Initial $Z = D$

Prima trecere prin elementele lui ρ:

Pentru ABC: $Z = D \cup ((D \cap ABC)^+ \cap ABC) = D \cup \varnothing = D$

Pentru CD: $Z = D \cup ((D \cap CD)^+ \cap CD) = D \cup (ABCD \cap CD) = CD$

A doua trecere prin elementele lui ρ:

Pentru ABC: $Z = CD \cup ((CD \cap ABC)^+ \cap ABC) = CD \cup (ABCD \cap ABC) = ABCD$. Stop. Am obtinut ca $A \subseteq Z$, deci dependenta $D \to A$ este pastrata, deci ρ pastreaza dependentele.

4.4.3. Algoritmi de descompunere

Algoritmii de testare al pastrarii dependentelor si a joinului fara pierderi pot fi aplicati atunci cand descompunerea unei scheme de relatie se face 'de mana', pe baza experientei pe care o are proiectantul bazei de date. Exista insa si niste algoritmi simpli care, pornind de la o schema de relatie si multimea de dependente functionale asociata ne duc direct la o descompunere care este in FN3 sau FNBC si in plus au proprietatea de join fara pierderi (deci nu se pierd date prin descompunere) si/sau de pastrare a dependentelor.

Algoritm de descompunere in FN3 cu pastrarea dependentelor

Fie R o schema de relatie si F multimea de dependente functionale asociata, cu

$$F = \{ X_1 \to Y_1,\ X_2 \to Y_2,\ \dots X_n \to Y_n \}$$

Atunci descompunerea $\rho = (X_1Y_1,\ X_2Y_2,\ \dots X_nY_n)$ este o descompunere in FN3 cu pastrarea dependentelor.

Se observa din definitia de mai sus a descompunerii ρ ca:

- Toate dependentele sunt pastrate: dependenta $X_i \to Y_i$ este in proiectia lui F pe X_iY_i
- Pentru a minimiza numarul de elemente din descompunere se aplica regula reuniunii: daca avem mai multe dependente care au aceeasi parte stanga le reunim intr-una singura.
- Daca in descompunere exista doua elemente X_iY_i si X_jY_j astfel incat $X_iY_i \subseteq X_jY_j$ atunci X_iY_i se elimina.

In literatura de specialitate exista demonstratia faptului ca fiecare schema din descompunere este in FN3.

Exemplul 1: R = ABCDE, $F = \{ A \to B, A \to C, A \to D, D \to E \}$. Rescriem prin reuniune multimea de dependente functionale: $F = \{ A \to BCD, D \to E \}$. Rezulta din algoritm descompunerea ρ = (ABCD, DE)

Exemplul 2: Fie relatia Produse = IdP, NumeP, Qty, IdF, NumeF, AdresaF avand dependentele functionale:

F = { IdP→NumeP, IdP→Qty, IdP→ IdF, IdF→NumeF, IdF→AdresaF }

Rescriem multimea de dependente. Raman doar doua dependente:

F = { IdP→NumeP, Qty, IdF; IdF→NumeF, AdresaF }

Descompunerea in FN3 cu pastrarea dependentelor va fi:

ρ = ((IdP, NumeP, Qty, IdF), (IdF, NumeF, AdresaF))

Algoritm de descompunere in FN3 cu pastrarea dependentelor si join fara pierderi

Daca la descompunerea obtinuta prin algoritmul anterior adaugam o cheie a relatiei (ca element al descompunerii) vom obtine o descompunere care are atat proprietatea de join fara pierderi cat si pe cea a pastrarii dependentelor. Formal putem scrie algoritmul astfel:

Fie R o schema re relatie si F multimea de dependente functionale asociata, cu

$$F = \{ X_1 \rightarrow Y_1,\ X_2 \rightarrow Y_2,\ \dots X_n \rightarrow Y_n \}$$

si X o cheie pentru R

Atunci descompunerea $\rho = (X, X_1Y_1,\ X_2Y_2,\ \dots X_nY_n)$ este o descompunere in FN3 cu pastrarea dependentelor si join fara pierderi.

Pastrarea dependentelor este evidenta, ca mai sus. Demonstratia faptului ca descompunerea are si proprietatea de join fara pierderi se gaseste in literatura de specialitate.

Observatie: Daca vreunul dintre elementele de forma X_iY_i contin deja o cheie a lui R atunci nu este necesara adaugarea unui element suplimentar in descompunere.

