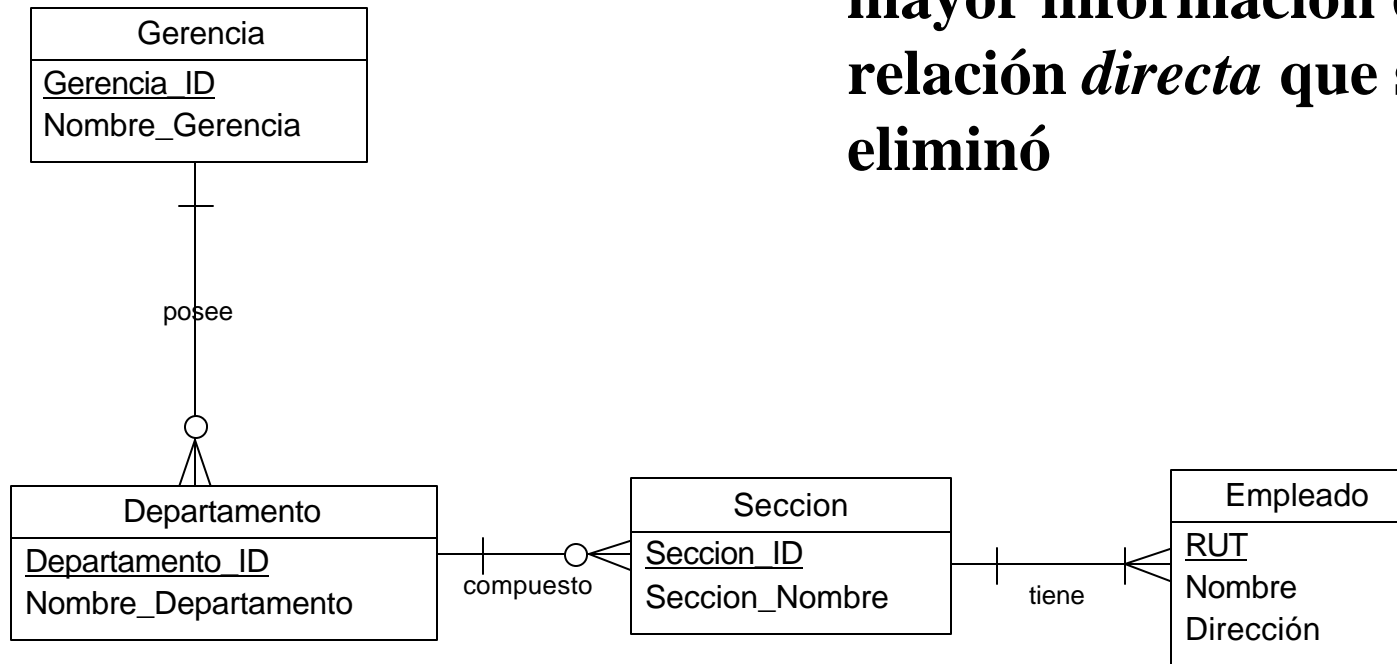
Redundancia de relaciones 2 de 3



- Todas las relaciones son de pertenencia
- Hay una sola relación de M:N normalizada (EMPLEADO es NUB natural)

