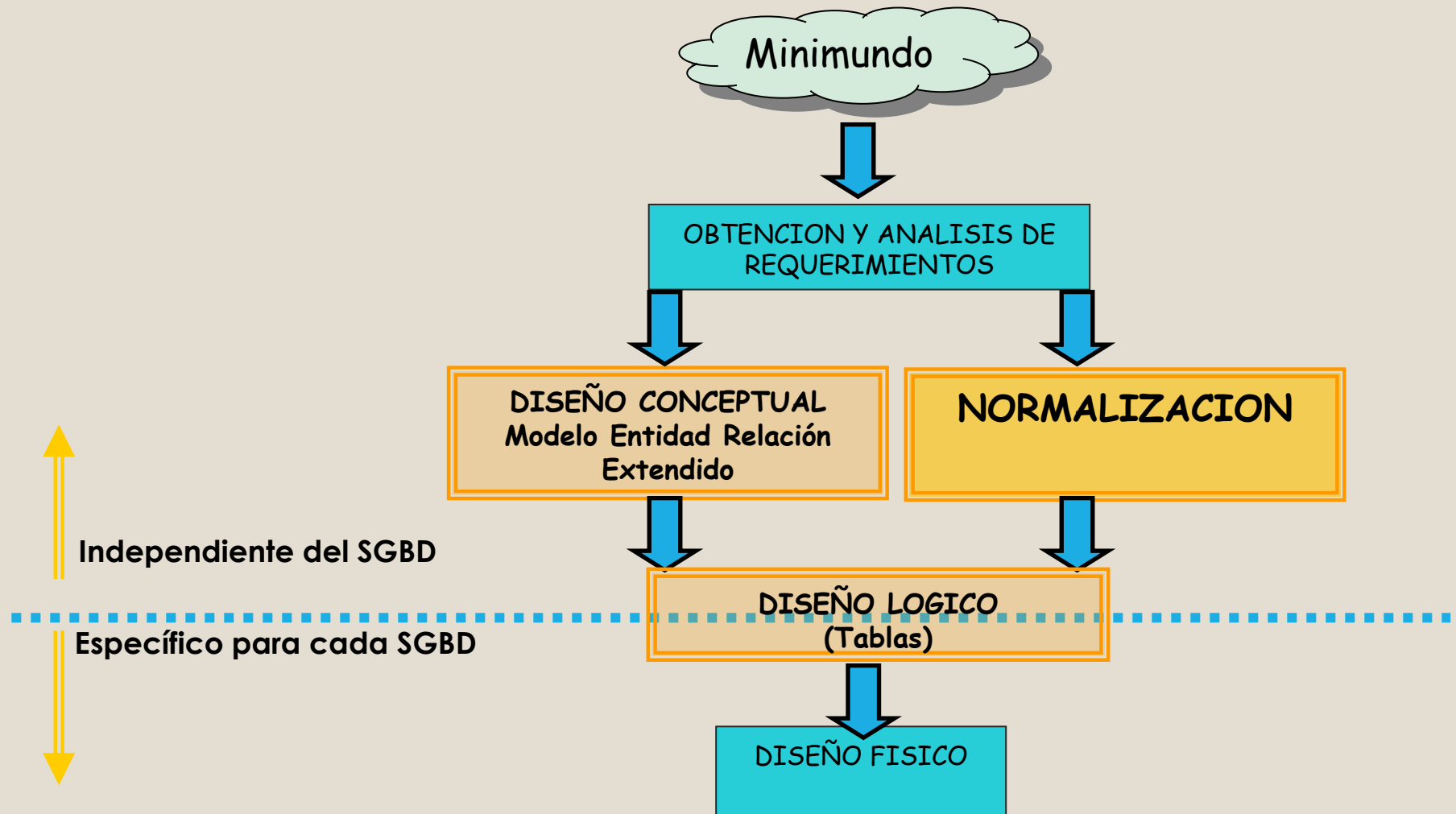




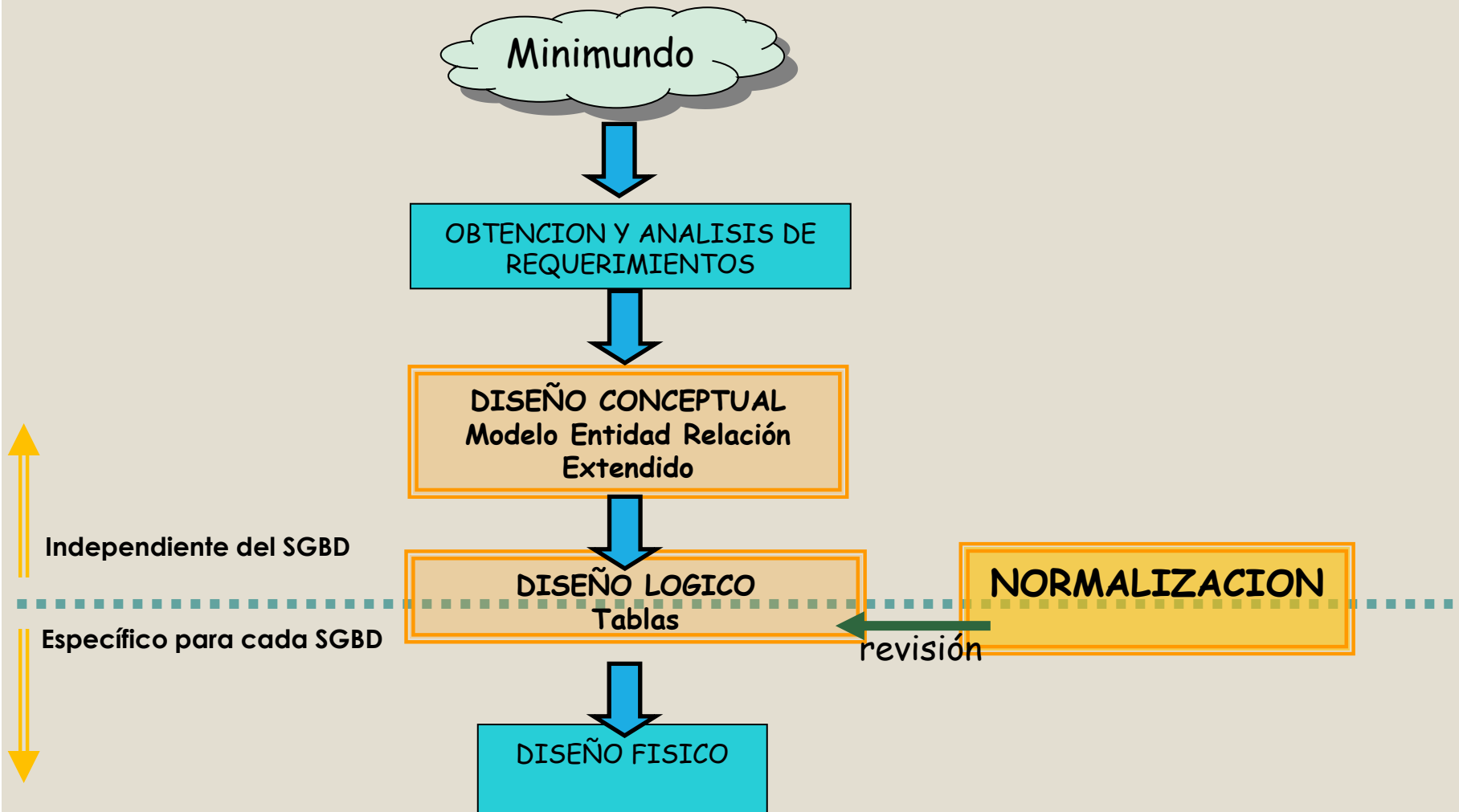
Unidad 3: Normalización

Diseño lógico de una
base de datos

Proceso de Construcción de una base de datos



¿Cómo utilizaremos nosotros la Normalización?



