



BASES DE DATOS

Tema 7

Teoría de la Normalización

(parte 2)

UCLM- E.S. de Informática

Coral Calero, Marcela Genero, Francisco Ruiz



Objetivos

- Conocer bien los **objetivos de la normalización**.
- Aprender a trabajar con las **formas normales básicas** que están basadas en el concepto de dependencia funcional.
- Presentar las **formas normales avanzadas** y los tipos de dependencias en que se basan.
- Aprender los **procedimientos de análisis y de síntesis** para normalizar relaciones.
- Presentar las diversas técnicas de **estructuración y reestructuración** para optimizar los esquemas relacionales.



Contenido

- Objetivos de la normalización
 - Conservación de la información
 - Conservación de las dependencias
 - Eliminar problemas
- Formas normales básicas
 - 1FN
 - 2FN
 - 3FN
 - Forma normal de Boyce-Codd
- Formas normales avanzadas
 - Dependencias multivaluadas y 4FN
 - Dependencias de combinación y 5FN
- Planteamientos distintos en la normalización
- Método de análisis
 - Descomposición sin pérdidas
 - Procedimiento de descomposición
- Método de síntesis
- Mejora de la calidad de esquemas relacionales
 - Estructuración y reestructuración de relaciones

ESI UCLM-BDa

7 Parte 2.3



Bibliografía

- Básica
 - Piattini et al. (2006)
 - Cap. 10
- Complementaria
 - Batini et al. (1994)
 - Cap. 12
 - Elmasri y Navathe (2002)
 - Cap. 14-15
 - Conolly y Begg (2005)
 - Cap. 13 y 15
 - Date (2004)
 - Caps. 11 y 12

ESI UCLM-BDa

7 Parte 2.4



Objetivos de la Normalización

- 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).



Objetivos de la Normalización

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

ESTUDIANTE_SOLICITA_BECA

Cód_Stud	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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



Objetivos de la Normalización

- 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):

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

tales que cumplan unas determinadas condiciones.



Objetivos de la Normalización

- 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)



Objetivos de la Normalización

- Si se cumplen a) y b), es decir, la transformación de R en $\{R_i\}$ se hace sin pérdida de información ni de dependencias, se dice que $\{R_i\}$ 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.**
 - Estos criterios **no son contemplados** por la teoría de la normalización.



Objetivos de la Normalización

- Algunos autores proponen otros **objetivos adicionales**:
 - Minimización de dependencias**, que incluye no sólo minimizar el número de dependencias, sino también el número de atributos contenidos en ellas.
 - 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.



Objetivos de la Normalización Conservación de la Información

- 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:

$$\bigstar_{i=1}^n r_i = r$$

ESI UCLM-BDa

7 Parte 2.11



Objetivos de la Normalización Conservación de la Información – tuplas espurias

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**

ESI UCLM-BDa

7 Parte 2.12



Objetivos de la Normalización

Conservación de las Dependencias

- 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 en el tema anterior, 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 visto, de equivalencia de dos conjuntos de dependencias.



Formas Normales Básicas

- 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

• Formas Normales Básicas:

- Inicialmente, Codd propuso en 1970 tres formas normales basadas en las **Dependencias Funcionales**: **primera** (1FN), **segunda** (2FN) y **tercera** forma normal (3FN).
- 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 forma 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



Formas Normales Básicas

1FN

- 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 GONZ ALEZ	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.



Formas Normales Básicas

2FN

- La **segunda forma normal** 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.



Formas Normales Básicas

2FN

- 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.



Formas Normales Básicas 2FN

- **Ejemplo:**

Sea el esquema de relación $R(AT, DEP)$ donde:

$AT = \{A, B, C, D\}$ $DEP = \{AB \rightarrow C, A \rightarrow D\}$ y $PK = (A, B)$

El atributo D no es un hecho (una información) acerca de la totalidad de la clave, sino acerca de parte de ella (en este caso del atributo A). Por tanto, la relación no está en 2FN.

