7. PROIECTAREA BAZELOR DE DATE

O categorie de probleme care pot sa apara in dezvoltarea unei aplicatii continand o baza de date este cea a proiectarii incorecte a schemelor de relatie. In acest caz pot sa apara o serie de ***anomalii*** care pot complica procesul de programare. Testarea corectitudinii unei scheme de relatie poate fi facuta cu ajutorul ***dependentelor functionale*** (sau de alt tip) atasate acelei scheme.
Acestea modeleaza corelatii care exista intre datele din lumea reala stocate in baza de date si, asa cum am mentionat anterior, in cadrul teoriei bazelor de date relationale, ele reprezinta criterii de corectitudine ale datelor incarcate in baza de date.
In cazul in care o relatie nu are o schema corespunzatoare ea trebuie inlocuita cu doua sau mai multe relatii (operatia este numita si ***descompunerea unei scheme de relatie***), fiecare relatie rezultata avand o schema corecta – aflata in ***forma normala*** dorita.
In cadrul acestui capitol vom prezenta elementele de proiectare a structurii unei baze de date subliniate mai sus.

4.1. Anomalii

Exemplificarea anomaliilor rezultate din proiectarea defectuoasa a schemei unei relatii va fi facuta folosind tabela Produse din paragraful 3.1.3:

Produse

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

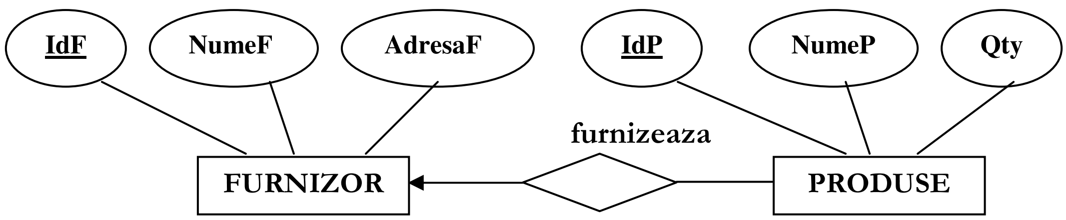
a. Redundanta: Redundanta reprezinta stocarea in mod nejustificata a unei aceleiasi informatii de mai multe ori in baza de date. Observam de exemplu ca pentru fiecare produs este stocat numele si adresa furnizorului, desi ele sunt unic determinate de codul acestuia.

b. Anomalia de actualizare: In cazul actualizarii unei informatii redundante, se poate intampla ca operatia sa modifice unele aparitii ale acesteia iar altele sa ramana cu vechea valoare.

c. **Anomalia de inserare:** Nu putem insera date despre un furnizor (numele si adresa sa) decat daca exista in stoc un produs furnizat de acesta.

d. **Anomalia de stergere:** La stergerea din relatie a ultimului produs al unui furnizor se pierd automat si datele despre acesta.

Aceste anomalii apar deoarece intr-o aceeasi tabela au fost stocate date despre doua clase diferite de obiecte. In cazul proiectarii cu ajutorul modelului entitate-asociere diagrama corecta este urmatoarea:



Prin transformarea acestei diagrame se obtin urmatoarele scheme de relatie:

Furnizor(IdF, NumeF, AdresaF)
Produse(IdP, NumeP, Qty, IdF)

Rezulta ca informatia din relatia Produse trebuie stocata de fapt in doua relatii astfel:

Furnizori

IdF	NumeF	AdresaF
20	Xerox	Str. Daniel Danielopolu 4-6, Sector 1, Bucureşti
23	IBM	Bd. D.Cantemir nr.1, Bucuresti

Produse

IdP	NumeP	Qty	IdF
101	Imprimanta laser	30	20
105	Calculator PC	20	23
124	Copiator	10	20

Procesul de 'spargere' a unei tabele care are o structura incorecta in doua sau mai multe tabele se numeste ***descompunerea schemei de relatie***. Pentru detectarea relatiilor care trebuiesc descompuse exista o serie de reguli de corectitudine, numite si ***forme normale***. Definirea acestor forme normale se bazeaza pe notiunea de ***dependenta (functionala sau multivalorica)*** prezentata in continuare.