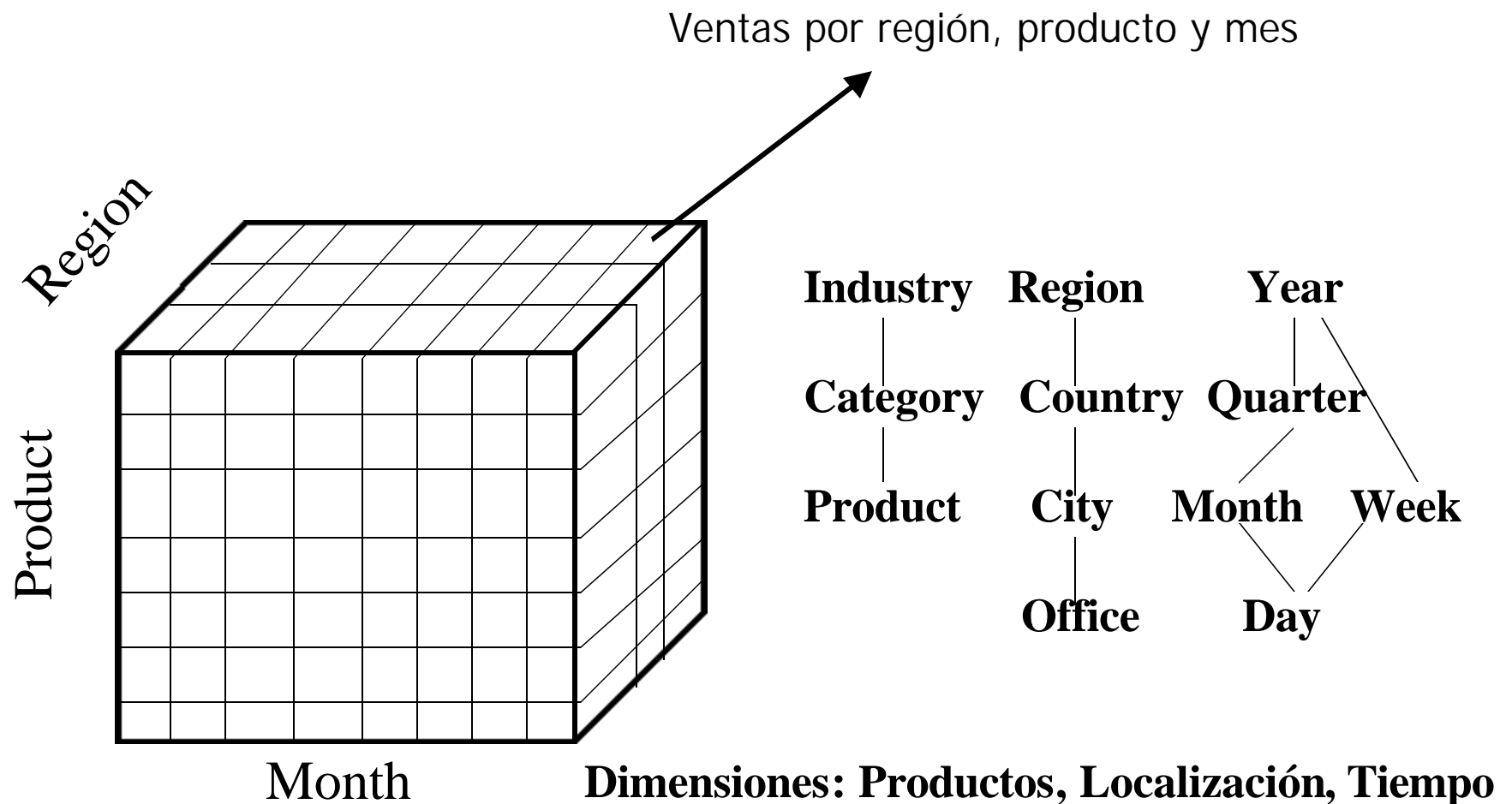
Redundancia de relaciones 3 de 3

La relación *indirecta* tiene mayor información que la relación *directa* que se eliminó



