



BASES DE DADES

Grau en Enginyeria en Informàtica

Universitat de Barcelona

Curs 2017/18

TEMA 5

NORMALITZACIÓ

Objectius

- 1) **Conèixer la normalització**
- 2) Identificar redundància de dades i anomalies d'actualització
- 3) Distingir entre dependències funcionals completes i parcials
- 4) Identificar dependències funcionals
- 5) Passar una relació de 1FN a 3FN

Normalització

- Al tema 4 hem vist el disseny ER i el seu corresponent pas a taules
- Vàrem acabar amb la combinació de taules...però ho podem fer millor? O comprovar que és un bon disseny?
- **Normalització:** és una tècnica de disseny de bases de dades que examina les relacions (concretament, les dependències funcionals) entre els atributs de les taules

...normalització

- La normalització utilitza una sèrie de formes normals que ens ajuden a agrupar òptimament els atributs
 - Mínim número d'atributs necessaris per assolir els objectius
 - Els atributs que tenen una relació lògica forta (dependència funcional) estan a la mateixa relació
 - Mínima redundància d'atributs – excepte les claus foranes

TEMA 5

NORMALITZACIÓ

Objectius

- 1) Conèixer la normalització
- 2) **Identificar redundància de dades i anomalies d'actualització**
- 3) Distingir entre dependències funcionals completes i parcials
- 4) Identificar dependències funcionals
- 5) Passar una relació de 1FN a 3FN

Redundància de dades

- A les bases de dades tenim redundància de dades
- Un exemple són les claus foranes – que són còpies de les claus primàries
- Aquesta redundància és necessària
- El que no volem és redundància innecessària

...redundància de dades

```
STAFF (staffNo, name, position, salary, branchNo)
```

```
BRANCH (branchNo, address)
```

```
STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
```

La taula STAFFBRANCH té dades redundants.

La clau primària és staffNo.

Si dos empleats treballen a la mateixa oficina, la relació té les dades de l'oficina duplicades

Les relacions que tenen dades redundants presenten anomalies d'actualització (inserció i eliminació)

Anomalies d'inserció

```
STAFF (staffNo, name, position, salary, branchNo)
```

```
BRANCH (branchNo, address)
```

```
STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
```

- Considerem la relació STAFFBRANCH
- Una anomalia d'inserció pot succeir en el moment d'inserir un nou treballador, ja que hem d'escriure totes les dades de l'oficina – i si ens equivoquem?
- Si aquestes dades estiguessin en una taula separada, llavors únicament ens caldria afegir una referència a l'oficina

Anomalies d'inserció

```
STAFF (staffNo, name, position, salary, branchNo)
```

```
BRANCH (branchNo, address)
```

```
STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
```

- Considerem la relació STAFFBRANCH
- Si volem introduir una oficina nova sense treballadors, no podem, perquè la clau primària (staffNo) no pot ser NULL

Anomalies d'eliminació

`STAFF (staffNo, name, position, salary, branchNo)`

`BRANCH (branchNo, address)`

`STAFFBRANCH (staffNo, name, position, salary, branchNo, address)`

- Considerem la relació STAFFBRANCH
- Si eliminem l'últim treballador d'una oficina concreta, les dades de l'oficina les perdem
- Si les dades de l'oficina estan a la taula BRANCH, llavors aquesta anomalia no la tindríem...

Comentari

`STAFF (staffNo, name, position, salary, branchNo)`

`BRANCH (branchNo, address)`

`STAFFBRANCH (staffNo, name, position, salary, branchNo, address)`

- Podeu pensar (i argumentar) que aquests anomalies passen perquè la relació / taula STAFFBRANCH està mal dissenyada
- Teniu raó. La normalització ens ajuda a comprovar que tenim un “bon” disseny, i a millorar-lo

TEMA 5

NORMALITZACIÓ

Objectius

- 1) Conèixer la normalització
- 2) Identificar redundància de dades i anomalies d'actualització
- 3) **Identificar dependències funcionals**
- 4) Distingir entre dependències funcionals completes i parcials
- 5) Passar una relació de 1FN a 3FN

Dependències funcionals

Una **dependencia funcional** describe la relación entre los atributos de una relación. Si A y B son atributos de una relación R, B es funcionalmente dependiente de A ($A \rightarrow B$) si cada valor de A está asociado con exactamente un valor de B. A y B pueden ser atributos o conjunto de atributos.