Objetivo de la Normalización

Elegir “buenas” estructuras de relaciones
(tablas)



permitiendo

Expresar formalmente las razones
por las que una agrupación de
atributos es mejor que otra

Aspecto fundamental a considerar a la hora de diseñar

Reducción de valores redundantes

Dentro de los *principales objetivos* en el diseño de relaciones



Minimizar el espacio de almacenamiento que ocupan las relaciones base (archivos)



Evitar anomalías de actualización

Reducción de valores redundantes

Cientes_Sucursal

DNI_Cli	Nombre_Cli	Tel_Cli	Direccion_Cli	Nro_Suc	Nombre_Suc	Direccion_Suc
17.343.232	Ana Perez	4223465	25 de Mayo 504 Este	1	Sacramento	Toranzo 350 Norte
7.432.567	Juan Flores	4223312	Av. Cordoba 107 Oeste	1	Sacramento	Toranzo 350 Norte
6.987.675	Sergio Alba	4221674	Aberastain 34 Sur	3	Espigas	Aberastain 333 Sur
14.878.234	Miguel Gomez	4340023	H. Yrigoyen 1171 Sur	5	Santa Rita	Av.Libertador 1230 Oeste
20.333.675	Silvia Gomez	4233495	Av. Libertador 3300 Oeste	3	Espigas	Aberastain 333 Sur
6.967.908	Javier Lima	4231100	H. Yrigoyen 11 Sur	6	Castillo	Ig. de la Roza 671

Restricción: *Una persona es cliente de sólo una sucursal.*

Clave Primaria = DNI__CLI

Analicemos:

- ✓ ¿Cuántas veces aparece la información correspondiente a cada sucursal?
- ✓ ¿Qué problemas puede acarrear?

Reducción de valores redundantes

Analicemos las inserciones:

- ✓ ¿Qué sucede si queremos ingresar una sucursal nueva que todavía no tiene clientes, por ejemplo la 2? ¿Es posible?

Nombre_Cli	Tel_Cli	Direccion_Cli	Nro_Suc	Nombre_Suc	Direccion_Suc
Ana Perez	4223465	25 de Mayo 504 Este	1	Sacramento	Toranzo 350 Norte
Juan Flores	4223312	Av. Cordoba 107 Oeste	1	Sacramento	Toranzo 350 Norte
Sergio Alba	4221674	Aberastain 34 Sur	3	Espigas	Aberastain 333 Sur
Miguel Gomez	4340023	H. Yrigoyen 1171 Sur	5	Santa Rita	Av.Libertador 1230 Oeste
Silvia Gomez	4233495	Av. Libertador 3300 Oeste	3	Espigas	Aberastain 333 Sur
Javier Lima	4231100	H. Yrigoyen 11 Sur	6	Castillo	Ig. de la Roza 671
null	null	null	2	Higueras	C.Cabot 3350 Oeste

Reducción de valores redundantes

Analicemos las supresiones:

✓ ¿Qué sucede si eliminamos el cliente Miguel Gómez?

DNI_Cli	Nombre_Cli	Tel_Cli	Direccion_Cli	Nro_Suc	Nombre_Suc	Direccion_Suc
17.343.232	Ana Perez	4223465	25 de Mayo 504 Este	1	Sacramento	Toranzo 350 Norte
7.432.567	Juan Flores	4223312	Av. Cordoba 107 Oeste	1	Sacramento	Toranzo 350 Norte
6.987.675	Sergio Alba	4221674	Aberastain 34 Sur	3	Espigas	Aberastain 333 Sur
14.878.234	Miguel Gomez	4340023	H. Yrigoyen 1171 Sur	5	Santa Rita	Av.Libertador 1230 Oeste
20.333.675	Silvia Gomez	4233495	Av. Libertador 3300 Oeste	3	Espigas	Aberastain 333 Sur
6.967.908	Javier Lima	4231100	H. Yrigoyen 11 Sur	6	Castillo	Ig. de la Roza 671

Reducción de valores redundantes

Analicemos las modificaciones:

- ✓ ¿Qué sucede si queremos registrar que la sucursal Espigas cambio su dirección?
 - De Aberastain 333 Sur, cambió a Estrada 122 Este