NUB natural

Modelo Multidimensional





Hechos y Dimensiones

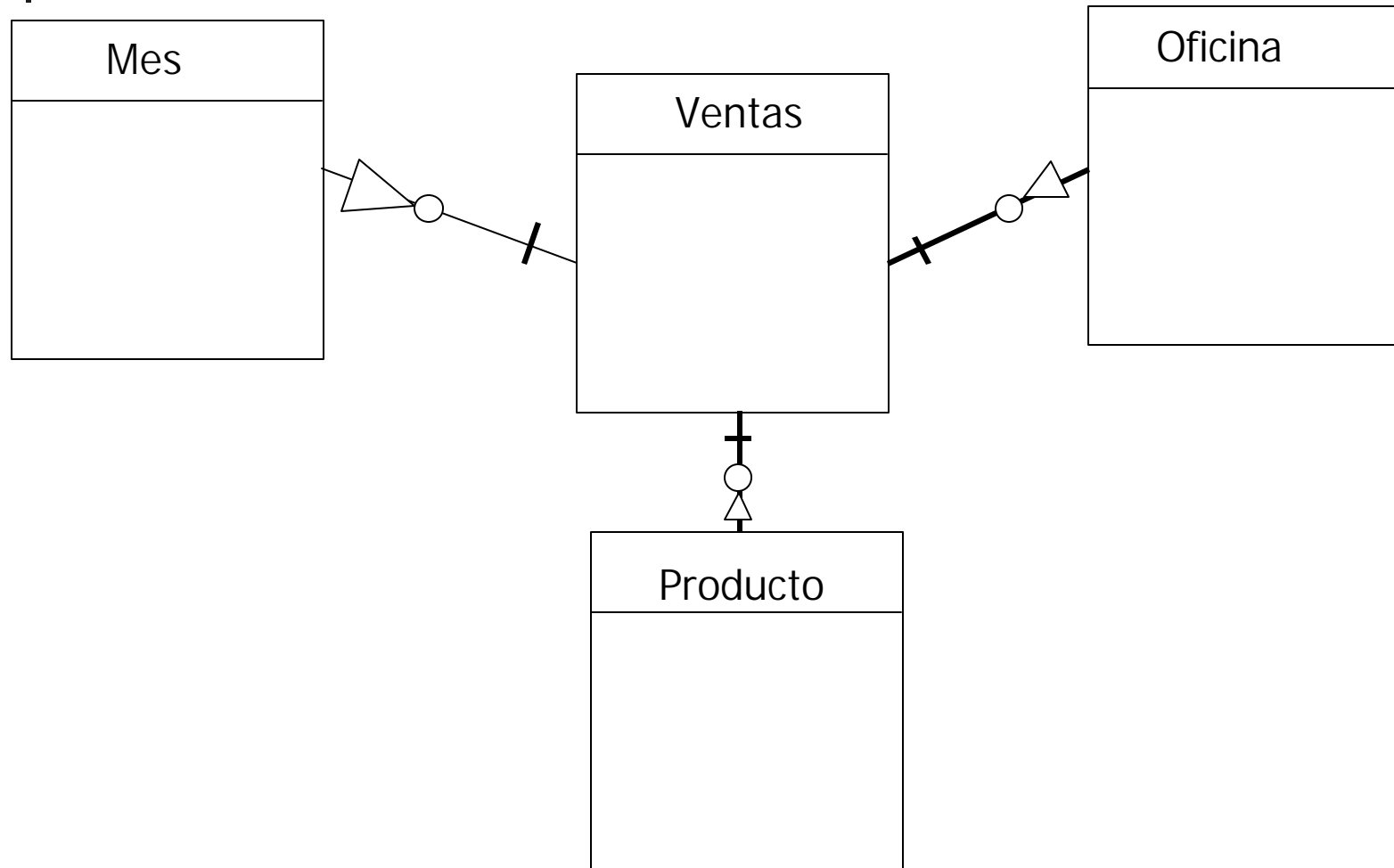
- Los valores almacenados en el cubo son denominados hechos o medidas y corresponden, normalmente a valores
- Los ejes del cubo son denominadas dimensiones, y son normalmente variables descriptivas con alguna escala