Assumim que cada atribut a la base de dades té el mateix nom.

Si tenim atributs amb el mateix nom, podem fer com al producte cartesià:

Relació.atribut per a distingir-los

Dependències funcionals

- La dependència funcional és una propietat del **significat** dels atributs d'una relació
- Si tenim dos atributs, A i B, i B depèn funcionalment de l'atribut A, significa que cada valor A està associat amb únicament un valor de B (i l'ordre és important)

A	C	D	B
2	3
2	3
...
2	3

Dependència funcional: A → B

A	C	D	B
2	1	4	3
2	1	5	3
...
2	3

No dependència funcional (C → D)
D → C: No ho sabem

Dependències funcionals

- Si B depèn funcionalment de A, també podem dir que A **determina** funcionalment B
- El determinant es refereix a un atribut, o a un conjunt d'atributs, ubicat al costat esquerra de la dependència funcional

Una **dependencia funcional** describe la relación entre los atributos de una relación. Si A y B son atributos de una relación R, B es funcionalmente dependiente de A ($A \rightarrow B$) si cada valor de A está asociado con exactamente un valor de B. A y B pueden ser atributos o conjunto de atributos.

Dependències funcionals

STAFF (*staffNo*, name, position, salary, branchNo)

BRANCH (branchNo, address)

Podem dir que *staffNo* determina funcionalment *position*. Assumim que cada treballador té un càrrec.

NO podem dir que *position* determina funcionalment *staffNo*. Per mateixos valors a *position*, podem tenir N valors a *staffNo*

Observació: la relació entre *staffNo* i *position* és 1 a 1. La relació entre *position* i *staffNo* és 1:N.

staffNo → *position*

Dependències funcionals

En normalizació, nos interessa identificar dependències funcionals que son independents de la instància o exemplar de la base de dades. **Una dependència funcional es una propietat del disseny de la relació y no del valor de los atributos en un momento particular.** Esto es muy importante. Te recomiendo que lo tengas presente.

Dependències funcionals

`STAFF (staffNo, name, position, salary, branchNo)`

- Donat un *staffNo*, podem identificar el nom del treballador
- Donat *name*, sembla que també podem trobar el seu *staffNo*
- Per tenir en compte l'aspecte temporal (no del valor dels atributs en un moment concret), hem d'entendre el propòsit de cada atribut. **El seu significat**

Dependències funcionals

`STAFF (staffNo, name, position, salary, branchNo)`

- L'objectiu de l'atribut *staffNo* és identificar unívocament a cada treballador
- L'objectiu de l'atribut *name* és guardar el nom d'un treballador
- Per tant, *name* no determina funcionalment *staffNo* (1 *name* : N *staffNo*) – quants empleats es poden dir Jordi a una oficina en Catalunya?

Exercici 1

- Tenim la següent relació
- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- Ens diuen el següent
 - una editorial pot publicar N llibres, però un llibre és publicat únicament per una editorial
 - El camp edició és per indicar l'edició de llibres amb mateix títol (edició 1, 2...)
 - La clau primària és ISBN
- Contesta a les següents preguntes:

Exercici 1

- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- ISBN → títol?
- ISBN → codi_editorial?
- títol → ISBN?
- codi_editorial → ISBN?

Exercici 1

- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- ISBN → títol? Sí. Un valor ISBN està relacionat amb únicament un valor a títol.
- ISBN → codi_editorial? Sí. Un ISBN (llibre) és publicat per una editorial (i no més d'una)
- títol → ISBN? No. Un llibre amb mateix títol pot tenir diferent ISBN (tenim el camp edició)
- codi_editorial → ISBN? No. Una editorial pot publicar N llibres

Comentari

- Les respostes a les dependències funcionals són per aquest cas, exercici, concret
- És a dir, depenen de la situació que ens diguin
- Les anteriors respostes es podrien haver escrit “Si / No en aquest cas concret”, però per simplificar, i ja que assumim que són per aquest exercici concret, les he escrit com “Si / No”

Exercici 2

- Considera la següent relació:
- STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
- És una dependència funcional?

