
Normalizarea Bazelor de Date Relationale (partea 2)

Ioana Ciuciu

oana@cs.ubbcluj.ro

<http://www.cs.ubbcluj.ro/~oana/>

Planificare

Saptamana	Curs	Seminar	Laborator
S1	1. Concepte fundamentale ale bazelor de date	1. Modelul Entitate-Relatie. Modelul relational	1. Modelarea unei BD in modelul ER si implementarea ei in SQL Server
S2	2. Modelare conceptuala		
S3	3. Modelul relational de organizare a bazelor de date	2. Limbajul SQL – definirea si actualizarea datelor	2. Interogari SQL
S4	4. Gestiunea bazelor de date relationale cu limbajul SQL		
S5-6	5-6. Dependente functionale, forme normale	3. Limbajul SQL – regasirea datelor	3. Interogari SQL avansate
S7	7. Interogarea bazelor de date relationale cu operatori din algebra relationala	4. Proceduri stocate	
S8	8. Structura fizica a bazelor de date relationale		
S9-10-11	9-10-11. Indeksi. Arbori B. Fisiere cu acces direct	5. View-uri. Functii definite de utilizator. Executie dinamica	4. Proceduri stocate. View. Trigger
S12	12. Evaluarea interogarilor in bazele de date relationale	6. Formele normale ale unei relatii. Indeksi	
S13	13. Extensii ale modelului relational si baze de date NoSQL	7. Probleme	Examen practic
S14	14. Aplicatii		

Planul cursului

▶ Curs 5:

- ▶ Normalizarea bazelor de date relationale
- ▶ Prima forma normala (1NF)
- ▶ 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)



Formele normale ale unei relatii

1NF

2NF

3NF

BCNF

4NF

5NF

Definitie (1NF). O relatie este de **prima forma normala** daca ea **nu contine attribute repetitive.**

Definitie (2NF). O relatie este de **a doua forma normala** daca:

- **Este de prima forma normala si**
- Orice **atribut** (simplu sau compus) **neprim** (deci care nu este inclus intr-o cheie) **este complet dependent functional de oricare cheie a relatiei.**

Definitie (3NF): O relatie este de **a treia forma normala** daca:

- **Este de 2NF**
- **orice atribut neprim nu este tranzitiv dependent de oricare cheie a relatiei.**

Descompunerea unei relatii: operatorul de **proiectie**

Compunerea relatiilor: operatorul de **join natural**

Exemple 1NF, 2NF, 3NF

▶ Exemplul I:

Fie relatia:

Student(Nume, AnulNasterii, Grupa, Disciplina, Nota)

- ▶ Cheia relatiei: {Nume}
- ▶ *Atribut compus repetitiv*: {Disciplina, Nota}

Nume	AnulNasterii	Grupa	Disciplina	Nota
Pop Ioana	1996	321	Baze de Date	10
			Sisteme de Operare	8
			Probabilitati	9
Muresan Silviu	1997	322	Baze de Date	9
			Sisteme de Operare	7
			Probabilitati si Statistica	8
			Proiect Individual	10



Exemple 1NF, 2NF, 3NF

- ▶ Exemplul I:
- ▶ Relatia Student se descompune in urmatoarele 2 relatii:

DateGenerale(Nume,AnulNasterii,Grupa)

Rezultate(Nume, Disciplina, Nota)

Nume	AnulNasterii	Grupa
Pop Ioana	1996	321
Muresan Silviu	1997	322

- ▶ Relatia este in 1NF
- ▶ 2NF,3NF?

