

Tema 4: Normalització



Dependències Functionals

- Una **dependència funcional** es dona quan a partir d'un o varis atributs d'un conjunt d'entitats es pot determinar el valor de qualsevol altre atribut.
- Existeix una dependència funcional entre l'atribut "a1" i "a2" que pertanyen a una mateixa entitat "A", si coneixent el valor de l'atribut "a1" podem saber sempre el valor de l'atribut "a2".
- $a1 \rightarrow \{a2\}$: a1 determina a a2 o a2 té una dependència funcional d'a1
- $a2 \rightarrow \{a1\}$: a2 determina a a1 o a1 té una dependència funcional d'a2

Dependències Funcionals

- **No sempre son tan clares** com per exemple podrien ser-ho els atributs codi postal i província (el prefix del codi postal indica la província), així que per tal d'identificar les dependències funcionals caldrà disposar d'un sistema ben definit. Per a explicar-ho, suposarem un conjunt d'entitats "A" amb els atributs "a1", "a2" i "a3".
- Comprovarem **quin atribut o atributs no tenen repeticions**.
- Si hi ha un **atribut que no té cap repetició, la resta d'atributs tindran una dependència funcional respecte d'aquest atribut**. → Els camps identificadors o claus primàries, fan que la resta d'atributs tinguin una dependència funcionals respecte d'ells al no tenir repeticions.

Dependències Functionals: Exemple

DNI	Nom	Cognom	Telefon
11111111G	Marc	González	666112233
11111111H	Maria	Sanchez	666112244
11111111V	Marc	Hernández	666112255
11111111B	Anna	Sanchez	666112266

- $\text{DNI} \rightarrow \{\text{Nom}, \text{Cognom}, \text{Telèfon}\}$
- $\text{Nom} \rightarrow \{\}$
- $\text{Cognom} \rightarrow \{\}$
- $\text{Telèfon} \rightarrow \{\text{DNI}, \text{Nom}, \text{Cognom}\}$

Dependències Functionals: Exercici genèric

Suposem que les entitats que agafem com a mostra són les que es mostren en la taula:

Conjunt entitats: A

atrib1 atrib2 atrib3

va11 va21 va31

va12 va22 va31

va13 va23 va31

va14 va24 va32

va15 va25 va33

va16 va26 va33

Dependències functionals

Observem com l'atribut 1 i 2 («atrib1» i «atrib2») no tenen cap valor repetit per les entitats de mostra. Si em diuen un valor de «l'atrib1» o de l'atrib2 sempre se sap a quin valor correspon per la resta d'atributs. Per tant, puc afirmar que:

atrib1 → {atrib2, atrib3} i **atrib2** → {atrib1, atrib3}

L'atrib3 no determina cap atribut, ja que per exemple, pel valor va31 no puc dir un valor de l'atrib1 únic ni tampoc de l'atrib2.

atrib3 → {}

Dependències Functionals: Exercici genèric

Si el nostre conjunt d'entitats tingués com a mostra de valors (s'ha sombrejat l'única entitat en la qual s'han fet canvis:

Conjunt entitats: A

atrib1 atrib2 atrib3

va11 va21 va31

va12 va22 va31

va13 va22 va31

va14 va24 va32

va15 va25 va33

va16 va26 va33

Dependències functionals

Els valors d'«atrib1» no han canviat i són tots diferents. Per tant, la dependència es manté:

atrib1 → {**atrib2, atrib3**}

Per contra, «atrib2» representa un valor duplicat «va22». Per aquest valor, no puc determinar un únic valor de l'atrib1, pot ser «va12» o «va13». Per tant, «atrib2» no em determina «atrib1». «atrib2» sí que continua determinant «atrib3», ja que pel valor repetit «va22», el valor d'«atrib3» és el mateix «va31»

