

Normalización

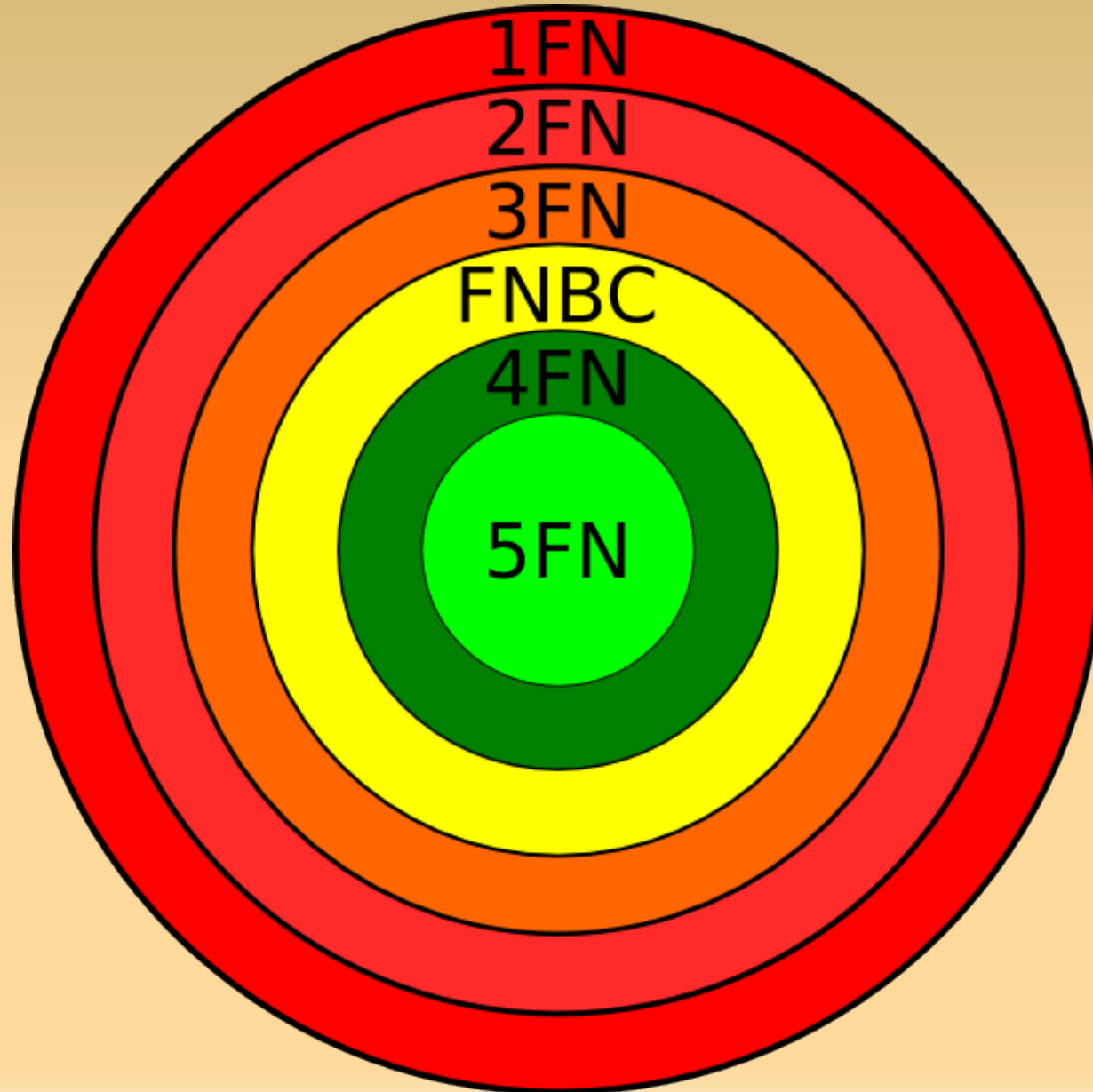


Universidad Autónoma
de Entre Ríos

Introducción

- En esta presentación vamos a ver como formalizar el diseño lógico de bases de datos relacionales.
- Para esto vamos a hacer uso de los conceptos ya dados como dependencia funcional, axiomas de Armstrong y conjunto irreducible de dependencias funcionales.
- En el proceso de normalización, empezaremos identificando la clave primaria de la relación.
- Luego construiremos el diagrama de dependencias funcionales y comenzaremos el proceso de análisis de las formas normales.