Nume	Disciplina	Nota
Pop Ioana	Baze de Date	10
Pop Ioana	Sisteme de Operare	8
Pop Ioana	Probabilitati	9
Muresan Silviu	Baze de Date	9
Muresan Silviu	Sisteme de Operare	7
Muresan Silviu	Probabilitati si Statistica	8
Muresan Silviu	Proiect Individual	10

Exemple 1NF, 2NF, 3NF

▶ Exemplul 2:

Fie relatia urmatoare care memoreaza contractele de studiu:

Contracte(Nume, Prenume, CNP, CodDisciplina, DenumireDisciplina)

- ▶ Cheia relatiei: {CNP, CodDisciplina}
- ▶ Exista doua DF: {CNP}->{Nume, Prenume} si {CodDisciplina}->{DenumireDisciplina}

- ▶ Pentru eliminarea acestor dependente, se descompune relatia in urmatoarele trei relatii:

Studenti(CNP, Nume, Prenume)

Discipline(CodDisciplina, DenumireDisciplina)

Contracte(CNP, CodDisciplina)

- ▶ Relatia este acum in 3NF



Exemple 1NF, 2NF, 3NF

▶ Exemplul 3:

Fie următoarea relație, care reține adresele unui grup de persoane:

Adrese(CNP, Nume, Prenume, CodPostal, LocalitateDomiciliu, Strada, Nr)

- ▶ Cheia relației: CNP
- ▶ Există următoarea DF: {CodPostal} → {LocalitateDomiciliu}
 - ▶ Deoarece la unele localități codul postal se stabilește la nivel de stradă, sau chiar porțiuni de stradă
- ▶ Relația nu este în 3NF, deci este necesară descompunerea ei

Domiciliu(CNP, Nume, Prenume, CodPostal, Strada, Nr)

Localitate(CodPostal, LocalitateDomiciliu)



Boyce-Codd Normal Form (BCNF)

Notiuni Preliminare

Exemplu:

- ▶ Sa consideram urmatoarea relatie care memoreaza o eventuala planificare a studentilor pentru examene

Planificare_Ex(Data, Ora, Cadru_did, Sala, Grupa)

- ▶ cu cateva **restrictii** care se transpun prin **definirea de chei** sau prin respectarea unei/unor **dependente functionale**
 1. O grupa de studenti da maximum un examen intr-o zi, deci **{Data, Grupa} este cheie**
 2. Un cadru didactic are examen cu o singura grupa la o anumita data si ora, deci **{Cadru_did, Data, Ora} este cheie**
 3. La un moment dat (zi, ora) intr-o sala este planificat cel mult un examen, deci **{Sala, Data, Ora} este cheie**
 4. **Intr-o zi cadrul didactic nu schimba sala**, in sala respectiva pot fi planificate si alte examene, dar la alte ore, deci exista urmatoarea **dependenta functionala**:
{Cadru_did, Data} → {Sala}



Boyce-Codd Normal Form (BCNF)

Notiuni Preliminare

Exemplu:

- ▶ Sa consideram urmatoarea relatie care memoreaza o eventuala planificare a studentilor pentru examene

Planificare_Ex(Data, Ora, Cadru_did, Sala, Grupa)

- ▶ Toate attributele din aceasta relatie apar in cel putin o cheie, deci **nu exista attribute neprime**
- ▶ Putem deci spune ca relatia este in **3NF**, conform definitiilor precizate anterior
- ▶ Pentru a elimina si dependenta functionala amintita mai sus, s-a introdus o noua forma normala



Boyce-Codd Normal Form (BCNF)

O relatie este in **3NF Boyce-Codd** (sau **BCNF**) daca **orice determinant** (pentru o dependenta functionala) **este cheie**, deci **nu exista dependente functionale** $X \rightarrow Y$ **astfel incat** X **sa nu fie cheie**

Pentru a elimina dependenta functionala din exemplul de mai sus, $\{\text{Cadru_did}, \text{Data}\} \rightarrow \{\text{Sala}\}$, trebuie sa facem urmatoarea descompunere pentru relatia $\text{Planificare_Ex}(\text{Data}, \text{Ora}, \text{Cadru_did}, \text{Sala}, \text{Grupa})$:

$\text{Planificare_Ex}(\underline{\text{Data}}, \underline{\text{Cadru_did}}, \text{Ora}, \text{Grupa})$

$\text{Repartizare_Sali}(\underline{\text{Cadru_did}}, \underline{\text{Data}}, \text{Sala})$

Dupa aceasta descompunere nu mai exista DF, deci relatiile sunt de tipul **BCNF**, **dar a disparut cheia asociata restrictiei precizate la punctul 3** de mai sus: $\{\text{Sala}, \text{Data}, \text{Ora}\}$

