



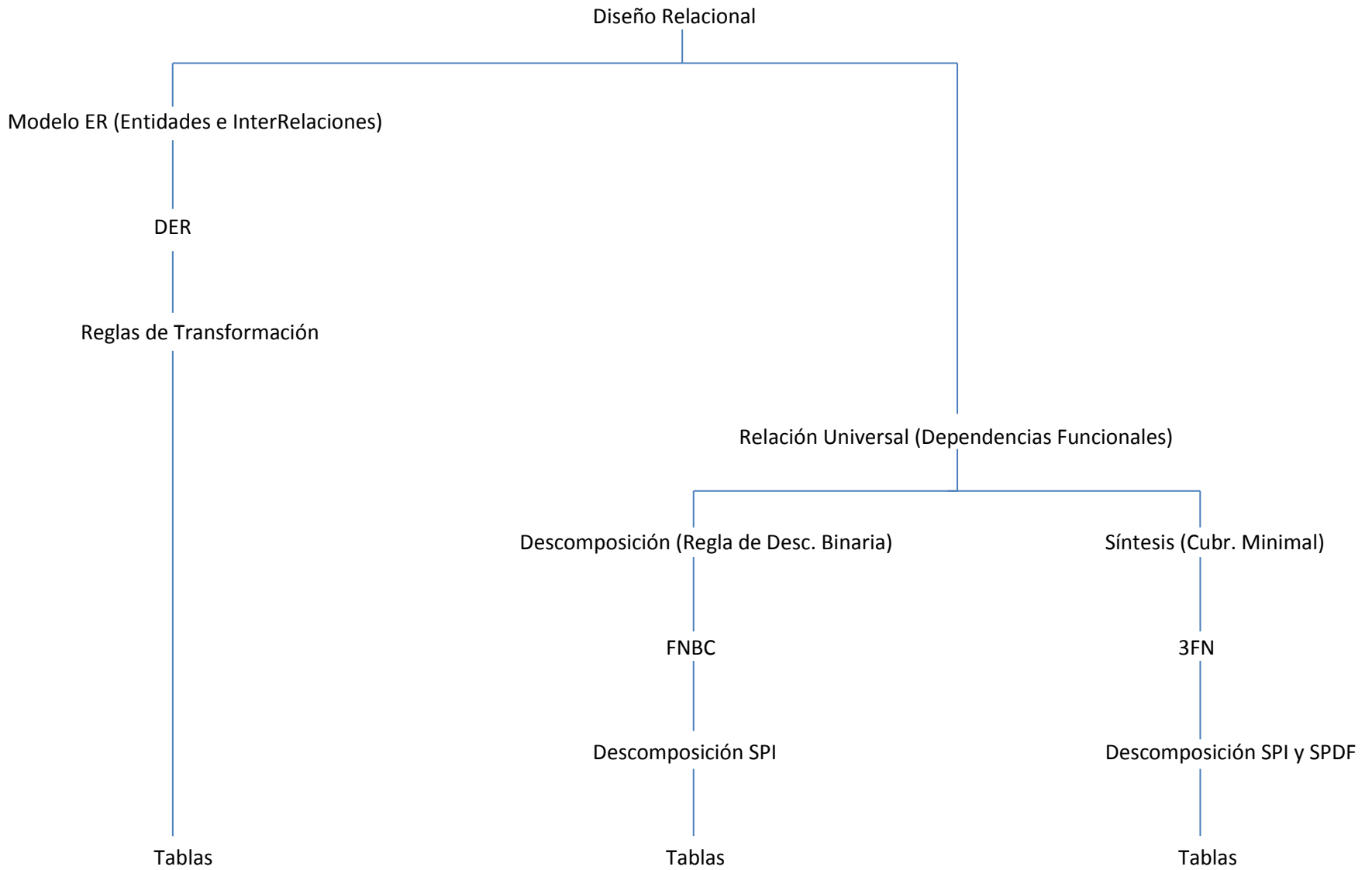
NORMALIZACION

**Primera Clase Práctica
2C-2014**

**Docente: Alejandro Eidelsztein
Base de Datos
FCEyN - UBA**

Hay dos aproximaciones básicas al problema del *Diseño Relacional*:

- La primera son los **Modelos Semánticos** –que son intuitivos y poco rigurosos- Uno de los más populares es el de Entidad Relación (ER) que utiliza los conceptos de Entidades e Inter-Relaciones y sus propiedades en un proceso natural de diseño. Consiste en la especificación inicial de un diagrama de ER (DER) que es luego “mapeado” al esquema relacional.
- La segunda es la de **Relación Universal** (RU) –más rigurosa y menos intuitiva que la de ER- En esta se asume que toda la semántica es capturada por un conjunto de Dependencias definidas sobre un conjunto universal de Atributos sin tener en cuenta -inicialmente- como agruparlos en relaciones. Deriva luego en la conocida metodología de Normalización la cual agrupa los atributos en relaciones eliminando las redundancias.