Distintas Formas Normales



Concepto Normalización

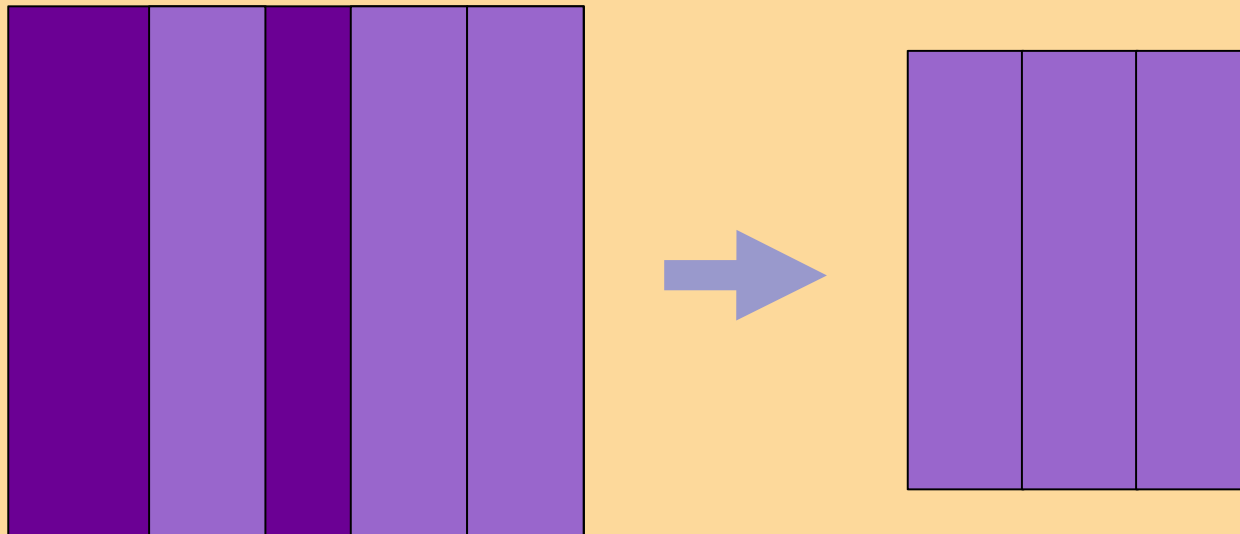
- Resulta poco práctico que en una base de datos tengamos una única relación R (relación universal), por la cantidad de redundancia y las anomalías ya vistas que debería controlar esta variable relacional.
- Entonces, en el proceso de normalización se trata de buscar un conjunto de variables relacionales R_i que sean equivalentes a R (relación universal). Para esto el conjunto R_i tienen que cumplir tres propiedades:
 - Conservación de la información.
 - Conservación de las dependencias.
 - Mínima redundancia de los datos.

Descomposición sin Pérdida

- La normalización se hace descomponiendo la relación original en un conjunto de proyecciones.
- La proyección (Π) es una operación del álgebra relacional.
- Si al recomponer las relaciones resultantes se obtiene nuevamente la relación original, esta descomposición se llama sin pérdida.
- Para recomponer se usa una operación del álgebra relacional llamada join.
- Para normalizar solo nos interesan la descomposición sin pérdida.

Proyección

- La proyección devuelve una relación que tiene como encabezado el conjunto de atributos indicado en la operación.
- El cuerpo de la relación será el conjunto de tuplas sin repetirse (para que siga siendo una relación).
- Se dice que la proyección opera sobre las columnas.