atrib2 → {**atrib3**} i **atrib3** → {}

Dependències Functionals: Exercici Curses de Cavalls

Cavall							
Cavall	Hípica	Data	Temps	Club	Llicència	Campió	Núm.
Amor	Camil	11/10/12	01:34:00	8474	2976	Fruit	2978
Diversió	Camil	11/10/12	01:49:00	8474	2977	Fruit	2978
Fruit	Camil	11/10/12	01:32:00	8474	2978	Fruit	2978
Gaudir	Camil	11/10/12	01:54:00	8474	2979	Fruit	2978
Amor	Camil	13/10/12	01:34:00	8474	2976	Diversio	2977
Diversió	Camil	13/10/12	01:39:00	8474	2977	Diversio	2977
Fruit	Camil	13/10/12	01:12:00	8474	2978	Diversio	2977
Gaudir	Camil	13/10/12	01:25:00	8474	2979	Diversio	2977
Amor	Rial	15/10/12	01:34:00	7000	2976	Diversio	2977
Diversió	Rial	15/10/12	01:39:00	7000	2977	Diversio	2977
Fruit	Rial	15/10/12	01:12:00	7000	2978	Diversio	2977
Gaudir	Rial	15/10/12	01:25:00	7000	2979	Diversio	2977

Dependències Functionals: Exercici Curses de Cavalls

- Cavall \rightarrow {Llicència}
- Hípica \rightarrow {Club}
- Data \rightarrow {Hípica, Club, Campió, Núm}
- Temps \rightarrow {Cavall, Llicència, Campió}
- Club \rightarrow {Hípica}
- Llicència \rightarrow {Cavall}
- Campió \rightarrow {Núm.}
- Núm. \rightarrow {Campió}

Hi ha algún atribut que determini tota la resta?

No. Per tant, agafarem parelles d'atributs fins a trobat l'atribut identificador.

Dependències Functionals: Exercici Curses de Cavalls

- data+cavall \rightarrow Tots
- data+hipica \rightarrow {club, campió, núm}
- data+temps \rightarrow Tots
- data+club \rightarrow {hípica, campió,núm}
- data+llicencia \rightarrow Tots
- data+campio \rightarrow {hípica, club, núm}
- Data+núm \rightarrow {}

Amb els camps «Data» i «Cavall» tota la resta de camps de la taula queden determinats. **La clau primària de la nostra taula podria ser {data,cavall}**, encara que no és la única opció que tenim.

Dependències Functionals: Exercici Llibres

- Determina les dependències functionals de la següent taula fins a trobar l'atribut identificador:

<u>AUTOR</u>	<u>NACIONALIDAD</u>	<u>COD LIBRO</u>	<u>TITULO</u>	<u>EDITORIAL</u>	<u>AÑO</u>
Date, C.	Norteamericana	98987	Database	Addison, W.	1990
Date, C.	Norteamericana	97777	SQL Stan	Addison, W.	1986
Date, C.	Norteamericana	98987	Guide for	Addison, W.	1988
Codd,E.	Norteamericana	7890	Relational	Addison,W.	1990
Gardarin	Francesa	12345	Basi Dati	Paraninfo	1986
Gardarin	Francesa	67890	Comp BD	Eyrolles	1984
Valduriez	Francesa	67890	Comp BD	Eyrolles	1984
Kim,W.	Norteamericana	11223	BD OO	ACM	1989
Lochovsky	Canadiense	11223	BD OO	ACM	1989

Normalització

- La normalització d'una base de dades és un procés que consisteix en aplicar un conjunt de normes i regles a les relacions obtingudes en el procés de passar del model **entitat-relació** al **model relacional**.
- L'objectiu és minimitzar la redundància de dades i facilitar la gestió posterior de la BDs.
- Aquest procés acostuma a implicar la fragmentació d'algunes de les taules obtingudes en el model relacional.

Normalització: Anomalies de disseny

Suposem un disseny on la BDs consta d'una única taula per a desar la informació dels estudiants i de les unitats formatives que estan cursant.

Estudiant UFs							
DNI	NOM	TELF.	POB	CP	ADREÇA	UF	Nom UF
40958733	Anna	934232775	Granollers	08402	C/Girona,23,7è 1a	UF1	BD Intro
40958733	Anna	665765431	Granollers	08402	C/Girona,23,7è 1a	UF2	SQL
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL
76543124	Maria	623903213	St. Celoni	08470	C/Major,15	UF2	SQL
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL
40958736	Maria	934232776	Granollers	08402	C/Major,15	UF2	SQL

Normalització: Anomalies de disseny

Redundància de dades

- Cada vegada que un estudiant cursa una unitat formativa diferent, és necessari informar de totes les dades de l'estudiant.
- La probabilitat d'error s'incrementa ja que hi ha un major volum d'informació.