Transformamos la relación R en las relaciones $R1$ y $R2$ que ya sí se encuentran en 2FN:

$R1(AT1, DEP1)$ donde:

$AT1 = \{A, B, C\}$ $DEP1 = \{AB \rightarrow C\}$

$R2(AT2, DEP2)$ donde:

$AT2 = \{A, D\}$ $DEP2 = \{A \rightarrow D\}$



Formas Normales Básicas 3FN

- La **tercera forma normal** 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.



Formas Normales Básicas

3FN

- 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.



Formas Normales Básicas

3FN

- **Ejemplo:**

Sea el esquema de relación $R(AT, DEP)$ donde:

$AT = \{A, B, C\}$ $DEP = \{B \rightarrow C, A \rightarrow B\}$ y $PK = (A)$

La única clave del esquema de relación es el atributo A . El atributo C es un hecho acerca del atributo B , atributo que no forma parte de la clave. Por lo tanto, este esquema de relación no está en 3FN (aunque sí en 2FN).

Se puede transformar la relación R en las relaciones $R1$ y $R2$ que ya sí se encuentran en 3FN:

$R1(AT1, DEP1)$ donde:

$AT1 = \{A, B\}$ $DEP1 = \{A \rightarrow B\}$

$R2(AT2, DEP2)$ donde:

$AT2 = \{B, C\}$ $DEP2 = \{B \rightarrow C\}$



Formas Normales Básicas

FNBC

- 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.



Formas Normales Básicas

FNBC

• Ejemplo:

Dado el esquema de relación R(AT, DEP) con :

$AT = \{A, B, C, D\}$ $DEP = \{A \leftrightarrow B, AC \rightarrow D\}$

R tendría dos claves candidatas: (A, C) y (B, C) .

Esta relación está en 3FN, sin embargo tiene anomalías de actualización, ya que se repetirían los valores de A y B por cada valor de C.

El problema es debido a que R no se encuentra en FNBC, ya que tanto A como B son determinantes, pero no son claves candidatas de la relación.



Formas Normales Básicas FNBC

- 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.



Formas Normales Básicas FNBC

- **Ejemplo:**

Dado el esquema de relación $R(AT, DEP)$ con :

$AT = \{A, B, C, D\}$ $DEP = \{ABC \rightarrow D, BCD \rightarrow A\}$

las claves candidatas de esta relación son (A, B, C) y (B, C, D) , que se solapan ya que comparten los atributos B y C ; sin embargo, debido a que los únicos determinantes son los dos descriptores anteriores, que son claves candidatas, la relación sí se encuentra en FNBC.



Formas Normales Básicas FNBC

- **Ejemplo:**

Dado el esquema de relación $R(AT, DEP)$ con :

$AT = \{A, B, C\}$ $DEP = \{AC \rightarrow B, B \rightarrow C\}$

Este esquema de relación tiene dos claves candidatas: (A, C) y (A, B) .

La relación así definida está en 3FN –todos sus atributos son principales- pero no está en FNBC, puesto que el determinante B no es una clave candidata de la relación.

Se puede transformar la relación R en las relaciones $R1$ y $R2$ que ya sí se encuentran en FNBC:

$R1(AT1, DEP1)$ donde:

$AT1 = \{A, B\}$ $DEP1 = \{\}$

$R2(AT2, DEP2)$ donde:

$AT2 = \{B, C\}$ $DEP2 = \{B \rightarrow C\}$

La dependencia $AC \rightarrow B$ se ha perdido en la transformación anterior, ya que no es posible deducirla del conjunto de dependencias de los esquemas resultantes. A pesar de ello, ésta es la mejor descomposición de las tres posibles, ya que en las otras dos se pierde también información.

