Bases de Datos

Tema 7 (parte 2)

Teoría de la Normalización

Francisco Ruiz

may-2001

documentación preparada con ayuda de Esperanza Marcos (Universidad Rey Juan Carlos) y Mario Piattini (Universidad de Castilla-La Mancha)

Índice (2)

7. Necesidad de un método formal de optimización del diseño relacional.
8. Objetivos de la teoría de la normalización.
    1. Conservación de la información.
    2. Conservación de las dependencias.
9. Formas normales básicas.
    1. Primera forma normal (1FN).
    2. Segunda forma normal (2FN).
    3. Tercera forma normal (3FN).
    4. Forma normal de Boyce-Codd (FNBC).
10. Enfoques de diseño relacional.
    1. Análisis.
        1. Descomposición en proyecciones independientes.
        2. Pérdidas al pasar a FNBC.
        3. Procedimiento de descomposición.
    2. Síntesis.

Índice (2bis)

11. Formas normales avanzadas.
    1. Dependencias multivaluadas (DM).
    2. Cuarta forma normal (4FN).
    3. Dependencias de combinación (DC).
    4. Quinta forma normal (5FN).
12. Organización de relaciones.
    1. Estructuración y reestructuración de relaciones.

- Los esquemas de relación obtenidos durante la etapa de diseño lógico, según la metodología ya explicada, pueden presentar algunos **problemas derivados de fallos** en:
  - la percepción del UD,
  - el diseño del esquema E/R, o
  - el paso al modelo relacional.
- Entre estos problemas cabe destacar los siguientes:
  - Incapacidad para almacenar ciertos hechos.
  - Redundancias y, por tanto, posibilidad de inconsistencias.
  - Ambigüedades.
  - Pérdida de información (aparición de *tuplas espurias*).
  - Pérdida de DF, es decir, de ciertas restricciones de integridad que dan lugar a interdependencias entre los datos.
  - Existencia de valores nulos (inaplicables).
  - Aparición en la BD de estados que no son válidos en el mundo real (anomalías de inserción, borrado y modificación).

- Ejemplo de diseño inadecuado: problemas de redundancia; y anomalías de modificación, inserción y borrado.

ESTUDIANTE_SOLICITA_BECA

Cód_Estud	*Nombre_E*	*Apellido*	*DNI*	*Dirección*	*Cód_Beca*	*Nombre*	*Requisito*	*Fecha*
012323	Roberto	Hens	456367	Antonio López 43	A22321	METRICA	Ing. Téc.	10/10/98
763476	Luis	García	345347	Av. Ciudades 29	B56784	ERASMU	Ing. Téc.	12/11/98
763476	Luis	García	345347	Av. Ciudades 29	A22321	METRICA	Ing. Téc .	14/10/98
763476	Luis	García	345347	Av. Ciudades 29	G65434	HIMMPA	Ingenie.	15/09/98
012323	Roberto	Hens	456367	Antonio López 43	G65434	HIMMPA	Ingenie.	17/09/98
987765	Gregorio	Celada	885764	Pl. Paises 67	G65434	HIMMPA	Ingenie.	21/09/98
012323	Roberto	Hens	456367	Antonio López 43	B56784	ERASMU	Ing. Téc.	11/11/98
987765	Gregorio	Celada	885764	Pl. Paises 67	B56784	ERASMU	Ing. Téc.	10/10/98
012323	Roberto	Hens	456367	Antonio López 43	A22321	METRICA	Ing. Téc.	12/11/98
232457	Mercedes	García	809234	Rio Miño 2	A22321	METRICA	Ing. Téc.	17/09/98

- Formalmente plantearemos la teoría de la normalización en los siguientes términos:

  *Dado **un esquema de relación** R(A, D) (esquema origen), siendo A el conjunto de atributos y D el conjunto de dependencias existentes entre ellos, se trata de **transformar**, por medio de sucesivas proyecciones, este esquema de partida en un **conjunto de n esquemas de relación** (esquemas resultantes):*

$$\left\{R_i(A_i, D_i)\right\}_{i=1}^{n}$$

  tales que cumplan unas determinadas condiciones.

- Se trata, por tanto, de buscar un conjunto de esquemas R_i que sean **equivalentes** a R (para lo cual será preciso definir el concepto de equivalencia de esquemas) y que sean también **mejores** (en el sentido que enunciaremos) que el esquema origen.
- Son tres las propiedades que han de cumplir los esquemas de relación R_i para ser **equivalentes** a R y **mejores** que R:
  - a) *Conservación de la información*.
  - b) *Conservación de las dependencias*.
  - c) *Mínima redundancia* de los datos (normalización de las relaciones).
- Si se cumplen a) y b), es decir, la transformación de R en {Ri} se hace sin pérdida de información ni de dependencias, se dice que {Ri} es **equivalente** a R.
- y si las relaciones resultantes $\{R_i\}$ están en formas normales más avanzadas que el esquema origen R, se dice que $\{R_i\}$ es **mejor** que R.

- Además del grado de normalización, existen otros criterios para calificar la **bondad** de un esquema relacional:
  - eficiencia frente a las consultas, y
  - captar mejor la semántica del mundo real.

- Algunos autores proponen otros objetivos adicionales:
  a) Minimización de dependencias, que incluye no sólo minimizar el número de dependencias, sino también el número de atributos contenidos en ellas.
  b) Minimización de los esquemas resultantes, que, al igual que el caso anterior, incluye no sólo minimizar el número de relaciones, sino también el número de sus atributos.

- No siempre es posible conseguir que se cumplan a la vez todos estos objetivos.

- La información contenida en la relación origen debe ser la misma que la contenida en el conjunto $\{R_i\}$ de esquemas resultantes. También se llama equivalencia de datos. Para que se cumpla esta propiedad es necesario satisfacer dos condiciones:

1) Conservación de los atributos.

- El conjunto de atributos de $\{R_i\}$ ha de ser igual al conjunto de atributos del esquema origen R:

$$\bigcup_{i=1}^{n} A_i = A$$

2) Conservación del contenido (de las tuplas).

- Para toda extensión r de R, la combinación (*join*) de las relaciones resultantes r_i ha de producir la relación origen r:

$$*_{i=1}^{n} r_i = r$$

ESTUDIANTE_NUEVA

Cód_Est	*Nombre_E*	*Apellidos*	*DNI*	*Dirección*	*Fecha*
012323	Roberto	Hens	456367	Antonio López 43	10/10/98
763476	Luis	García	345347	Av. Ciudades 29	12/11/98
763476	Luis	García	345347	Av. Ciudades 29	14/10/98
763476	Luis	García	345347	Av. Ciudades 29	15/09/98
012323	Roberto	Hens	456367	Antonio López 43	17/09/98
987765	Gregorio	Celada	885764	Pl. Países 67	21/09/98
012323	Roberto	Hens	456367	Antonio López 43	11/11/98
987765	Gregorio	Celada	885764	Pl. Países 67	10/10/98
012323	Roberto	Hens	456367	Antonio López 43	12/11/98
232457	Mercedes	García	809234	Río Miño 2	17/09/98

BECA_NUEVA

Cód_Beca	*Nombre*	*Requisito*	*Fecha*
A22321	METRICA	Ing. Téc.	10/10/98
B56784	ERASMUS	Ing. Téc.	12/11/98
A22321	METRICA	Ing. Téc.	14/10/98
G65434	HIMMPA	Ingeniería	15/09/98
G65434	HIMMPA	Ingeniería	17/09/98
G65434	HIMMPA	Ingeniería	21/09/98
B56784	ERASMUS	Ing. Téc.	11/11/98
B56784	ERASMUS	Ing. Téc.	10/10/98
A22321	METRICA	Ing. Téc.	12/11/98
A22321	METRICA	Ing. Téc.	17/09/98