DNI_Cli	Nombre_Cli	Tel_Cli	Direccion_Cli	Nro_Suc	Nombre_Suc	Direccion_Suc
17.343.232	Ana Perez	4223465	25 de Mayo 504 Este	1	Sacramento	Toranzo 350 Norte
7.432.567	Juan Flores	4223312	Av. Cordoba 107 Oeste	1	Sacramento	Toranzo 350 Norte
6.987.675	Sergio Alba	4221674	Aberastain 34 Sur	3	Espigas	Estrada 122 Este
14.878.234	Miguel Gomez	4340023	H. Yrigoyen 1171 Sur	5	Santa Rita	Av.Libertador 1230 Oeste
20.333.675	Silvia Gomez	4233495	Av. Libertador 3300 Oeste	3	Espigas	Estrada 122 Este
6.967.908	Javier Lima	4231100	H. Yrigoyen 11 Sur	6	Castillo	Ig. de la Roza 671

Reducción de valores redundantes

Analicemos los mismos interrogantes con las siguientes relaciones:

Sucursales

Nro_Suc	Nombre	Direccion
1	Sacramento	Toranzo 350 Norte
2	Higueras	C.Cabot 3350 Oeste
3	Espigas	Aberastain 333 Sur
5	Santa Rita	Av. Libertador 1230 Oeste
6	Castillo	Ig. de la Roza 671

Clientes

Nombre_Cli	Tel_Cli	Direccion_Cli	Nro_Suc
Ana Perez	4223465	25 de Mayo 504 Este	1
Juan Flores	4223312	Av. Cordoba 107 Oeste	1
Sergio Alba	4221674	Aberastain 34 Sur	3
Miguel Gomez	4340023	H. Yrigoyen 1171 Sur	5
Silvia Gomez	4233495	Av. Libertador 3300 Oeste	3
Javier Lima	4231100	H. Yrigoyen 11 Sur	6

- ✓ ¿Qué sucede si queremos **ingresar una sucursal nueva** que todavía no tiene clientes, por ejemplo la 2?
- ✓ ¿Qué sucede si **eliminamos el cliente Miguel Gómez**?
- ✓ ¿Qué sucede si queremos registrar que la sucursal **Espigas cambio su dirección**?

Reducción de valores redundantes

El diseño de las relaciones (o tablas) de manera de evitar las anomalías de inserción, eliminación y modificación en las relaciones



**EVITAR LA INCONSISTENCIA
DE LA BASE DE DATOS**

Vamos ahora a algunos
conceptos necesarios
antes de entrar en las
Formas Normales...



NOTA: En esta unidad mencionaremos RELACIONES, asociándolas al **mismo concepto** de TABLAS. Sin embargo, estrictamente hablando, no son iguales.

Diseño
lógico
de una
base de
datos

Dependencia Funcional

$$X \rightarrow Y$$

X determina funcionalmente a Y, o dicho en el otro sentido,
Y depende funcionalmente de X, si:

- ✓ Para cada valor de X le corresponde **sólo un valor de Y** en todo estado de R
... o expresado de otra manera
- ✓ Si para 2 tuplas cualesquiera $t1$ y $t2$ de R tales que:
 $t1(X) = t2(X)$ entonces $t1(Y) = t2(Y)$ en todo estado de R

Dependencia Funcional

Analicemos la siguiente relación (mostrada como tabla)

1. ¿ $X \rightarrow Y$?

2. ¿ $M \rightarrow X$?

A	X	M	Y
100	a	12	100
11	b	120	100
13	a	12	104

Nota: Para asegurar una DF debemos conocer la semántica de los atributos, no en base a un estado de la relación. Esto es sólo un ejercicio.

Clave Candidata

Dada una relación R , un subconjunto de atributos \mathbf{K} de R **es clave candidata o simplemente clave** si cumple:

- ✓ **Unicidad**: No existe un par de tuplas que tengan el mismo valor para \mathbf{K} , es decir, es superclave.
- ✓ **Minimalidad**: Si se quita algún atributo de la misma deja de cumplir la unicidad.