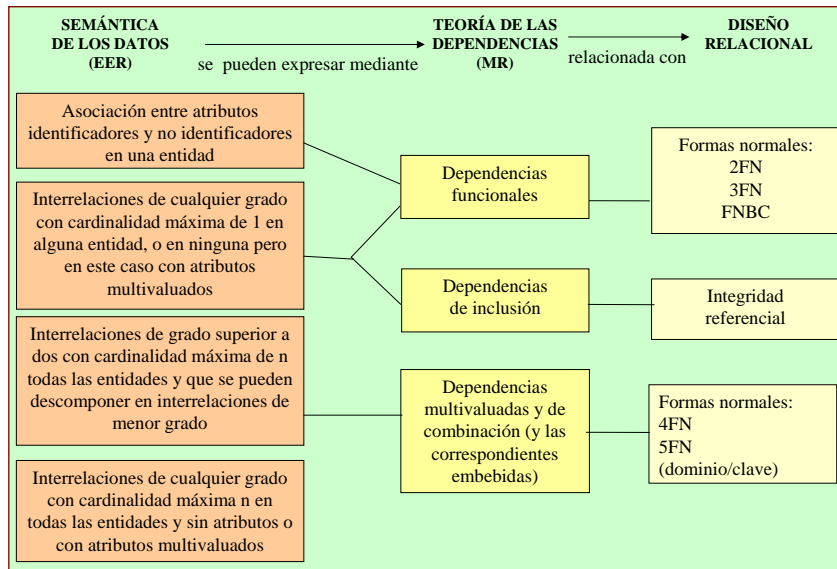


Formas Normales Avanzadas

- 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.



Formas Normales Avanzadas



ESI UCLM-BDa

7 Parte 2.29



Formas Normales Avanzadas

Dependencias Multivaluadas y 4FN

- 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 $X \twoheadrightarrow Y$ (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 $X \twoheadrightarrow Y$, al mismo tiempo habrá de cumplirse $X \twoheadrightarrow \{A - X - Y\}$.
 - Se representa como $X \twoheadrightarrow Y \mid \{A - X - Y\}$

ESI UCLM-BDa

7 Parte 2.30



Dependencias Multivaluadas y 4FN

- Una DM $X \twoheadrightarrow Y$ 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 $X \twoheadrightarrow Y \mid Z$, 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.



Dependencias Multivaluadas y 4FN

ASIGNATURAS

<i>Nom_Asignatura</i>	<i>Profesor</i>	<i>Texto</i>
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



Formas Normales Avanzadas Dependencias Multivaluadas y 4FN

ASIGNATURAS

<i>Nom_Asignatura</i>	<i>Profesor</i>	<i>Texto</i>
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 →→ *Profesor* *Nom_Asignatura* →→ *Texto*

Dependencias multivaluadas generadas al poner en 1FN

ESI UCLM-BDa

7 Parte 2.33



Formas Normales Avanzadas Dependencias Multivaluadas y 4FN

- 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. $R1 \cap R2 \rightarrow R1 - R2$
 2. $R1 \cap R2 \rightarrow R2 - R1$
- Se demuestra 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 $A \twoheadrightarrow B \mid C$.

ESI UCLM-BDa

7 Parte 2.34



Formas Normales Avanzadas
Dependencias de Combinación y 5FN

- 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

<i>Editorial</i>	<i>Idioma</i>	<i>Tema</i>
RA-MA	Inglés	BD
RA-MA	Español	CASE
AddisonWesley	Español	BD
RA-MA	Español	BD