ESTUDIANTE_NUEVA * BECA_NUEVA

Cód_Est	*Nombre_E*	*Apellidos*	*DNI*	*Dirección*	*Cód_Beca*	*Nombre*	*Requisito*	*Fecha*
012323	Roberto	Hens	456367	Antonio López 43	A22321	METRICA	Ing. Téc.	10/10/98
987765	**Gregorio**	**Celada**	**885764**	**Pl. Países 67**	**A22321**	**METRICA**	**Ing. Téc.**	**10/10/98**
763476	Luis	García	345347	Av. Ciudades 29	B56784	ERASMUS	Ing. Téc.	12/11/98
012323	**Roberto**	**Hens**	**456367**	**Antonio López 43**	**B56784**	**ERASMUS**	**Ing. Téc.**	**12/11/98**
763476	Luis	García	345347	Av. Ciudades 29	A22321	METRICA	Ing. Téc.	14/10/98
763476	Luis	García	345347	Av. Ciudades 29	G65434	HIMMPA	Ingeniería	15/09/98
012323	Roberto	Hens	456367	Antonio López 43	G65434	HIMMPA	Ingeniería	17/09/98
232457	**Mercedes**	**García**	**809234**	**Río Miño 2**	**G65434**	**HIMMPA**	**Ingeniería**	**17/10/98**
987765	Gregorio	Celada	885764	Pl. Países 67	G65434	HIMMPA	Ingeniería	21/09/98
012323	Roberto	Hens	456367	Antonio López 43	B56784	ERASMUS	Ing. Téc.	11/11/97
987765	Gregorio	Celada	885764	Pl. Países 67	B56784	ERASMUS	Ing. Téc.	10/10/97
012323	**Roberto**	**Hens**	**456367**	**Antonio López 43**	**B56784**	**ERASMUS**	**Ing. Téc.**	**10/10/97**
012323	Roberto	Hens	456367	Antonio López 43	A22321	METRICA	Ing. Téc.	12/11/98
763476	**Luis**	**García**	**345347**	**Av. Ciudades 29**	**A22321**	**METRICA**	**Ing. Téc.**	**12/11/98**
232457	Mercedes	García	809234	Río Miño 2	A22321	METRICA	Ing. Téc.	17/09/98
012323	**Roberto**	**Hens**	**456367**	**Antonio López 43**	**A22321**	**METRICA**	**Ing. Téc.**	**17/09/98**

TUPLAS ESPURIAS

*Aparición de **tuplas espurias** al no cumplirse la condición 2ª*

- El conjunto de DF de partida debe ser equivalente al conjunto de DF de los esquemas resultantes (**equivalencia de dependencias**).
- Teniendo en cuenta el concepto de equivalencia de dependencias definido anteriormente, podemos decir que se han conservado las dependencias si se cumple:

$$\left(\bigcup_{i=1}^{n} DF_i \right)^+ = DF^+$$

- Para verificar esta condición se puede utilizar el algoritmo, ya explicado, de equivalencia de dos conjuntos de dependencias.

- La tercera propiedad que debe cumplir el conjunto $\{R_i\}$ de esquemas resultantes en un proceso de descomposición, es que estas relaciones alcancen un nivel de normalización superior al del esquema origen R, a fin de eliminar en lo posible las redundancias y, por tanto, las anomalías de actualización.
- Se dice que un esquema de relación está en una determinada forma normal, si satisface un cierto conjunto específico de restricciones.
- Cuanto más alta sea la forma normal en la que se encuentran los esquemas de relación, menores serán los problemas que aparecen en el mantenimiento de la BD.

- Formas Normales Básicas:
  - Inicialmente, Codd propuso en 1970 tres formas normales basadas en las DF: **primera** (1FN), **segunda** (2FN) y **tercera** forma normal (3FN), CODD (1970).
  - Debido a que en 3FN aún persisten algunos problemas en las relaciones, en 1974 Codd y Boyce introdujeron una definición más restrictiva de la tercera formal normal, que se denominó Forma Normal de **Boyce-Codd** (FNBC).
- Formas Normales Avanzadas:
  - En 1977 y 1979 Fagin introduce la **cuarta** (4FN) y **quinta** (5FN) formas normales respectivamente. Ambas están basadas en otro tipo de dependencias distintas de las funcionales: las dependencias multivaluadas (4FN) y las dependencias de combinación (5FN).
- Cuando un esquema de relación está en una Forma Normal, implícitamente también está en las formas normales inferiores.

- La primera forma normal (1FN) es una restricción inherente al modelo relacional, por lo que su cumplimiento es obligatorio:
  - para que una tabla pueda ser considerada una relación no debe admitir grupos repetitivos, esto es, debe estar en 1FN.

ESTUDIANTE (*Código, Nombre, Cursos*)

CODIGO	NOMBRE	CURSOS
178263782	PEDRO PERALES	ERASMUS HIMMPA
031928733	ALBERTO GONZALEZ	METRICA ERASMUS
763459374	FRANCISCO VIDAL	HIMMPA METRICA

Es una tabla pero no una relación

NO ESTA EN 1FN

HAY GRUPOS REPETITIVOS

CODIGO	NOMBRE	CURSO
178263782	PEDRO PERALES	ERASMUS
178263782	PEDRO PERALES	HIMMPA
031928733	ALBERTO GONZALEZ	METRICA
031928733	ALBERTO GONZALEZ	ERASMUS
763459374	FRANCISCO VIDAL	HIMMPA
763459374	FRANCISCO VIDAL	METRICA

ESTA EN 1FN

Definición:

Una relación está en 1FN cuando cada atributo sólo toma un valor del dominio simple subyacente.

- La 2FN está basada en el concepto de dependencia plena y en las interrelaciones existentes entre los atributos principales (los que se encuentran en alguna de las claves) y no principales (los que no se encuentran en ninguna clave).

Definición:

- Se dice que una relación está en 2FN si:
  - Está en 1FN.
  - Cada atributo no principal tiene DF plena respecto de cada una de las claves.

- Siempre es posible transformar un esquema de relación que no esté en 2FN, en esquemas de relación en 2FN, sin que se produzca pérdida de información o de dependencias.

- La 2FN no se cumple cuando:
  - Algún atributo no principal, depende funcionalmente de algún subconjunto de una clave.

- Siempre están en 2FN las relaciones:
  - Binarias.
  - Con todas las claves simples, es decir, que contienen un sólo atributo.
  - Que todos los atributos son **principales**, es decir, forman parte de alguna clave.

- La 3FN está basada en el concepto de dependencia transitiva.

Definición:

- Un esquema de relación R, está en 3FN si, y sólo si:
  - Está en 2FN.
  - No existe ningún atributo no principal que dependa transitivamente de alguna clave de R.

- Siempre es posible transformar un esquema de relación que no esté en 3FN, en esquemas de relación en 3FN, sin que se produzca pérdida de información o de dependencias.

- La 3FN no se cumple cuando:
  - Existen atributos no principales que dependen funcionalmente de otros atributos no principales.

- Siempre están en 3FN las relaciones:
  - Binarias.
  - En las que todos los atributos son principales.
  - Que tienen un único atributo no principal.

- La 3FN mantiene ciertos problemas en relaciones que presentan varias claves candidatas compuestas que se solapan.
- Por ello, en 1974, **Boyce y Codd** definieron la llamada forma normal que lleva su nombre (**FNBC**). Se trata de una redefinición más estricta de la 3FN.

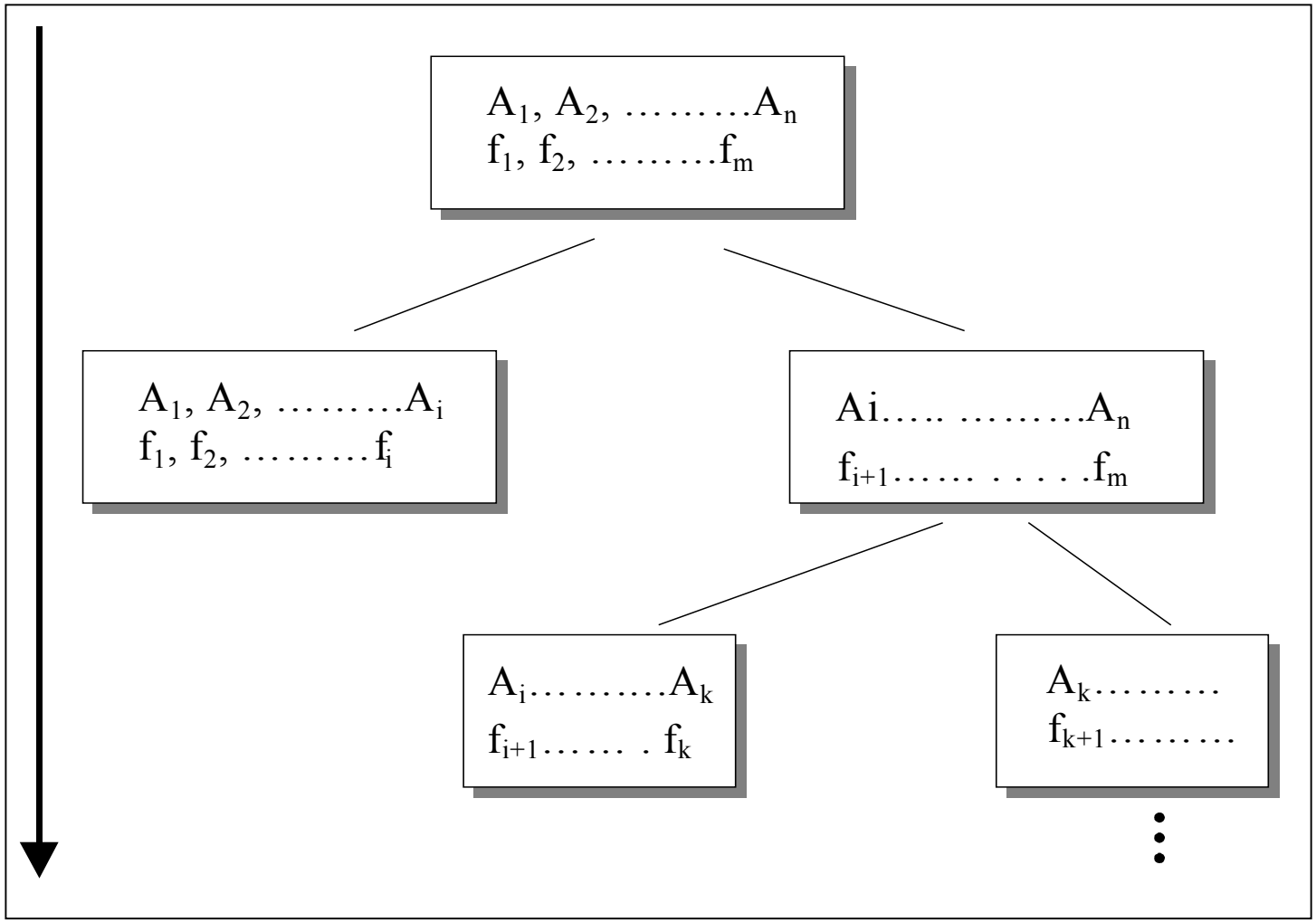
Definición:

  – Una relación se encuentra en FNBC si, y sólo si, todo determinante es una clave candidata.

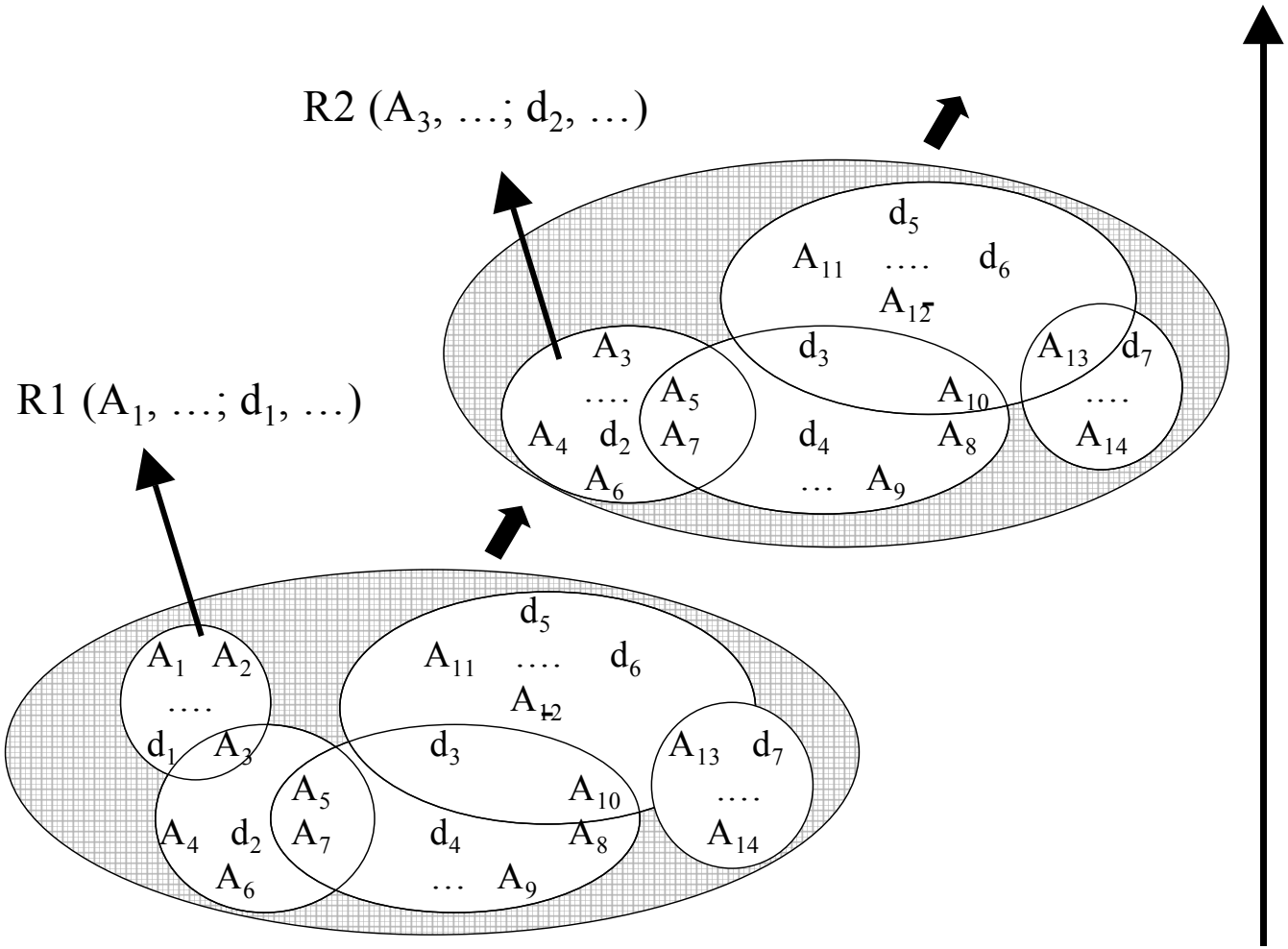
- No siempre es posible transformar un esquema de relación que no está en FNBC en esquemas de relación en FNBC sin que se produzca pérdida de dependencias funcionales. Sí se puede asegurar que la transformación se puede producir siempre sin pérdida de información.

- La FNBC no se cumple cuando:
  - Existe algún determinante que no es clave candidata.
- Siempre están en FNBC las relaciones:
  - Binarias.
  - Que están en 3FN y sus claves no se solapan.
- La existencia de claves candidatas solapadas no siempre supone que la relación no esté en FNBC.
- Puede ocurrir que ciertas relaciones que se encuentran en FNBC presenten todavía redundancias y anomalías, pero éstas ya no se deben a las DF, y por tanto, para evitarlas hay que recurrir a las formas normales avanzadas basadas en otro tipo de dependencias.

- Existen dos enfoques diferentes a la hora de aplicar la teoría de la normalización:
  - **Análisis** o **descomposición**:
    - El esquema de una relación se va **descomponiendo** por sucesivas proyecciones, que han de cumplir los principios de conservación de la información y de las dependencias; los esquemas resultantes serán cada vez de menor grado y estarán en formas normales cada vez más avanzadas, es decir, se irán reduciendo las anomalías.
  - **Síntesis**:
    - Se realiza el camino inverso al anterior, obteniendo (**sintetizando**) relaciones a partir de un conjunto de atributos y de dependencias funcionales.
- La síntesis busca agrupar atributos a fin de tener en una relación toda la información correspondiente a un objeto (entidad o interrelación) del mundo real, mientras que en el análisis se pretende separar la información referente a objetos distintos.



Análisis: árbol de descomposición.



Representación del método de síntesis

- La conservación de la información y de las DF depende del método utilizado (análisis o síntesis), de los tipos de dependencias de partida y de la forma normal que se quiere satisfacer:

	SÍNTESIS		**ANÁLISIS**		
Dependencias de partida	DF		DF		DF, DM, DC
Forma normal alcanzada	3FN	FNBC	3FN	FNBC	5FN
Conservación de la información	Sí	no se asegura	Sí	Sí	Sí
Conservación de las DF	Sí	Sí	Sí	no se asegura	no se asegura

- Formalmente, la descomposición de un esquema de relación R(A,DF) consiste en la sustitución de dicho esquema por un conjunto de proyecciones del mismo:

$$R_1, R_2, ..., R_p \quad \text{donde} \quad R_i(A_i,DF_i)$$

  tales que el conjunto resultante sea **equivalente** y **mejor** que el esquema origen.

- Para que la descomposición se lleve a cabo sin pérdida de información se ha de cumplir que:

$$R = R_1 * R_2 * ... * R_p$$

  para toda extensión de R (descomposición SPI).

- Asimismo, la descomposición debe hacerse sin pérdida de dependencias funcionales:

$$(\cup DF_i)^+ = DF^+$$

- RISSANEN propuso en 1976 un método para saber si una determinada descomposición es correcta, es decir, si conserva la información y las DF; para ello se introduce el concepto de *proyecciones independientes*.
- Sea R una relación y R_1 y R_2 dos de sus proyecciones, se dice que dichas proyecciones son independientes si, y sólo si:
  1) Sus atributos comunes son la clave candidata de, al menos, una relación.
  2) Cada DF en R puede deducirse de las de R_1 y R_2.

