

Normalización

Explicación de Práctica
BBDD1 - 2024

Inicialmente, algunas definiciones...



Dependencia funcional trivial

Es una DF de la forma:

$$A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$$

tal que:

$$\{B_1, B_2, \dots, B_m\} \subseteq \{A_1, A_2, \dots, A_n\}$$

Clave de una relación

Los atributos $\{A_1, A_2, \dots, A_n\}$ son la **clave** de una relación R si cumplen:

- $\{A_1, A_2, \dots, A_n\}$ determinan funcionalmente a todos los restantes atributos de la relación R .
- No existe un subconjunto de $\{A_1, A_2, \dots, A_n\}$ que determine funcionalmente a todos los atributos de R .

Esto implica que una clave es un **conjunto minimal**.

Claves candidatas

En caso de existir dos o más conjuntos de atributos $\{A1, A2, \dots, An\}$, $\{B1, B2, \dots, Bk\}, \dots$, $\{N1, N2, \dots, Nm\}$ en una relación R tales que:

- $\{A1, A2, \dots, An\}$ determinan funcionalmente a todos los restantes atributos de la relación R.
- $\{B1, B2, \dots, Bk\}, \dots$ y $\{N1, N2, \dots, Nm\}$ también por sí mismos determinan al resto de los atributos de R.
- No existe un subconjunto de $\{A1, A2, \dots, An\}$ o $\{B1, B2, \dots, Bk\}, \dots$ o $\{N1, N2, \dots, Nm\}$ que determine funcionalmente a todos los atributos de R.

Entonces $\{A1, A2, \dots, An\}, \{B1, B2, \dots, Bk\}, \dots \{N1, N2, \dots, Nm\}$ son **CLAVES CANDIDATAS** para la relación R.

Superclave

“Superconjunto” de una clave.

Los atributos $\{A_1, A_2, \dots, A_n\}$ son la **superclave** de una relación R si cumplen:

- $\{A_1, A_2, \dots, A_n\}$ determinan funcionalmente a todos los restantes atributos de la relación R .
 - Notar que:
 - Una clave está contenida en una superclave.
 - Una superclave no necesariamente es minimal (como lo es la clave por la segunda condición de su definición).
-

BCNF (Forma Normal de Boyce y Codd)

Un esquema de relación está en BCNF si, siempre que una dependencia funcional de la forma $X \rightarrow A$ es válida en R , entonces se cumple que:

- X es superclave de R
o bien
 - $X \rightarrow A$ es una dependencia funcional trivial
-

Proceso de Normalización

1. Obtener todas las dependencias funcionales no triviales válidas sobre el esquema.
2. Obtener la(s) clave(s) candidata(s).
3. Nos preguntamos si el esquema cumple con la definición de BCNF.

Para toda dependencia funcional válida en el esquema se cumple que:

- X es superclave de R , o bien
- $X \rightarrow A$ es una dependencia funcional trivial

4. Si no cumple justificamos el motivo y particionamos el esquema. Si cumple, solamente justificamos.
-

Ejercicio: Instalaciones

INSTALACIONES (idCuidador, nyAp, cuilCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, nombreEspecie, quimicoPlanta, consultorVivero, patente_vehiculo)

- El idVivero es un identificador único que no se repite para diferentes viveros. Del vivero se conoce su nombre (diferentes viveros pueden tener el mismo nombre), los metros cuadrados que ocupa, la temperatura promedio que debe mantener y el cuidador responsable del mismo.
 - Un mismo cuidador (idCuidador) puede cuidar diversos viveros. Tener en cuenta que un vivero tiene solamente un cuidador responsable asignado.
 - Del cuidador se conoce su nombre y apellido y el cuil. El idCuidador no se repite para diferentes cuidadores.
 - El idPlanta es único. Por ejemplo, una planta es el helecho. De cada planta se conoce el nombre de la planta y la especie a la que pertenece
-