Coherència de dades

- No es garanteix que les dades siguin coherents, ja que si per un mateix estudiant s'han introduït dues adreces diferents, el sistema no pot controlar-ho.

Normalització: Procés

Identificar les dependències funcionals

En el nostre exemple és fàcil observar que hi ha 4 dependències funcionals diferents:

- $DNI \rightarrow \{NOM, TLF, POBL., CP, ADREÇA\}$
- $UF \rightarrow \{NOM UF\}$
- $Nom UF \rightarrow \{UF\}$
- $CP \rightarrow \{POB.\}$

Cap de les dependències funcionals indica una clau primaria.

Formes normals



Primera forma normal (1FN)

Per tal que una taula estigui en 1FN cal que es compleixin els següents punts:

- No hi ha registres **duplicats**
- No hi hagi atributs derivats/**calculats**.
- No hi ha atributs **compostos**.
- El **nom** del camp és **únic** en tota la taula.
- Tots els valors de cada camp són atòmics: **no hi ha multivalors**.

Primera forma normal (1FN): Duplicats

Estudiant UFs									
DNI	Nom	Telf.	POBLACIO		Adreça	UF		Nota	Total UFs
			NOM	CP		Codi	Nom		
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF1	BD	8	2
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF2	SQL	7	2
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8	2
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8	2
75343788	Joan	934232775	Granollers	08402	C/Girona,23	UF2	SQL	7	1
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL	7	1
40958736	Maria	934232776	St. Celoni	08470	C/Major,15	UF2	SQL	7	1

Primera forma normal (1FN): Duplicats [Corregit]

Estudiant UFs									
DNI	Nom	Telf.	POBLACIO		Adreça	UF		Nota	Total UFs
			Nom	CP		Codi	Nom		
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF1	BD	8	2
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF2	SQL	7	2
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8	1
75343788	Joan	934232775	Granollers	08402	C/Girona,23	UF2	SQL	7	1
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL	7	1
40958736	Maria	934232776	St. Celoni	08470	C/Major,15	UF2	SQL	7	1

Primera forma normal (1FN): Atributs derivats/calculats

Estudiant UFs									
DNI	Nom	Telf.	POBLACIO		Adreça	UF		Nota	Total UFs
			Nom	CP		Codi	Nom		
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF1	BD	8	2
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF2	SQL	7	2
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8	1
75343788	Joan	934232775	Granollers	08402	C/Girona,23	UF2	SQL	7	1
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL	7	1
40958736	Maria	934232776	St. Celoni	08470	C/Major,15	UF2	SQL	7	1

Primera forma normal (1FN): Atributs compostos [Corr.]

Estudiant UFs								
DNI	Nom	Telf.	Nom	CP	Adreça	Codi	Nom	Nota
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF1	BD	8
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF2	SQL	7
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8
75343788	Joan	934232775	Granollers	08402	C/Girona,23	UF2	SQL	7
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL	7
40958736	Maria	934232776	St. Celoni	08470	C/Major,15	UF2	SQL	7

Primera forma normal (1FN): Camps duplicats [Corr.]

Estudiant UFs								
DNI	Nom	Telf.	Població	CP	Adreça	Codi	UF	Nota
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF1	BD	8
40958733	Anna	934232775,61543121	Granollers	08402	C/Girona,23	UF2	SQL	7
76543123	Maria	623903212	Granollers	08403	C/Balmes,15	UF2	SQL	8
75343788	Joan	934232775	Granollers	08402	C/Girona,23	UF2	SQL	7
40958735	Pere	934232776	Granollers	08402	C/Major,15	UF2	SQL	7
40958736	Maria	934232776	St. Celoni	08470	C/Major,15	UF2	SQL	7

Primera forma normal (1FN): Multivalors [Corregit]