$\text{staffNo} \rightarrow (\text{name}, \text{position}, \text{salary}, \text{branchNo}, \text{address})$

Exercici 2

- Considera la següent relació:
- STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
- És una dependència funcional? Sí, donat l'identificador d'un empleat, sabem la resta de dades. Relació 1:1

`staffNo` → (name, position, salary, branchNo, address)

Exercici 2

- Considera la següent relació:
- STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
- És una dependència funcional?

branchNo → address

Exercici 2

- Considera la següent relació:
- STAFFBRANCH (staffNo, name, position, salary, branchNo, address)
- És una dependència funcional? Sí. El codi de l'oficina determina el valor de l'atribut direcció (suposarem que la direcció inclou nom del carrer, pis, porta...). També es pot pensar en una segona dependència funcional

branchNo → address

address → branchNo

Exercici 3

- En aquest cas no tenim l'esquema de la relació, sinó una instància. Ens diuen que els valors són representatius dels atributs, i ens demanen trobar les dependències funcionals

A	B	C	D	E
A	B	Z	W	Q
E	B	R	W	P
A	D	Z	W	T
E	D	R	W	Q
A	F	Z	S	T
E	F	R	S	T

Exercici 3

A	B	C	D	E
A	B	Z	W	Q
E	B	R	W	P
A	D	Z	W	T
E	D	R	W	Q
A	F	Z	S	T
E	F	R	S	T

DF1: Cuando el valor 'A' aparece en la columna A, tenemos el valor 'Z' en la columna C. Cuando el valor 'E' aparece en la columna A, tenemos el valor R en la columna C. Por tanto, tenemos una relación 1 a 1, y los valores de C son consistentes con los valores de A. Siempre que tenemos una 'A' en A, tenemos una 'Z' en C. Siempre que tenemos una 'E' en A, tenemos una 'R' en C. Hemos encontrado una dependencia funcional.

DF1: A → C

Exercici 3

A	B	C	D	E
A	B	Z	W	Q
E	B	R	W	P
A	D	Z	W	T
E	D	R	W	Q
A	F	Z	S	T
E	F	R	S	T

DF2: C → A

Exercici 3

A	B	C	D	E
A	B	Z	W	Q
E	B	R	W	P
A	D	Z	W	T
E	D	R	W	Q
A	F	Z	S	T
E	F	R	S	T

DF3: B → D

Exercici 3

A	B	C	D	E
A	B	Z	W	Q
E	B	R	W	P
A	D	Z	W	T
E	D	R	W	Q
A	F	Z	S	T
E	F	R	S	T

Hasta ahora hemos ido columna a columna. El siguiente paso es considerar combinaciones de pares de columnas. (C, D), (B, D), (A, C), (A, D) no determinan E. Sin embargo, (A,B), y (B, C) determinan E. Hemos encontrado dos dependencias funcionales más.

DF4: (A, B) \rightarrow E

DF5: (B, C) \rightarrow E

TEMA 5

NORMALITZACIÓ

Objectius

- 1) Conèixer la normalització
- 2) Identificar redundància de dades i anomalies d'actualització
- 3) Identificar dependències funcionals
- 4) **Distingir entre dependències funcionals completes i parcials**
- 5) Passar una relació de 1FN a 3FN

Dependència funcional completa i parcial

Una dependencia funcional **completa** indica que si A y B son atributos de una relación, B depende funcionalmente de A si y sólo si B depende funcionalmente de A pero no de cualquier subconjunto de A. En otras palabras, el determinante es mínimo.