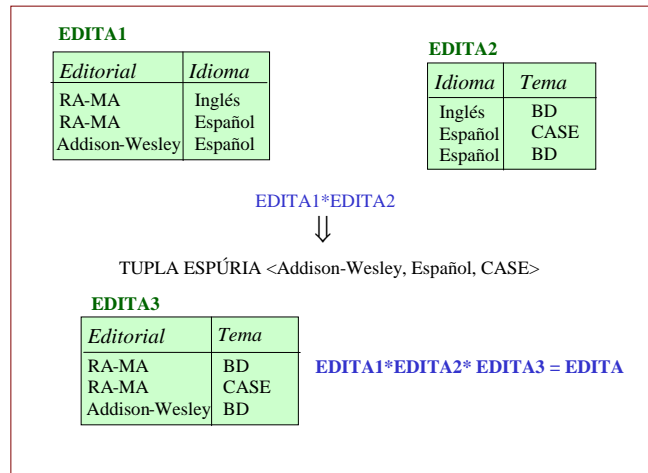
Formas Normales Avanzadas
Dependencias de Combinación y 5FN

- 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)



Formas Normales Avanzadas
Dependencias de Combinación y 5FN

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



ESI UCLM-BDa

7 Parte 2.37



Formas Normales Avanzadas
Dependencias de Combinación y 5FN

- 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 una DC respecto de sus proyecciones R_1, R_2, \dots, R_j , si:
 - $R = R_1 * R_2 * \dots * R_j$
se denota por: $DJ * (R_1, \dots, R_j)$

ESI UCLM-BDa

7 Parte 2.38



Formas Normales Avanzadas
Dependencias de Combinación y 5FN

- 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 $DC^*(R_1, R_2 \text{ y } R_3)$, entonces se satisface la restricción:
 - $\langle A, B, C \rangle \in R$ si y sólo si
 - $\langle A, B \rangle \in R_1$,
 - $\langle B, C \rangle \in R_2$, y
 - $\langle C, A \rangle \in R_3$

ESI UCLM-BDa

7 Parte 2.39



Formas Normales Avanzadas
Dependencias de Combinación y 5FN

- 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, \dots, X_j puede ser descompuesta **sin pérdida de información** en j proyecciones independientes:
$$R = R_1(X_1) * R_2(X_2) * \dots * R_j(X_j)$$

ESI UCLM-BDa

7 Parte 2.40

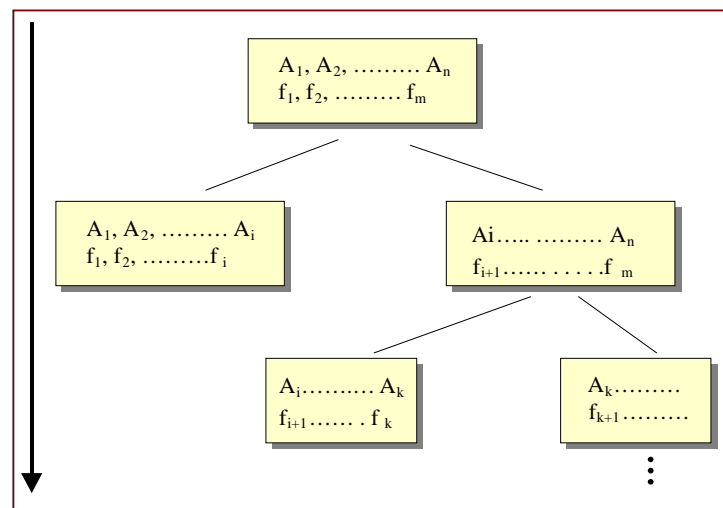


Planteamientos Distintos en la Normalización

- 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.



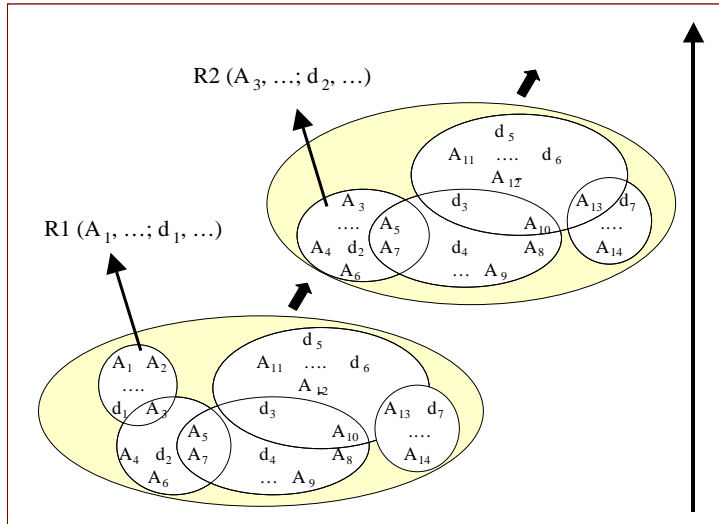
Planteamientos Distintos en la Normalización