Ejercicio: Instalaciones -cont.-

INSTALACIONES (idCuidador, nyAp, cuilCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, nombreEspecie, quimicoPlanta, consultorVivero, patente_vehiculo)

- El idEspecie es único. Un ejemplo de especie es el árbol. Cada planta pertenece a una única especie y a una especie pertenecen diversas plantas.
 - A cada planta en un vivero (por ejemplo: helecho en el vivero 1) se le aplica un conjunto de químicos. Los mismos químicos se pueden aplicar a plantas de diferentes viveros y a diferentes plantas en el mismo vivero.
 - Se cuenta con el registro de los vehículos de los viveros que se utilizan para traslados y repartos.
 - Cada vivero tiene diversos consultores de viveros (consultorVivero), que son quienes asesoran ante dudas eventuales. El mismo consultor puede asesorar en diversos viveros.
-

Resolución

Dependencias funcionales:

1. idCuidador -> nyAp, cuilCuidador
2. cuilCuidador -> nyAp, idCuidador
3. idVivero -> nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idCuidador
4. idVivero -> nombreVivero, mtrCuadradosVivero, tempPromedioVivero, cuilCuidador
5. idEspecie -> nombreEspecie
6. idPlanta -> nombrePlanta, idEspecie

CC: {idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero}

Resolución

Nos preguntamos: INSTALACIONES cumple con la definición de BCNF?

NO, dado que existe al menos el determinante de la df5 que NO es superclave del esquema INSTALACIONES.

Por lo tanto, particionamos por la DF5, creando dos nuevas relaciones:

- **I1** (idEspecie, nombreEspecie)
- **I2** (idCuidador, nyAp, cuilCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, quimicoPlanta, patente_vehiculo, consultorVivero)

I1 está en BCNF ya que {idEspecie} es superclave del esquema y sólo vale la DF5 en el esquema.

No se pierde información ya que I1 ∪ I2 es {idEspecie} clave de I1

En I2 valen DF1, DF2, DF3, DF4, DF6.

I2 no está en BCNF ya que existe, al menos, {idCuidador} de la DF1 que NO es superclave del esquema. Por lo tanto, particionamos I2 por la DF1.

- **I3** (idCuidador, nyAp, cuilCuidador)
- **I4** (idCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, quimicoPlanta, patente_vehiculo, consultorVivero)

En I3 valen 2 DFs: DF1 y DF2. I3 está en BCNF ya que tanto {idCuidador} como {cuilCuidador} son superclaves del esquema.

No se pierde información ya que $I3 \cap I4$ es {idCuidador}, clave en I3.

Dependencias funcionales equivalentes

Decimos que dos dependencias funcionales son equivalentes si producen el mismo resultado, independientemente del conjunto de atributos que se utilice como determinante.

La DF1 y la DF2 son equivalentes, ya que producen el mismo conjunto de atributos.

1. idCuidador \rightarrow nyAp, cuilCuidador
2. cuilCuidador \rightarrow nyAp, idCuidador

Lo mismo sucede con DF3 y DF4.

- **I4** (idCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, quimicoPlanta, patente_vehiculo, consultorVivero)

En I4 valen DF3 y DF6. I4 no está en BCNF ya que {idVivero} de la DF3 que NO es superclave del esquema.

Qué pasó con la DF4?

DF4: idVivero -> nombreVivero, mtrCuadradosVivero, tempPromedioVivero, cuilCuidador

Se perdió?

La DF4 NO se pierde, ya que es posible acceder con su determinante a todos sus determinados, ya sea directa o indirectamente.

- *nombreVivero*, *mtrCuadradosVivero*, *tempPromedioVivero* pueden accederse directamente ya que están en I4
- *cuilCuidador* no está en I4, sin embargo la dependencia funcional no se pierde ya que existe una relación indirecta a través de la dependencia funcional *idVivero* -> *idCuidador* y la dependencia *idCuidador* -> *cuilCuidador*.