Entonces, sea $\mathbf{K} = \{A_1, A_2, \dots, A_k\}$, si \mathbf{K} es clave entonces $\mathbf{K} - \{A_i\}$ no es clave para $1 \leq i \leq k$

Clave Candidata

Analicemos algunas relaciones, identifiquemos CC para cada una:

1. Facturas (El nro. de facturas es único para la empresa)

NroFac	FechaFac	Cliente	Sucursal
100	12/01/2023	19.555.987	1
101	02/02/2023	45.555.876	1
102	02/02/2023	20.555.987	2
103	03/02/2023	23.555.987	3

2. Facturas (El nro. de facturas no es único dentro de la empresa, es único por cada sucursal)

NroFac	FechaFac	Cliente	Sucursal
100	12/01/2023	19.555.987	1
101	02/02/2023	45.555.876	1
100	02/02/2023	20.555.987	2
100	03/02/2023	23.555.987	3

Claves

Una relación puede contener más de una **clave candidata**:

- ✓ Se elige una como **CLAVE PRIMARIA**
- ✓ A las restantes, se las denomina **CLAVES SECUNDARIAS o ALTERNATIVAS**

Ejemplo:

Vehiculo = {*nroPat, nroChasis, modelo, marca, ...*}

- 1) *Clave Candidata = nroPat*
- 2) *Clave Candidata = nroChasis*

Elegimos la 1) como clave primaria (nroPat). Asumiendo que en ese minimundo no se van a presentar vehículos 0km

Claves

¿Qué consideraciones tener en cuenta para elegir la clave primaria cuando existe más de una candidata?

Algunas:

- ✓ Tamaño (afecta al tamaño de la base de datos en su conjunto.
¿Por qué?
- ✓ Presencia del valor en todas las filas o tuplas
- ✓ Atributo más usado/común en el minimundo

Atributo Primo

Un atributo de R se denomina **atributo primo** de R si es miembro de alguna clave de R .



*“Un atributo de R se denomina **atributo no primo** de R si no es atributo primo”*

¿Repasamos los conceptos que vimos hasta el momento?

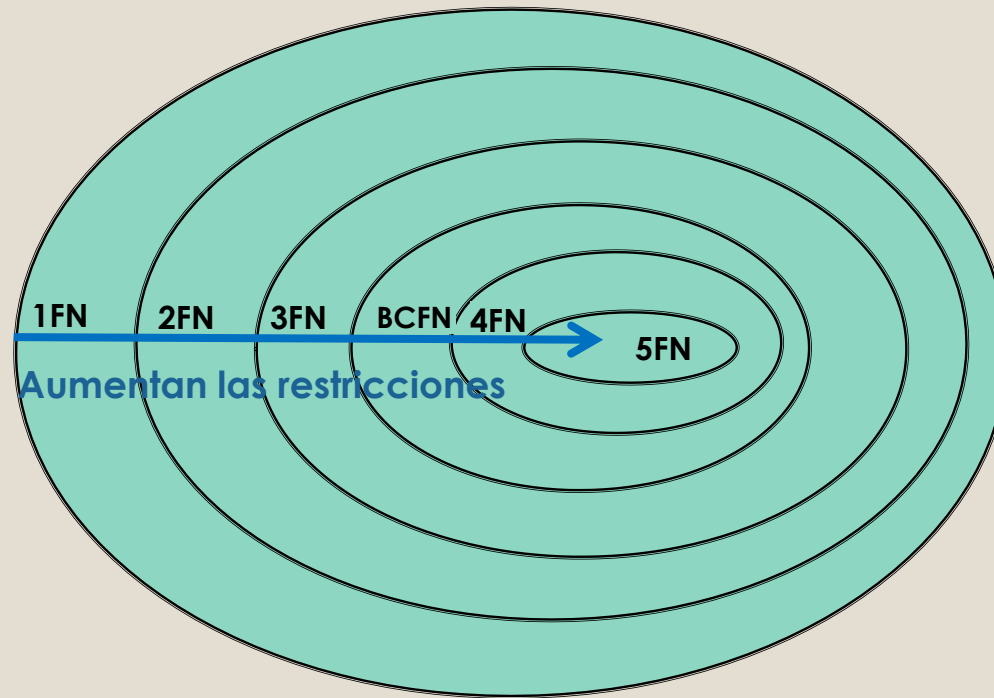
- ✓ **Redundancia**
- ✓ **Inconsistencia**
- ✓ **Anomalías de actualización**
- ✓ **Claves candidatas**
- ✓ **Claves primarias**
- ✓ **Atributos primos**
- ✓ **Atributos no primos**
- ✓ **Dependencia funcional**



Nos
relajamos
20 minutos?



