Normalizarea Bazelor de Date Relationale (partea 1)

Ioana Ciuciu <u>ioana.ciuciu@ubbcluj.ro</u> http://www.cs.ubbcluj.ro/~oana/



Planificare

Saptama na	Curs	Seminar	Laborator	
SI	I. Concepte fundamentale ale bazelor de date. Modelare conceptuala	I. Modelul Entitate-Relatie. Modelul relational	I. Modelarea unei BD in modelul ER si implementarea ei in SQL Server	
S2	2. Modelul relational de organizare a bazelor de date. Modelare conceptuala			
S3	3. Gestiunea bazelor de date relationale cu limbajul SQL (DDL)	2. Limbajul SQL – definirea si actualizarea datelor	2. Interogari SQL	
S4	4. Gestiunea bazelor de date relationale cu limbajul SQL (DML)			
S5-6	5-6. Dependente functionale, forme normale	3. Limbajul SQL – regasirea datelor	3. Interogari SQL avansate	
S7	7. JDBC (Java Database Connectivity)	4. Proceduri stocate	4. Proceduri stocate. View. Trigger	
\$8	8. Interogarea bazelor de date relationale cu operatori din algebra relationala			
S9	9. Structura fizica a bazelor de date relationale	5. View-uri. Functii definite de utilizator. Trigger		
S10-11	10-11. Indecsi. Arbori B. Fisiere cu acces direct	6. Formele normale ale unei relatii. Indecsi		
S12	12. Evaluarea interogarilor in bazele de date relationale			
\$13	13. Extensii ale modelului relational si baze de date NoSQL	7. Probleme	Examen practic	
\$14	14. Aplicatii			

Planul cursului

Curs 5:

- Normalizarea bazelor de date relationale
- Prima forma normala (INF)
- ▶ A 2-a forma normala (2NF)
- ▶ A 3-a forma normala (3NF)

Curs 6:

- Forma normala Boyce-Codd (BCNF)
- A 4-a forma normala(4NF)
- ► A 5-a forma normala (5NF)



- Normalizarea bazelor de date este procesul organizarii relatiilor (si atributelor) unei baze de date relationale cu scopul de a reduce redundanta si de a imbunatati integritatea datelor
- Normalizarea garanteaza faptul ca dependentele intre atribute sunt corect impuse de catre constrangerile de integritate ale bazei de date
- Normalizarea se realizeaza prin procesul de descompunere
- **Scopul**: acela de a face posibile actualizarea si interogarea bazei de date intr-o maniera cat mai simpla



- ▶ De ce este necesara normalizarea?
 - Fie urmatoarea relatie:

PRODUCT	CLIENT	ADDRESS	QUANTITY	AMOUNT
Lotion	Martin	Paris	10	250
Polish	Martin	Paris	250	2500
Cream	Martin	Paris	20	1000
Lotion	John	London	30	750
Cream	Dupont	Lyon	20	1000
Lotion	Dubois	Paris	20	500

Singura cheie posibila pentru relatie este: PRODUCT-CLIENT



De ce este necesara normalizarea?

- O astfel de relatie, chiar daca corecta din punctul de vedere al unui SGBD, va pune probleme de mentenanta si ar putea determina anomalii
 - Adresa clientului se repeta pentru fiecare comanda
 - Actualizarea unei adrese impune accesarea BD de mai multe ori
 - ☐ Anomalii posibile

```
Exemplu: La inserarea urmatoarei linii
```

<Cream, John, New-York, 10, 500>

Sistemul nu va detecta faptul ca John are doua adrese (ceea ce nu este

realist)



De ce este necesara normalizarea?

- O astfel de relatie, chiar daca corecta din punctul de vedere al unui SGBD, va pune probleme de mentenanta si ar putea determina anomalii
 - Adresa clientului se repeta pentru fiecare comanda
 - Actualizarea unei adrese impune accesarea BD de mai multe ori
 - ☐ Anomalii posibile

Exemplu: La inserarea urmatoarei linii

<Cream, John, New-York, 10, 500>

Sistemul nu va detecta faptul ca John are doua adrese (ceea ce nu este

realist)

PRODUCT	CLIENT	ADDRESS	QUANTITY	AMOUNT
Lotion	Martin	Paris	10	250
Polish	Martin	Paris	250	2500
Cream	Martin	Paris	20	1000
Lotion	John	London	30	750
Cream	Dupont	Lyon	20	1000
Lotion	Dubois	Paris	20	500
Cream	John	New-York	10	500

De ce este necesara normalizarea?