Exemplul 1: Pentru relatiile din exemplele de mai sus descompunerea ramane aceeasi deoarece:

- In cazul relatiei R = ABCDE cheia este A, deja inclusa in ABCD, deci descompunerea ramane ρ = (ABCD, DE).
- In cazul relatiei PRODUSE de asemenea cheia este IdP, inclusa deja intr-unul din elementele descompunerii.

Exemplul 2: Fie R = ABCDE, F = $\{ A \rightarrow B, B \rightarrow A, A \rightarrow C, D \rightarrow E \}$. Cheile relatiei sunt AD si BD.

- Rescriem multimea de dependente: F = $\{ A \rightarrow BC, B \rightarrow A, D \rightarrow E \}$.
- Rezulta descompunerea cu pastrarea dependentelor: ρ = (ABC, AB, DE). Cum AB e inclus in ABC rezulta in final ρ = (ABC, DE).
- Cum elementele descompunerii nu contin vreo cheie a lui R, o adaugam. Obtinem in final descompunerea ρ = (AD, ABC, DE).

Algoritm de descompunere in FNBC cu join fara pierderi

Fie R o schema de relatie si F multimea de dependente functionale asociata, F in forma canonica: $F = \{ X_1 \rightarrow A_1,\ X_2 \rightarrow A_2,\ \dots X_n \rightarrow A_n \}$. Putem calcula descompunerea in FNBC cu join fara pierderi iterativ:

- Initial ρ = (R)
- La fiecare pas se alege o schema T care contine o dependenta de forma $X \rightarrow A$ care violeaza conditiile de FNBC. Schema respectiva este inlocuita in ρ prin T_1 si T_2 unde $T_1 = XA$ si $T_2 = T - \{A\}$
- Procesul se opreste cand in ρ nu mai exista elemente care nu sunt in FNBC

Exemplu: Fie relatia R = ABCD cu F = $\{ AB \rightarrow C, AB \rightarrow D, D \rightarrow A \}$. Cheia relatiei este AB. Relatia este in FN3 dar nu este in FNBC din cauza dependentei $D \rightarrow A$ care nu are in partea stanga o supercheie a lui R.

- Initial: $\rho = (R) = (ABCD)$
- Alegem dependenta $D \rightarrow A$ care violeaza conditia de FNBC.
- Inlocuim T = ABCD cu T_1 = DA si T_2 = ABCD – A = BCD.
- T_1 mosteneste de la T dependenta $D \rightarrow A$, cheia va fi D si T_1 e in FNBC
- T_2 mosteneste de la T dependenta $\{ DB \rightarrow C \}$. Cheia va fi DB si T_2 e in FNBC.
- Rezulta ca descompunerea in FNBC cu join fara pierderi este $\rho = (AD, BCD)$.

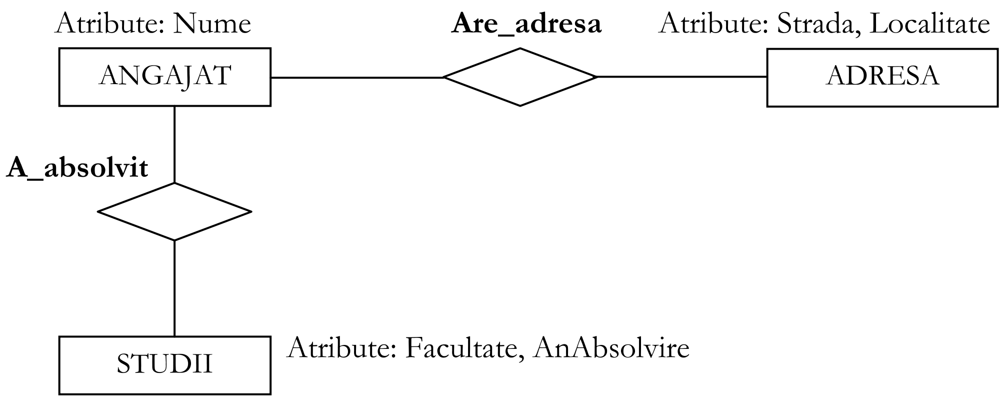
Observatii:

1. Dependenta mostenita de T_2 este din F^+. Ea se deduce astfel: Din $D \rightarrow A$ prin augmentare cu B obtinem $DB \rightarrow AB$ si impreuna cu dependenta $AB \rightarrow C$, prin tranzitivitate obtinem $DB \rightarrow C$. Analog din $AB \rightarrow D$ se deduce $DB \rightarrow D$ dar aceasta este o dependenta triviala (partea dreapta e inclusa in cea stanga).
2. In multe cazuri este nevoie de mai multe iteratii, relatiile de tip T_2 (egale in algoritm cu T – A) nefiind uneori in FNBC. Ele se descompun din nou in acelasi fel.