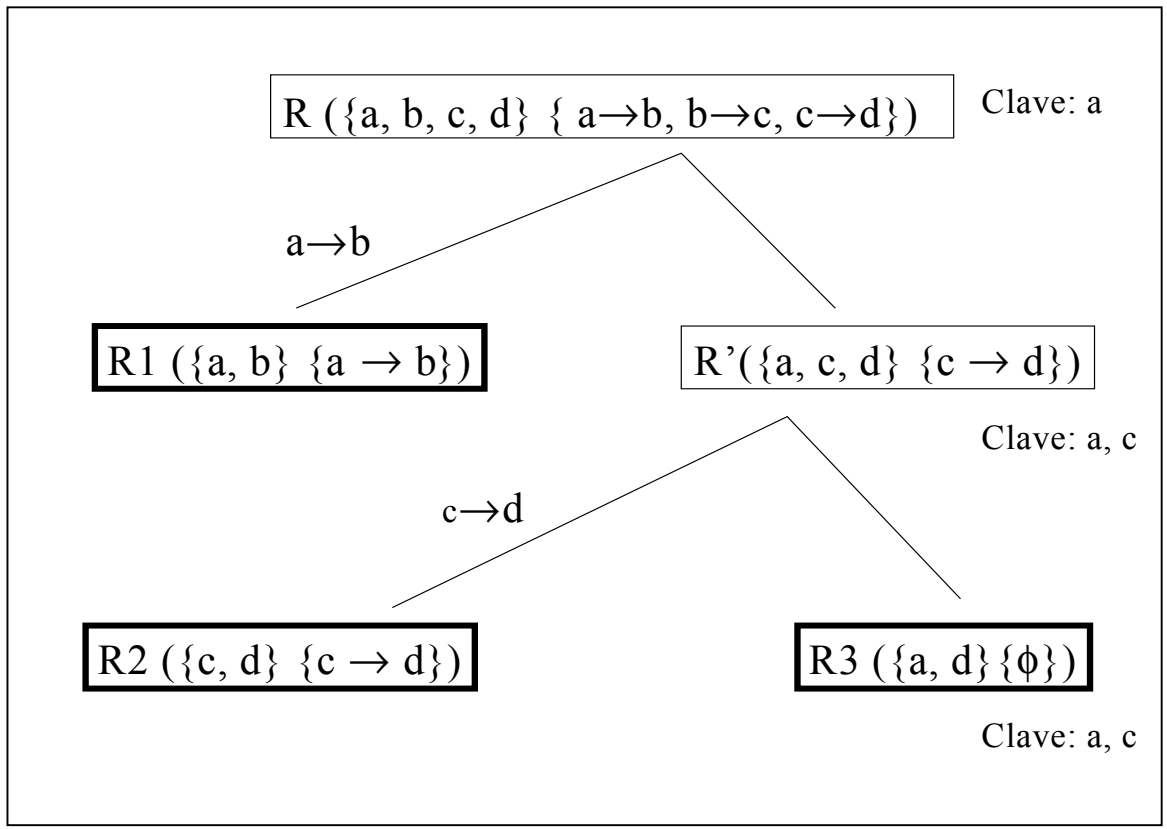
- Para ser correcta, es preciso que la descomposición de relaciones se efectúe en proyecciones independientes.
- Para las tres primeras formas normales, esto siempre es posible.
- Pero no siempre lo es si se desea llegar hasta la forma normal del Boyce y Codd, ya que en ese caso, puede ser preciso perder dependencias o información.

- Dado el esquema de relación R(A, DF), donde A es un conjunto de atributos y DF un conjunto de dependencias funcionales. Los pasos a seguir en un proceso de descomposición para aumentar el nivel normalización, son:

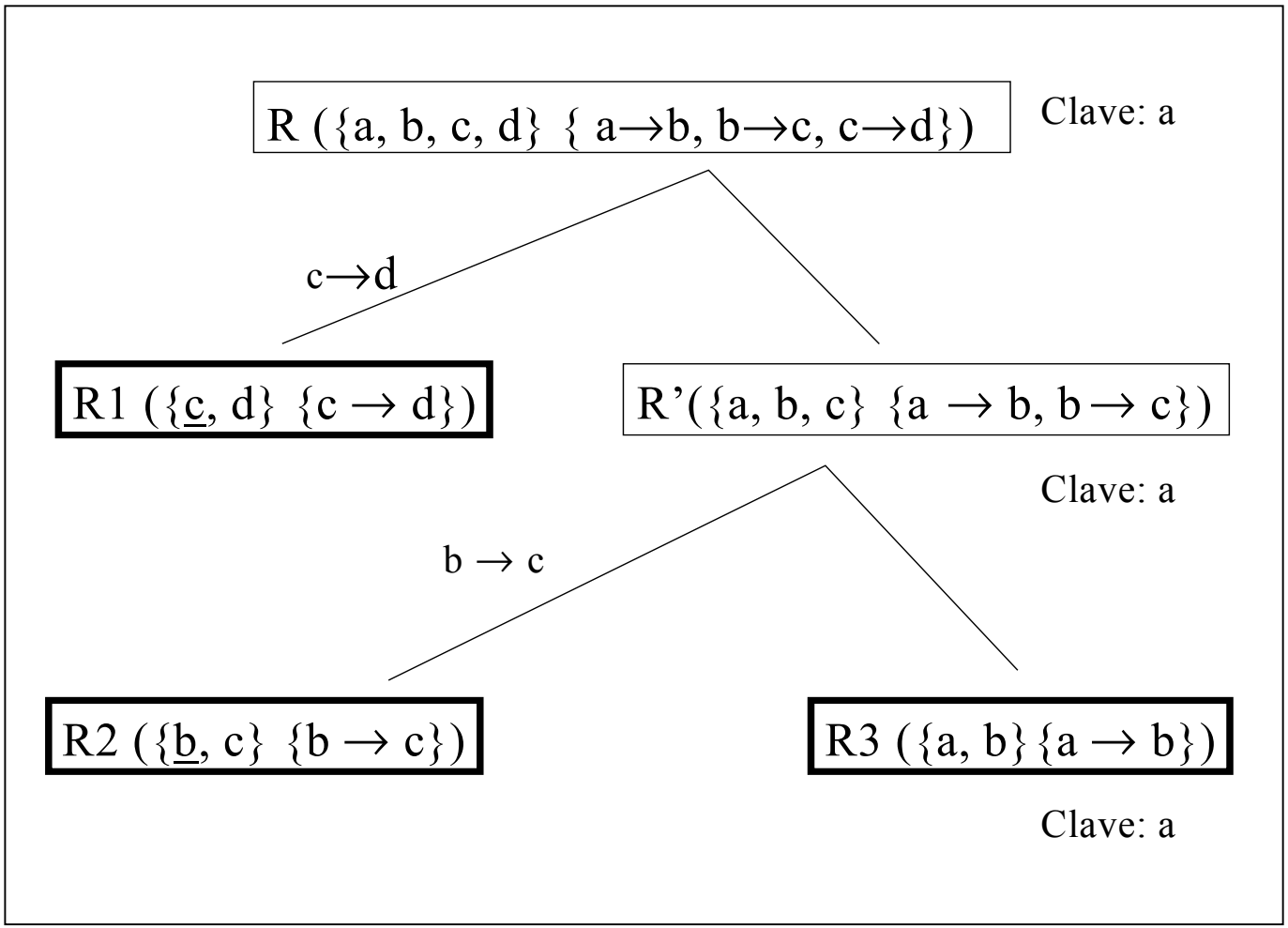
  1. Hallar un recubrimiento minimal DF^m.
  2. Determinar la(s) clave(s), así como los atributos principales y no principales.
  3. Identificar la FN en que se encuentra la relación. Si dicha FN es la buscada acabar, en caso contrario continuar.
  4. Agrupar las DF que tengan el mismo implicante.
  5. Obtener proyecciones independientes sobre cada uno de los grupos de DF, de forma que los atributos que aparecen en el correspondiente grupo constituyen una nueva relación, y el implicante del grupo desaparezca de la relación origen.
  6. Repetir el paso 5 hasta que no pueda continuarse porque todas las dependencias estén implicadas por una clave (en este caso, para llegar hasta FNBC hay que perder dependencias; y es decisión del diseñador parar el proceso en la 3FN o avanzar hasta FNBC con el inconveniente señalado).

- Es relevante el orden en el que se van tomando los grupos de dependencias funcionales: los atributos que desaparecen de la relación origen deben ser aquellos que no entran a formar parte de ninguna otra dependencia (en caso contrario se perderían dependencias).