Daca se mai doreste pastrata o astfel de restrictie, atunci ea trebuie verificata altfel (de exemplu, prin program)



Consecinte metodologice

- ▶ **Fiecare DF** trebuie sa se traduca printr-o **relatie**
- ▶ Intr-un SGBD relational, singura modalitate de a lua in considerare o **DF este de a declara determinantul** acesteia ca si **cheie** a unei relatii
- ▶ Se interzice astfel duplicarea valorilor identice si se asigura coerenta valorilor stocate in BD



Consecinte metodologice

- ▶ Cand o relatie contine cel putin o DF, aceasta nu poate fi decat de 1NF, 2NF, 3NF sau de BCNF



Exercitii

- ▶ Gasiti formele normale ale relatiilor urmatoare si propuneti o descompunere daca e necesar:
 - ▶ R1(nr_client,nr_produc, cantit_comandata, nume_produc)
 - ▶ R2(nr_comanda,nr_produc, cantit_comandata)
 - ▶ R3(nr_client, nume_client, nume_reprezentant)
 - ▶ R4(nr_produc, nume_produc, nr_atelier, nume_sef_atelier)
 - ▶ R5(nr_client, nume_client, nr_reprezentant, nume_reprezentant)
 - ▶ R6(nr_produc, nr_furnizor, pret, nume_furnizor)



A 4-a forma normala (4NF)

Notiuni Preliminare

- ▶ Fie relatia:

DCS(Departament, CadruDidactic, DataSedinta),
unde attributele CadruDidactic si DataSedinta sunt repetitive

- ▶ O inregistrare in aceasta relatie ar putea sa fie de tipul dat in tabelul alaturat (sus)
- ▶ Pentru a nu avea attribute repetitive (pentru ca relatia sa fie cel putin in 1NF) trebuie sa memoram datele din aceasta inregistrare ca in tabelul de jos (s-au trecut date numai pentru o inregistrare din primul tabel)
- ▶ In acest tabel, fiecare cadru didactic are aceleasi date pentru sedinta; deci la adaugarea, modificarea sau stergerea liniilor din tabel **trebuie facute unele verificari suplimentare**

Informatica	CDI1 CDI2 . . . CDIm	DII DI2 . . . DIn
Matematica	CDM1 CDM2 . . . CDMp	DMI DM2 . . . DMq

Informatica	CDI1	DII
Informatica	CDI1	DI2
.....
Informatica	CDI1	DIn
Informatica	CDI2	DII
Informatica	CDI2	DI2
.....

A 4-a forma normala (4NF)

Notiuni Preliminare

- ▶ Dupa cum s-a precizat in cursul anterior, **dependenta functionala simpla** $X \rightarrow Y$ inseamna ca fiecarei valori a lui **X** i se asociaza **o valoare unica** a lui **Y**
- ▶ **Dependenta functionala multipla** $X \twoheadrightarrow Y$ (**Y este multiplu dependent functional de X**) presupune ca o valoare u a lui X are asociata o multime de valori v pentru Y: $X(u) = \{v_1, v_2, \dots, v_n\}$, iar aceasta asociere este valabila indiferent de valorile din $Z = A - Y - X$ (unde A este multimea de attribute din relatie, deci $A = X \cup Y \cup Z$)
- ▶ **Proprietate: daca** $X \twoheadrightarrow Y$ **si** $A = X \cup Y \cup Z$, **atunci** $X \twoheadrightarrow Z$



A 4-a forma normala (4NF)

Notiuni Preliminare

- ▶ Pentru relatia DCS din ex. precedent avem

$\{\text{Departament}\} \implies \{\text{CadruDidactic}\}$, dar si

$\{\text{Departament}\} \implies \{\text{DataSedinta}\}$

- ▶ Daca R are in schema numai attributele X si Y, deci $R[X,Y]$, atunci avem numai o **dependenta functionala multipla triviala**
- ▶ Daca $R[X,Y,Z]$, $X \implies Y$ si Z nu este vida, deci exista o dependenta functionala multipla netriviala, atunci se pune problema eliminarii acestei DF (pentru a evita necesitatea verificarilor suplimentare)