Ejemplo Proyección

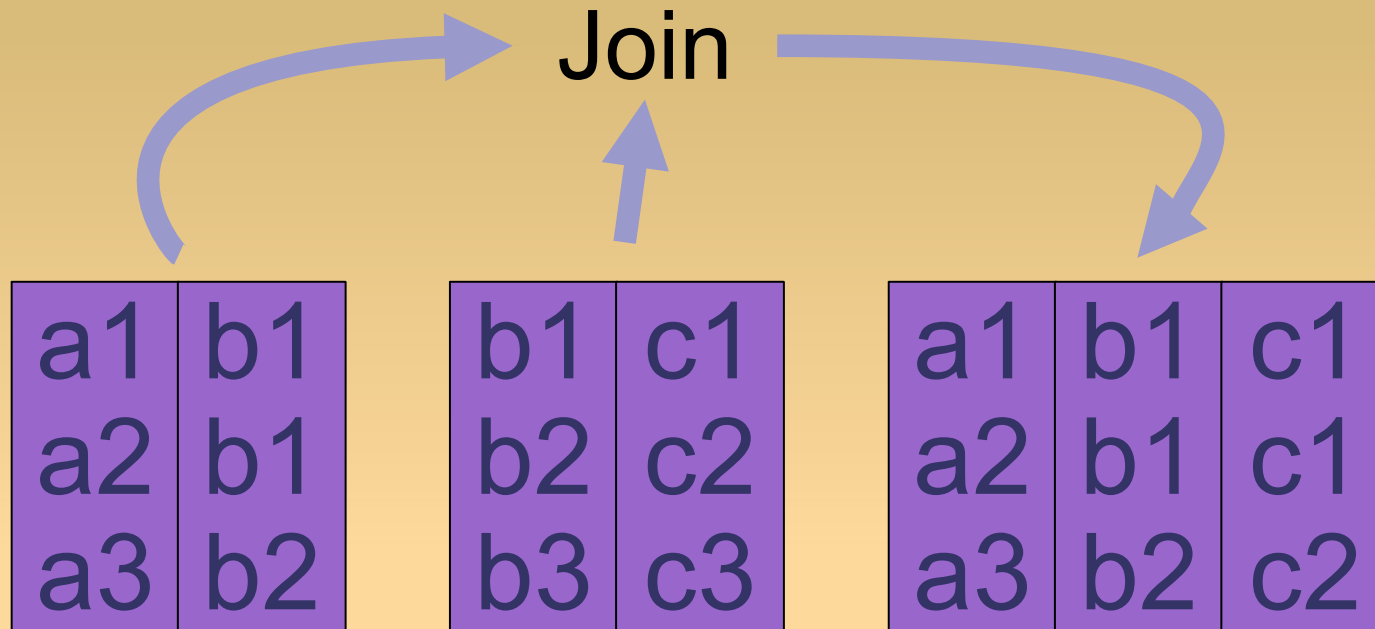
NOTA_ALUMNO				
DNI	Nombre	Telefono	Materia	Nota
20.000.000	Juan	44444	Algebra I	8
23.230.230	Pablo	55555	Algebra II	7
20.000.000	Juan	43333	Algebra II	7

- Dada la relación NOTA_ALUMNO, proyectar sobre los atributos DNI y Nombre.

$\Pi_{(DNI, Nombre)} \text{NOTA_ALUMNO}$

RESULTADO	
DNI	Nombre
20.000.000	Juan
23.230.230	Pablo

Equi Natural Join



Devuelve una relación que contiene todas las tuplas posibles que son combinación de dos tuplas provenientes de sendas relaciones, tal que las tuplas que contribuyen a una combinación dada tienen valores comunes para los atributos en común, y esos valores comunes aparecen una vez (y no dos) en las tuplas del resultado.

Ejemplo Equi Natural Join

NOMBRE_ALUMNO	
DNI	Nombre
20.000.000	Juan
23.230.230	Pablo

njoin

TEL_ALUMNO	
DNI	Telefono
20.000.000	44444
23.230.230	55555

=

ALUMNO		
DNI	Nombre	Telefono
20.000.000	Juan	44444
23.230.230	Pablo	55555

Ejemplo Descomposición sin Pérdida

Dado

ALUMNO		
DNI	Nombre	Telefono
20.000.000	Juan	44444
23.230.230	Pablo	55555

NOMBRE_ALUMNO	
DNI	Nombre
20.000.000	Juan
23.230.230	Pablo

TEL_ALUMNO	
DNI	Telefono
20.000.000	44444
23.230.230	55555

njoin

ALUMNO		
DNI	Nombre	Telefono
20.000.000	Juan	44444
23.230.230	Pablo	55555

Ejemplo Descomposición con Pérdida

Dado

ALUMNO		
DNI	Nombre	Telefono
20.000.000	Juan	44444
23.230.230	Pablo	55555

njoin

DNI_ALUMNO
DNI
20.000.000
23.230.230

NOMBRE_TEL	
Nombre	Telefono
Juan	44444
Pablo	55555

Tuplas
espúreas

RESULTADO		
DNI	Nombre	Telefono
20.000.000	Juan	44444
20.000.000	Pablo	55555
23.230.230	Juan	44444
23.230.230	Pablo	55555

¿Pero como descomponemos sin pérdida?