EJERCICIO:

En una facultad se desea registrar los datos personales de los alumnos, las materias en las que están inscriptos (con la fecha en que lo hicieron) y los exámenes que rindieron (con su correspondiente fecha y nota). Los alumnos se identificarán por el número de libreta. Cada alumno se puede inscribir una sola vez en cada materia y puede rendir varias veces cada una hasta aprobarla.

SE DESEA REGISTRAR:

Datos personales de los alumnos

Datos de las materias

Inscripciones

Exámenes

CON LAS SIGUIENTES RESTRICCIONES:

No puede haber dos alumnos con el mismo número de libreta

No hay dos materias con el mismo código de materia

Se puede inscribir una sola vez en cada materia

A cada examen le corresponde una nota

Una materia se puede rendir varias veces hasta aprobarla

SE PIDE:

- a) Diseñar un esquema de *relación universal*.
- b) Determinar las *dependencias funcionales*.
- c) Hallar las *claves*.
- d) Indicar las *anomalías* que podría presentar este esquema.
- e) Indicar si el esquema cumple con la *tercera forma normal (3FN)*
- f) Para eliminar las anomalías indicadas, hallar por *síntesis* una descomposición en 3FN que sea sin pérdida de información (SPI) y sin pérdida de dependencias funcionales (SPDF)
- g) Idem anterior pero hallar una descomposición en *forma normal de Boyce Codd (FNBC)* que sea SPI.
- h) ¿La descomposición hallada en el punto anterior es SPDF?

a) Construimos el esquema:

FACULTAD (Lu, Nom,	(datos personales)
Mat, Desc	(datos de las materias)
Mat, Desc, IFec,	(materias en las que se inscribió)
EFec, Nota)	(exámenes que rindió)

Observar que en este esquema hay varias entidades juntas...



b) Agregamos las dependencias funcionales:

Lu --> Nom	(no puede haber dos alumnos con el mismo Lu)
Mat --> Desc	(no puede haber dos materias con el mismo código)
Lu, Mat --> IFec	(se puede inscribir una sola vez en cada materia)
Lu, Mat, EFec --> Nota	(hay una sola nota por examen)

¿Sería correcto agregar esta otra?:

Lu, Mat --> EFec

No, porque no se podría rendir varias veces la misma materia

c) Hallamos las claves:

FACULTAD (Lu, Nom, Mat, Desc, IFec, EFec, Nota)

Lu --> Nom

Mat --> Desc

Lu, Mat --> IFec

Lu, Mat, EFec --> Nota

Primero computamos los atributos que no están en ningún lado derecho:

{Lu, Mat, EFec}

Luego calculamos la clausura:

{Lu, Mat, EFec}⁺ = FACULTAD,

Luego {Lu, Mat, EFec} es *Superclave*

Vemos que también es *Clave* y es la única, ¿Por qué?

d) Analizamos los problemas que podría tener el esquema:

FACULTAD (Lu, Nom, Mat, Desc, IFec, EFec, Nota)

- 1) **Redundancia de información:** El nombre del alumno se repite por cada materia en la que se inscribe y por cada examen que rinda. Lo mismo pasa con el nombre de la materia y la fecha de inscripción si la rinde varias veces.
- 2) **Anomalías de actualización:** Si una alumna se casa y decide cambiar por su apellido de casada deberíamos actualizar varias tuplas y podríamos cometer errores u omisiones.
- 3) **Anomalías de inserción:** No podemos dar de alta a un alumno hasta que se haya inscripto en la primer materia o aún peor hasta que haya rendido el primer examen. Podríamos hacerlo pero tendríamos que poner valores nulos en campos que forman la clave como Materia, IFec y Efec.

FACULTAD (lu1, nom1, - , - , -, -, -)

Si se toma un examen y no se presenta nadie también tendríamos que poner valores nulos en algunos campos de la clave.

FACULTAD (-, -, mat1, desc1 , -, efec1, -)

- 4) **Anomalías de bajas:** Si en algún momento decidimos borrar los datos de los exámenes rendidos por un alumno perderemos los datos personales del mismo.

Definición de FNBC:

R está en FNBC si para toda dependencia no trivial $X \twoheadrightarrow A$ sobre R,

X es superclave de R

Definición de 3FN:

R está en 3FN si para toda dependencia no trivial $X \twoheadrightarrow A$ sobre R,

X es superclave de R o

A es primo

Atributo Primo: A es primo si pertenece al alguna de la claves

Si R está en FNBC \implies también está en 3FN

e) Verificamos la FN del esquema inicial:

FACULTAD (Lu, Nom, Direc, Mat, IFec, EFec, Nota)

LU --> Nom, Direc

Mat --> Desc

Lu, Mat --> IFec

Lu, Mat, EFec --> Nota

Clave: {Lu, Mat, EFec}

No está en FNBC, ¿Por qué?

Tampoco está en 3FN, ¿Por qué?

En realidad no está ni en 2FN

ALGORITMO PARA OBTENER 3FN SPI SPDF

Se obtiene una descomposición ρ de R por Síntesis a partir de un Cubrimiento Minimal del conjunto de dependencias F. El ρ resultante es SPI y SPDF.