Incorrecto:
se pierde
b→c

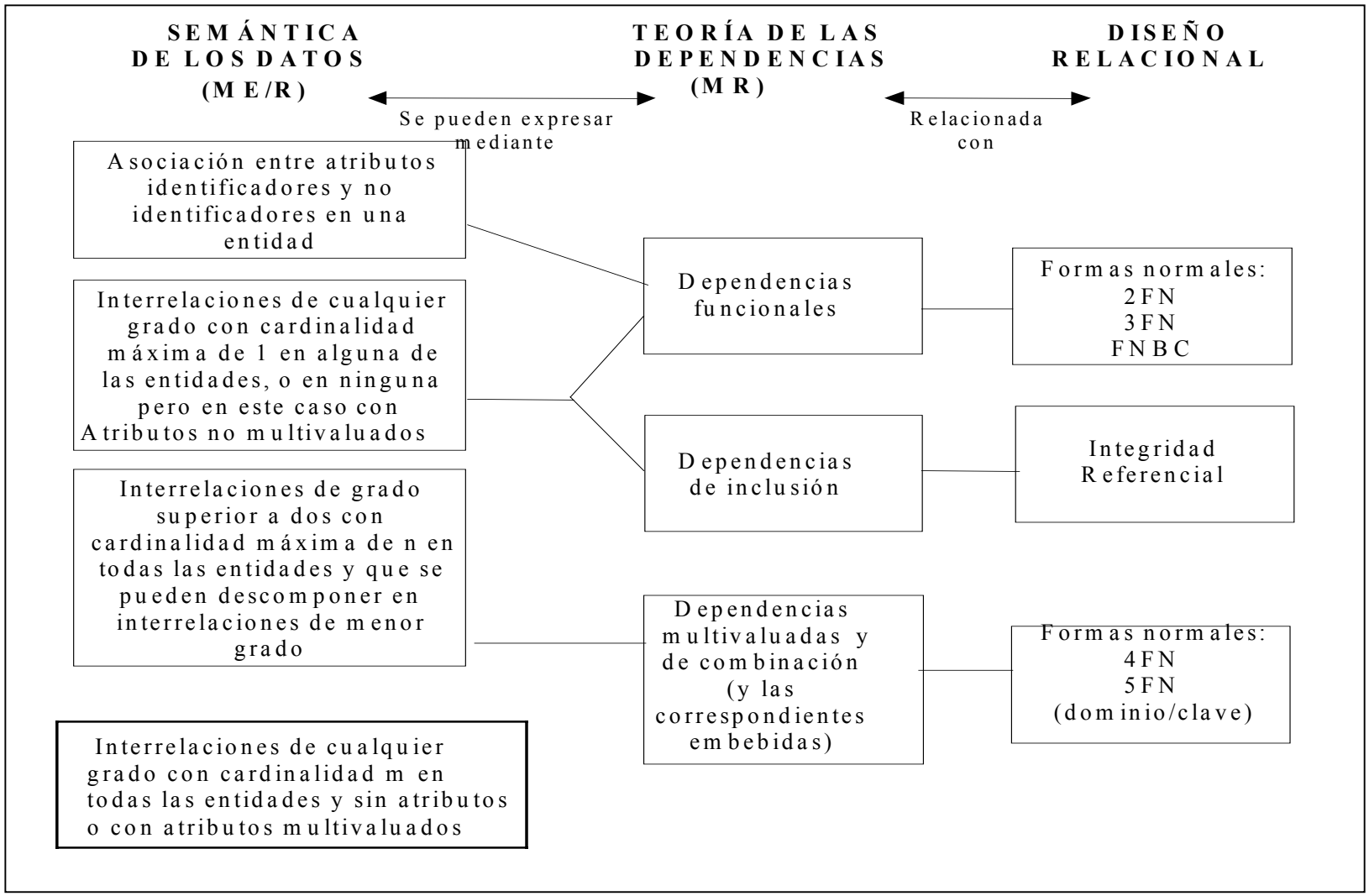


Correcto: se debe comenzar separando c→d



- Dado el esquema de relación R(A, DF), donde A es el conjunto de atributos de R y DF el conjunto de dependencias funcionales existentes entre dichos atributos. Los pasos a seguir en un proceso de **síntesis** son:
  1. Se busca un recubrimiento minimal DF^m del conjunto de dependencias funcionales DF.
  2. Se agrupan las dependencias de DF^m en particiones que tengan el mismo implicante (considerar a los descriptores equivalentes como el mismo implicante).
  3. Se forma un esquema de relación R_i para cada partición, el cual tendrá como atributos todos los que aparezcan en la partición así como las dependencias funcionales implicadas.
  4. Si existen atributos que no son implicantes ni implicados en DF^m, se forma un esquema de relación con ellos y sin dependencias funcionales. En caso contrario, se añade la clave de la relación inicial como otra relación.

- Hasta ahora se ha estudiado un tipo especial de restricciones semánticas sobre los datos, las dependencias funcionales, que permiten diseñar esquemas en 2FN, 3FN y FNBC.
- Pero **aún en FNBC**, a veces siguen presentándose **redundancias** que provocan anomalías de actualización. La razón de ello es que existen otros tipos de dependencias entre los datos, distintas de las funcionales, que también hay que tener en cuenta en el diseño relacional.
- Esto conduce a una generalización de las dependencias funcionales, apareciendo **nuevos tipos de dependencias** como las **multivaluadas** y las de **combinación**. Las dependencias funcionales son un caso especial de las multivaluadas y éstas, a su vez, de las de combinación.



Vinculación existente entre la semántica de los datos (expresada en el ME/R), la teoría de las dependencias y el diseño relacional

- Las dependencias multivaluadas (DM) son una generalización de las DF:
  - Mientras que en una DF el valor de un descriptor (implicante) determina un **sólo** valor de otro descriptor (implicado) ...
  - En las DM a un valor del implicante le corresponde un conjunto de valores del implicado (en lugar de uno solo).
- Se originan cuando en una tabla aparecen atributos multivaluados independientes entre sí, y se normaliza la tabla para que esté en 1FN.
- Dada la relación R(A), se cumple la DM $X \rightarrow\rightarrow Y$ (leer X "multidetermina" a Y) si, para cada valor de X, hay un conjunto de cero o más valores de Y, independientemente de los valores del resto de atributos $\{A - X - Y\}$.
- Las dependencias multivaluadas siempre se producen por parejas:
  - Si en el esquema R(A) existe la dependencia $X \rightarrow\rightarrow Y$, al mismo tiempo habrá de cumplirse $X \rightarrow\rightarrow \{A - X - Y\}$.
  - Se representa como $X \rightarrow\rightarrow Y \mid \{A - X - Y\}$

- Una DM es **trivial** si se da alguno de los casos siguientes:
  - $X = Y$
  - Y es un subconjunto de X ($Y \subset X$)
  - $X \cup Y = A$ (conjunto de atributos de la relación)
- **DM Embebidas**:
  - Son un tipo especial de DM que no existen en un cierto esquema de relación pero que, en cambio, si aparecen en una proyección de dicha relación. También se llaman jerárquicas.
  - Una relación R(A) satisface una DM embebida $X \rightarrow\rightarrow Y \mid Z$, donde $\{X \cup Y \cup Z\} \subset A$, si dicha DM no existe en R y, en cambio, si se cumple para cualquier extensión de un esquema $R'(X \cup Y \cup Z)$, siendo R' una proyección de R sobre los atributos $\{X \cup Y \cup Z\}$ involucrados en la dependencia embebida.

ASIGNATURAS

Nom _Asignatura	*Profesor*	*Texto*
Ficheros y BD	Sr. Sanchez Sra. Hidalgo	Concepción y Diseño de BD Fundamentos de BD
BD avanzadas	Sra. Hidalgo Sr. Martín	Diseño de BD avanzadas

Ejemplo de tabla no normalizada con atributos multivaluados

ASIGNATURAS

Nom _Asignatura	*Profesor*	*Texto*
Ficheros y BD	Sr. Sanchez	Concepción y Diseño de BD
Ficheros y BD	Sr. Sanchez	Fundamentos de BD
Ficheros y BD	Sra. Hidalgo	Concepción y Diseño de BD
Ficheros y BD	Sra. Hidalgo	Fundamentos de BD
BD avanzadas	Sra. Hidalgo	Diseño de BD avanzadas
BD avanzadas	Sr. Martín	Diseño de BD avanzadas

Nom _Asignatura $\rightarrow\!\!\rightarrow$ *Profesor* *Nom _Asignatura* $\rightarrow\!\!\rightarrow$ *Texto*

Dependencias multivaluadas generadas al poner en 1FN la tabla anterior

- Una relación se encuentra en **4FN** si, y sólo si, las únicas DM no triviales son aquellas en las cuales una clave multidetermina un atributo, es decir, toda DM viene determinada por una clave candidata.
- La descomposición sin pérdida de información para el caso de DF tiene su equivalencia cuando se trata de DM:
  - Dado el esquema de relación R(A,D), donde D es un conjunto de dependencias (funcionales y multivaluadas), dos proyecciones R1 y R2 cumplen la propiedad de descomposición sin pérdida de R, si se cumplen las condiciones:
    1. $R1 \cap R2 \rightarrow\rightarrow R1 - R2$
    2. $R1 \cap R2 \rightarrow\rightarrow R2 - R1$
- Se demuestra matemáticamente que una relación R(A,B,C) se puede descomponer sin pérdida en dos proyecciones R1(A,B) y R2(A,C), si, y sólo si, existe en R la DM $A \rightarrow\rightarrow B \mid C$.