- La forma de descomposición debe hacerse siguiendo el principio de conservación de las dependencias.
- Pero ya vimos que es un procedimiento bastante tedioso y costoso hacer el cálculo de la cobertura de un conjunto de DF.
- Por eso es que Heath enunció un teorema mediante el cual es posible hacer proyecciones y conservar la dependencia.

Teorema de Heath:

Sea $R\{A, B, C\}$. Si R satisface $A \rightarrow B$, entonces R es igual al join de sus proyecciones sobre $\{A, B\}$ y $\{A, C\}$

Primera Forma Normal

Una relación esta en 1FN si y solo si, en cada valor válido de esa relación contiene exactamente un valor para cada atributo.

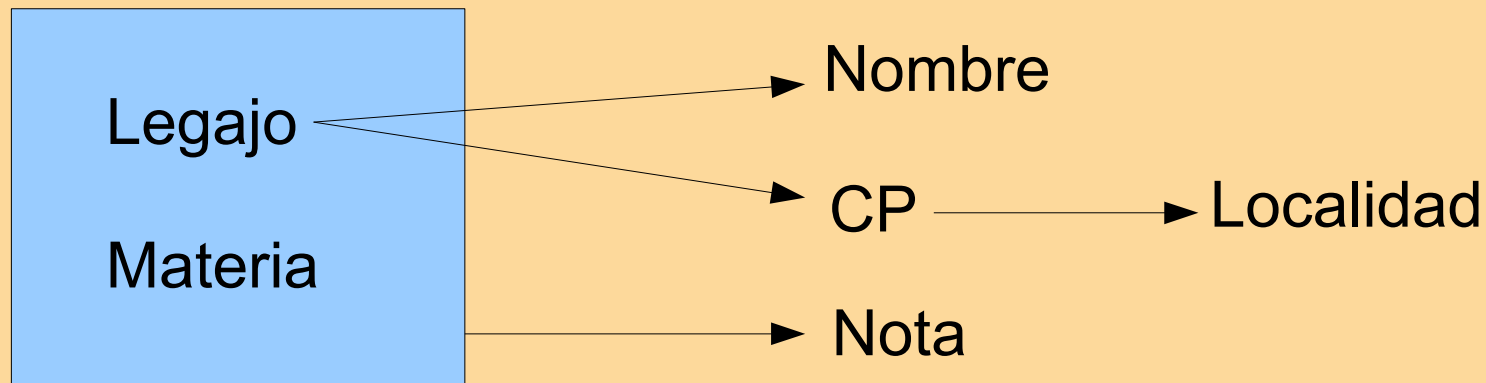
- Por lo tanto toda relación está en 1FN

Ejemplo

CURSA					
Legajo	Nombre	Materia	Nota	CP	Localidad
1	Brau	G. Datos	4	3260	C. del Uruguay
2	Zabalegui	G. Datos	7	3260	C. del Uruguay
3	Colombo	Álgebra	9	3280	Colón

Problemas 1FN

- Fácilmente podemos ver que la relación CURSA tiene redundancia y sufre de las anomalías de inserción, eliminación y modificación.
- Por eso vamos a construir el diagrama de DFs que debería coincidir con el conjunto irreducible de dependencias funcionales si lo construimos correctamente.



Segunda Forma Normal

Una relación esta en 2FN si y solo si, está en 1FN y todos los atributos que no sean clave dependen irreduciblemente de la/s clave/s candidata/s

- La relación cursa no está en 2FN, ya que hay atributos que no dependen irreduciblemente de la clave candidata.
- Estos atributos son Nombre, CP y Localidad.
- Vamos a aplicar el teorema de Heath para obtener 2 proyecciones llamadas CURSA2 y ALUMNO.

Aplicando Teorema de Heath

CURSA2		
Legajo	Materia	Nota
1	G. Datos	4
2	G. Datos	7
3	Álgebra	9

ALUMNO			
Legajo	Nombre	CP	Localidad
1	Brau	3260	C. del Uruguay
2	Zabalegui	3260	C. del Uruguay
3	Colombo	3280	Colón

- Podemos ver que CURSA2 está en 2FN ya que todo atributo no clave (Nota) depende irreduciblemente de la clave primaria (Legajo, Materia).
- De la misma manera ALUMNO está en 2FN ya que todo atributo no clave (CP y Localidad) dependen irreduciblemente de la clave primaria (Legajo).
- Gráficamente nos podemos dar cuenta que una relación no está en 2FN, porque salen flechas desde adentro de la/s clave/s candidatas.