- 1) Cada dependencia funcional se convierte en un esquema (las que tienen igual lado izquierdo se juntan)
- 2) Si ninguno de los esquemas resultantes contiene una clave se agrega uno con los atributos de alguna clave
- 3) Eliminar esquemas redundantes: Si alguno de los esquemas resultantes esta contenido totalmente en otro, eliminarlo.

Cubrimiento Mininal:

Dado F buscamos un F_m tal que $F_m \equiv F$ y además F_m tiene:

- 1) Todo lado derecho tiene un único atributo (regla de descomposición)
- 2) Todo lado izquierdo es reducido (no tiene atributos redundantes, $B \subset X$ es redundante para $X \twoheadrightarrow A$ si $A \in (X - \{B\})^+$)
- 3) No contiene dependencias funcionales redundantes (en general las que se obtienen por transitividad, $X \twoheadrightarrow A$ es redundante si $(F - \{X \twoheadrightarrow A\}) \equiv F$)

f) Hallamos una descomposición por *Síntesis* que esté en 3FN que sea SPI y SPDF

Primero debemos verificar que el conjunto de dependencias dado no tenga redundancias. Debe ser un *Cubrimiento Minimal*.

Luego convertimos cada dependencia en un subesquema:

Lu --> Nom == >> ALUMNO (Lu, Nomb)

Mat --> Desc == >> MATERIA (Mat, Desc)

Lu, Mat --> Ifec == >> INSCRIPTO (Lu, Mat, Ifec)

Lu, Mat, EFec --> Nota == >> EXAMEN (Lu, Mat, Efec, Nota)

El determinante de cada dependencia se convierte en la clave de cada subesquema.

Vemos que EXAMEN contiene los atributos de la clave de FACULTAD.

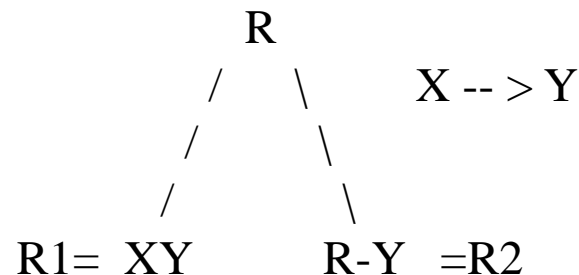
Si ninguno de los subesquemas hubiera contenido alguna de las claves, tendríamos que haber agregado uno con los atributos de una de las claves a la descomposición.

Esta descomposición preserva la información y también preserva las dependencias.

ALGORITMO PARA OBTENER FNBC SPI

Se obtiene una descomposición ρ de R partiendo R aplicando la propiedad de descomposición binaria. El ρ resultante es siempre SPI, pero a veces no es SPDF.

Si hay una dependencia funcional $X \twoheadrightarrow Y$ que viola la FNBC se parte R en R_1 y R_2 de la siguiente forma:



Vemos que esta descomposición es SPI porque aplica la regla de descomposición binaria y por lo tanto:

$$R_1 \cap R_2 = X$$

$$R_1 - R_2 = Y$$

Luego se aplica la misma regla a R_1 y R_2 y así sucesivamente hasta que todas las hojas queden en FNBC.

g) Hallamos una descomposición que esté en FNBC y sea SPI:

Lu Nom Mat Desc IFec EFec Nota

/ \

/ \

Lu --> Nom (viola la FNBC)

/ \

Lu Nom Lu Mat Desc IFec EFec Nota

/ \

/ \

Mat--> Desc (viola la FNBC)

/ \

Mat Desc

Lu Mat IFec EFec Nota

/ \

/ \

Lu Mat--> IFec (viola la FNBC)

/ \

Lu Mat IFec

Lu Mat EFec Nota

De la descomposición ρ en FNBC SPI hallada nos quedan los siguientes subesquemas:

R1= ALUMNO (Lu, Nomb)

R2= MATERIA (Mat, Desc)

R3= INSCRIPTO (Lu, Mat, IFec)

R4= EXAMEN (Lu, Mat, Efec, Nota)

La junta de los subesquemas debe dar la misma información que teníamos en el esquema inicial:

FACULTAD= ALUMNO |><| INSCRIPTO |><| MATERIA |><| EXAMEN

O sea que la descomposición hallada *preserva la información*

Más Formalmente:

$\rho = \{R_1, R_2, \dots, R_k\}$ es SPI, si se cumple:

$$\bigcup_{i=1,k} R_i = R$$

$$r = \bigcap_{i=1,k} \Pi_{R_i}(r)$$

R: Esquema de relación

r: Instancia de R

h) Verificamos si se preservan las dependencias funcionales:

$$\rho = (\text{R1= Lu Nom,} \\ \text{R2= Mat Desc,} \\ \text{R3= Lu Mat IFec,} \\ \text{R4= Lu Mat Efec Nota})$$

Rápidamente vemos que:

Lu-->Nom se proyecta en R1

Mat-->Desc se proyecta en R2

Lu,Mat-->IFec se proyecta en R3

Lu,Mat,Efec-->Nota se proyecta en R4

O sea que se proyectan todas las dependencias y luego se cumple:

$F = (U_{i=1,k} \Pi_{R_i} (F))$ y por lo tanto:

ρ es SPDF

¿Dudas?

¿?