- Las DF y DM analizadas hasta ahora permiten la descomposición sin pérdida de una relación en **dos** proyecciones; sin embargo, existen relaciones en las que no se puede llevar a cabo tal descomposición binaria sin pérdida de información:

EDITA

Editorial	*Idioma*	*Tema*
RA-MA	Inglés	BD
RA-MA	Español	CASE
Addison Wesley	Español	BD
RA-MA	Español	BD

- En la tabla EDITA no existen DF o DM, pero se puede observar que existen redundancias y sus correspondientes anomalías, producidas por un nuevo tipo de dependencias, llamadas **dependencias de combinación** (DC).
- Estas anomalías son debidas a que los atributos de la relación EDITA no son independientes, existiendo entre ellos unas ciertas restricciones que podemos expresar de la siguiente forma:
  - Si existen las tres tuplas (E,_,T) (E,I,_) (_,I,T)
    entonces también tiene que existir la tupla (E,I,T)

- Pero EDITA sí puede descomponerse en tres proyecciones sin pérdida de información:

EDITA1

Editorial	*Idioma*
RA-MA	Inglés
RA-MA	Español
Addison-Wesley	Español

EDITA2

Idioma	*Tema*
Inglés	BD
Español	CASE
Español	BD

EDITA1*EDITA2

⇓

TUPLA ESPÚRIA <Addison-Wesley, Español, CASE>

EDITA3

Editorial	*Tema*
RA-MA	BD
RA-MA	CASE
Addison-Wesley	BD

EDITA1*EDITA2*EDITA3=EDITA

- Por tanto, la relación EDITA tiene una nueva restricción, que se ha de cumplir para todas sus extensiones (es independiente del tiempo), a la que se denomina **dependencia de combinación**.
- Las DC son, al igual que las otras dependencias, una restricción sobre una relación y constituyen una generalización de las anteriores, de forma que siempre una dependencia funcional es también multivaluada y de combinación, y una multivaluada es también de combinación. La afirmación inversa no es cierta.
- Una relación R tiene la DC respecto de sus proyecciones R_1, R_2, ..., R_j, si:
  - $R = R_1 * R_2 * ... * R_j$

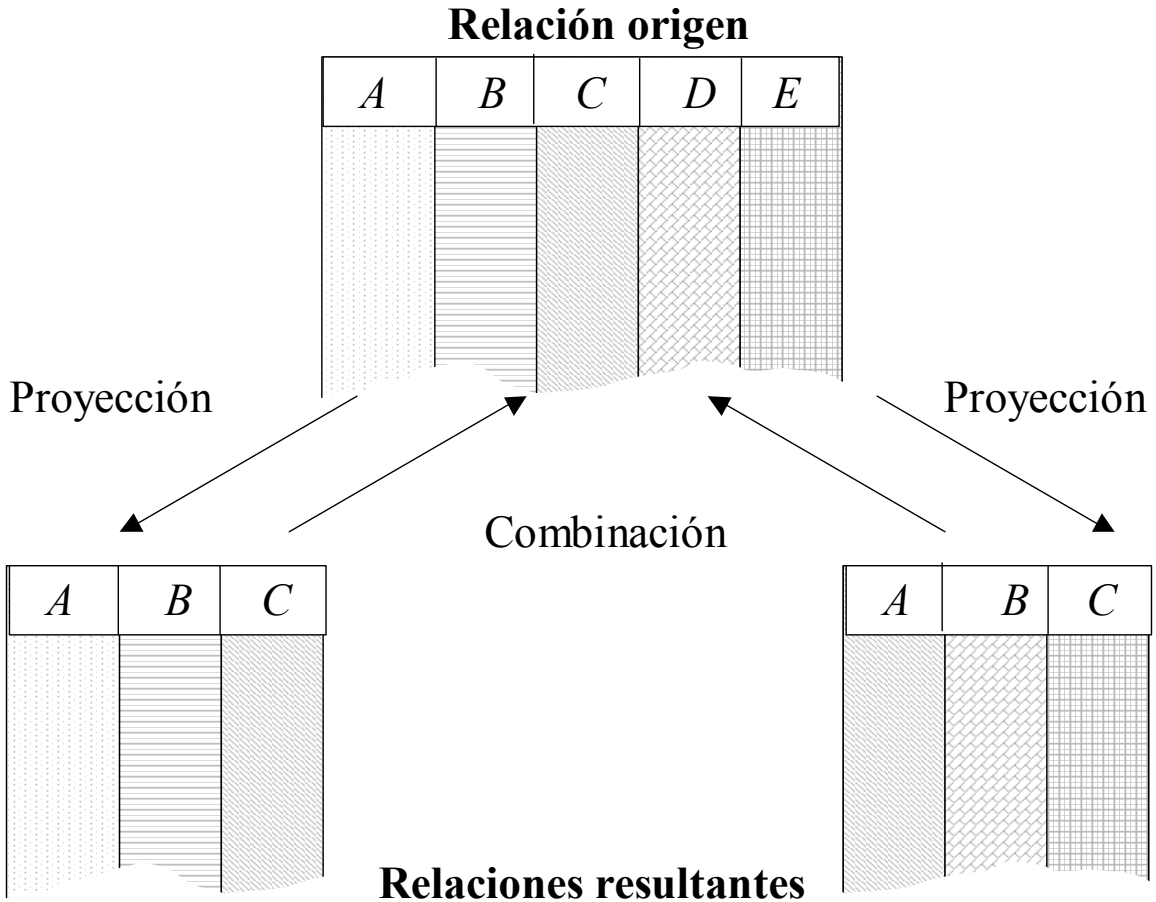
    se denota por: $DJ * (R_1, ..., R_j)$

- Una DM es un caso especial de una DC en la que j = 2.
- Una DC es trivial si cualquiera de los esquemas de relación R_i es igual a R.
- Si R satisface DJ*(R_1, R_2 y R_3), entonces se satisface la restricción:
  - <A, B, C> $\in$ R si y sólo si
    <A, B> $\in$ R_1,
    <B, C> $\in$ R_2, y
    <C, A> $\in$ R_3