4.5. Dependente multivalorice. Forma normala 4

Exista situatii in care, desi o relatie este in forma normala Boyce Codd, instantele sale contin date redundante. Acest fapt se datoreaza unei proiectari defectuoase in care in aceeasi relatie sunt stocate date care apartin mai multor entitati si a cel putin doua asocieri multi-multi.

Sa luam urmatorul exemplu:



Ambele asocieri sunt multi-multi: un angajat poate sa fie absolvent al mai multor facultati si in acelasi timp poate avea mai multe adrese (de exemplu una pentru domiciliul stabil si alta pentru rezidenta temporara la un moment dat).

In cazul in care toate datele din aceasta diagrama sunt stocate intr-o singura tabela putem avea urmatoarea incarcare cu date corecte:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	**Viitorului**	**Ploiesti**	**Automatica**	**2000**
Vasile	**Viitorului**	**Ploiesti**	**Comert**	**2004**
Vasile	**Dreapta**	**Bucuresti**	**Automatica**	**2000**
Vasile	**Dreapta**	**Bucuresti**	**Comert**	**2004**
Mariana	**Revolutiei**	**Timisoara**	**Constructii**	**1998**
Mariana	**Revolutiei**	**Timisoara**	**Drept**	**2003**
Mariana	**Revolutiei**	**Timisoara**	**Master ASE**	**2006**
Mariana	**Calea Vitan**	**Bucuresti**	**Constructii**	**1998**
Mariana	**Calea Vitan**	**Bucuresti**	**Drept**	**2003**
Mariana	**Calea Vitan**	**Bucuresti**	**Master ASE**	**2006**

Putem observa ca nu exista nici o dependenta functionala netriviala valida pentru aceasta relatie, deci nu exista dependente care sa violeze conditiile FNBC. Ca urmare relatia este in FNBC avand ca singura cheie posibila multimea tuturor atributelor relatiei: din axioma de reflexivitate (A1) putem obtine dependenta:

Nume,Strada,Localitate,Facultate,AnAbsolvire →
Nume,Strada,Localitate,Facultate,AnAbsolvire

Desi relatia este in FNBC adresa si facultatea absolvita de un angajat sunt prezente repetat in relatie: adresa pentru fiecare facultate absolvita iar facultatea pentru fiecare adresa a angajatului.

Exemplul de mai sus sugereaza faptul ca seturile de atribute {Strada, Localitate} si {Facultate, AnAbsolvire} sunt independente unele de altele, in sensul ca fiecare adresa apare cu fiecare facultate absolvita de un angajat si reciproc. Astfel de situatii sunt modelate cu un nou tip de dependente numite ***dependente multivalorice***.

4.5.1. Dependente multivalorice (DMV)

Definitie: Fie o relatie R si doua multimi de atribute X si Y incluse in R. Se spune ca X multidetermina Y sau ca exista dependenta multivalorica X →→ Y daca si numai daca ori de cate ori avem doua tupluri ale relatiei t1 si t2 cu t1[X] = t2[X] atunci exista in relatie un tuplu t3 pentru care:

- t3[X] = t1[X] = t2[X]
- t3[Y] = t1[Y] si t3[R-X-Y] = t2[R-X-Y]

	X	**Y**	**R – X – Y**
t1	**AAA**	**BBB**	**CCC**
t2	**AAA**	**DDD**	**EEE**
t3	**AAA**	**BBB**	**EEE**

O consecinta interesanta a acestei definitii este ca, daca inversam tuplurile t1 si t2, rezulta ca exista si un tuplu t4 pentru care

- t4[X] = t1[X] = t2[X]
- t4[Y] = t2[Y] si t4[R-X-Y] = t1[R-X-Y]

Tot din aceasta definitie rezulta ca daca in R exista dependenta multivalorica X $\rightarrow\rightarrow$ Y atunci exista si dependenta X $\rightarrow\rightarrow$ R – X – Y (acest fapt va fi prezentat in paragraful urmator ca axioma de complementare a dependentelor multivalorice).