- O astfel de relatie, chiar daca corecta din punctul de vedere al unui SGBD, va pune probleme de mentenanta si ar putea determina anomalii
 - Adresa clientului se repeta pentru fiecare comanda
 - Actualizarea unei adrese impune accesarea BD de mai multe ori
 - ☐ Anomalii posibile

Exemplu: La inserarea urmatoarei linii

<Cream, John, New-York, 10, 500>

Sistemul nu va detecta faptul ca John are doua adrese (ceea ce nu este realist)

Daca, de ex., Dupont isi anuleaza comanda, numele si adresa sa sunt pierdute

PRODUCT CLIENT ADDRESS QUANTIT AMOUNT

OUANTIT PRODUCT CLIENT ADDRESS AMOUNT Martin 10 250 Lotion Paris **Polish** Martin Paris 250 2500 Cream Martin Paris 20 1000 Lotion John London 30 750 1000 **Dupont** 20 Cream Lyon Lotion Dubois Paris 500

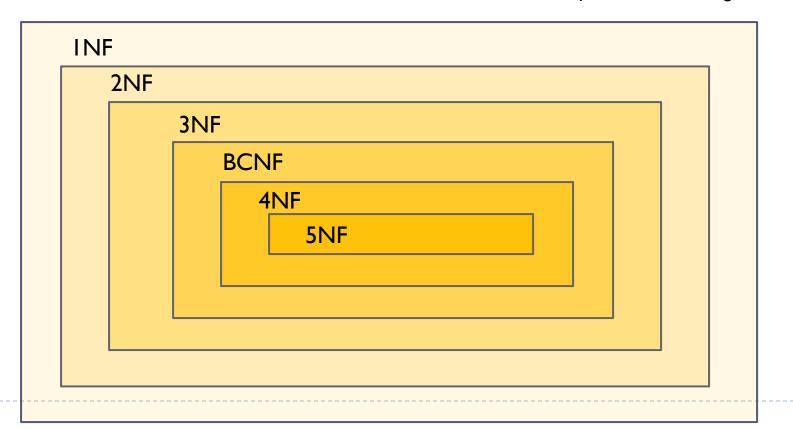


- Se cunosc mai multe forme normale pentru relatii
- Cele mai cunoscute sunt: INF, 2NF, 3NF, BCNF, 4NF, 5NF

Codd

Boyce-Codd

Fagin



Notiuni Preliminare

Pentru descompunerea unei relatii se foloseste operatorul de proiectie

Fie
$$R[A_1,A_2,...,A_n]$$
 o relatie si $\alpha=\{A_{i_1},A_{i_2},...,A_{i_p}\}$ o multime de atribute, unde $\alpha\subset\{A_1,A_2,...,A_n\}$. Prin proiectia relatiei R pe se intelege relatia:
$$R'\Big\lfloor A_{i_1},A_{i_2},...,A_{i_p}\Big\rfloor=\prod_{\alpha}(R)=\prod_{\{A_{i_1},A_{i_2},...,A_{i_p}\}}(R)=\{r[\alpha]\mid r\in R\}$$
 Unde:
$$\forall r=(a_1,a_2,...,a_n)\in R\Rightarrow \prod_{\alpha}(r)=r[\alpha]=(a_{i_1},a_{i_2},...,a_{i_p})\in R'$$

Pentru compunerea relatiilor se foloseste operatorul de join natural

Fie $R[\alpha, \beta], S[\beta, \gamma]$ doua relatii peste multimile de atribute precizate, $\alpha \cap \gamma = \emptyset$ Prin joinul natural al relatiilor R si S se intelege relatia:

$$R * S[\alpha, \beta, \gamma] = \left\{ \left(\prod_{\alpha}(r), \prod_{\beta}(r), \prod_{\gamma}(s) \right) | r \in R, s \in S \text{ and } \prod_{\beta}(r) = \prod_{\beta}(s) \right\}$$



Notiuni Preliminare

▶ O relatie R se poate descompune in mai multe relatii noi RI, R2, ..., Rn. Aceasta descompunere este buna daca R=RI*R2*...*Rn

Deci datele din R se pot regasi din datele memorate in relatiile R I, R2,...,Rn; nu apar si nu se pierd date cu aceste operatii de compunere (join natural)