Una dependència funcional és parcial si quan eliminem algun atribut del determinant la dependència funcional es manté

Dependència funcional completa i parcial

`STAFF (staffNo, name, position, salary, branchNo)`

- La següent dependència funcional, és completa o parcial?
Raona la teva resposta

`(staffNo, name) → branchNo`

Dependència funcional completa i parcial

`STAFF (staffNo, name, position, salary, branchNo)`

- La següent dependència funcional, és completa o parcial?
És una dependència funcional parcial. L'atribut *name* del determinant no és necessari

`(staffNo, name) → branchNo`

Dependència funcional completa i parcial

`STAFF (staffNo, name, position, salary, branchNo)`

- La següent dependència funcional, és completa o parcial?
És una dependència funcional parcial. L'atribut *name* del determinant no és necessari

`(staffNo, name) → branchNo`

La dependència funcional és

`staffNo → branchNo`

Dependència funcional completa i parcial

- Si la relació entre BRANCH i STAFF fos 1 a 1, llavors branchNo -> staffNo seria una dependència funcional completa (però sense gaire sentit)
- Les dependències funcionals que utilitzem en normalització tenen 3 característiques:
 - Relació 1:1
 - Es mantenen al llarg del temps (significat dels atributs)
 - El determinant és mínim

Exercici

- Tenim la següent relació
- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- Ens diuen el següent
 - una editorial pot publicar N llibres, però un llibre és publicat únicament per una editorial
 - El camp edició és per indicar l'edició de llibres amb mateix títol (edició 1, 2...)
- Contesta a les següents preguntes:

Exercici

- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- (ISBN, títol, edició, gènere) → codi_editorial
 - Completa o parcial?
 - És una dependència funcional PARCIAL. Amb ISBN tenim prou – un llibre és publicat per una única editorial
- (títol, edició) → gènere
 - Completa o parcial?
 - És una dependència funcional PARCIAL. Amb el títol tenim prou.
 - Amb edició, segur que NO; i és poc habitual que un llibre amb diferents edicions no pertanyi el mateix gènere

Exercici

- T_{LLIBRES} (ISBN, títol, edició, gènere, codi_editorial)
- (títol, edició, gènere, codi_editorial) → ISBN
 - Completa o parcial?
 - És una dependència funcional PARCIAL. Necessitem gènere?
 - El codi_editorial no és necessari ja que dos editorials no publiquen el mateix llibre (en aquest cas concret)
- (ISBN) → (títol, edició, gènere, codi_editorial)
 - Completa o parcial?
 - És una dependència funcional COMPLETA
 - A més, **el determinant és mínim**

TEMA 5

NORMALITZACIÓ

Objectius

- 1) Conèixer la normalització
- 2) Identificar redundància de dades i anomalies d'actualització
- 3) Identificar dependències funcionals
- 4) Distingir entre dependències funcionals completes i parcials
- 5) **Passar una relació de 1FN a 3FN**

Formes (nivells) normals

- **1FN:** Primera Forma Normal
- **2FN:** Segona Forma Normal
- **3FN:** Tercera Forma Normal
- **Altres**
 - Una variant de la 3FN és la BCNF (Boyce-Codd Normal Form)
 - 4FN
 - 5FN

1FN

- És **obligatòria** per a que un model relacional sigui vàlid
- Les altres formes normals són opcionals, però (molt) recomanables

Una relación está en 1FN si la intersección de cada fila y columna contiene únicamente un valor.

2FN

Diremos que una relación está en 2FN si está en 1FN y cada atributo que no forma parte de la clave primaria tiene dependencia funcional completa de la clave primaria.

Observació: Si una relació està en 1FN i la seva clau primària està formada per un únic atribut, llavors automàticament està en 2FN

2FN

- Si tenim una relació $R(\underline{A}, \underline{B}, N, O, P)$, amb clau primària $(\underline{A}, \underline{B})$
- llavors
- cap dels atributs que no són part de la clau primària (N, O, P) pot estar determinat per únicament A o per B

2FN

- Passar de 1FN a 2FN
 - Pas 1) Identificar clau primària si no ens la donen
 - Pas 2) Trobar dependències funcionals parcials respecte a la clau primària
 - Pas 3) Crear una taula per cada dependència funcional parcial respecte a la clau primària del pas 2.
 - Eliminar atributs que no siguin part de la clau primària de la taula original i afegir-los a una nova relació
 - Afegir una còpia del determinant de la DF en aquesta nova relació
 - Pas 4) Com a resultat del pas 3, la taula amb la clau primària es fa petita