Análisis: árbol de descomposición.



Planteamientos Distintos en la Normalización



Síntesis: agrupando datos

ESI UCLM-BDa

7 Parte 2.43



Planteamientos Distintos en la Normalización

- 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		
	DF	DF	DF	DF, DM, DC	DF, DM, DC
Dependencias de partida	DF	DF	DF	DF, DM, DC	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

ESI UCLM-BDa

7 Parte 2.44



Descomposición sin Pérdidas

- 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:

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

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:

$$R = R_1 * R_2 * \dots * R_p$$

para toda extensión de R (descomposición SPI).

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

$$(U DF_i)^+ = DF^+$$



Descomposición sin Pérdidas

- 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:
 - Sus atributos comunes son la clave candidata de, al menos, una relación.**
 - Cada DF en R puede deducirse de las de R_1 y R_2 .**



Descomposición sin Pérdidas

- **Ejemplo:**

Dada la relación $R(\{A, B, C\}, \{A \rightarrow B, B \rightarrow C\})$ que no está en 3FN, existen tres posibilidades de descomposición:

1)

$R1(\{A, C\}, \{A \rightarrow C\})$ y $R2(\{B, C\}, \{B \rightarrow C\})$

Que atenta contra el primer principio de Rissanen.

2)

$R3(\{A, B\}, \{A \rightarrow B\})$ y $R4(\{A, C\}, \{A \rightarrow C\})$

Que atenta contra el segundo principio.

3)

$R5(\{A, B\}, \{A \rightarrow B\})$ y $R6(\{B, C\}, \{B \rightarrow C\})$

Que cumple los dos principios de Rissanen.



Descomposición sin Pérdidas

- 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-Codd**, ya que en ese caso, puede ser preciso perder dependencias.



Método de Análisis

Descomposición sin Pérdidas

- **Ejemplo en el que siempre hay pérdidas para llegar a FNBC:**

Dada la relación $R(\{A, B, C\}, \{AC \rightarrow B, B \rightarrow C\})$, poner en FNBC.

Esta relación se encuentra en 2FN y 3FN, ya que todos los atributos son principales. Las claves candidatas de la relación son (A, C) y (A, B) , mientras que los determinantes son (A, C) y (B) por lo que no todo determinante es clave candidata, y por tanto, R no está en FNBC.

La relación puede descomponerse de tres formas distintas:

1)

$R1(\{A, C\}, \{\})$ y $R2(\{B, C\}, \{B \rightarrow C\})$

En esta descomposición se pierden DF e información.

2)

$R3(\{A, B\}, \{\})$ y $R4(\{A, C\}, \{\})$

En esta descomposición también se pierden DF e información.

3)

$R5(\{A, B\}, \{\})$ y $R6(\{B, C\}, \{B \rightarrow C\})$

Esta descomposición resulta la mejor de las tres, pero se pierde la DF $AC \rightarrow B$.



Método de Análisis

Procedimiento de Descomposición

- Dado el esquema $R(A, DF)$, 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 los implicados del grupo no aparecen en la relación restante.