En este caso, esto puede demostrarse mediante la utilización del algoritmo, o bien, explicarse adecuadamente.

La DF4 no quedó válida pero no se pierde.

Entonces, en I4 valen las DF3 y la DF6.

I4 no está en BCNF ya que existe, al menos, {idPlanta} de la DF6 que NO es superclave del esquema. Por lo tanto, particionamos I4 por la DF6.

- **I5** (idPlanta, nombrePlanta, idEspecie)
- **I6** (idCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

En I5 vale la DF6. I5 está en BCNF ya que {idPlanta} es superclave del esquema.

No se pierde información ya que $I5 \cap I6$ es {idPlanta}, clave en I5.

En I6 valen la DF3

I6 no está en BCNF ya que existe, {idVivero} de la DF3 que NO es superclave del esquema. Por lo tanto, particionamos I6 por la DF3.

- **I7** (idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idCuidador)
- **I8** (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

En I7 vale la DF3. I7 está en BCNF ya que {idVivero} es superclave del esquema.

No se pierde información ya que $I7 \cap I8$ es {idPlanta}, clave en I7.

Notar que en I8 quedaron los atributos que conforman la clave.

Particiones en BCNF

- **I1** (idEspecie, nombreEspecie)
- **I3** (idCuidador, nyAp, cuilCuidador)
- **I5** (idPlanta, nombrePlanta, idEspecie)
- **I7** (idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idCuidador)
- **I8** (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

Clave: (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

Dependencia multivaluada

$X \twoheadrightarrow Y$

si dado un valor de X, hay un conjunto de valores de Y asociados y este conjunto de valores de Y NO está relacionado (ni funcional ni multifuncionalmente) con los valores de R - X -Y (donde R es el esquema).

Es decir Y es **independiente** de los atributos de R-X-Y.

Dependencia multivaluada trivial

Sea R un esquema de relación, una dependencia multivaluada de la forma

$$X \twoheadrightarrow Y,$$

que vale en R es trivial si:

el conjunto de atributos $X \cup Y$ son todos los atributos del esquema R .

4FN (Cuarta forma normal)

Un esquema R está en 4NF con respecto a un conjunto de dependencias multivaluadas D,

si para toda dependencia multivaluada de la forma

$X \twoheadrightarrow Y$

se cumple que:

$X \twoheadrightarrow Y$ es una dependencia multivaluada trivial en R

o bien, no tiene DMs.

RETOMAMOS el ejercicio: Instalaciones

INSTALACIONES (idCuidador, nyAp, cuilCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, nombreEspecie, quimicoPlanta, consultorVivero, patente_vehiculo)

- El idVivero es un identificador único que no se repite para diferentes viveros. Del vivero se conoce su nombre (diferentes viveros pueden tener el mismo nombre), los metros cuadrados que ocupa, la temperatura promedio que debe mantener y el cuidador responsable del mismo.
 - Un mismo cuidador (idCuidador) puede cuidar diversos viveros. Tener en cuenta que un vivero tiene solamente un cuidador responsable asignado.
 - Del cuidador se conoce su nombre y apellido y el cuil. El idCuidador no se repite para diferentes cuidadores.
 - El idPlanta es único. Por ejemplo, una planta es el helecho. De cada planta se conoce el nombre de la planta y la especie a la que pertenece
-

RETOMAMOS el ejercicio: Instalaciones -cont.-

INSTALACIONES (idCuidador, nyAp, cuilCuidador, idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idPlanta, nombrePlanta, idEspecie, nombreEspecie, quimicoPlanta, consultorVivero, patente_vehiculo)