A 4-a forma normala (4NF)

O relatie R este in **4NF** daca este in **BCNF** si **nu contine dependente functionale multiple netriviale**

NB: o relatie de forma 4NF este redusa la cheia sa (toate attributele sunt incluse in cheia)

- ▶ Daca $R[X,Y,Z]$ si exista $X \twoheadrightarrow Y$, atunci R se descompune astfel:

$$R_1[X,Y] = \prod_{X \cup Y} (R)$$

$$R_2[X,Z] = \prod_{X \cup Z} (R)$$

- ▶ Pentru relatia DCS din exemplul dat se obtine descompunerea:

DC(Departament, CadruDidactic)

DS(Departament, DataSedinta)

- ▶ DC si DS sunt deci in 4NF
-



A 5-a forma normala (5NF)

Notiuni Preliminare

- ▶ Cand apare o DF (simpla, multipla) ea se elimina prin descompunerea relatiei in relatii noi
- ▶ Totusi, **exista relatii fara DF care inca mai contin informatii redundante**, deci care **pot genera erori in BD**
- ▶ Exemplu: fie relatia

PrSeDi(Profesor, Sectie, Disciplina)

care memoreaza sectiile si disciplinele la care predau cadrele didactice

- ▶ Aceasta relatie nu are DF; cheia relatiei este {Profesor, Sectie, Disciplina}



A 5-a forma normala (5NF)

Notiuni Preliminare

► Consideram urmatoarele date in relatie:

Pr	Se	Di
P1	S1	D2
P1	S2	D1
P2	S1	D1
P1	S1	D1

Exista unele date redundante (apar unele asocieri intre mai multe inregistrari):

- Profesorul P1 preda la sectia S1
- Profesorul P1 preda disciplina D1
- La sectia S1 se preda disciplina D1

Daca se doreste sa se schimbe valorile din relatie, ca de exemplu:

- "profesorul P1 sa predea disciplina D3 în locul disciplinei D1",
atunci **trebuie facute mai multe modificari, fara sa se stie în câte înregistrari.**

La fel se întâmpla pentru modificarile

- "la sectia S1 în locul disciplinei D1 sa se predea disciplina D3",
- "profesorul P1 preda la sectia S3 în locul sectiei S1".



A 5-a forma normala (5NF)

Notiuni Preliminare

- ▶ Relatia anterioara nu se poate descompune in doua relatii (prin proiectie), pentru ca prin join se introduc date noi.
- ▶ Justificam aceasta afirmatie prin cele trei proiectii posibile pe doua attribute

PrSe	Pr	Se
	P1	S1
	P1	S2
	P2	S1

SeDi	Se	Di
	S1	D2
	S2	D1
	S1	D1

PrDi	Pr	Di
	P1	D2
	P1	D1
	P2	D1

- ▶ Daca se evalueaza $\text{PrSe} * \text{SeDi}$ se obtin datele:
- ▶ In aceasta relatie rezultat se obtine o linie suplimentara fata de relatia initiala

$R' = \text{PrSe} * \text{SeDi}$	Pr	Se	Di
	P1	S1	D2
	P1	S1	D1
	P1	S2	D1
	P2	S1	D2
	P2	S1	D1

A 5-a forma normala (5NF)

Notiuni Preliminare

- ▶ La fel se intampla si pentru alte combinatii de join: **PrSe * PrDi** si **SeDi * PrDi**
- ▶ Daca insa se calculeaza **R' * PrDi**, deci se evalueaza **PrSe * SeDi * PrDi**, atunci se obtine relatia initiala **PrSeDi**
- ▶ **Concluzie: PrSeDi** nu se descompune in doua proiectii, dar se poate descompune in trei proiectii, deci **PrSeDi** este **3-decompozabila**:

$$\mathbf{PrSeDi} = \mathbf{PrSe} * \mathbf{SeDi} * \mathbf{PrDi}, \text{ sau } \mathbf{PrSeDi} = *(\mathbf{PrSe}, \mathbf{SeDi}, \mathbf{PrDi})$$



A 5-a forma normala (5NF)