Normalización (Formas Normales)



Primera Forma Normal

Una relación R está en 1FN si los dominios de todos sus atributos son atómicos



Toda relación R está en 1FN, por definición

Segunda Forma Normal

Se basa en el concepto de dependencia
funcional completa o total:

- ✓ Una dependencia funcional $X \rightarrow Y$ es **completa** si la eliminación de cualquier atributo A de X hace que la dependencia no sea válida. Es decir, **X-A** no determina funcionalmente a Y.
- ✓ Una dependencia funcional $X \rightarrow Y$ es **parcial** si es posible eliminar algún atributo A de X y la dependencia sigue siendo válida.

Analicemos: NroPat + Modelo \rightarrow Marca?

Segunda Forma Normal

Una relación **R está en 2FN** si todo atributo no primo A de R depende funcionalmente en forma completa de la clave de R.

Analicemos el siguiente ejemplo:

ventas = (nroFac, nroProd, nombreProd, rubro, precioActual, cantVen, precioVta, fechaFac)

1. ¿Dependencias Funcionales?
2. ¿Claves candidatas y primaria?

Segunda Forma Normal

ventas = {nroFac, nroProd, nombreProd, rubro, precioActual, cantVen, precioVta, fechaFac}

Dependencias Funcionales

DF1: nroFac -> fechaFac

DF2: nroProd -> nombreProd, rubro, precioActual

DF3: nroFac+nroProd -> cantVen, precioVta

Claves candidatas y primaria

cc = nroFac+nroProd = cp

Recordando... **R está en 2FN** si todo atributo no primo A de R depende funcionalmente en forma completa de la clave de R

Atributos no primos = fechaFac, nombreProd, rubro, precioActual, cantVen, precioVta

La relación **compra no está en 2FN** puesto que en DF1 y DF2 se verifica que existen atributos no primos que dependen en forma parcial de la clave

Segunda Forma Normal

Solución:

Si una relación no está en 2FN, descomponerla y crear una nueva relación para cada clave parcial con su/s atributo/s dependiente/s.

Ventas = (nroFac, nroProd, nombreProd, rubro, precioActual, cantVen, precioVta, fechaFac)

Ventas1 = {nroFac, fechaFac}

Ventas2 = {nroProd, nombreProd, rubro, precioActual}

Ventas3 = { nroFac, nroProd, cantVen, precioVta}

NOTA: El join (unión) de las 3 relaciones debe generar la relación original.

Descomposiciones válidas

Toda descomposición debe cumplir una restricción
(no sólo para la 2FN, sino para todas)



No generar pérdida de información, es decir, el join (lo veremos mas adelante) de las proyecciones genera la relación original



Inclusive, provocan ganancia de información.

Permiten registrar información que en la relación original era imposible. *(Recordar lo explicado en problemas de actualización, al inicio de la clase)*

Tercera Forma Normal

- ✓ Una relación R está en 3FN si:
 - ✓ Está en 2FN
 - ✓ Los **atributos no primos de R son mutuamente independientes**, es decir, **no existen dependencias funcionales entre ellos**.

Consecuentemente, también puede expresarse:

- ✓ Está en 2FN
- ✓ **Ningún atributo no primo de R depende en forma transitiva de la clave.**

Tercera Forma Normal

Analicemos el siguiente ejemplo:

producto = (nroProd, nombreProd, rubro, cuitProveedor *, direProveedor)

***Nota:** Cada producto es provisto solamente por un proveedor

- **Dependencias Funcionales**

DF1: cuitProveedor -> direProveedor

DF2: nroProd -> nombreProd, rubro, cuitProveedor

- **Claves candidatas**

cc = nroProd

La relación **producto no está en 3FN** puesto DF1 muestra que los atributos no primos cuitProveedor y direProveedor no son mutuamente independientes

Tercera Forma Normal

Solución:

Si una relación no está en 3FN, descomponerla creando otra relación que contenga el/los atributo/s no claves que determinen funcionalmente a otro/s atributo/s no clave.