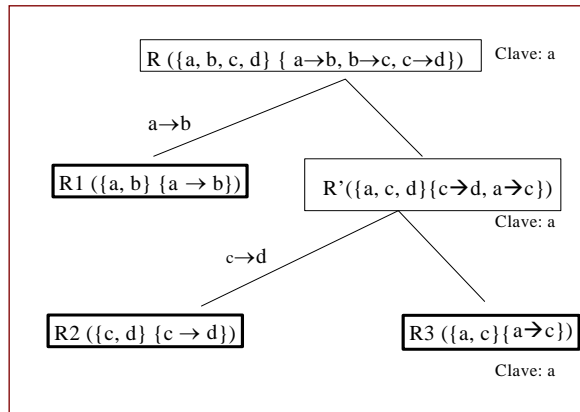
6. **Repetir el paso 5** hasta que no pueda continuarse porque todas las dependencias estén implicadas por una clave (en este caso, para llegar hasta FNBC hay que perder dependencias; y es decisión del diseñador parar el proceso en la 3FN o avanzar hasta FNBC con el inconveniente señalado).



Procedimiento de Descomposición

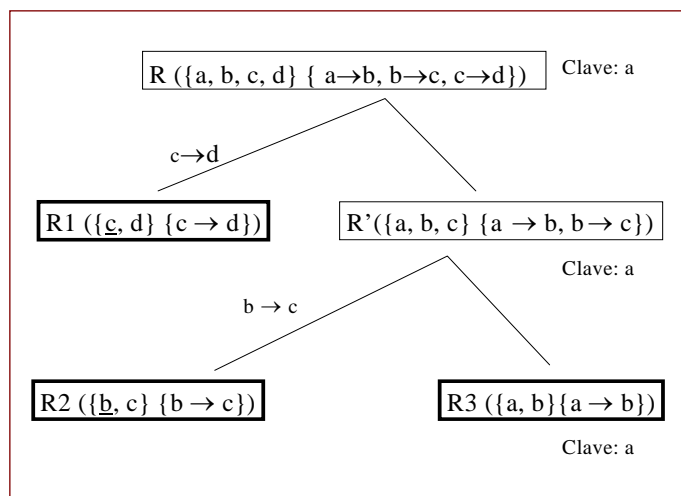
- Es relevante el orden en el que se van tomando los grupos de dependencias funcionales: los atributos que desaparecen de la relación origen deben ser aquellos que no entran a formar parte de ninguna otra dependencia (en caso contrario se perderían dependencias).

Incorrecto:
se pierde
 $b \rightarrow c$



Procedimiento de Descomposición

Correcto: se debe comenzar separando $c \rightarrow d$





Método de Síntesis

- 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.