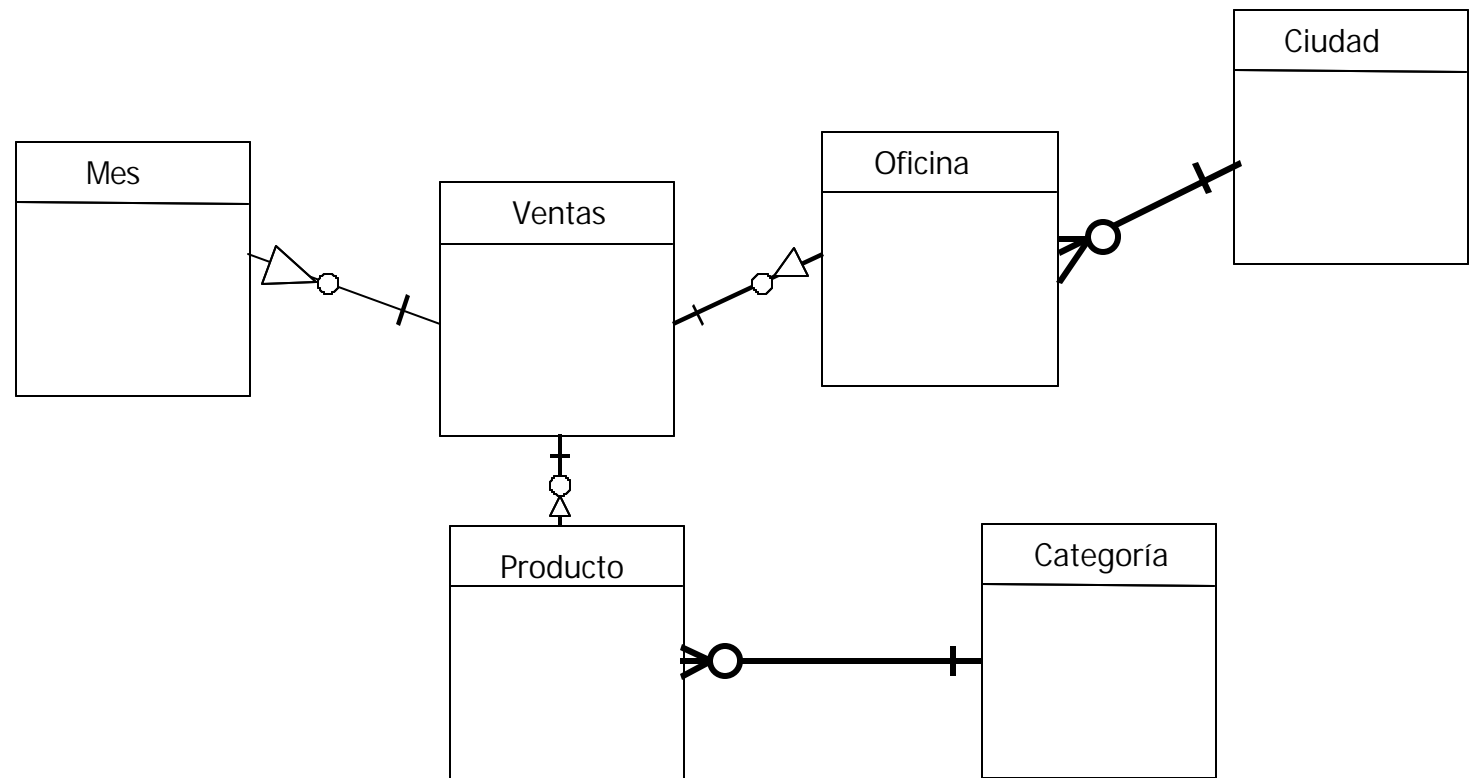
Modelo Conceptual Multidimensional

- Esquema de Estrella (Star schema): Un hecho está en medio del conjunto de dimensiones
- Esquema de Copo de Nieve (Snowflake schema): Un refinamiento del anterior, donde alguna jerarquía dimensional es normalizada en un conjunto de pequeñas dimensiones.
- Constelación de Hechos: Tablas de múltiples hechos, vistas como una colección de estrellas.