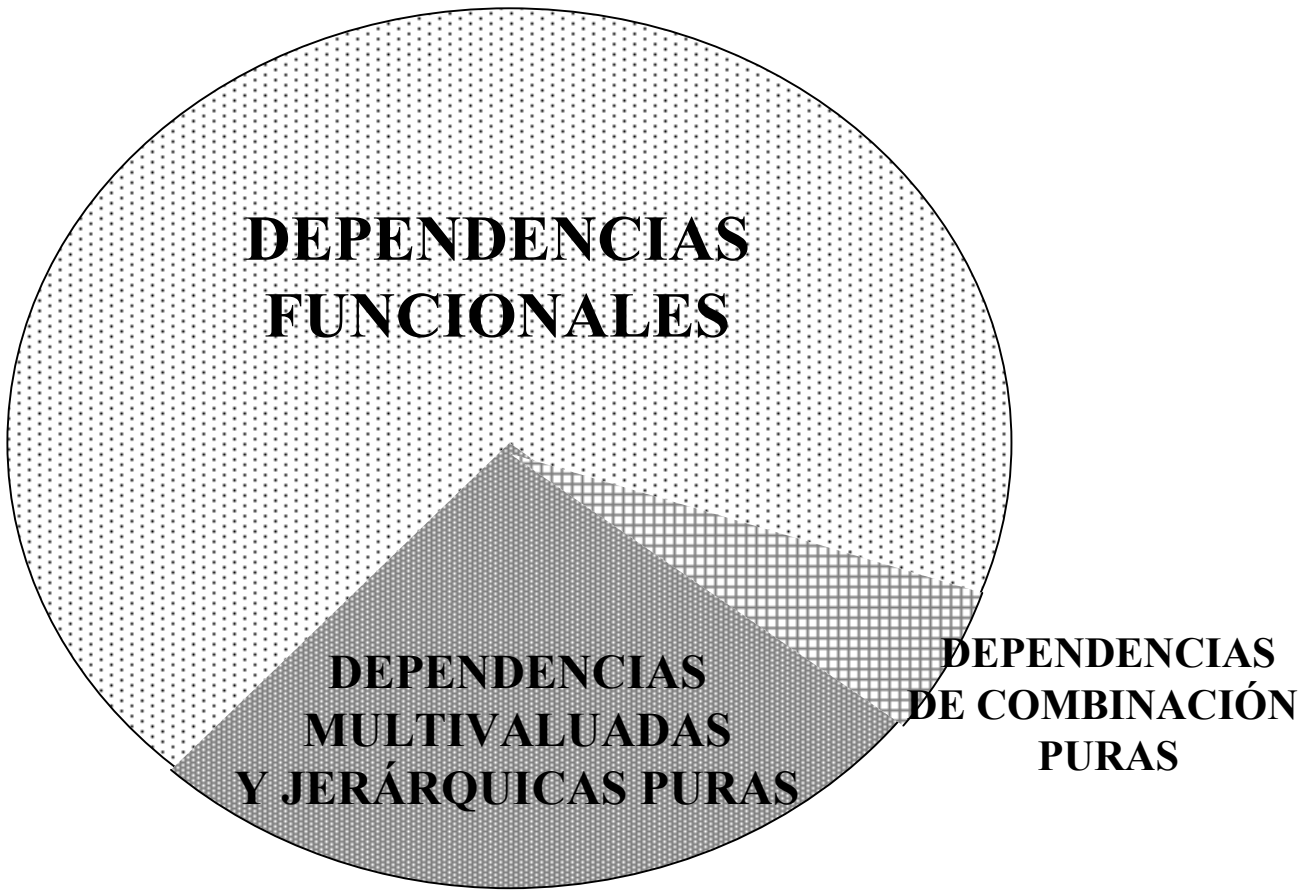
- Una relación R está en 5FN si, y sólo si, está en todas las anteriores formas normales y toda DC está implicada por una clave candidata.
- Otra definición equivalente que no se basa en las formas normales anteriores:
  - *Una relación está en 5FN si, y sólo si, toda DF, DM o DC no trivial es consecuencia de las claves candidatas*.
- Una relación que no se encuentre en 5FN por tener una DC sobre sus descriptores X_1, X_2, ..., X_j puede ser descompuesta **sin pérdida de información** en j proyecciones independientes:

  $R = R_1(X_1) * R_2(X_2) * ... * R_j(X_j)$

- Las descomposiciones de relaciones se realizan utilizando el **operador de proyección**. Siempre ha de ser posible reconstruir la relación original aplicando el **operador de combinación** ("*join*") a las relaciones resultantes.

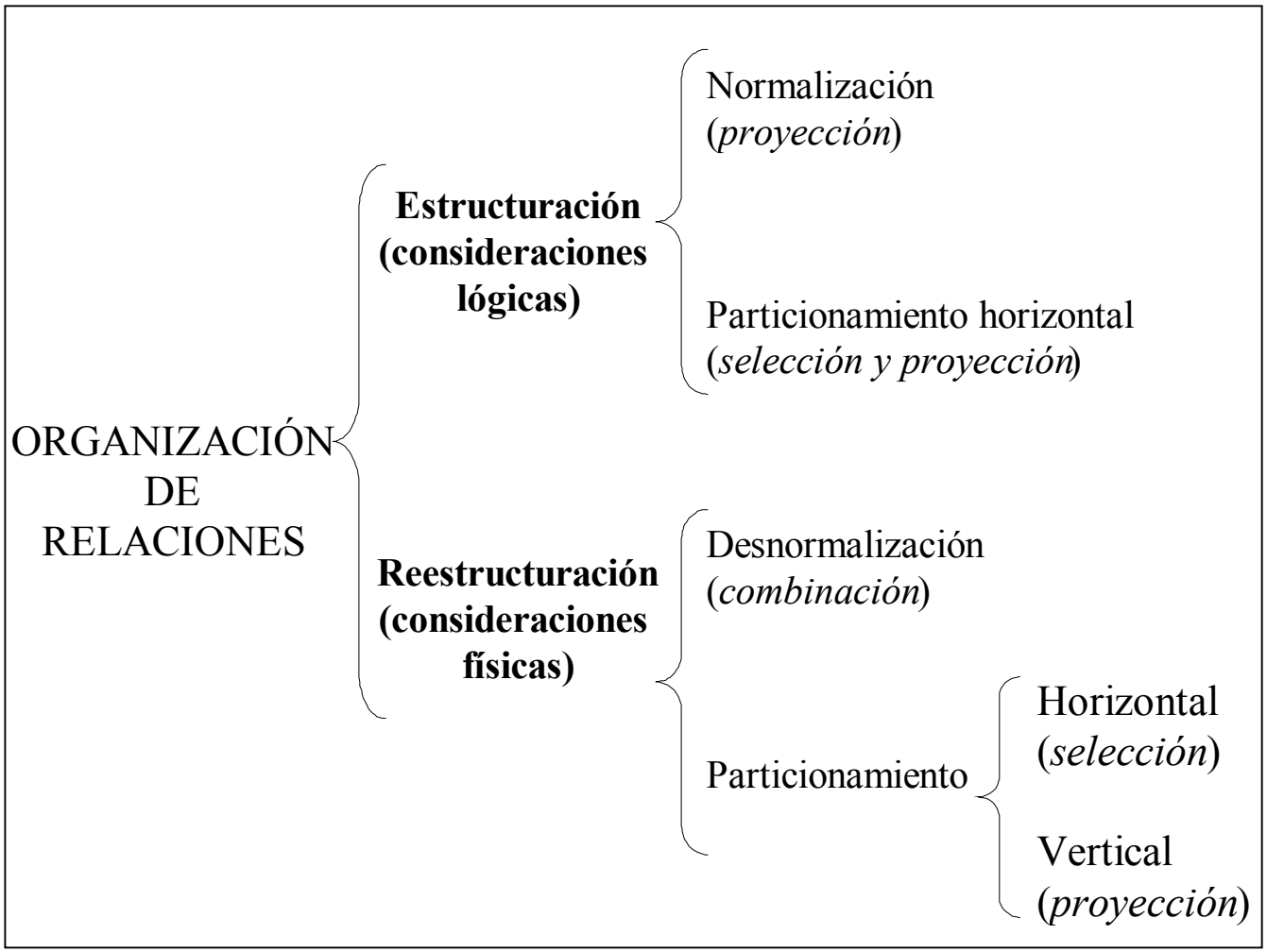


- En la realidad, se dan pocos casos de relaciones que encontrándose en FNBC no estén también en 4FN y menos aún las que estando en 4FN no estén también en 5FN. Esto se debe a la distribución de los tipos de dependencias en las aplicaciones reales:



- Es recomendable normalizar hasta 5FN (siempre que no se pierdan dependencias funcionales), aún cuando sea necesario un posterior proceso de desnormalización; de esta forma se pueden documentar las causas de tener que desnormalizar y el esquema en 5FN teóricamente perfecto.
- Si se sigue una metodología que parta de un esquema E/R y se aplican correctamente las reglas de transformación del tema 6, se llega, salvo contadas excepciones, directamente a un esquema relacional en 5FN.
- Sin embargo, la normalización se muestra insuficiente para alcanzar todos los objetivos de diseño lógico y físico, por lo que, en la práctica, muchas veces es preciso proceder a un proceso posterior de **organización de las relaciones**.

