

Tema 6: Teoria de la Normalització en BD relacionals

Bases de Dades

Marta Tarrés-Puertas

Enginyeria de Sistemes TIC
Universitat Politècnica de Catalunya
<http://itic.cat>

27 d'abril de 2015

1 Consultes amb valors nuls

2 Teoria de la Normalització

- Primera Forma Normal (1FN)
- Segona Forma Normal (2FN)
- Tercera Forma Normal (3FN)
- Forma Normal de Boyce-Codd
- Quarta Forma Normal
- Quarta Forma Normal
- Cinquena forma normal
- Cinquena forma normal

Penseu en la consulta de valors que pren un atribut d'una taula de mil tuples de les quals 750 tenen valor nul en aquest camp: es processaran mil tuples per obtenir només 250 valors resultants.

- Els valors nuls poden ser un problema, no tant per l'espai ocupat, com pel temps invertit en les consultes.
- Per solucionar aquest punt es pot crear una nova taula que no contindrà les tuples amb valors nuls. Aquesta solució és cara tant en disseny com en l'execució per part del SGBD (representa una nova taula), i només té sentit d'aplicar-la quan el nombre de valors nuls sigui elevat respecte dels valors possibles.

Associació binària 1:N (estudi sobre valors nuls)
Modelar la relació de pertanyença

- Un empleat pot pertànyer a 0..1 departaments
- Un departament pot tenir N empleats

- Solució 1:
departament(dep,...)
empleats(emp,...,dep) on dep pot tenir valors nuls

- Solució 2:
departament(dep,...)
pertanyença(emp,dep,...) dep sense valors nuls
empleats(emp,...)

Associació binària 1:1 (estudi sobre valors nuls)
Modelar la relació de Manager

- Un empleat pot tenir de 0..1 manager
- Un departament pot tenir 0..1 empleats empleats
- Solució 1:
departament(dep,...)
empleats(emp,... ,dep) on dep pot tenir valors nuls
- Solució 2:
departament(dep,... ,emp) on emp pot tenir valors nuls
empleats(emp,...)
- Solució 3:
departament(dep,...)
cap(emp,dep,...) dep sense valors nuls
empleats(emp,...)

Dependències funcionals

- Una **dependència funcional** és una restricció sobre una relació amb esquema $R(A_1, A_2, \dots, A_n)$ denotada com a $\{X\} \rightarrow \{Y\}$, on $\{X\}$ i $\{Y\}$ són subconjunts de $\{A_1, A_2, \dots, A_n\}$, que garanteix que donat un valor de $\{X\}$, aquest determina de manera única el valor de $\{Y\}$.
- Cada valor de $\{X\}$ té associat un, i només un, valor de $\{Y\}$. En aquest cas, diem que $\{Y\}$ depèn funcionalment de $\{X\}$ o, alternativament, $\{X\}$ determina funcionalment $\{Y\}$.
- La clau primària d'una relació sempre determina funcionalment la resta d'atributs de la relació. Aquesta conclusió es pot estendre a totes les claus alternatives que la relació pugui tenir.
- Una dependència funcional $\{X\} \rightarrow \{Y\}$ és **dependència funcional completa** quan cap subconjunt propi de $\{X\}$ determina funcionalment $\{Y\}$.

Subministraments			
CodiProv	CodiArticle	Quantitat	CiutatProv
1	a1	100	Manresa
1	a2	150	Manresa
2	a1	200	Barcelona
2	a2	250	Barcelona
3	a2	100	Barcelona

Subministraments(CodiProv,CodiArticle,Quantitat,CiutatProv)

Detecció de dependències funcionals

Anomalies de disseny

- 1** Anomalies de **modificació**: obliguen a modificar totes les tuples que guarden un fet determinat, si aquest canvia. Exemple. El proveïdor 2 es trasllada a Vic.
- 2** Anomalies **d'esborrament**: Pèrdua de dades. El proveïdor 3 només proporciona un codi de producte i deixa de subministrar-lo. Caldria esborrar-lo de la taula de Subministraments. Pèrdua de dades (CodiProv i CiutatProv)
- 3** Anomalies **d'inserció**: No es poden inserir fets elementals de manera independent. No podem emmagatzemar dades personals d'un proveïdor, exemple, CodiProv=5 i CiutatProv='Lleida' si no sabem els articles que subministra, donat que hauriem d'afegir un CodiArticle=null i no es compliria la regla d'integritat d'entitat de la clau primària.

Subministraments			
CodiProv	CodiArticle	Quantitat	CiutatProv
1	a1	100	Manresa
1	a2	150	Manresa
2	a1	200	Barcelona
2	a2	250	Barcelona
3	a2	100	Barcelona

Tota relació que no representa un concepte (o fet elemental) únic del món real està subjecta a presentar redundàncies, anomalies de manteniment i inconsistències potencials.

Tota relació ha de descriure un concepte semàntic únic.
Estudiarem les 6 formes normals més importants.