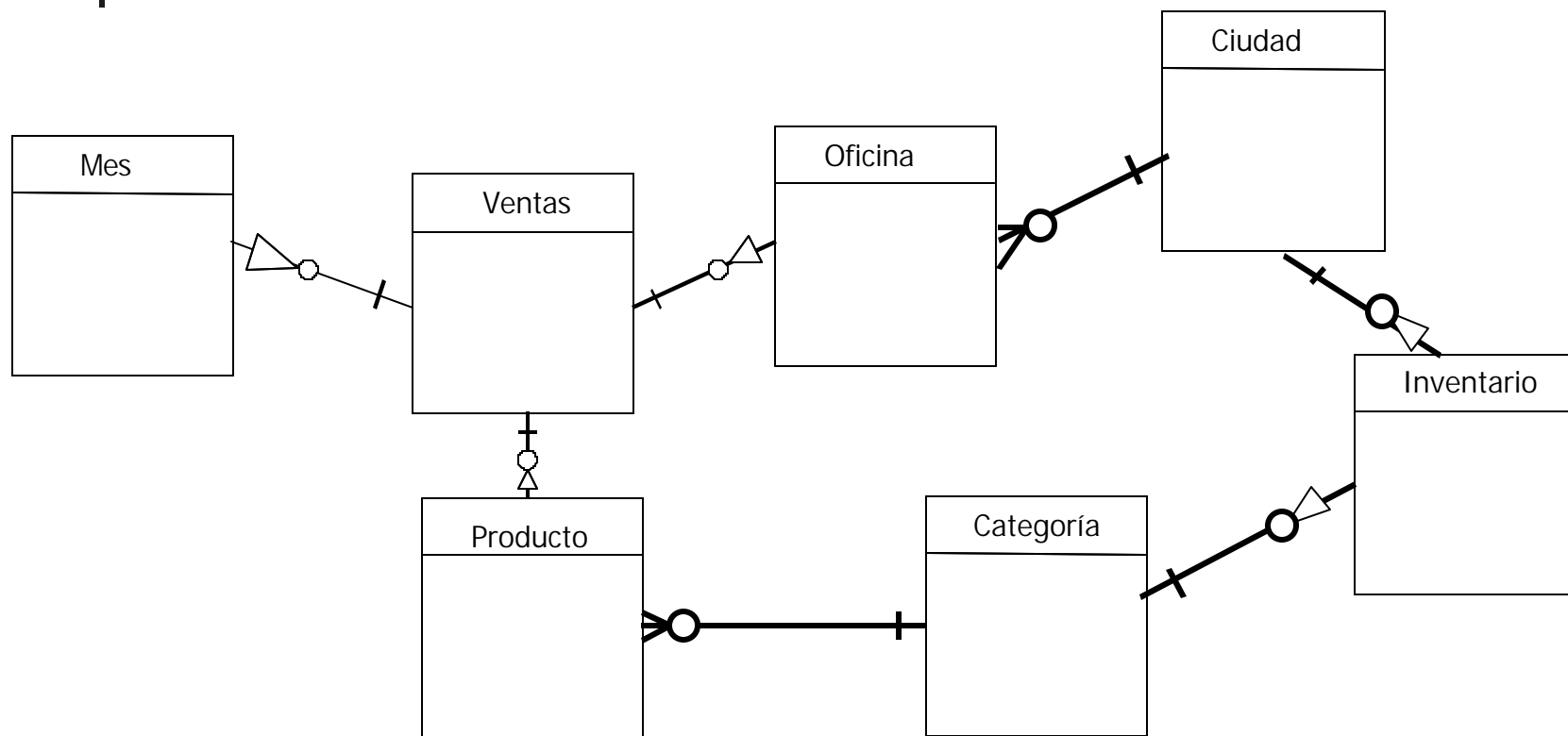
Ejemplo de Esquema en Estrella



Esquema en Snowflake



Esquema en Constelación





Conclusiones

- El modelo conceptual canónico busca proveer un modelo de datos con criterios de calidad dependientes de la redundancia y acceso a los datos.
- La redundancia se puede dar a nivel de atributos y relaciones.
- El objetivo es poder llegar a la Tercera Forma Normal (a lo menos en este curso).
- El modelo multidimensional es un caso especial del modelo general en el cual se distinguen medidas y dimensiones