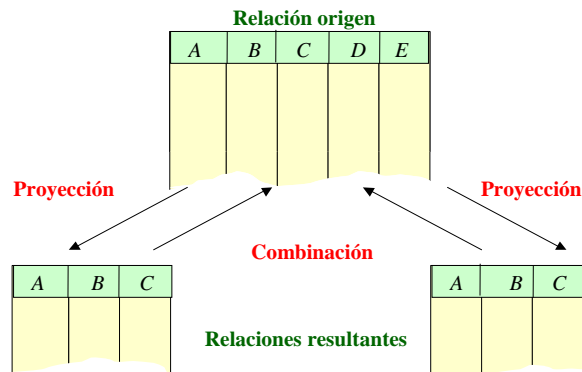
Método de Síntesis

- **Ejemplo:**
 Poner en 3FN $R(AT, DF)$, siendo $AT = \{P, E, N, A, H, L, G, T, D\}$ y
 $DF = \{HE \rightarrow L, HP \rightarrow L, HL \rightarrow A, D \rightarrow T, HE \rightarrow A, EA \rightarrow N, P \rightarrow T, P \rightarrow D, T \rightarrow P, D \rightarrow P\}$
 Recubrimiento minimal: $DF^m = \{HE \rightarrow L, HP \rightarrow L, HL \rightarrow A, EA \rightarrow N, P \rightarrow D, D \rightarrow T, T \rightarrow P\}$
 Claves: $K1 = \{H, E, P, G\}$ $K2 = \{H, E, D, G\}$ $K3 = \{H, E, T, G\}$
 Como N depende de parte de una clave ($EA \rightarrow N$) y no de su totalidad, **la relación no está en 2FN**, y por tanto, sólo **está en 1FN**.
 Aplicación del método de síntesis:
 Paso 1
 El recubrimiento minimal ya ha sido calculado
 Paso 2
 Dividimos DF^m en particiones con igual implicante:
 $DF1 = \{HE \rightarrow L\}$ $DF2 = \{HP \rightarrow L\}$ $DF3 = \{HL \rightarrow A\}$
 $DF4 = \{EA \rightarrow N\}$ $DF5 = \{P \rightarrow D, D \rightarrow T, T \rightarrow P\}$
 Paso 3
 Creamos una relación para cada partición:
 $R1(\underline{H, E}, L); \{HE \rightarrow L\}$
 $R2(\underline{H, P}, L); \{HP \rightarrow L\}$
 $R3(\underline{H, L}, A); \{HL \rightarrow A\}$
 $R4(\underline{E, A}, N); \{EA \rightarrow N\}$
 $R5(\underline{P, D, T}); \{P \rightarrow D, D \rightarrow T, T \rightarrow P\}$ también $\{D\}$ y $\{T\}$ son claves candidatas
 Paso 4
 Se forma otra relación con la clave de la relación original R :
 $R6(\underline{H, E, P, G}); \{\}$



Mejora de la Calidad de Esquemas Relacionales

- 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.



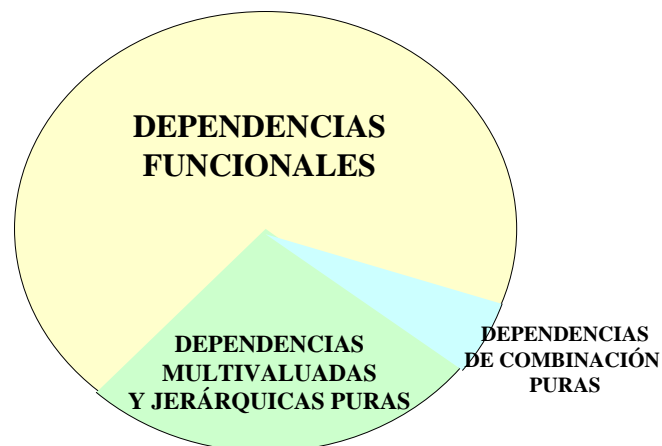
ESI UCLM-BDa

7 Parte 2.55



Mejora de la Calidad de Esquemas Relacionales

- 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:



ESI UCLM-BDa

7 Parte 2.56



Mejora de la Calidad de Esquemas Relacionales

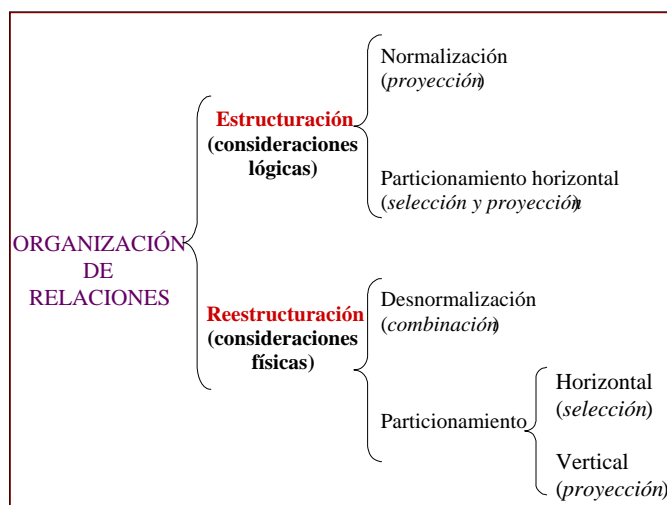
- Es **recomendable normalizar hasta 5FN** (siempre que no se pierdan dependencias funcionales), aún cuando sea necesario un **posterior proceso de desnormalización**; de esta forma se pueden documentar las causas de tener que desnormalizar el esquema en 5FN teóricamente perfecto.
- Si se sigue una metodología 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**.



