8. Proiectia unei multimi de DF

Asa cum s-a mentionat in paragraful 4.2.4. inchiderea unei multimi de dependente functionale F^+ a fost introdusa si pentru a putea defini setul de dependente functionale mostenite de o schema de relatie obtinuta prin descompunerea unei scheme incorect proiectata. Sa luam cazul relatiei anterioare continand produsele dintr-un depozit:

Produse = IdP, NumeP, Qty, IdF, NumeF, AdresaF

avand asociata multimea de dependente:

$$F = \{ IdP \rightarrow NumeP, IdP \rightarrow Qty, IdP \rightarrow IdF, IdF \rightarrow NumeF, IdF \rightarrow AdresaF \}$$

Prin spargerea acestei relatii in doua obtinem relatiile:

Produse = IdP, NumeP, Qty, IdF

Furnizori= IdF, NumeF, AdresaF

Atributele relatiei initiale se regasesc fie doar intr-una dintre schemele rezultate fie in amandoua. Se pune insa si problema: ce dependente mostenesc cele doua relatii de la relatia initiala? Solutia este de a defini proiectia unei multimi de dependente pe o multime de atribute.

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

$$\pi_S(F) = \{X \rightarrow Y \in F^+ \mid X, Y \subseteq S \}$$

Pentru exemplul de mai sus proiectiile sunt urmatoarele:

$F_{PRODUSE} = \pi_{PRODUSE}(F) = \{ IdP \rightarrow NumeP, IdP \rightarrow Qty, IdP \rightarrow IdF\}$

$F_{FURNIZORI} = \pi_{FURNIZORI}(F) = \{ IdF \rightarrow NumeF, IdF \rightarrow AdresaF \}$

Observatie: Atunci cand descompunem o schema se poate intampla ca unele dintre dependentele schemei initiale sa se piarda.

Exemplu: Fie R = ABCD si $F = \{ A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A \}$. In cazul in care descompunem R in R_1 = AB si R_2 = CD atunci:

$F_{R1} = \pi_{R2}(F) = \{ A \rightarrow B, B \rightarrow A \}$

$F_{R2} = \pi_{R1}(F) = \{ C \rightarrow D, D \rightarrow C \}$

A doua dependenta din fiecare multime nu este in F dar este in F^+ (obtinuta prin tranzitivitate).

Observam insa ca dependentele $B \rightarrow C$ si $D \rightarrow A$ nu mai pot fi obtinute nici din F_{R1} nici din F_{R2} nici din reuniunea lor. In subcapitolul 4.4. va fi prezentata o metoda prin care se poate testa daca prin descompunere dependentele initiale sunt pastrate sau nu.

4.2.7. Inchiderea unei multimi de atribute

Fie R o schema de relatie, F multimea de dependente asociata si $X \subseteq R$. Se poate defini ***inchiderea multimii de atribute X in raport cu F*** (notata X^+) astfel:

$$X^+ = \{ A \mid X \rightarrow A \in F^+ \}$$

X^+ contine deci toate atributele care apar in partea dreapta a unei dependente din F sau care se poate deduce din F folosind regulile si axiomele.

Pentru calculul lui X^+ exista un algoritm simplu, prezentat in continuare.

Algoritm de calcul pentru X^+

Intrare: R o schema de relatie, F multimea de dependente asociata si $X \subseteq R$

Iesire: X^+

Metoda: se procedeaza iterativ astfel:

- Se porneste cu $X^{(0)} = X$
- Pentru $i \geq 1$, $X^{(i)} = X^{(i-1)} \cup \{ A \mid (\exists) Y \rightarrow A \in F \text{ cu } Y \subseteq X^{(i-1)} \}$
- Oprirea se face atunci cand $X^{(i)} = X^{(i-1)}$

Exemplu: Fie R = ABCDE si F = $\{ A \rightarrow B, A \rightarrow C, D \rightarrow E \}$. Pentru a calcula A^+ si AD^+ procedam astfel:

Calcul A^+ :

- $X^{(0)} = \{A\}$
- Din $A \rightarrow B$ si $A \rightarrow C$ rezulta ca $X^{(1)} = X^{(0)} \cup \{ B, C \} = \{ A \} \cup \{ B, C \} =$ ABC
- Singurele dependente care au partea dreapta in $X^{(1)}$ sunt tot primele doua deci $X^{(2)} = X^{(1)} \cup \{ B, C \} = \{ A, B, C \} \cup \{ B, C \} = ABC$
- Oprire deoarece $X^{(2)} = X^{(1)}$

Rezulta ca $(A)^+ = ABC$

Calcul AD^+ :

- $X^{(0)} = \{A, D\}$
- Din $A \rightarrow B$, $A \rightarrow C$ si $D \rightarrow E$ rezulta ca $X^{(1)} = X^{(0)} \cup \{ B, C, E \} = \{ A, D \} \cup \{ B, C , E\} = ABCDE$
- Oprire deoarece $X^{(1)} = R$ deci oricate iteratii am face nu mai pot sa apara noi atribute.