¿Pero es un buen diseño llegar a 2FN?

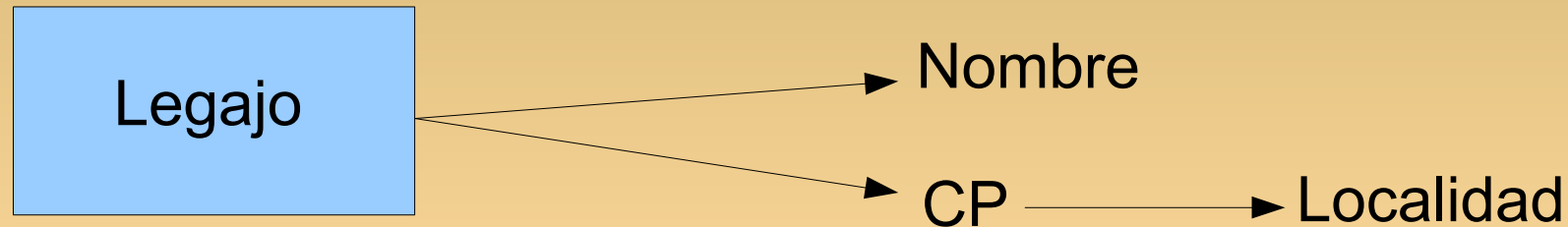
- ¿Que pasa si insertamos una tupla de otro alumno que cursa en Concepción del Uruguay?

ALUMNO			
Legajo	Nombre	CP	Localidad
1	Brau	3260	C. del Uruguay
2	Zabalegui	3260	C. del Uruguay
3	Colombo	3280	Colón
4	Gonzales	3260	Concepción del Uruguay

- Lo mismo sucede si actualizamos.
- O que no disponemos del nombre de una localidad hasta que no tenemos un alumno de dicha localidad.

Analizando el Diagrama de DFs

- Armamos el diagrama DFs de la relación ALUMNO



- Vimos que el problema se da con los valores del atributo Localidad.
- Gráficamente podemos ver que existe una flecha que no sale de la clave primaria. Pero todos los atributos dependen irreduciblemente de la clave.
- Gráficamente podemos ver que la solución es eliminar las flechas que no salen de la/s clave/s candidata/s (transitividades).

Tercera Forma Normal

Una relación esta en 3FN si y solo si, está en 2FN y todos los atributos que no sean clave dependen en forma no transitiva de la/s clave/s candidata/s

- Analizando la relación alumno, podemos ver que Localidad depende transitivamente de Legajo, por lo que ALUMNO no se encuentra en 3FN.

Normalizamos ALUMNO

- Aplicamos el Teorema de Heath sobre la DF problemática $CP \rightarrow Localidad$ y obtenemos dos relaciones ALUMNO2 y LOCALIDAD.

ALUMNO2		
Legajo	Nombre	CP
1	Brau	3260
2	Zabalegui	3260
3	Colombo	3280

LOCALIDAD	
CP	Localidad
3260	C. del Uruguay
3280	Colón

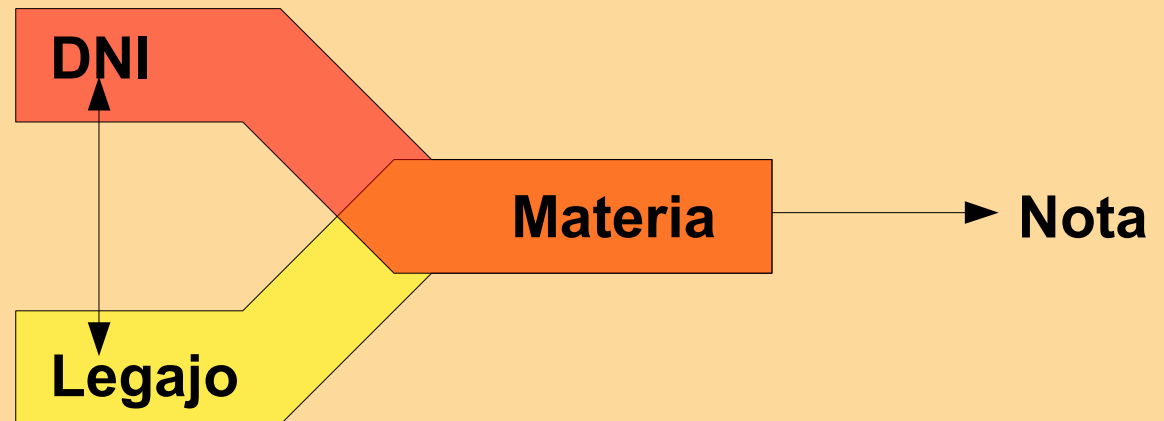
- Analizando ALUMNO2, está en 2FN y todos los atributos no clave dependen en forma no transitiva de la clave primaria Legajo, por lo tanto está en 3FN.
- Analizando LOCALIDAD, está en 2FN y todos los atributos no clave dependen en forma no transitiva de la clave primaria CP.