Mejora de la Calidad de Esquemas Relacionales

Estructuración y Reestructuración de Relaciones

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





Estructuración y Reestructuración de Relaciones

- 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**.

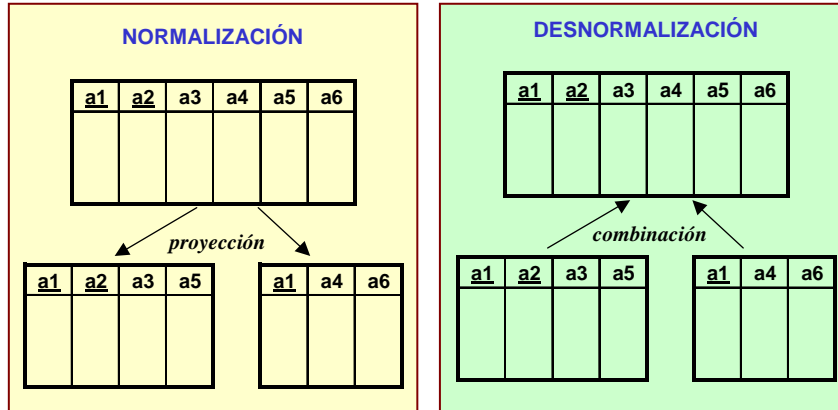


Estructuración y Reestructuración de Relaciones

- La **reestructuración** se rige por consideraciones de tipo físico:
 - tasa de actualizaciones respecto a la de recuperaciones,
 - veces que se accede conjuntamente a los atributos,
 - longitud de los atributos,
 - tipo de proceso (en línea/por lotes),
 - prioridad de los procesos,
 - 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.



Mejora de la Calidad de Esquemas Relacionales Estructuración y Reestructuración de Relaciones

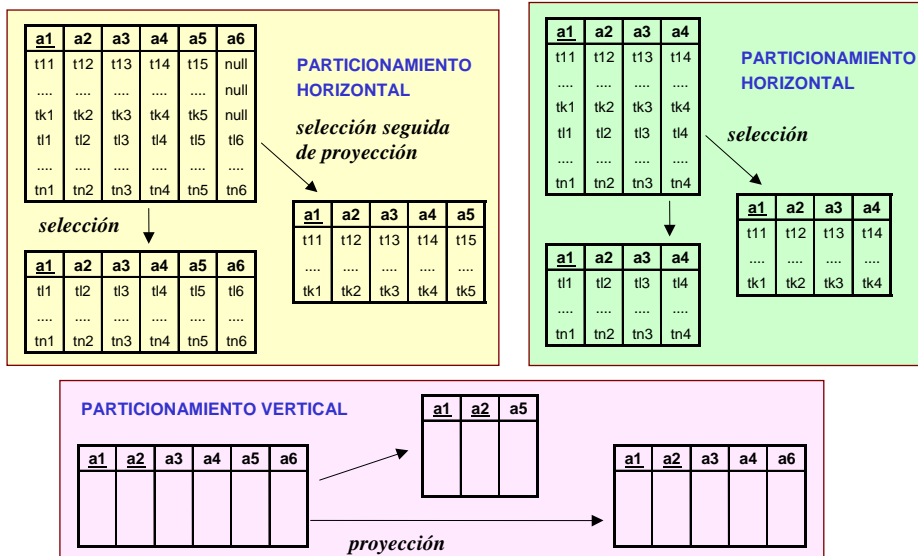


ESI UCLM-BDa

7 Parte 2.61



Mejora de la Calidad de Esquemas Relacionales Estructuración y Reestructuración de Relaciones



ESI UCLM-BDa

7 Parte 2.62