Rezulta ca $(AD)^+ = ABCDE$

Scopul introducerii acestei notiuni este si cel de a putea ocoli in alti algoritmi si definitii calculul lui F^+ . Avem urmatorul rezultat teoretic:

Propozitie: Fie R o schema de relatie, F multimea de dependente asociata si X, Y $\subseteq$ R. Atunci X $\rightarrow$ Y se poare deduce din F daca si numai daca Y $\in$ X^+

Demonstratia acestei propozitii se gaseste in literatura de specialitate.

4.2.8. O alta definitie a cheii

Pe baza propozitiei din paragraful anterior se poate da o alta definitie pentru cheia sau supercheia unei relatii, bazata nu pe F^+ ca in paragraful 4.2.5 ci pe inchiderea unei multimi de atribute.

Definitie: Fie R o schema de relatie, F multimea de dependente functionale asociata si X $\subseteq$ R. Atunci X este cheie pentru R daca si numai daca:

- $X^+ = R$

si

- X este minimala: oricare ar fi $Y \subset X$, $Y \neq X$ atunci $Y^+ \neq R$ (deci orice submultime stricta a lui X nu mai indeplineste conditia anterioara).

Daca numai prima conditie este indeplinita atunci X este supercheie pentru R

Echivalenta acestei definitii cu cea anterioara este evidenta:

- $X^+ = R$ inseamna cf. propozitiei ca $X \rightarrow R$
- minimalitatea este de asemenea definita echivalent: $\neg(F \Rightarrow Y \rightarrow R)$ este echivalenta cu $\neg(Y^+ = R)$ adica $Y^+ \neq R$

Folosind aceasta definitie se poate defini o euristica de gasire a cheilor unei relatii:

Euristica de gasire a cheilor unei relatii

Pentru gasirea cheilor unei relatii pornim de la observatia ca atributele care nu sunt in partea dreapta a nici unei dependente nu pot sa apara in procesul de inchidere a unei multimi de atribute si deci ele apartin oricarei chei a relatiei.

Intrare: R o schema de relatie si F multimea de dependente functionale asociata (F in forma canonica).
Iesire: Cheia unica sau cheile alternative ale lui R
Metoda:

1. Se porneste de la multimea de atribute X $\subseteq$ R care nu apar in partea dreapta a nici unei dependente
2. Se calculeaza X^+. Daca X^+ = R atunci X este cheia unica minimala a relatiei R si calculul se opreste aici. Pasii urmatori se efectueaza doar daca $X^+ \neq R$
3. Se adauga la X cate un atribut din R - X^+ obtinandu-se o multime de chei candidat.
4. Se calculeaza X^+pentru fiecare dintre candidate. Daca se obtin toate atributele lui R atunci acel X este o cheie a lui R.
5. Se repeta pasii 3 si 4 pornind de la acele multimi candidat X care nu sunt gasite ca si chei la pasul anterior. Intre multimile candidat nu luam niciodata in considerare pe cele care contin o cheie gasita anterior.
6. Procesul se opreste cand nu se mai pot face augmentari.

Exemplul 1: Fie $R = ABCDE$ si $F = \{ A \to B, A \to C, D \to E \}$.

- Multimea atributelor care nu apar in partea dreapta a nici unei dependente este $X = AD$.
- Calculam $(AD)^+$. Obtinem $(AD)^+ = ABCDE = R$.
- Procesul se opreste. Rezulta ca AD este cheie unica pentru R

Exemplul 2: Fie $R = ABCDE$ si $F = \{ A \to B, B \to A , A \to C, D \to E \}$.

- Multimea atributelor care nu apar in partea dreapta a nici unei dependente este $X = D$.
- Calculam $(D)^+$. Obtinem $(D)^+ = DE$. Rezulta ca D nu este cheie unica pentru R
- Calculam multimea de candidate: augmentam D cu atribute din $R – D^+ = ABCDE – DE = ABC$. Obtinem AD, BD si CD
- Calculam inchiderile lor. Obtinem $(AD)^+ = R$, $(BD)^+ = R$ si $(CD)^+ = CDE \neq R$. Rezulta ca AD si BD sunt chei ale lui R dar CD nu e cheie.
- Calculam o noua multime de candidate pornind de ca CD. Putem augmenta CD cu atribute din $R – (CD)^+ = ABCDE – CDE = AB$. Nici una dintre augmentari nu este insa posibila pentru ca atat ACD cat si BCD contin o cheie gasita anterior (AD respectiv BD).
- Procesul se opreste. Singurele chei ale lui R raman AD si BD.