Notiuni Preliminare

- ▶ Concluzia anterioara (**PrSeDi** este **3-decompozabila**) este valabila pentru datele precizate in relatie
- ▶ Pentru a fi adevarata indiferent de date, trebuie sa fie indeplinite anumite restrictii
- ▶ O astfel de restrictie ar fi:
if (PI,SI) apartine **PrSe** and (PI,DI) apartine **PrDi** and (SI,DI) apartine **SeDi** then (PI,SI,DI) apartine **PrSeDi**
- ▶ Aceasta restrictie se poate inlocui cu alta:
if (PI,SI,D2) apartine PrSeDi and (PI,S2,DI) apartine PrSeDi and (P2,SI,DI) apartine PrSeDi then (PI,SI,DI) apartine PrSeDi
- ▶ Consideram urmatoarele date in relatie, cu respectarea restrictiei de mai sus:

Pr	Se	Di
PI	SI	D2
PI	S2	DI

Daca (P2,SI,DI) se adauga la relatie, atunci trebuie adaugata si inregistrarea (PI,SI,DI)

A 5-a forma normala (5NF)

Notiuni Preliminare

- Consideram urmatoarele date in relatie, cu respectarea restrictiei de mai sus:

Pr	Se	Di
PI	SI	D2
PI	S2	D1

Daca (P2,SI,DI) se adauga la relatie, atunci trebuie adaugata si inregistrarea (PI,SI,DI)

Pr	Se	Di
PI	SI	D2
PI	S2	D1
P2	SI	DI
PI	SI	DI

Daca se sterge (PI,SI,DI) din relatie, atunci trebuie sterse si alte date, cel putin (P2,SI,DI) pentru ca restrictia sa fie indeplinita

A 5-a forma normala (5NF)

Notiuni Preliminare

- ▶ **Dependentă join:** Fie $R[A]$ o relație și $R_i[X_i]$, $i=1, \dots, m$, proiecțiile relației R pe X_i . În R există o **dependentă join** dacă $R = R_1 * \dots * R_m$, $m \geq 2$
- ▶ Exemplu: În PrSeDi există o dependentă join deoarece $\text{PrSeDi} = \text{PrSe} * \text{SeDi} * \text{PrDi}$
- ▶ **Dependentă join trivială:** dependentă join $R = R_1 * \dots * R_m$, $R_i[X_i]$, este **trivială** dacă X_i este cheie pentru R , $i=1, \dots, m$



A 5-a forma normala (5NF)

O relatie R este in **5NF** daca este **4NF** si **nu exista dependente join netriviale**.

- ▶ Daca intr-o relatie R exista o dependenta join $R = R_1 * \dots * R_m$, atunci se descompune in relatiile R_i , $i = 1, \dots, m$



Recapitulare

- ▶ Fie $R(\underline{C1}, \underline{C2}, \underline{C3})$

- ▶ I caz: 4NF

- ▶ A. Exista o DF multipla intre $C1$ si $C2/C3$, atunci putem descompune relatia R :

$$R(\underline{C1}, \underline{C2}, \underline{C3}) = R1(\underline{C1}, \underline{C2}) * R2(\underline{C1}, \underline{C3})$$

- ▶ B. Nu exista DF multipla

Atunci $R(\underline{C1}, \underline{C2}, \underline{C3})$ nu se descompune, si este in 4NF