1FN: Valors atòmics

- Una relació està en primera forma normal (1FN) si, i només si, cap atribut de la relació és en si mateix una relació. És a dir, si tot atribut de la relació és atòmic (no hi ha d'haver més d'un valor), no descomponible (exemple: nom i cognoms en un mateix camp), no grup repetitiu (no grups dades repetides).
- El model relacional sempre garanteix que les relacions estan en 1FN, atès que només hi ha una estructura per a representar les dades (la relació) i, a més, cada dada es representa uniformement mitjançant valors atòmics.

Subministraments			
CodiProv	CodiArticle	Quantitat	CiutatProv
1	a1	100	Manresa
	a2	150	
2	a1	200	Barcelona
	a2	250	
3	a2	100	Barcelona

1FN: Aplanar

Subministraments			
CodiProv	CodiArticle	Quantitat	CiutatProv
1	a1	100	Manresa
1	a2	150	Manresa
2	a1	200	Barcelona
2	a2	250	Barcelona
3	a2	100	Barcelona

2FN: Dependències funcionals no completes

- 1 Una relació està en **segona forma normal (2FN)** si, i només si, està en 1FN i tot atribut no clau depèn funcionalment en forma completa de la clau primària.
 - 2 Excepció: un atribut pot dependre funcionalment de part de la clau primària, si aquest atribut és part d'una clau alternativa.
 - 3 Les relacions amb clau primària simple (formada per un únic atribut) sempre estaran en 2FN.
- Una relació que no està en 2FN té **redundància**.
 - Tots els atributs que participen en la dependència funcional no completa s'hauran de projectar en una nova relació que correspon al concepte semàntic que representen.

Subministraments(CodiProv,CodiArticle,Quantitat,CiutatProv)

Subministraments(CodiProv,CodiArticle,Quantitat)

CiutatsProveïdors(CodiProv,Ciutat)

3FN: Dependències funcionals

- Una relació està en tercera forma normal (3FN) si, i només si, està en 2FN i cap atribut no clau depèn funcionalment de cap altre conjunt d'atributs no clau.
- Els atributs que participen en la dependència funcional s'hauran de projectar en una nova relació que correspon al concepte semàntic que representen.

Clients(CodiProv,Carrer,Numero,Ciutat,Provincia)

Clients(CodiProv,Carrer,Numero,Ciutat)

Ciutats(Ciutat,Provincia)

FNBC: introducció intuïtiva

Notes(DniAlumne,CodiAssig,CodiMatr,Nota)

- Està en 1FN
- Està en 2FN (hi ha dependència funcional no completa de part de la clau cap a un atribut no clau. Aquesta dependència involucra un atribut no clau que forma part d'una clau alternativa de la relació, s'aplica l'excepció)
- Està en 3FN

Notes			
DniAlumne	CodiAssig	CodiMatr	Nota
23232323	05001	123	A
23232323	04002	123	B
45454545	05002	312	B
45454545	05001	312	B
45454545	04002	312	C

- Redundàncies
- Anomalies de modificació

Donada una dependència funcional $\{X\} \rightarrow \{Y\}$, un **determinant** és el conjunt $\{X\}$.

Una relació està en **forma normal de Boyce-Codd (FNBC)** si, i només si, està en 1FN, i si tots els determinants són clau candidata de la relació.

Notes(DniAlumne,CodiAssig,CodiMatr,Nota)

- 1 Notes(DniAlumne,CodiAssig,Nota) i Alumnes(DniAlumne,CodiMatr)
- 2 Notes(DniAlumne,CodiAssig,Nota) i Alumnes(CodiMatr,DniAlumne)
- 3 Notes(CodiMatr,CodiAssig,Nota) i Alumnes(CodiMatr,DniAlumne)
- 4 Notes(CodiMatr,CodiAssig,Nota) i Alumnes(DniAlumne,CodiMatr)

Les opcions més naturals a triar són la primera o la segona, atès que la tercera i darrera opció són mixtes. Des del punt de vista de la secretaria del centre, possiblement la segona opció seria la més correcta.

- Sempre és possible normalitzar fins a FNBC
- El procés de normalització no és únic
- Com a conseqüència de la normalització, el model lògic final és millor que l'inicial atès que
 - Elimina redundàncies i anomalies
 - Separa fets semànticament diferents

4FN: dependència multivaluada independent

ConeixProg		
DNI	LlengProg	Idiomes
74123492	Java	Anglès
	Python	Català
11234561	C++	Català
	Java	Castellà
33234123	Python	Anglès
		Castellà
		Francès

4FN: dependència multivaluada independent

ConeixProg(DNI,LlengProg,Idiomes)

ConeixProg		
DNI	LlengProg	Idiomes
74123402	Java	Anglès
74123402	Java	Català
74123402	Python	Anglès
74123402	Python	Català
11234561	C++	Català
11234561	C++	Castellà
11234561	Java	Català
11234561	Java	Castellà
33234123	Python	Anglès
33234123	Python	Castellà
33234123	Python	Francès