4.3. Forme normale

Exista cateva seturi de conditii care ne arata ca o schema de relatie este corect proiectata in sensul ca ea nu permite aparitia anomaliilor prezentate la inceputul capitolului.
Daca schema indeplineste cerintele unui anumit set de conditii se spune ca este in ***forma normala*** asociata acelui set. In continuare sunt prezentate formele normale Boyce-Codd si forma normala 3. In finalul acestul capitol va fi prezentata si forma normala 4 care se defineste in functie de alt tip de dependente, si anume dependentele multivalorice.

4.3.1. Forma normala Boyce-Codd (FNBC)

Definitie. R o schema de relatie si F multimea de dependente functionale asociata. Se spune ca R este in forma normala Boyce-Codd daca si numai daca oricare ar fi o dependenta netriviala $X \to Y$ din F atunci X este supercheie pentru R

Rezulta ca o schema de relatie este in FNBC daca si numai daca fiecare dependenta din F are in partea stanga o supercheie. Nu este necesar ca F sa fie in forma canonica dar nu trebuie sa contina dependente triviale (obtinute din prima axioma - de reflexivitate, de tipul $AB \to A$ sau $AB \to AB$)

Exemple:

1. Relatia $R = ABCDE$ avand $F = \{ A \to B, B \to A , A \to C, D \to E \}$ nu este in forma normala Boyce-Codd deoarece are cheile AD si BD dar nici o dependenta nu are in partea stanga o supercheie a lui R

2. Relatia Produse = IdP, NumeP, Qty, IdF avand asociata multimea de dependente functionale $F_{PRODUSE} = \pi_{PRODUSE}(F)$ = { IdP→NumeP, IdP→Qty, IdP→ IdF} este in forma normala Boyce-Codd: cheia unica a relatiei este IdP si toate dependentele au in partea stanga o supercheie (asa cum s-a mentionat orice cheie este in acelasi timp si supercheie)

3. Relatia Produse = IdP, NumeP, Qty, IdF, NumeF, AdresaF avand dependentele:
F = { IdP→NumeP, IdP→Qty, IdP→ IdF, IdF→NumeF, IdF→AdresaF }
nu este in forma normala Boyce-Codd: cheia unica este IdP dar exista dependente care nu au in partea stanga o dupercheie: IdF→NumeF, IdF→AdresaF

4.3.2. Forma normala 3 (FN3)

Pentru definitia formei normale 3 este necesara definirea notiunii de ***atribut prim***:

Definitie. R o schema de relatie si F multimea de dependente functionale asociata. Un atribut $A \in R$ se numeste atribut prim daca el apartine unei chei a lui R.

Exemplu: R = ABCDE avand F = { A → B, B → A , A → C, D → E }. Cum cheile relatiei sunt AD si BD rezulta ca in R sunt trei atribute prime: A, B si D.

Definitie. R o schema de relatie si F multimea de dependente functionale asociata. Se spune ca R este in forma normala 3 daca si numai daca oricare ar fi o dependenta netriviala X → A din F atunci

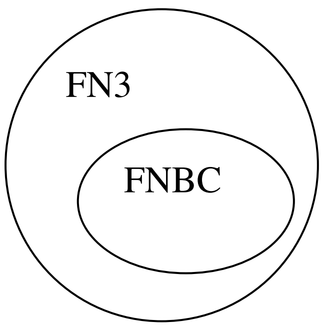
- X este supercheie pentru R

sau

- A este atribut prim

De remarcat ca daca in F avem dependente care contin mai multe atribute in partea dreapta putem aplica regula de descompunere pentru a obtine dependente care in partea dreapta au cate un singur atribut.

Observatie: Conditia de FNBC este inclusa in definitia FN3. Din acest motiv orice relatie care este in FNBC este implicit si in FN3. Reciproca nu este adevarata.



Rezulta de asemenea ca daca o schema de relatie nu este in FN3 ea nu poate fi nici in FNBC.

Exemple:

1. Relatia R = ABCD avand F = { $AB \to C$, $AB \to D$, $D \to A$ } are cheia unica AB.
   - Relatia este in FN3 deoarece primele doua dependente au in partea stanga o supercheie (AB) iar a treia dependenta are in partea dreapta atributul prim A.
   - Relatia nu este in FNBC deoarece a treia dependenta violeaza definitia pentru aceasta forma normala (nu are in partea stanga o supercheie.

2. Relatia R = ABCDE avand F = { $A \to B$, $B \to A$, $A \to C$, $D \to E$ } are cheile AD si BD. Rezulta ca:
   - R nu este in FN3 deoarece dependentele 3 si 4 nu au nici supercheie in partea stanga nici atribut prim in partea dreapta
   - R nu e in FNBC deoarece nu e in FN3

3. Relatia R = ABCD avand F = { $A \to B$, $B \to C$, $C \to D$, $D \to A$ } are cheile A, B, C si D. Rezulta ca:
   - R este in FNBC deoarece in partea stanga a dependentelor sunt numai superchei
   - R este in FN3 deoarece este in FNBC