Estudiant UFs							
DNI	Nom	Població	CP	Adreça	Codi	UF	Nota
40958733	Anna	Granollers	08402	C/Girona,23	UF1	BD	8
40958733	Anna	Granollers	08402	C/Girona,23	UF2	SQL	7
76543123	Maria	Granollers	08403	C/Balmes,15	UF2	SQL	8
75343788	Joan	Granollers	08402	C/Girona,23	UF2	SQL	7
40958735	Pere	Granollers	08402	C/Major,15	UF2	SQL	7
40958736	Maria	St. Celoni	08470	C/Major,15	UF2	SQL	7

Telèfon	
DNI	Telf.
40958733	934232775
40958733	61543121
76543123	623903212
75343788	934232775
40958735	934232776
40958736	934232776

Segona forma normal (2FN)

Per tal que una taula estigui a 2FN cal que es compleixin els següents punts:

- Ha de complir amb la 1FN.
- No existeix cap atribut NO identificador que tingui dependència d'un dels atributs identificadors.
- Si la clau identificadora està compostat per 1 únic atribut: Està en 2FN.
- Si la clau identificadora està formada per 2 o més atributs, caldrà comprovar que no hi hagi cap atribut NO clau que depengui d'una part de la clau identificador. No poden dependre parcialment de la clau identificadora

Segona forma normal (2FN)

- La taula telèfon està en 1a FN.
- La seva clau principal està formada pels camps «DNI» i «Telf.».

$\{\text{DNI}, \text{Telf.}\} \rightarrow \{\}$

- Aquests dos camps no en determinen cap altre, ja que no hi ha cap altre camp a la taula.

Telèfon	
<u>DNI</u>	<u>Telf.</u>
40958733	934232775
40958733	61543121
76543123	623903212
75343788	934232775
40958735	934232776
40958736	934232776

Segona forma normal (2FN)

S'han identificat les següents dependències funcionals:

- $\text{DNI} \rightarrow \{\text{Nom}, \text{Població}, \text{CP}, \text{Adreça}\};$
- $\text{Nom} \rightarrow \{\text{DNI}\}$
- $\text{Població} \rightarrow \{\}$
- $\text{Codi} \rightarrow \{\text{UF}\}$
- $\text{UF} \rightarrow \{\text{Codi}\}$
- $\text{Adreça} \rightarrow \{\}$
- $\text{Nota} \rightarrow \{\}$
- $\text{CP} \rightarrow \{\text{Població}\}$

Estudiant UFs							
DNI	Nom	Població	CP	Adreça	Codi	UF	Nota
40958733	Anna	Granollers	08402	C/Girona,23	UF1	BD	8
40958733	Anna	Granollers	08402	C/Girona,23	UF2	SQL	7
76543123	Maria	Granollers	08403	C/Balmes,15	UF2	SQL	8
75343788	Joan	Granollers	08402	C/Girona,23	UF2	SQL	7
40958735	Pere	Granollers	08402	C/Major,15	UF2	SQL	7
40958736	Maria	St. Celoni	08470	C/Major,15	UF2	SQL	7

Segona forma normal (2FN)

- Podem veure que la combinació dels camps «DNI» i «Codi» ens determinen a tota la resta de camps de la taula;

$$\{\mathbf{DNI}, \mathbf{Codi}\} \rightarrow \{\text{Nom, Població, CP, Adreça, UF, nota}\}$$

- Clau identificadora: DNI i Codi
- Atributs NO claus: Nom, Telf., Població, CP., Adreça, UF

Segona forma normal (2FN)

- Utilitzant només el camp DNI es pot determinar algun dels camps: «nom», «població», «codi postal», «adreça», «UF» o «nota»?
- Utilitzant només el camp «Codi» es pot determinar algun dels camps: «nom», «població», «codi postal», «adreça», «UF» o «nota»?
- A l'inici del desenvolupament s'han obtingut les següents dependències funcionals:
 - {DNI} → {Nom, Població., CP., Adreça}
 - {Codi} → {UF}
- I, per tant, és obvi que la resposta a les dues preguntes és «**Sí**» i «**Sí**»
- Hi ha camps de la clau primària que determinen camps que no formen part de la clau primària. La taula «Estudiant UFs» **NO** està en 2a FN

Segona forma normal (2FN)

