

Com normalitzar una base de dades i no morir en l'intent

El procés de normalització d'una base de dades relacional consisteix a aplicar una sèrie de regles per evitar en un futur realitzar *queries*, o consultes innecessàriament complexes. En altres paraules estan enfocades a eliminar redundàncies i inconsistències de dependència en el disseny de les taules.

Les bases de dades es normalitzen per:

- Evitar la redundància de dades (repetició de dades en un sistema).
- Protegir la integritat de les dades (inconsistències de les dades com a resultat de dades redundants, pèrdues no intencionades de dades pel fet que s'han esborrat altres dades, etc.)
- Evitar problemes d'inserció (impossibilitat d'afegir dades a la base de dades a causa de l'absència d'altres dades) i d'actualització de les dades en les taules

Per poder dir que la nostra base de dades està normalitzada s'han de respectar 3 nivells de normalització.

En la pràctica, si la BD s'ha dissenyat fent ús de models semàntics com el model E / R no sol ser necessària la normalització. D'altra banda si ens proporcionen una base de dades creada sense realitzar un disseny previ, és molt probable que necessitem normalitzar.

Abans de veure les formes normals, coneixerem unes definicions prèvies:

- **Dependència funcional:** $A \rightarrow B$, representa que B és funcionalment dependent d'A. Per a un valor de A sempre apareix un valor de B.
Exemple: Si A és el DNI, i B el Nom, està clar que per a un número de DNI, sempre apareix el mateix nom de titular.
- **Dependència funcional completa:** Té sentit plantejar-se aquest tipus de dependència quan A està compost per més d'un atribut. $A \rightarrow B$, si B depèn de A en la seua totalitat (de tots els atributs que formen A).
Exemple: suposem que A correspon a l'atribut compost: DNI_Empleat + Codi_Dep i B és Nom_Dep. En aquest cas B depèn del Codi_Dep, Però no del DNI_Empleat. Per tant, no hi hauria dependència funcional completa.
- **Dependència transitiva:** $A \rightarrow B \rightarrow C$. Si $A \rightarrow B$ i $B \rightarrow C$ (i no $B \rightarrow A$), llavors diem que C depèn de forma transitiva d'A

- *Exemple:* A és el DNI d'un alumne, B la localitat en què viu i C la província. És un cas de dependència transitiva $A \rightarrow B \rightarrow C$.

Formes normals

Primera Forma Normal: 1FN

Una Relació està en 1FN si i només si cada atribut és atòmic (no pren més d'un valor, és indivisible).

Cal seguir una sèrie de passos per poder dir que la nostra taula està en 1FN:

- Eliminar els grups repetitius de la taules individuals.
- Crear una taula separada per cada grup de dades relacionades.
- Identificar cada grup de dades relacionades amb una clau primària

Per identificar si ho hem fet de manera correcta revisarem:

- Tots els atributs són atòmics?
- La taula conté una clau primària única?
- La clau primària no conté atributs nuls?
- Els camps no clau han d'identificar-per la clau (Dependència Funcional).
- Una taula no pot tenir múltiples valors en cada columna.

Exemple: Suposem que tenim que tenim la següent taula amb dades d'alumnat d'un centre. La Clau Principal està formada pels atributs DNI i Curs:

Alumnos							
DNI	Nombre	Curso	FechaMatrícula	Tutor	LocalidadAlumno	ProvinciaAlumno	Teléfonos
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Écija	Sevilla	660111222
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Écija	Sevilla	660222333 660333444 660444555
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Écija	Sevilla	
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	El Villar	Córdoba	
55555555E	José	2ESO-A	02-Julio-2016	Federico	El Villar	Córdoba	661000111 661000222

Com es pot observar, aquesta taula no està en 1FN ja que el camp Telèfons conté diverses dades dins d'una mateixa cel·la i per tant no és un camp on els valors siguin atòmics.

Com ho solucionem?

DNI	Nombre	Curso	Alumnos		LocalidadAlumno	ProvinciaAlumno
			FechaMatrícula	Tutor		
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Écija	Sevilla
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Écija	Sevilla
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Écija	Sevilla
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	El Villar	Córdoba
55555555E	José	2ESO-A	02-Julio-2016	Federico	El Villar	Córdoba

Teléfonos

DNI	Teléfono
11111111A	660111222
22222222B	660222333
22222222B	660333444
22222222B	660444555
55555555E	661000111
55555555E	661000222

Segona Forma Normal: 2FN

Una Relació està en 2FN si i només si està en 1FN i tots els atributs que no formen part de la Clau Principal tenen dependència funcional completa d'ella. És a dir que no existeixen dependències parcials, o dit d'altra manera, si la nostra Clau Principal està formada per més d'un atribut, els atributs que no formen part de la Clau Principal depenen de tota la Clau Principal, no només d'un dels atributs.