Nota: “Tener cuidado con las posibles descomposiciones”

Tercera Forma Normal

producto = {nroProd, nombreProd, rubro, cuitProveedor, direProveedor}

Dependencias Funcionales y claves

DF1: cuitProveedor -> direProveedor

DF2: nroProd -> nombreProd, rubro, cuitProveedor

cc = nroProd

Solución1:

Producto1 = {nroProd, nombreProd, rubro, cuitProveedor}

Producto2 = {cuitProveedor, direProveedor}

Solución2 :

Producto1 = {nroProd, nombreProd, rubro, cuitProveedor}

Producto2 = {nroProd, direProveedor}

Analicemos las 2 alternativas, ¿cuál será mejor? ¿ por qué?

Normalización

Bibliografía:

- Date - Introducción a los Sistemas de Bases de Datos
- Elmasri - Fundamentos de Bases de Datos

¿Dudas?



N
O
R
M
A
L
I
Z
A
C
I
Ó
N

Forma Normal de Boyce y Codd

Una relación R está en FNBC si todo determinante es clave candidata

Nota:

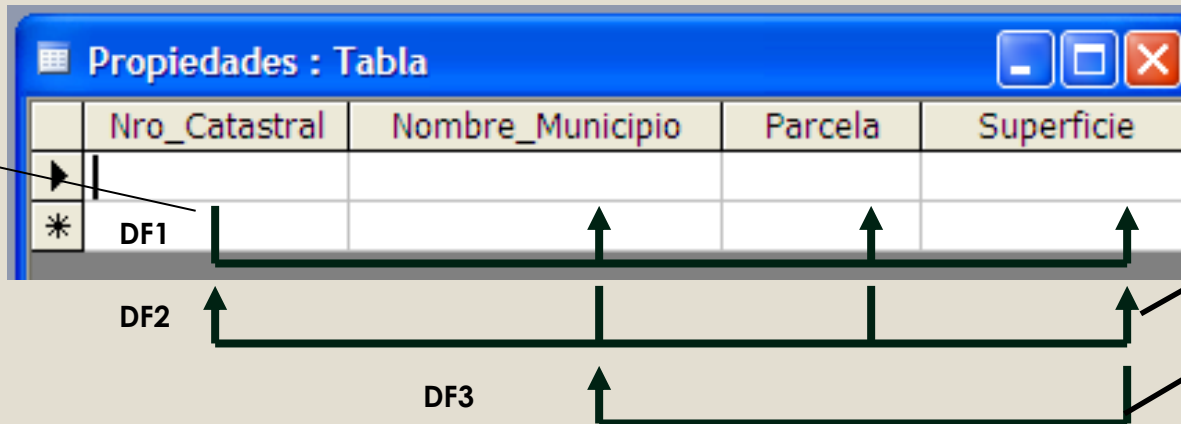
- No refiere a una FN anterior



Forma Normal de Boyce y Codd

Veamos un ejemplo:

Cada parcela es identificada por su nro. catastral



Cada parcela es identificada dentro de un municipio, por su nro. de parcela

Las sup. de las parcelas de cada municipio son diferentes, pero, no existe una sup. que corresponda a más de un municipio.

Determinantes:

- Nro_Catastral
- Nombre_Municipio+Parcela

• Superficie

No es clave candidata

Claves candidatas:


- Nro_Catastral
- Nombre_Municipio+Parcela

~~FNBC~~

Forma Normal de Boyce y Codd

Al descomponer:

	Nro_Catastral	Superficie	Parcela



Determinantes:

- Nro_Catastral


Claves candidatas:

- Nro_Catastral



BCNF

	Superficie	Nombre_Municipio



Determinantes:

- Superficie

Claves candidatas:

- Superficie



BCNF