Intorcandu-ne la exemplul anterior rezulta ca in relatia continand date despre angajati, studii si adrese avem urmatoarele dependentele multivalorice (a doua fiind obtinuta din prima prin complementare):

Nume $\rightarrow\rightarrow$ Strada, Localitate
Nume $\rightarrow\rightarrow$ Facultate, AnAbsolvire

Intradevar, daca luam in considerare pentru t1 si t2 tuplurile 2 si 3 din relatie:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Viitorului	Ploiesti	Comert	2004
Vasile	Dreapta	Bucuresti	Automatica	2000

gasim in relatie pe prima pozitie si tuplul t3 de forma:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Viitorului	Ploiesti	Automatica	2000

In acelasi timp gasim pe pozitia 4 si tuplul t4:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Dreapta	Bucuresti	Comert	2004

Sa facem o alta alegere pentru t1 si t2:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Viitorului	Ploiesti	Automatica	2000
Vasile	Viitorului	Ploiesti	Comert	2004

Atunci t3 si t4 vor fi:

t3:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Viitorului	Ploiesti	Comert	2004

t4:

Nume	Strada	Localitate	Facultate	AnAbsolvire
Vasile	Viitorului	Ploiesti	Automatica	2000

Observam ca t3 = t2 si t4 = t1 ceea ce este corect pentru ca in definitia dependentelor multivalorice nu se cere ca t3 sa fie diferit de t1 si t2.

Consecinta importanta: orice dependenta functionala este in acelasi timp si o dependenta multivalorica:

Fie relatia R si o dependenta functionala X → Y pentru R. Atunci daca doua tupluri t1 si t2 au aceleasi valori pe atributele X vor avea aceleasi valori si pe atributele Y. Rezulta ca t2 indeplineste conditiile pentru t3 din definitia dependentelor multivalorice:

	X	**Y**	**R – X – Y**
t1	**AAA**	**BBB**	**CCC**
t2	**AAA**	**BBB**	**DDD**
t3 este t2	**AAA**	**BBB**	**DDD**

Rezulta ca daca X → Y avem si dependenta multivalorica X →→ Y

Exemplu: Fie relatia Produse anterioara:

IdP	**NumeP**	**Qty**	**IdF**	**NumeF**	**AdresaF**
101	Imprimanta laser	30	20	Xerox	Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti
105	Calculator PC	20	23	IBM	Bd. D.Cantemir nr.1, Bucuresti
124	Copiator	10	20	Xerox	Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti

In aceasta avem dependenta functionala:

IdF → NumeF, AdresaF

Avem doua tupluri cu aceleasi valori pe IdF:

	IdP	**NumeP**	**Qty**	**IdF**	**NumeF**	**AdresaF**
t1	101	Imprimanta laser	30	**20**	**Xerox**	**Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti**
t2	124	Copiator	10	20	Xerox	Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti

In acest caz putem forma tuplul t3 astfel:

- Pe IdF valoarea 20
- Pe NumeF si adresaF valorile din primul tuplu
- Pe restul atributelor valorile din al doilea tuplu.

Obtinem t3 identin cu t2:

	IdP	**NumeP**	**Qty**	**IdF**	**NumeF**	**AdresaF**
t3	124	Copiator	10	**20**	**Xerox**	**Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti**

Ca si in cazul dependentelor functionale exista si in cazul DMV o serie de axiome si reguli care ne permit ca, pornind de la un set de dependente sa obtinem alte dependente.

4.5.2. Axiome si reguli pentru DMV

Urmatoarele axiome sunt specifice DMV. Le numerotam incepand cu A4 deoarece intr-o schema de relatie pot fi atat dependente functionale (carora li se aplica axiomele A1-A3 descrire anterior) cat si dependente multivalorice.

A4. Complementare: Fie R o schema de relatie si X, Y $\subseteq$ R.

Daca X $\rightarrow\rightarrow$ Y atunci si X $\rightarrow\rightarrow$ (R – X – Y)

A5. Augmentare pentru DMV: Fie R o schema de relatie si X, Y, Z, W $\subseteq$ R.