Significa, que en una relació només s'ha d'emmagatzemar informació sobre un tipus d'entitat, i que els atributs que no aporten informació directa sobre la clau principal s'han d'emmagatzemar en una relació (taula) separada.

Quan la clau primària es compon d'un sol atribut, automàticament està en 2FN.

Hem de seguir els següents passos:

- Crear taules separades per a aquells grups de dades que s'apliquen a diversos registres.
- Relacionar aquestes taules mitjançant una clau externa (per seguir mantenint la informació).

Exemple: Seguim amb l'exemple anterior. Treballarem amb la següent taula:

DNI	Nombre	Curso	Alumnos		LocalidadAlumno	ProvinciaAlumno
			FechaMatrícula	Tutor		
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Écija	Sevilla
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Écija	Sevilla
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Écija	Sevilla
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	El Villar	Córdoba
55555555E	José	2ESO-A	02-Julio-2016	Federico	El Villar	Córdoba

Anem a examinar les dependències funcionals:

Sempre que apareix un DNI apareixerà el Nom corresponent i la Localitat de l'alumne corresponent. Per tant $DNI \rightarrow Nom$ i $DNI \rightarrow LocalidadAlumno$.

D'altra banda sempre que apareix un Curs apareixerà el Tutor corresponent. Per tant $Curs \rightarrow Tutor$.

Els atributs Nom i LocalidadAlumno no depenen funcionalment de Curs, ja que només depenen de DNI; i l'atribut Tutor no depèn funcionalment de DNI, ja que només depèn de Curs.

L'únic atribut que sí depèn de forma completa de la clau composta DNI i Curs és FechaMatricula: $(DNI, Curs) \rightarrow FechaMatricula$.

IMPORTANT! *A l'hora d'establir la Clau Primària d'una taula hem de triar un atribut o conjunt d'ells dels quals depenguen funcionalment la resta d'atributs. A més ha de ser una dependència funcional completa.*

Si escollim DNI com a clau primària, tenim un atribut (Tutor) que no depèn funcionalment d'ell. Si escollim Curs com a clau primària, tenim altres atributs que no depenen d'ell.

Si escollim la combinació (DNI, Curs) com a clau primària, llavors sí tenim tota la resta d'atributs amb dependència funcional respecte a aquesta clau. Però és una dependència parcial, no total (llevat FechaMatricula, que sí que té dependència completa). Per tant aquesta taula no està en 2FN.

Com ho solucionem? Dividim en xicotetes taules de forma que cadascun dels atributs que no formen part de la Clau Principal depenguen de la Clau Principal.

Alumnos				Matrículas		
DNI	Nombre	Localidad	Provincia	DNI	Curso	FechaMatrícula
11111111A	Eva	Écija	Sevilla	11111111A	1ESO-A	01-Julio-2016
22222222B	Ana	Écija	Sevilla	22222222B	1ESO-A	09-Julio-2016
33333333C	Susana	El Villar	Córdoba	33333333C	1ESO-B	11-Julio-2016
44444444D	Juan	El Villar	Córdoba	44444444D	2ESO-A	05-Julio-2016
55555555E	José	Écija	Sevilla	55555555E	2ESO-A	02-Julio-2016

Cursos	
Curso	Tutor
1ESO-A	Isabel
1ESO-B	Roberto
2ESO-A	Federico

Tercera Forma Normal: 3FN

Una Relació està en 3FN si i només si està en 2FN i no hi ha dependències transitives. Totes les dependències funcionals han de ser respecte a la clau principal.

És a dir, no hi ha atributs que no pertanyen a la clau primària que puguin ser coneguts mitjançant un altre atribut que no forme part de la clau primària.

Hem de considerar els següents casos:

- Eliminar aquells camps que no depenguen de la clau.
- Cap columna pot dependre d'una columna que no siga una clau.
- No hi pot haver dades derivades.