Notiuni Preliminare

- Exemplu de descompunere care nu este buna
- Fie relatia ContracteStudiu(Student, Cadru didactic, Disciplina), si doua noi relatii obtinute prin proiectia acestei relatii:

SC(Student, Cadru didactic) and CD(Cadru didactic, Disciplina)

Presupunem ca in relatia initiala avem urmatoarele valori:

Student	Cadru didactic	Disciplina
SI	TI	DI
S2	T2	D2
SI	T2	D3



Notiuni Preliminare

Folosind definitia proiectiei, se obtin urmatoarele valori pentru cele doua relatii obtinute din R:

Student	Cadru didactic
SI	TI
S2	T2
SI	T2

Cadru didactic	Disciplina
TI	DI
T2	D2
T2	D3

Si pentru joinul natural al acestor relatii (SC*CD)

			_
Student	Cadru didactic	Disciplina	
SI	ΤI	DI	
S2	T2	D2	L
S2	T2	D3	
SI	T2	D2	
SI	T2	D3	

Se observa ca in relatia SC*CD se obtin inregistrari suplimentare fata de relatia initiala, deci descompunerea sugerata nu este buna



- ▶ Atribut simplu: un atribut oarecare din relatie
- Atribut compus: o multime de atribute (cu cel putin doua atribute) din relatie
- ▶ Atribut repetitiv: atribute (simple sau compuse) care iau mai multe valori (vectori) pentru o inregistrare din relatie
- ▶ Atribut atomic: atribut care nu se descompune in alte elemente
- Atribut scalar: atribut care nu se poate memora in tablouri



O relatie este in prima forma normala (INF) daca ea contine doar atribute atomice si nu contine atribute repetitive

Exemplu:

Contract(Sid, Snume, Cnume)

Obiectiv: gasiti toti studentii
care urmeaza cursul de C

(SQL+Aplicatie/program)

Sid	Snume	Cnume
I	Α	C,C++
2	В	C, Java, C#
•••	•••	



O relatie este in prima forma normala (INF) daca ea contine doar atribute atomice si nu contine atribute

repetitive

Exemplu:

Contract(Sid, Snume, Cnume)

Obiectiv: gasiti toti studentii
care urmeaza cursul de C

(SQL+ Aplicatie/program)

Solutia??

Sid	Snume	Cnume
I	Α	C,C++
2	В	C, Java, C#
•••	•••	•••

Sid	Snume	Cnume
I	Α	С
ı	Α	C++
2	В	С
2	В	Java
2	В	C#



O relatie este in prima forma normala (INF) daca ea contine doar atribute atomice si nu contine atribute

repetitive

Exemplu:

Contract(Sid, Snume, Cnume)

Obiectiv: gasiti toti studentii

care urmeaza cursul de C

(SQL+ Aplicatie/program)

Solutia??

REDUNDANTA-

Sid	Snume	Cnume
I	Α	C,C++
2	В	C, Java, C#
	•••	•••

	Sid	Snume	Cnume
	I	Α	С
	1	Α	C++
	2	В	С
>	2	В	Java
	2	В	C#



O relatie este in prima forma normala (INF) daca ea contine doar atribute atomice si nu contine atribute repetitive

Exemplu:

Contract(Sid, Snume, Cnume)

Obiectiv: gasiti toti studentii care urmeaza cursul de C

(SQL+ Aplicatie/program)

Sid	Snume	Cnume
I	Α	C,C++
2	В	C, Java, C#
•••	•••	

Solutia2: descompunerea tabelei in doua tabele:

Student(Sid, Snume)

Contract(Sid, Cnume)

Sid	Snume
I	Α
2	В

Sid	Cnume	
I	С	
1	C++	
2	С	
2	Java	
2	C#	

Notiuni Preliminare

Dependenta functionala (DF): Fie R o relatie si fie X si Y doua submultimi arbitrare de atribute din R. Atributul (simplu sau compus) Y este dependent functional de atributul X (simplu sau compus), notatie:

$$X \rightarrow Y$$

daca si numai daca fiecare valoare a lui X din R are asociata o valoare precisa si unica pentru Y.