Daca X $\rightarrow\rightarrow$ Y si Z $\subseteq$ W atunci si XW $\rightarrow\rightarrow$ YZ

A6. Tranzitivitate pentru DMV: Fie R o schema de relatie si X, Y, Z $\subseteq$ R.

Daca X $\rightarrow\rightarrow$ Y si Y $\rightarrow\rightarrow$ Z atunci si X $\rightarrow\rightarrow$ (Z – Y)

Ultimele doua axiome leaga dependentele multivalorice cu cele functionale:

A7. Fie R o schema de relatie si X, Y $\subseteq$ R.

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

A8. Fie R o schema de relatie si X, Y, Z, W $\subseteq$ R. cu W $\cap$Y = $\emptyset$

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

Observatie importanta: orice dependenta functionala este in acelasi timp si o dependenta multivalorica insa reciproca nu este adevarata: exista dependente multivalorice pentru care in schema relatiei nu avem o dependenta functionala corespunzatoare. Exemplu pentru acest fapt este dependenta multivalorica existenta in tabela de angajati din paragraful anterior:

Nume $\rightarrow\rightarrow$ Strada, Localitate

In relatie nu exista insa si o dependenta functionala echivalenta de tipul:

Nume $\rightarrow$ Strada, Localitate

Rezulta ca:

- Putem folosi si axiomele A1-A3 dar doar pentru dependente multivalorice care sunt in acelasi timp si dependente functionale.
- Pentru restul dependentelor multivalorice putem folosi doar A4-A6.

Exista de asemenea o serie de reguli care se pot deduce din axiome. Toate considera existenta unei scheme de relatie R iar X, Y, Z, W sunt submultimi ale lui R:

R1. Reuniune: Daca X $\rightarrow\rightarrow$ Y si X $\rightarrow\rightarrow$ Z atunci

X $\rightarrow\rightarrow$ YZ

R2. Pseudotranzitivitate: Daca X $\rightarrow\rightarrow$ Y si WY $\rightarrow\rightarrow$ Z atunci

WX $\rightarrow\rightarrow$ Z – WY

R3. Pseudotranzitivitate mixta: Daca X $\rightarrow\rightarrow$ Y si XY $\rightarrow\rightarrow$ Z atunci

X $\rightarrow\rightarrow$ Z – Y

R4. Diferenta: Daca $X \rightarrow\rightarrow Y$ si $X \rightarrow\rightarrow Z$ atunci:

$X \rightarrow\rightarrow Y - Z$

$X \rightarrow\rightarrow Z - Y$

R5. Intersectie: Daca $X \rightarrow\rightarrow Y$ si $X \rightarrow\rightarrow Z$ atunci:

$X \rightarrow\rightarrow Y \cap Z$

R6. Eliminare atribute comune: Daca $X \rightarrow\rightarrow Y$ atunci:

$X \rightarrow\rightarrow Y - X$

R7. Toate atributele: Daca $X \cup Y = R$ atunci

$X \rightarrow\rightarrow Y$ si $Y \rightarrow\rightarrow X$

R8. Reflexivitate: Daca $Y \subseteq X$ atunci

$X \rightarrow\rightarrow Y$

Aceste axiome si reguli se pot folosi pentru calculul inchiderii unei multimi de dependente functionale si multivalorice. Definitia inchiderii este aceeasi ca la dependentele functionale:

Definitie: Fie R o schema de relatie si G multimea de dependente functionale si multivalorice asociata. Atunci ***inchiderea multimii de dependente*** G, notata G^+, este o multime de dependente (DF si DMV) care sunt in G sau se pot deduce din G folosind axiomele si regulile.

Analog cu cazul dependentelor functionale se poate defini si proiectia unei multimi de dependente functionale si multivalorice pe o multime de atribute:

Definitie. Fie o relatie R, o multime asociata de dependente functionale si multivalorice G si o submultime de atribute $S \subseteq R$. ***Proiectia multimii de dependente G pe S***, notata cu $\pi_S(G)$ este multimea dependentelor din G^+ care au si partea stanga si pe cea dreapta incluse in S.

In momentul in care o schema de relatie se descompune in doua sau mai multe subscheme, fiecare subschema va mosteni o multime de dependente functionale si multivalorice obtinuta prin proiectia multimii initiale G pe atributele din subschema respectiva.