- El idEspecie es único. Un ejemplo de especie es el árbol. Cada planta pertenece a una única especie y a una especie pertenecen diversas plantas.
 - A cada planta en un vivero (por ejemplo: helecho en el vivero 1) se le aplica un conjunto de químicos. Los mismos químicos se pueden aplicar a plantas de diferentes viveros y a diferentes plantas en el mismo vivero.
 - Se cuenta con el registro de los vehículos de los viveros que se utilizan para traslados y repartos.
 - Cada vivero tiene diversos consultores de viveros (consultorVivero), que son quienes asesoran ante dudas eventuales. El mismo consultor puede asesorar en diversos viveros.
-

Dependencias multivaluadas

I8 (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

DM1) idVivero, idPlanta ->> quimicoPlanta

DM2) idVivero ->> consultorVivero

DM3) { } ->> patente_vehiculo

I8 (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

Teniendo en cuenta DM3:

- I9 (patente_vehiculo) **4FN**
- I10 (idVivero, idPlanta, quimicoPlanta, consultorVivero)

I9 está en 4NF porque no valen DMs que no sean triviales en ella.

I10 no está en 4NF porque DM2 vale en ella y no es trivial.

I10 (idVivero, idPlanta, quimicoPlanta, consultorVivero)

Considerando DM2:

- I11 (idVivero, consultorVivero) **4FN**
- I12 (idVivero, idPlanta, quimicoPlanta) **4FN**

Tanto I11 como I12 están en 4FN ya que no existen dependencias multivaluadas que no sean triviales en ellas.

Esquemas en 4FN

- I9 (patente_vehiculo)
- I11 (idVivero, consultorVivero)
- I12 (idVivero, idPlanta, quimicoPlanta)

Se llevó el esquema a 4ta forma normal.

Existen dependencias multivaluadas en los esquemas I1 a I7 ?

Esquemas resultantes en 4FN y que no son proyecciones de otros

- **I1** (idEspecie, nombreEspecie)
 - **I3** (idCuidador, nyAp, cuilCuidador)
 - **I5** (idPlanta, nombrePlanta, idEspecie)
 - **I7** (idVivero, nombreVivero, mtrCuadradosVivero, tempPromedioVivero, idCuidador)
 - **I9** (patente_vehiculo)
 - **I11** (idVivero, consultorVivero)
 - **I12** (idVivero, idPlanta, quimicoPlanta)
-

Orden de las dependencias multivaluadas

Al trabajar con dependencias funcionales el orden de selección es de vital importancia, ya que elegir un orden incorrecto puede provocar que perdamos DFs.

- En ese caso debemos volver atrás y particionar tomando otra DF.

Cuando trabajamos con dependencias multivaluadas el orden no hará que perdamos DMs pero sí hará que aparezcan nuevas.

- Es importante que en todo esquema existan DMs que cubran **todos** sus atributos.



I8 (idVivero, idPlanta, quimicoPlanta, patente_vehiculo, consultorVivero)

I8 no está en 4NF porque DM1 vale en ella y no es trivial. Entonces, particiono por DM1.

- I9 (idVivero, idPlanta, químicoPlanta)
- I10 (idVivero, idPlanta, patente_vehiculo, consultorVivero)

I9 está en 4NF porque no valen DMs que no sean triviales en ella.

I10 no está en 4NF porque DM2 vale en ella y no es trivial.

Considerando DM2:

- I11 (idVivero, consultorVivero)
- I12 (idVivero, idPlanta, patente_vehiculo)

Considerando DM3:

- I11 (patente_vehiculo)
 - I12 (idVivero, idPlanta) - **Qué pasa con I12? Está en 4FN?**
-

I12 no está en 4FN. Vemos que aparece una nueva DM:

DM4) idVivero ->> idPlanta

Si tomo esta DM para particionar I12 obtengo:

- I13 (idVivero, idPlanta) - Proyección de I9

Notar que I13 no es más que una proyección de I9 generado anteriormente, por lo que puede eliminarse del resultado. SIN EMBARGO ES NECESARIO INCLUIR ESTO EN EL PROCESO.