Exemple: Seguim amb l'exemple anterior. Treballarem amb la següent taula (recordem que la Clau Principal és DNI):

Alumnos			
DNI	Nombre	Localidad	Provincia
11111111A	Eva	Écija	Sevilla
22222222B	Ana	Écija	Sevilla
33333333C	Susana	El Villar	Córdoba
44444444D	Juan	El Villar	Córdoba
55555555E	José	Écija	Sevilla

Les dependències funcionals existents són les següents: $DNI \rightarrow Nom$

Com podem observar hi ha una dependència funcional transitiva: $DNI \rightarrow Localitat \rightarrow Província$

Perquè la taula estiga en 3FN, no poden existir dependències funcionals transitives. Per solucionar el problema haurem de crear una nova taula. El resultat és:

Alumnos		
DNI	Nombre	Localidad
11111111A	Eva	Écija
22222222B	Ana	Écija
33333333C	Susana	El Villar
44444444D	Juan	El Villar
55555555E	José	Écija

Localidades	
Localidad	Provincia
Écija	Sevilla
El Villar	Córdoba

La resta de taules ja estaria en 3FN

Resultat final de l'exercici:

Alumnos		
DNI	Nombre	Localidad
11111111A	Eva	Écija
22222222B	Ana	Écija
33333333C	Susana	El Villar
44444444D	Juan	El Villar
55555555E	José	Écija

Localidades	
Localidad	Provincia
Écija	Sevilla
El Villar	Córdoba

Teléfonos	
DNI	Teléfono
11111111A	660111222
22222222B	660222333
22222222B	660333444
22222222B	660444555
55555555E	661000111
55555555E	661000222

Matrículas		
DNI	Curso	FechaMatrícula
11111111A	1ESO-A	01-Julio-2016
22222222B	1ESO-A	09-Julio-2016
33333333C	1ESO-B	11-Julio-2016
44444444D	2ESO-A	05-Julio-2016
55555555E	2ESO-A	02-Julio-2016

Cursos	
Curso	Tutor
1ESO-A	Isabel
1ESO-B	Roberto
2ESO-A	Federico

Exercici resolt pas a pas

Per entendre molt millor tot això revisem la següent taula:

id	curso	slug	autor	materiales	fecha_creacion
1	Curso de Python	curso-de-python	David Aroesti	23	01/12/2017
2	Fundamentos de Bases de Datos	fundamentos-de-bases-de-datos	Pedro Díaz	15	02/12/2017
3	Curso de PostgreSQL	curso-de-postgresql	Mario Chávez	29	03/12/2017

Si ens fixem bé, aquesta taula no compleix amb el requisit que diu que tots els atributs han de ser atòmics, és a dir que hem de descompondre-la.

Seguint els aspectes a considerar podríem dir que la nostra taula inicial es transformaria en:

id	curso	slug	primer_nombre_profesor	segundo_nombre	materiales	fecha_creacion
1	Curso de Python	curso-de-python	David	Aroesti	23	01/12/2017
2	Fundamentos de Bases de Datos	fundamentos-de-bases-de-datos	Pedro	Díaz	15	02/12/2017
3	Curso de PostgreSQL	curso-de-postgresql	Mario	Chávez	29	03/12/2017

Amb això podríem dir que la nostra taula compleix amb la Primera forma Normal.

Ara, si ens fixem aquesta encara no compleix amb els requisits que demana la Segona forma Normal.

Podríem crear taules separades per Cursos i per a professors i cadascuna d'aquestes tindria la seva clau primària i estarien relacionades per una clau externa (Foreign Key).

Taula cursos

id	curso	materiales	autorId	fecha_creacion
1	Curso de Python	23	3	01/12/2017
2	Fundamentos de Bases de Datos	15	2	02/12/2017
3	Curso de PostgreSQL	29	1	03/12/2017

Taula autors

id	primer_nombre	segundo_nombre	fecha_creacion
1	Mario	Chávez	01/12/2017
2	Pedro	Díaz	01/12/2017
3	Mario	Aroesti	01/12/2017

Ara la nostra taula està en 2FN.

Si detallem la nostra taula de Cursos encara podríem simplificar, per això podríem extreure la columna de materials i crear una nova taula per això.

id	nombre material	slug	cursoId	fecha_creacion
1	Operadores Matemáticos con Python	operadores-matematicos-con-python	1	22/12/2017
2	Visión General de los Datos	vision-general-de-los-datos	2	22/12/2017
3	Uso de Roles en Postgresql	uso-de-roles-en-postgres	3	23/12/2017

Ara la nostra taula compliria amb els requisits de la 3FN i estaria a punt.