2FN

- Considera la següent relació
- ClientRental (clientNo, propertyNo, cName, pAddress, rentStart, rentFinish, rent, ownerNo, oName)
- Clau primària (clientNo, propertyNo)
- Exercici: passa aquesta relació a 2FN

2FN: abans de començar

- Ens diuen la clau primària – no l'hem de buscar
- La clau primària està formada per dos atributs. Llavors, la relació pot no estar en 2FN
- Si la clau estigués formada per un únic atribut, llavors la relació ja estaria en 2FN – per definició

2FN

- ClientRental (clientNo, propertyNo, cName, pAddress, rentStart, rentFinish, rent, ownerNo, oName)
 - **Trobar dependències funcionals parcials respecte a la clau primària**

DF1: (clientNo, propertyNo) \rightarrow (rentStart, rentFinish). La clave primaria de la relación CLIENTRENTAL es el determinante de la dependencia funcional.

DF2: clientNo \rightarrow cName. Es una dependencia funcional, porque dado el identificador de un cliente, puedo obtener el nombre del cliente, pero es una dependencia funcional parcial con respecto a la clave primaria, porque contiene uno de los dos atributos que la forman.

DF3: propertyNo \rightarrow (address, rent, ownerNo, oName). Es una dependencia funcional, porque dado el identificador de propiedad, puedo identificar unívocamente, y a lo sumo, la dirección, alquiler, código y nombre del propietario. Sin embargo, es una dependencia funcional parcial con respecto a la clave primaria, porque el determinante tiene uno de los atributos que la forman.

2FN

- ClientRental (clientNo, propertyNo, cName, pAddress, rentStart, rentFinish, rent, ownerNo, oName)
- Crear una taula per cada dependència funcional parcial respecte a la clau primària
 - Eliminar atributs que no siguin part de la clau primària de la taula original i afegir-los a una nova relació
 - Afegir una còpia del determinant de la DF en aquesta nova relació
- CLIENT (clientNo, cName)
- PROPERTYOWNER (propertyNo, address, rent, ownerNo, oName)
- CLIENTRENTAL(clientNo, propertyNo, rentStart, rentFinish)

2FN

- ClientRental (clientNo, propertyNo, cName, pAddress, rentStart, rentFinish, rent, ownerNo, oName)
 - Com a resultat, la taula principal es fa petita.
 - Com a resultat, els atributs que no són part de la clau primària tenen dependència funcional completa de la clau primària
- CLIENT (clientNo, cName)
- PROPERTYOWNER (propertyNo, address, rent, ownerNo, oName)
- CLIENTRENTAL(clientNo, propertyNo, rentStart, rentFinish)

2FN

- ClientRental (clientNo, propertyNo, cName, pAddress, rentStart, rentFinish, rent, ownerNo, oName)
 - Les noves taules (CLIENT, PROPERTYOWNER) estan en 2FN
 - La taula principal (CLIENTRENTAL) també està en 2FN
- CLIENT (clientNo, cName)
- PROPERTYOWNER (propertyNo, address, rent, ownerNo, oName)
- CLIENTRENTAL(clientNo, propertyNo, rentStart, rentFinish)

3FN

- La 3FN es centra en eliminar les dependències transitives

Si $A \rightarrow B$, y $B \rightarrow C$, entonces diremos que **C depende transitivamente** de A vía B.

- Una relació està en 3FN si està en 2FN i els atributs que no formen part de la clau primària no tenen dependència transitiva de la clau primària

3FN

- Pas de 2FN a 3FN
 - Significa eliminar dependències transitives.
 - Pas 1. Identificar dependències transitives
 - Pas 2. Per cada dependència transitiva
 - Eliminar atribut o atributs que depenen transitivament (és a dir, si $A \rightarrow B$, i $B \rightarrow C$, l'atribut a eliminar és C)
 - Aquest atribut (atributs) el/s col·loquem a una nova relació, amb còpia del determinant de la DF (en aquest cas, el determinant és B)