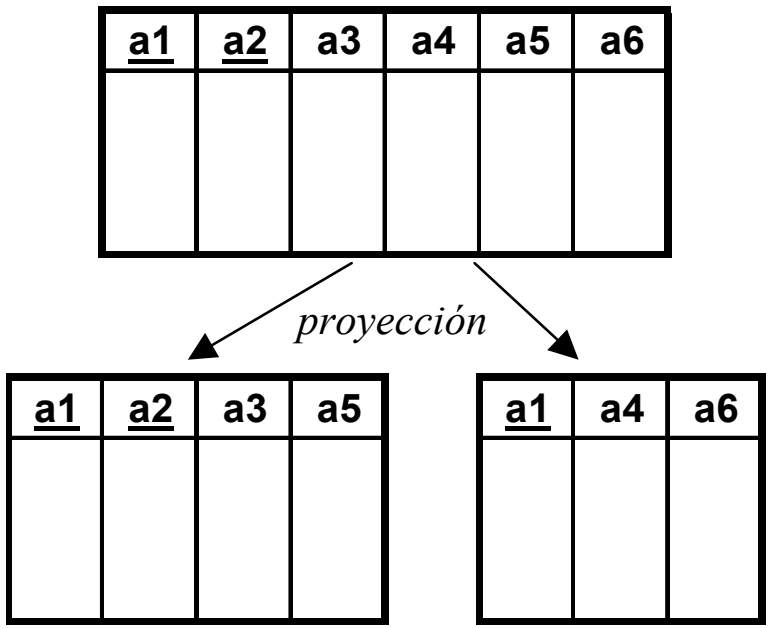
- Este proceso de organización se puede descomponer en dos etapas:



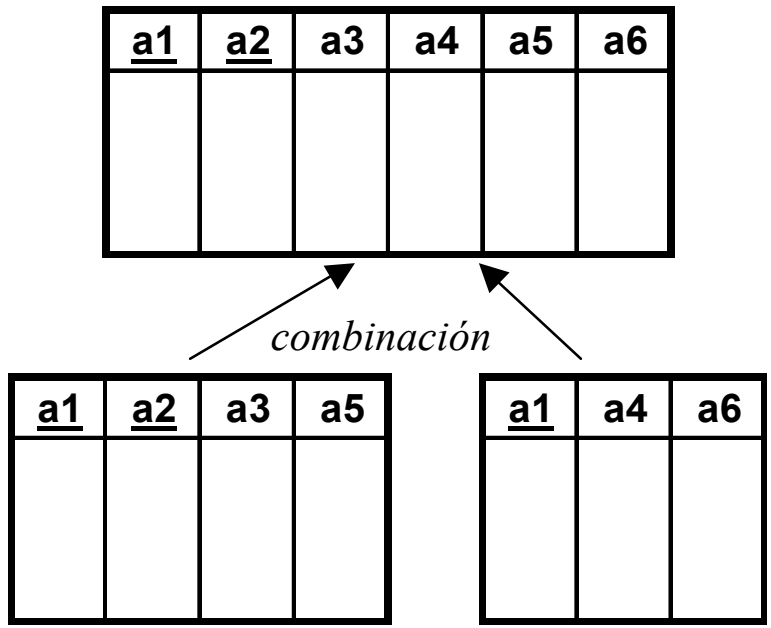
- La **estructuración** de relaciones atiende a consideraciones de tipo lógico. Utiliza dos técnicas:
  - Normalización (*proyección*), y
  - Particionamiento horizontal (*selección*+*proyección*).
    - Utilizada especialmente para eliminar valores nulos originados porque un atributo no es aplicable a todas las tuplas de una relación. Para ello:
      1. Se aplica el operador de *selección* para separar en una relación las tuplas con valor nulo y en otra relación las que no.
      2. Después, se realiza una *proyección* en la relación con nulos para quitar el atributo que los contiene.
    - La relación original se puede reconstruir mediante una *unión*.

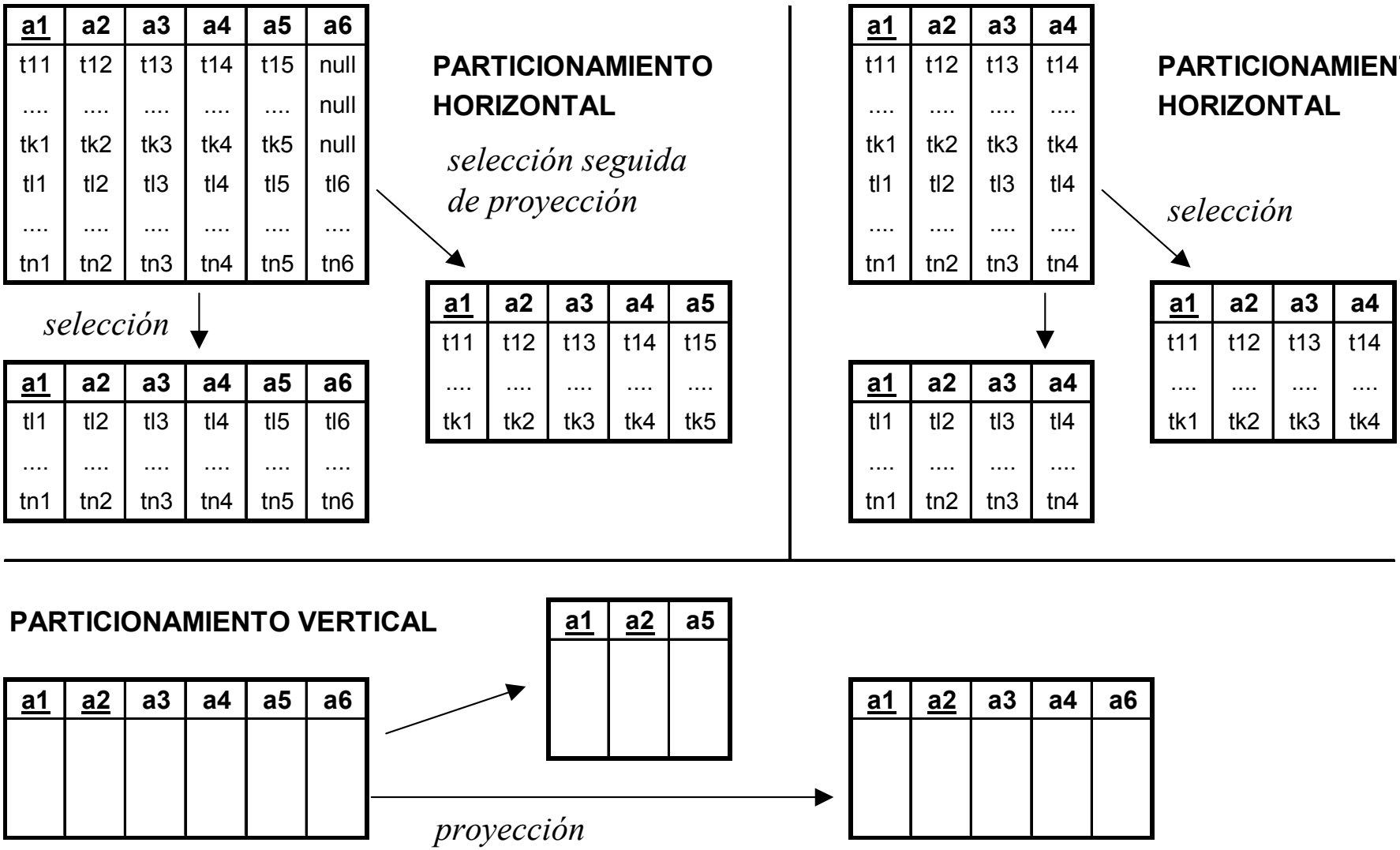
- La **reestructuración** se rige por consideraciones de tipo físico:
  - la tasa de actualizaciones respecto a la de recuperaciones,
  - las veces que se accede conjuntamente a los atributos,
  - la longitud de los mismos,
  - el tipo de proceso (en línea/por lotes),
  - la prioridad de los procesos,
  - el tamaño de las tablas,
  - etc ...
- Las técnicas utilizadas son tres:
  - Desnormalización: proceso contrario a la normalización, utiliza el operador de *combinación* para deshacer proyecciones.
  - Particionamiento horizontal: divide el conjunto de tuplas de una relación en dos o mas relaciones haciendo uso del operador de *selección*.
  - Particionamiento vertical: se basa en utilizar el operador de *proyección*, pero sin tener en cuenta las dependencias.

NORMALIZACIÓN



DESNORMALIZACIÓN




a1 a2 a3 a4 a5 a6
t11 t12 t13 t14 t15 null
.... null
tk1 tk2 tk3 tk4 tk5 null
tl1 tl2 tl3 tl4 tl5 tl6
....
tn1 tn2 tn3 tn4 tn5 tn6
PARTICIONAMIENTO HORIZONTAL
selección seguida de proyección
a1 a2 a3 a4 a5
t11 t12 t13 t14 t15
....
tk1 tk2 tk3 tk4 tk5
selección
a1 a2 a3 a4 a5 a6
tl1 tl2 tl3 tl4 tl5 tl6
....
tn1 tn2 tn3 tn4 tn5 tn6
a1 a2 a3 a4
t11 t12 t13 t14
....
tk1 tk2 tk3 tk4
tl1 tl2 tl3 tl4
....
tn1 tn2 tn3 tn4
PARTICIONAMIENTO HORIZONTAL
selección
a1 a2 a3 a4
t11 t12 t13 t14
....
tk1 tk2 tk3 tk4
a1 a2 a3 a4
tl1 tl2 tl3 tl4
....
tn1 tn2 tn3 tn4
PARTICIONAMIENTO VERTICAL
a1 a2 a3 a4 a5 a6
a1 a2 a5
a1 a2 a3 a4 a6
proyección