$\{\text{DNI}\} \rightarrow \{\text{Nom, Telf., Població., CP., Adreça}\}$

$\{\text{Codi}\} \rightarrow \{\text{UF}\}$

UF	
<u>Codi</u>	Nom
UF1	BD
UF2	SQL
UF2	SQL
UF2	SQL

Taula no normalitzada

La taula no està en primera forma normal, ja que hi ha registres duplicats. Després de traspasar-la a 1FN obtenim

Estudiant				
<u>DNI</u>	Nom	Població	CP	Adreça
40958733	Anna	Granollers	08402	C/Girona,23
76543123	Maria	Granollers	08403	C/Balmes,15
75343788	Joan	Granollers	08402	C/Girona,23
40958735	Pere	Granollers	08402	C/Major,15
40958736	Maria	St. Celoni	08470	C/Major,15

UF	
<u>Codi</u>	Nom
UF1	BD
UF2	SQL

Taula normalitzada

Segona forma normal (2FN)

- Finalment cal crear la taula per la dependència funcional:
- $\{\text{DNI}, \text{Codi}\} \rightarrow \{\text{Nom}, \text{Telf.}, \text{Població.}, \text{CP.}, \text{Adreça}, \text{UF}, \text{nota}\}$
- Aquesta taula contindrà els dos camps claus $\{\text{DNI}, \text{Codi}\}$ i tots els camps que no depenguin només de «DNI» ni de «Codi», en el nostre cas «nota».

Qualificació		
<u>DNI</u>	<u>UF</u>	Nota
40958733	UF1	8
40958733	UF2	7
76543123	UF2	8
75343788	UF2	7
40958735	UF2	7
40958736	UF2	7

Segona forma normal (2FN)

Taula	Compleix segona forma normal?	
Estudiant	Està en 1a FN. La seva clau primària està formada només per un atribut.	Sí
Telèfon	Està en 1a FN. No hi ha atributs no identificadors. Està en 2FN	Sí
UF	Està en 1a FN (Després d'eliminar un registre) La seva clau primària està formada només per un atribut.	Sí.
Qualificació	Està en 1a FN. El camp no clau «Nota» no depèn de «DNI» i d'«UF»	Sí

Tercera forma normal (3FN)

Per tal que una taula estigui a 3FN cal que es compleixin els següents punts:

- Ha de complir amb la 2FN.
- Si només hi ha un atribut NO identificador.
- No pot existir cap dependència funcional entre atributs no identificadors.

Tercera forma normal (3FN)

- Teniem les següents taules:
 - TELEFON: {DNI, Telefon}
 - ESTUDIANT: {DNI, Nom, Població, CP, Adreça}
 - UF: {Codi, Nom}
 - QUALIFICACIO: {DNI, UF, Nota}
- Les taules: Telèfon, UF i qualificació, només ténen un atribut NO identificador, per tant, no pot haver cap DF amb un altre.
- En la taula Estudiant, tenim 4 atributs NO Identificadors.
- Hem vist en les DF anteriors que $CP \rightarrow \{Població\}$
- Per tant **NO ESTÀ** EN 3ra FORMA NORMAL.

Tercera forma normal (3FN)

Tenim la DF de CP \rightarrow {Població}

- Es crea taula CP amb CP i Població.
- S'elimina Població de la taula ESTUDIANT.
 - ESTUDIANT: {DNI, Nom, ~~Població~~, CP, Adreça}

RESULTAT FINAL:

- TELEFON: {DNI, Telefon}
- ESTUDIANT: {DNI, Nom, CP, Adreça}
- UF: {Codi, Nom}
- QUALIFICACIO: {DNI, UF, Nota}
- CP: {CP, Població}

Altres formes normals

Malgrat en un inici sols es varen definir les regles relatives a les tres formes normals que ja hem tractat, més endavant la teoria de normalització es va ampliar amb les següents formes normals:

- La forma normal de Boyce-Codd (FNBC), una versió més estricta de la 3FN
- La quarta forma normal (4FN)
- La cinquena forma normal (5FN)
- La forma normal de domini/clau (DKNF)
- La sisena forma normal (6FN)

En tot cas, a la pràctica es considera com a bo qualsevol disseny que compleixi fins a la 3FN, essent la resta de formes més teòriques que pràctiques, forçant restriccions que sovint no es recomanen per entorns reals.