¿Se acabaron los problemas?

- Los ejemplos vistos hasta ahora no sufren mas las anomalías.
- Pero existen relaciones que siguen teniendo problemas, aunque se encuentren en 3FN.
- Esas relaciones poseen mas de una clave candidata, además esas claves son compuestas y además comparten al menos un atributo entre ellas.

NUEVACURSA			
Legajo	DNI	Materia	Nota
1	21	G. Datos	4
2	22	G. Datos	7
3	23	Álgebra	9

Diagrama de DFs



¿En que forma normal se encuentra NUEVACURSA?

- La relación NUEVACURSA está en 1FN.
- Todos los atributos que no son clave (Nota) dependen irreduciblemente de las claves candidatas {DNI, Materia} y {Legajo, Materia}, por lo tanto está en 2FN.
- Todos los atributos que no son clave (Nota) dependen en forma no transitiva de las claves candidatas, por lo tanto está en 3FN.

¿Entonces que problemas tiene NUEVACURSA?

- Si prestamos atención al diagrama de DFs vemos que hay una flecha entre DNI y Legajo.
- Porque conociendo el Legajo del alumno podemos conocer su DNI y conociendo el DNI podemos conocer el Legajo.
- ¿Pero qué sucede si el alumno con legajo 1 cursa otra materia?

NUEVACURSA			
Legajo	DNI	Materia	Nota
1	21	G. Datos	4
2	22	G. Datos	7
3	23	Álgebra	9
1	11	Álgebra	8

- Evidentemente existen problemas porque hay redundancia.

Forma Normal de Boyce-Codd

FNBC

- Boyce fue una de las personas que estudió estos problemas. Después de la publicación del artículo se redefinió la 3FN. Pero con el tiempo se la empezó a conocer como una forma por si misma.
- Generalmente en la jerga de los especialistas de sistemas se dice que la base de datos está normalizada cuando la llevaron a FNBC. Este término no es muy correcto porque una base de datos está normalizada con solo cumplir la 1FN.

Una relación esta en FNBC si y solo si toda DF no trivial, irreducible a la izquierda, tiene a una clave candidata como su determinante

Analizando NUEVACURSA por FNBC

- Vemos que la definición de FNBC es completa, no se basa en las anteriores formas normales para saber si una relación se encuentra en FNBC.
- En NUEVACURSA tenemos dos DFs no triviales que no dependen de las claves candidatas.
 - La DF $\text{DNI} \rightarrow \text{Legajo}$ no es trivial y es irreducible a izquierda. Si estuviera en FNBC el determinante debería ser la clave candidata {DNI, Materia}.
 - La DF $\text{Legajo} \rightarrow \text{DNI}$ no es trivial y es irreducible a izquierda. Si estuviera en FNBC el determinante debería ser la clave candidata {Legajo, Materia}.

Solucionando el problema de NUEVACURSA

- Para aplicar el Teorema de Heath, primero vamos a tener que seleccionar cual clave candidata va a ser nuestra **clave primaria**.
- Luego proyectamos y obtenemos dos nuevas relaciones, NUEVACURSA2 y ALUMNO2. Por ejemplo:

NUEVACURSA2		
Legajo	Materia	Nota
1	G. Datos	4
2	G. Datos	7
3	Álgebra	9

ALUMNO2	
Legajo	DNI
1	21
2	22
3	23

- NUEVACURSA2 está en FNBC porque todos sus DFs no triviales, irreducibles a izquierda tienen como determinante a la clave primaria (Legajo).
- ALUMNO2 está en FNBC porque su DF no trivial e irreducible a izquierda tiene como determinante a la clave primaria (Legajo).

Bibliografía

- Introducción a los Sistemas de Bases de Datos, C. J. Date, Séptima Edición
-