- valoarea X se numeste determinant
- valoarea Y se numeste determinat

Observatii

- Dependenta functionala se poate folosi ca proprietate (restrictie) pe care BD trebuie sa o indeplineasca: se pot adauga sau modifica valori in relatie numai daca DF este verificata
- Stabilirea DF este o sarcina a administratorului BD
- DF depind de semnificatia (semantica) datelor care se memoreaza in relatie
- Umatoarele trei forme normale (i.e., 2NF, 3NF, BCNF) utilizeaza notiunea de dependenta functionala
- Cand o relatie contine cel putin o DF, aceasta poate fi doar in INF, 2NF, 3NF, sau BCNF



Notiuni Preliminare

Proprietatile si axiomatizarea DF (axiomele lui Armstrong)

Fie X,Y, Z si W multimi de atribute din relatia R

Reflexivitatea

Daca Y este submultime a lui X, atunci $X \rightarrow Y$

Marirea:

Daca $X \rightarrow Y$, atunci $XZ \rightarrow YZ$

Tranzitivitatea:

Daca $X \rightarrow Y$ si $Y \rightarrow Z$, atunci $X \rightarrow Z$

Auto-determinarea

 $X \rightarrow X$

Reuniunea:

Daca $X \rightarrow Y$ si $X \rightarrow Z$, atunci $X \rightarrow YZ$

Decompunerea:

Daca $X \rightarrow YZ$, atunci $X \rightarrow Y$ si $X \rightarrow Z$

Pseudotranzitivitatea:

Daca $X \rightarrow Y$ si $WY \rightarrow Z$, atunci $WX \rightarrow Z$

Compunerea:

Daca $X \rightarrow Y$ si $Z \rightarrow W$, atunci $XZ \rightarrow YW$

- Inchiderea unei multimi de dependente functionale: multimea tuturor DF obtinute logic din multimea S de DF se numeste inchiderea multimii S, notata S+
 - (S⁺ se poate obtine folosind axiomele lui Armstrong)
- Inchiderea unei multimi de atribute (sub o multime de DF):
 - In practica, nu e necesar sa se calculeze inchiderea unei multimi de DF;
 - Ceea ce se doreste este calculul unei anume submultimi a inchiderii:
 i.e., acea submultime continand toate DF avand o multime precizata
 Z de atribute in partea stanga (determinant);
 - Mai precis, inchiderea unei multimi de atribute dandu-se o relatie R, o multime Z de atribute din R si o multime S de DF valide pentru R, determina multimea tuturor atributelor din R care sunt functional dependente de Z − inchiderea Z⁺ lui Z sub R.



- Inchiderea unei multimi de atribute:
 - Algoritm de determinare a inchiderii Z^+ a unei multimi Z sub R (fie S o multime de DF sub R).



- Acoperire minimala de dependente functionale:
 - O multime F de DF este acoperire minimala daca si numai daca indeplineste urmatoarele trei proprietati:
 - Partea dreapta a fiecarei DF din F este formata dintr-un singur atribut
 - 2. Partea stanga a fiecarei DF din F este la randul ei minimala
 - 3. Reducand oricare DF se va schimba continutul lui F



- ▶ Atribut prim: atribut inclus in cheie
- Atribut neprim: atribut care nu este inclus in nici o cheie
- Dependenta partiala
 - X → a este dependenta partiala
 - Daca X submultime proprie a unei chei candidate
 - a este atribut neprim



Anomalii cauzate de dependentele functionale

Fie urmatorul exemplu

Examen (Snume, Disciplina, Nota, Cadru didactic)

Unei discipline ii corespunde un singur profesor, dar unui profesor ii corespund mai multe discipline, deci {Disciplina} → {Cadru didactic}

Snume	Disciplina	Nota	Cadru didactic
Bob	Baze de date	10	Boyce
Alice	Sisteme de operare	9	Codd
Bob	Sisteme de operare	8	Codd
John	Baze de date	9	Boyce
Mike	Baze de date	10	Boyce

Daca pastram o astfel de DF, pot aparea urmatoarele probleme, care necesita efort de programare aditional pentru a o rezolva (SQL nu este suficient):



- Anomalii cauzate de dependentele functionale
 - Risipa de spatiu, redundanta: aceeasi informatie este stocata de mai multe ori (asocierea dintre Baze de date si Boyce e memorata de trei ori, si cea dintre Sisteme de operare si Codd de doua ori)

Snume	Disciplina	Nota	Cadru didactic
Bob	Baze de date	10	Воусе
Alice	Sisteme de operare	9	Codd
Bob	Sisteme de operare	8	Codd
John	Baze de date	9	Boyce
Mike	Baze de date	10	Воусе