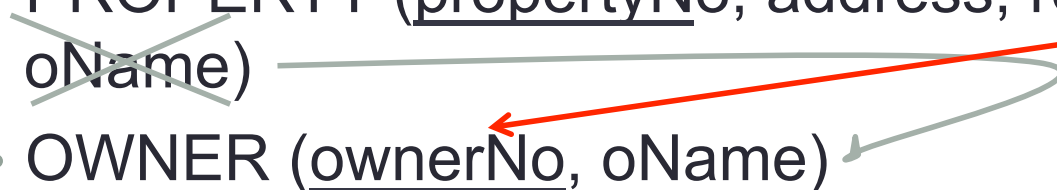
3FN

- CLIENT (clientNo, cName)
 - PROPERTYOWNER (propertyNo, address, rent, ownerNo, oName)
 - CLIENTRENTAL(clientNo, propertyNo, rentStart, rentFinish)
-
- Pas 1. Identificar dependències functionals transitives
 - CLIENT no té
 - CLIENTRENTAL no té
 - PROPERTYOWNER si que té
 - **propertyNo** -> address, rent, **ownerNo**, oName
 - **ownerNo** -> **oName**

3FN

- Pas 1. Identificar dependències functionals transitives
 - CLIENT no té
 - CLIENTRENTAL no té
 - PROPERTYOWNER si que té
 - **propertyNo** -> address, rent, **ownerNo**, oName
 - **ownerNo** -> **oName**
- L'atribut que depèn transitivament és oName.
- Apliquem el Pas 2

3FN

- Pas 2
 - **propertyNo** -> address, rent, **ownerNo**, oName
 - **ownerNo** -> oName
 - ~~PROPERTY (propertyNo, address, rent, ownerNo, oName)~~
 - OWNER (ownerNo, oName)
 - CLIENT (clientNo, cName)
 - CLIENTRENTAL(clientNo, propertyNo, rentStart, rentFinish)
- 

Exercici

- Passar a 3FN la següent relació
- **STUDENT_ACTIVITY** (studentID, Activity, ActivityFee)
- La clau primària és (studentID, activity)
- Un estudiant pot participar en N activitats (nadar, jugar al golf...)

Exercici

- **STUDENT_ACTIVITY** (studentID, Activity, ActivityFee)
- Està en 2FN? La clau primària està formada per dos atributs, llavors em toca examinar depedèndies funcionals respecte aquesta clau
- DF1: (studentId, activity) -> activityFee
- DF2: (activity) -> activityFee
- DF2 és una dependència funcional parcial respecte a la clau primària...

Exercici

- **STUDENT_ACTIVITY (studentID, Activity, ActivityFee)**
- DF1: (studentId, activity) -> activityFee
- DF2: (activity) -> activityFee
- STUDENT_ACTIVITY(studentId, activity)
- ACTIVITY (activity, activityFee)
- Ara les tenim en 2FN

Exercici

- **STUDENT_ACTIVITY(studentId, activity)**
- **ACTIVITY (activity, activityFee)**
- Està en 3FN? Tenim dependències funcionals transitives?

Exercici

- Passa a 3FN la següent relació
- **STUDENT_HOUSING** (studentID, building, housingfee)
- La clau primària és studentID
- HousingFee significa el “lloguer” per viure en un “building”
- Aquesta relació indica que 1 estudiant viu a 1 building. I a 1 building poden viure N estudiants

Exercici

- STUDENT_HOUSING (studentID, building, housingfee)
- Està en 2FN?
- La clau primària està formada per únicament un atribut, llavors ja la tenim en 2FN
- Us deixo el raonament a vosaltres

Exercici

- STUDENT_HOUSING (studentID, building, housingfee)
- Està en 3FN?
- Per contestar, hem de trobar dependències funcionals transitives
- DF1: studentId -> building
- DF2: building -> housingfee
- Hem trobat una DFT

Exercici

- STUDENT_HOUSING (studentID, building, housingfee)
- DF1: studentId -> building
- DF2: building -> housingfee
- **STUDENT_HOUSING (studentID, building)**
- **BUILDING (building, housingfee)**