- Hi ha redundàncies
- La relació recull dos fets multivaluats: els llenguatges de programació i els idiomes
- Els dos fets són independents entre sí

4FN: dependència multivaluada independent

Sigui R una relació amb esquema (A_1, A_2, \dots, A_n) i siguin $\{X\}, \{Y\}, \{Z\}$ atributs de R . Aleshores, la **dependència multivaluada independent** $\{X\} \twoheadrightarrow \{Y\}$ existeix si, i només si, el conjunt de valors possibles de $\{Y\}$ per a un parell $\{X, Z\}$ depèn únicament del valor de $\{X\}$ i és independent del valor de $\{Z\}$. Si existeix $\{X\} \twoheadrightarrow \{Y\}$, llavors també existirà la dependència multivaluada independent $\{X\} \twoheadrightarrow \{Z\}$.

- 1 Una relació només pot tenir dependències multivaluades independents si la relació té clau composta formada, com a mínim, per tres atributs
- 2 En una relació, les dependències multivaluades independents sempre es donen per parelles, és a dir, si $\{X\} \twoheadrightarrow \{Y\}$ independentment de $\{Z\}$, llavors també $\{X\} \twoheadrightarrow \{Z\}$, independentment de $\{Y\}$.

4FN: dependència multivaluada independent

- Una relació està en 4FN si, i només si, està en FNBC i no té dependències multivaluades independents.
- Per aconseguir normalitzar una relació a 4FN, caldrà evitar les dependències multivaluades independents. Per tant, els atributs que participen en cada dependència multivaluada independent s'hauran de projectar en relacions diferents.

ConeixProg(DNI,LlengProg,Idiomes)

LlengProg(DNI,LlengProg,...)

IdiomesProg(DNI,Idiomes,...)

5FN: dependències de projecció-combinació

ProfAssigCentre		
Professor	Assignatura	Centre
Pons	BD	FIB
Pons	Programació	FIB
García	BD	FIB
García	Àlgebra	FIB
García	Àlgebra	FME
Puig	Àlgebra	FIB

5FN: dependències de projecció-combinació

Descomposició 1

ProfAssig	
Professor	Assignatura
Pons	BD
Pons	Programació
García	BD
García	Àlgebra
García	Àlgebra
Puig	Àlgebra

AssigCentre	
Assignatura	Centre
BD	FIB
Programació	FIB
Àlgebra	FIB
Àlgebra	FME

Descomposició 2

ProfAssig	
Professor	Assignatura
Pons	BD
Pons	Programació
García	BD
García	Àlgebra
García	Àlgebra
Puig	Àlgebra

ProfCentre	
Professor	Centre
Pons	FIB
García	FIB
García	FME
Puig	FIB

5FN: dependències de projecció-combinació

Descomposició 3

AssigCentre	
Assignatura	Centre
BD	FIB
Programació	FIB
Àlgebra	FIB
Àlgebra	FME

ProfCentre	
Professor	Centre
Pons	FIB
García	FIB
García	FME
Puig	FIB

Descomposició 4

AssigCentre	
Assignatura	Centre
BD	FIB
Programació	FIB
Àlgebra	FIB
Àlgebra	FME

ProfCentre	
Professor	Centre
Pons	FIB
García	FIB
García	FME
Puig	FIB

ProfAssig	
Professor	Assignatura
Pons	BD
Pons	Programació
García	BD
García	Àlgebra
García	Àlgebra
Puig	Àlgebra

5FN: dependències de projecció-combinació

- Quan una relació es pot descompondre en més de dues sense pèrdua d'informació, es diu que la relació no està en 5FN.
- Els fets multivaluats als quals es pot reduir la relació original són conseqüència d'una llei de simetria i constitueixen un nou tipus de dependència denominada **dependència de projecció-combinació**. El nom d'aquest tipus de dependència té l'origen en el fet que la relació original es pot reconstruir a partir de la combinació de les noves relacions en les quals la relació s'ha descompost.
- Una relació està en 5FN si, i només si, està en 4FN i no té dependències de projecció-combinació sense variació d'informació.

5FN: dependències de projecció-combinació

- Dependències de projecció-combinació sense variació: la relació no està en 5FN i, per tant, presenta redundàncies.
- Dependències de projecció-combinació amb variació: un cop descomposta la relació original en relacions noves, ja no es pot reconstruir. La relació original podia tenir repeticions, però aquestes són inevitables i, per tant, no són redundàncies. La relació original representava un únic fet del món real i ja estava en 5FN.

5FN: dependències de projecció-combinació

- Una relació només pot no complir les formes normals 4FN i 5FN si s'ha obtingut com a conseqüència d'una interrelació ternària com a mínim. Per tant, la seva clau primària ha d'estar formada com a mínim per 3 atributs.
- Com que una relació només pot estar en 5FN si també està en 4FN, les dependències de projecció-combinació són un cas particular de les dependències multivaluades dependents.
- Una relació que està en 5FN pot presentar repeticions en el seu contingut, però aquestes són inevitables i, per tant, no són redundàncies.