- Anomalii cauzate de dependentele functionale
 - Anomalii la actualizare: actualizarea datelor dintro asociere implica aceeasi actualizare in toate asocierile corespunzatoare pentru a evita inconsistenta BD. De ex., daca schimbam valoarea din coloana Cadru didactic in prima inregistrare dar nu si din inregistrarile 4 si 5, actualizarea va introduce o eroare in relatie

Snume	Disciplina	Nota	Cadru didactic	
Bob	Baze de date	10	Meersman	
Alice	Sisteme de operare	9	Codd	
Bob	Sisteme de operare	8	Codd	inconsistenta
John	Baze de date	9	Boyce	
Mike	Baze de date	10	Boyce	



- Anomalii cauzate de dependentele functionale
 - Anomalii la inserare: la inserarea unei noi inregistrari, valorile atributelor trebuie cunoscute (nu se accepta valori nedefinite pentru atributele implicate in DF)
 - Anomalii la stergere: la stergerea inregistrarilor, se pot elimina in mod ireversibil asocieri. De ex., daca inregistrarile 2 si 3 sunt sterse, asocierea dintre Disciplina (Sisteme de operare) si Cadru didactic (Codd) e pierduta

Snume	Disciplina	Nota	Cadru didactic
Bob	Baze de date	10	Boyce
Alice	Sisteme de operare	9	Codd
Bob	Sisteme de operare	8	Codd
John	Baze de date	9	Boyce
Mike	Baze de date	10	Boyce



Solutia

- Anomaliile determinate de DF se pot elimina pastrand dependentele in tabele separate
- Aceasta necesita ca tabela initiala sa fie descompusa printr-o descompunere buna
- O astfel de descompunere se face in stadiul de design al BD, cand se stabilesc DF



O tabela este in 2NF daca si numai daca:

- Este in INF si
- Orice atribut neprim (simplu sau compus) (deci care nu este inclus intr-o cheie) este complet dependent functional de oricare cheie a relatiei (nu exista dependente partiale)

Exemplu:

```
Examen(Snume, Disciplina, Nota, Cadru didactic) {Disciplina} → {Cadru didactic}
```



O tabela este in 2NF daca si numai daca

- Este in INF si
- Orice atribut neprim (simplu sau compus) (deci care nu este inclus intr-o cheie) este complet dependent functional de oricare cheie a relatiei (nu exista dependente partiale)

Exemplu:



Examen(Snume, Disciplina, Nota, Cadru didactic) {Disciplina} → {Cadru didactic} – dependenta partiala

Nu e 2NF



O tabela este in 2NF daca si numai daca:

- Este in INF si
- Nu exista dependente partiale

Example:



Examen(Snume, Disciplina, Nota, Cadru didactic)
{Disciplina} → {Cadru didactic} – dependenta partiala

Nu e 2NF

Solutia: Se elimina DF prin urmatoarea descompunere a relatiei initiale:

Rezultate(Snume, Disciplina, Nota)

DiscCadre(Disciplina, Cadru didactic)



- Doua sau mai multe atribute sunt mutual independente daca nici unul dintre ele nu este functional dependent de nici o combinatie a celorlalte
- O astfel de independenta asigura posibilitatea ca fiecare atribut sa poata fi actualizat separat de restul atributelor
- ▶ **Dependenta tranzitiva**: Un atribut Z este tranzitiv dependent de atributul X daca exista Y astfel incat:
 - $\square \times \rightarrow Y$
 - \square Y \rightarrow Z,
 - \square Y \rightarrow X nu se verifica
 - □ Z nu e inclus in reuniunea multimilor X si Y



O tabela este in 3NF daca si numai daca:

- ▶ Este in 2NF si
- Orice atribut neprim nu este tranzitiv dependent de oricare cheie a relatiei

Observatii:

- O relatie R care este in 2NF dar nu in 3NF are o dependenta $X \rightarrow Y$, X e neprim; aceasta dependenta se poate elimina descompunand R in mod similar ca in cazul 2NF
- Intr-o relatie 3NF
 - Nu exista nici o DF determinata de o submultime a cheii
 - Nu exista nici o DF intre doua atribute neprime

Exemplu: fie relatia

Lucrare_Licenta(Snume, Nota, Responsabil, Departament), unde

{Responsabil}

The relatia is a second of the second of t



O tabela este in 3NF daca si numai daca:

- Este in 2NF si
- Orice atribut neprim nu este tranzitiv dependent de oricare cheie a relatiei

```
Exemplu: fie relatia

Lucrare_Licenta(Snume, Nota, Responsabil, Departament), unde

{Responsabil} → {Departament}

Nu e 3NF
```



O tabela este in 3NF daca si numai daca:

- Este in 2NF si
- Orice atribut neprim nu este tranzitiv dependent de oricare cheie a relatiei

Exemplu: fie relatia

Lucrare_Licenta(Snume, Nota, Responsabil, Departament), unde

{Responsabil} → {Departament}

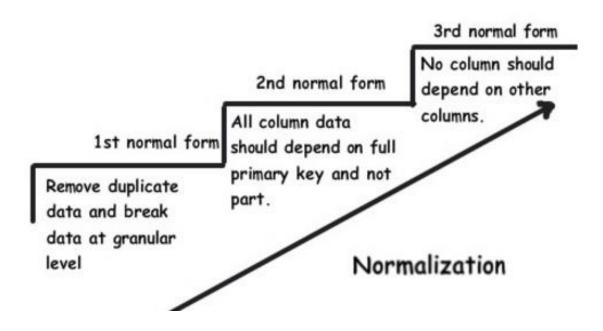
Nu e 3NF

Solutia: Eliminarea DF prin descompunerea urmatoare a relatiei initiale

Rezultate(Snume, Nota, Responsabil)
Cadre(Responsabil, Departament)



Concluzie



Sursa: http://www.xpode.com/ShowArticle.aspx?ArticleId=578



Exercitii

- I. Gasiti formele normale ale relatiilor urmatoare:
- ▶ RI (n°pacient, diagnostic, tratament)
- ▶ R2 (n°spital, n°pacient, nume spital, nume pacient)
- ▶ R3 (CNP, n°pacient, nume pacient, adresa pacient)
- ▶ R4 (n°pacient, n°reteta, nume pacient, tratament, cost)



Exercitii

- 2. Fie urmatoarea relatie:
- Cursuri (curs, cadru didactic, ora, sala, student, nota)

Fie urmatoarele DF pentru relatia Cursuri:

- ▶ curs → cadru didactic(cursul are un singur cadru did.)
- ora,sala → curs
 (un singur curs poate fi tinut intr-o anumita sala la o anumita ora)
- ▶ ora, cadru didactic → sala
 (un cadru did. poate fi intr-o singura sala la o anumita ora)
- ▶ ora, student → sala
 (un student poate fi intr-o singura sala la o anumita ora)
- curs, student → nota
 (un student primeste o singura nota la un curs)

- a) Care sunt cheile posibile ale relatiei Cursuri?
- b) Descompuneti relatia Cursuri in relatii care sa fie in 3NF



Exercitii

- 2. Fie urmatoarea relatie:
- Cursuri (curs, cadru didactic, ora, sala, student, nota)

Fie urmatoarele DF pentru relatia Cursuri:

- ▶ curs → cadru didactic(cursul are un singur cadru did.)
- ora,sala → curs
 (un singur curs poate fi tinut intr-o anumita sala la o anumita ora)
- ▶ ora, cadru didactic → sala
 (un cadru did. poate fi intr-o singura sala la o anumita ora)
- ▶ ora, student → sala
 (un student poate fi intr-o singura sala la o anumita ora)
- curs, student → nota
 (un student primeste o singura nota la un curs)

- a) Care sunt cheile posibile ale relatiei Cursuri?
- b) Descompuneti relatia Cursuri in relatii care sa fie in 3NF
 - ! SGBD nu permit
 existenta a doua elemente
 distincte intr-o relatie cu
 aceeasi valoare pentru
 oricare cheie (principala
 sau secundara)



Cursul 6

- Normalizarea bazelor de date relationale (partea 2)
 - Boyce-Codd Normal Form (BCNF)
 - A 4-a Forma Normala (4NF)
 - ▶ A 5-a Forma Normala (5NF)

References

- Date, C.J., An Introduction to Database Systems (8th Edition), Addison-Wesley, 2004, chapters 11, 12
- ▶ Garcia-Molina, H., Ullman, J., Widom, J., Database Systems: The Complete Book, Pearson Prentice Hall, 2008, chapter 3
- ▶ Leon Tambulea, UBB Cluj, Database course, course 5