- ▶ II-lea caz: 5NF

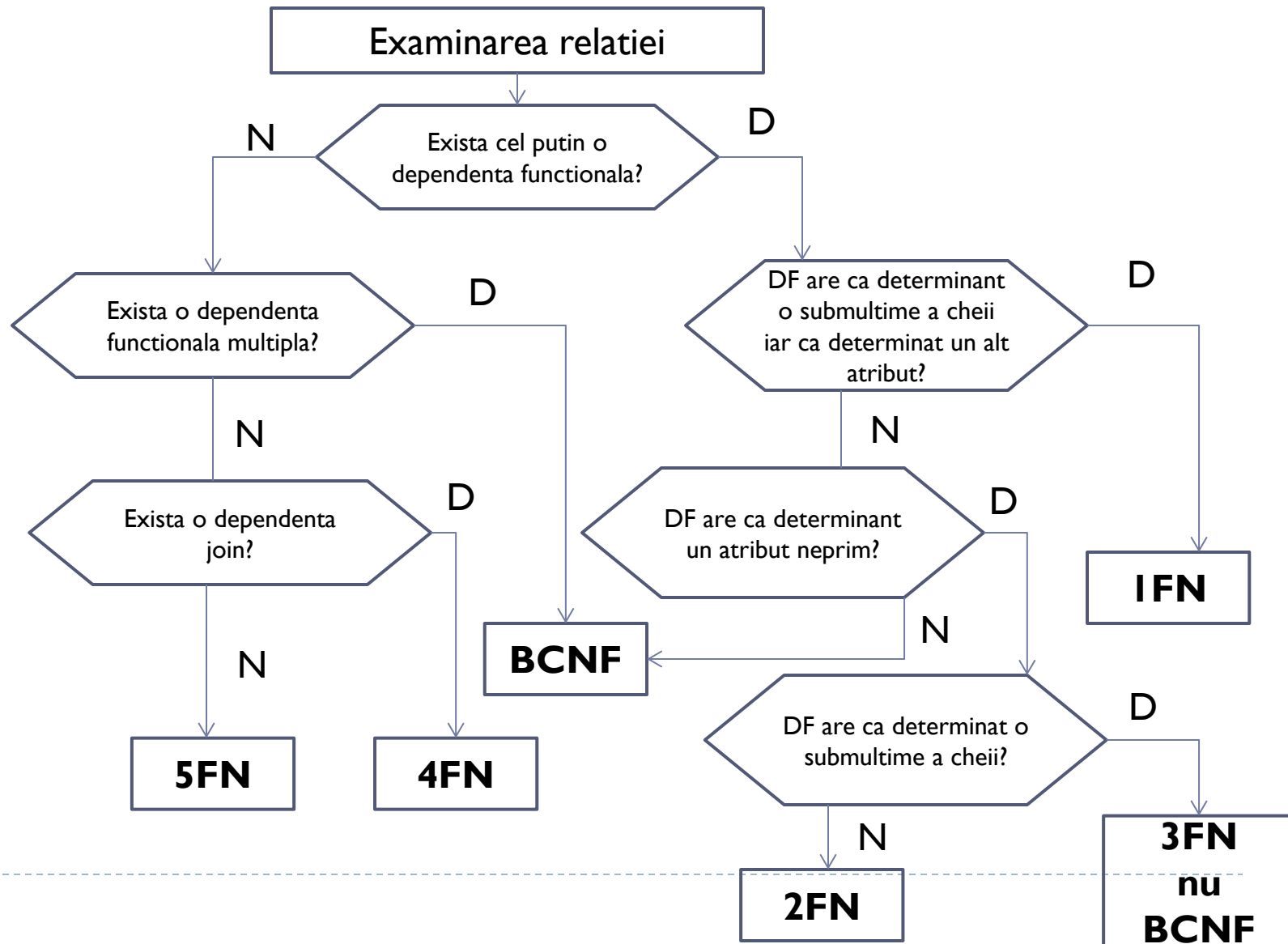
- ▶ A. exista o DJ intre $C1$, $C2$ si $C3$, atunci

$$R(\underline{C1}, \underline{C2}, \underline{C3}) = R1(\underline{C1}, \underline{C2}) * R2(\underline{C1}, \underline{C3}) * R3(\underline{C2}, \underline{C3})$$

- ▶ B. Nu exista DJ, si deci $R(\underline{C1}, \underline{C2}, \underline{C3})$ nu se poate descompune si este in 5NF



Algoritm de identificare a formei normale a unei relatii



Referinte

- ▶ 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*
- ▶ Ramakrishnan, R., *Database Management Systems*. McGraw-Hill, 2007, *cap. 15*
- ▶ Leon Tambulea, *Curs de baze de date*, UBB Cluj-Napoca
- ▶ Christine Verdier, *curs Systèmes d'Information et Bases de Données*, UJF Grenoble