4.5.3. Forma normala 4

Pentru a preintampina redundantele prezentate la inceputul paragrafului 4.5. este bine ca schemele de relatie sa fie intr-o forma normala superioara FNBC. Aceasta forma care considera si dependentele multivalorice se numeste **forma normala 4** (FN4).

Definitia ei este similara cu cea pentru FNBC dar conditia se pune pentru dependentele multivalorice ale relatiei respective:

Definitie: O schema de relatie R este in forma normala 4 daca orice dependenta multivalorica netriviala $X \rightarrow\rightarrow Y$ are in partea stanga o supercheie

Dependentele multivalorice triviale sunt de doua feluri:

1. Dependente provenite din R8, deci cele in care partea dreapta este inclusa in partea stanga: $X \rightarrow\rightarrow Y$ unde $Y \subseteq X$
2. Dependente provenite din regula R7: $X \rightarrow\rightarrow Y$ pentru $X \cup Y = R$

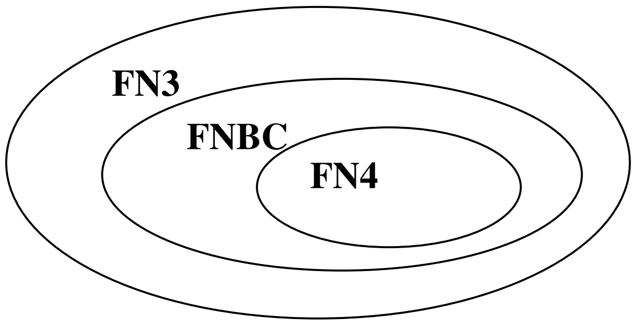
Conditia de FN4 spune deci ca orice DMV care nu intra in una din categoriile de mai sus are in partea stanga o supercheie.

Exemplu: Relatia angajati-studii-adrese are dependenta netriviala

Nume $\rightarrow\rightarrow$ Strada, Localitate

Cum cheia relatiei e multimea tuturor atributelor acesteia, rezulta ca relatia nu este in FN4 deoarece {Nume} nu e supercheie.

Relatia dintre formele normal2 FN3, FNBC si FN4 este una de includere, in aceasta ordine. Orice relatie in FN4 este in acelasi timp si in FNBC si FN3:



Algoritm de descompunere in FN4

Acest algoritm este similar cu cel de descompunere in FNBC dar ia in considerare dependentele multivalorice care violeaza FN4. Atentie: dependentele multivalorice ale unei relatii sunt atat cele care provin prin axioma A7 din dependente functionale cat si dependente multivalorice care nu au corespondent in multimea celor functionale.

Fie R o schema de relatie si G multimea de dependente multivalorice asociata (consideram ca din G au fost eliminate dependentele triviale). Putem calcula descompunerea in FN4 iterativ:

- Initial ρ = (R)
- La fiecare pas se alege o schema T care contine o dependenta de forma $X \rightarrow\rightarrow Y$ care violeaza conditia pentru FN4. Schema respectiva este inlocuita in ρ prin T_1 si T_2 unde $T_1 = XY$ si $T_2 = T - Y$
- Procesul se opreste cand in ρ nu mai exista elemente care nu sunt in FN4

Exemplu: Pentru relatia Angajati care nu era in FN4

- Initial ρ = ((Nume,Strada,Localitate,Facultate,AnAbsolvire))

- Alegem dependenta Nume $\rightarrow\rightarrow$ Strada, Localitate care violeaza conditia pentru FN4. Obtinem
    - T_1 = Nume, Strada, Localitate si
    - T_2 = Nume, Facultate, AnAbsolvire
- Obtinem ρ = ((Nume, Strada, Localitate), (Nume, Facultate, AnAbsolvire)). Fiecare subschema mosteneste cate o dependenta multivalorica:
    - T_1 dependenta Nume $\rightarrow\rightarrow$ Strada, Localitate
    - T_2 dependenta Nume $\rightarrow\rightarrow$ Facultate, AnAbsolvire
- Cum cele doua dependente multivalorice mostenite de T_1 si T_2 sunt triviale (contin toate atributele relatiei respective) rezulta ca cele doua relatii sunt in FN4 deoarece nu exista dependente care violeaza conditia de FN4. Procesul s-a incheiat.

Observatie: Asa cum s-a mentionat anterior, fiecare subschema T_i obtinuta la descompunere mosteneste de la relatia originala T proiectia multimii de dependente a lui T (DF si DMV) pe T_i.