4.2. Dependente functionale

4.2.1. Notatii

In paragrafele urmatoare vom folosi urmatoarea conventie de notare, intalnita in multe lucrari din literatura de specialitate a domeniului:

- R, S, T, …: scheme de relatii,
- r, s, …: instante ale relatiilor R respectiv S,
- A, B, C, D, … (litere mari de la inceputul alfabetului): atribute ale unei relatii,
- X, Y, Z, W, U, … (litere mari de la sfarsitul alfabetului): multimi de atribute dintr-o schema de relatie,
- $X \subseteq R$: Multimea de atribute X este inclusa in multimea atributelor relatiei R.
- $Y \subseteq X$: Multimea de atribute Y este inclusa in multimea de atribute X
- $A \in X$: Atributul A apartine multimii de atribute X
- t, t1, t2, … tupluri ale unei relatii,
- t[X]: valorile atributelor din X aflate in tuplul t,
- F, G, …: multimi de dependente functionale atasate unei scheme de relatie

In paragrafele urmatoare termenul generic de **relatie** semnifica atat ***schema relatiei*** (descrierea structurii acesteia) cat si ***o instanta a acesteia*** (continutul de date de la un moment dat al relatiei).

4.2.2. Definitia dependentelor functionale

Definitie: Fie:

- R o schema de relatie
- $X, Y \subseteq R$ doua multimi de atribute ale acesteia.

Spunem ca X ***determina functional pe Y*** (sau Y este determinata functional de X) daca si numai daca oricare ar fi doua tupluri t_1 si t_2 din orice instanta a lui R atunci:

$$t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y].$$

sau, in cuvinte, daca doua tupluri au aceleasi valori pe atributele X atunci ele au aceleasi valori si pe atributele Y.

Notatia pentru dependente functionale este o sageata de la stanga spre dreapta:

$$X \rightarrow Y$$

Exemplu: In relatia Produse din paragraful anterior putem scrie urmatoarele dependente functionale:

IdP → NumeP, Qty, IdF, NumeF, AdresaF,
IdF → NumeF, AdresaF

Aceste dependente arata ca

- daca doua produse au acelasi IdP, este vorba de fapt de acelasi produs
- daca doua produse au acelasi IdF (Id furnizor) atunci si valorile pentru numele si adresa acestuia trebuie sa fie aceleasi.

Observatie: Dependentele functionale nu se determina din inspectarea continutului de la un moment dat al relatiei ci din semnificatia atributelor acesteia. In exemplul prezentat, a doua dependenta functionala arata ca daca la doua produse apare acelasi Id furnizor atunci numele si adresa furnizorului sunt de asemenea aceleasi (deoarece nu pot sa existe doi furnizori diferiti cu acelasi Id).

4.2.3. Axiome si reguli

Pornind de la o multime de dependente functionale atasate unei scheme de relatie se pot deduce alte dependente functionale valide. Exista o multitudine de reguli de inferenta. Pentru a se putea face o prezentare formala a acestora, trei dintre ele au fost alese ca axiome iar restul se pot deduce pornind de la ele. Cele trei axiome (numite in literatura si ***Axiomele lui Armstrong***) sunt urmatoarele:

A1. Reflexivitate: Fie R o schema de relatie si $X \subseteq R$.

$$\text{Daca } Y \subseteq X \text{ atunci } X \rightarrow Y$$

Toate dependentele functionale care rezulta din aceast axioma sunt numite si ***dependente triviale***. Ele nu spun nimic in plus fata de setul de dependente initial dar sunt dependente functionale valide.

A2. Augmentare: Fie R o schema de relatie si $X, Y, Z \subseteq R$.

$$\text{Daca } X \rightarrow Y \text{ atunci si } XZ \rightarrow YZ$$

Aceasta axioma arata ca se poate reuni o aceeasi multime Z in stanga si in dreapta unei dependente functionale valide obtinand de asemenea o dependenta functionala valida.

A3. Tranzitivitate: Fie R o schema de relatie si $X, Y, Z \subseteq R$.

$$\text{Daca } X \rightarrow Y \text{ si } Y \rightarrow Z \text{ atunci si } X \rightarrow Z$$

Pe baza acestor axiome se pot demonstra o serie de reguli de inferenta pentru dependente functionale dintre care cele mai importante sunt urmatoarele:

R1. Descompunere: Fie R o schema de relatie si $X, Y, Z \subseteq R$.

$$\text{Daca } X \rightarrow Y \text{ si } Z \subseteq Y \text{ atunci si } X \rightarrow Z$$

Regula descompunerii ne permite sa rescriem un set de dependente functionale astfel incat sa obtinem doar dependente care au in partea dreapta doar un singur atribut. Sa presupunem ca avem o dependenta functionala de forma:

$$X \rightarrow A_1A_2A_3\ldots A_n$$

Atunci ea poate fi inlocuita cu urmatoarele **n** dependente functionale:

$$X \rightarrow A_1$$
$$X \rightarrow A_2$$
$$X \rightarrow A_3$$
$$\ldots$$
$$X \rightarrow A_n$$

R2. Reuniune: Fie R o schema de relatie si $X, Y, Z \subseteq R$.

$$\text{Daca } X \rightarrow Y \text{ si } X \rightarrow Z \text{ atunci si } X \rightarrow YZ$$

Rezulta si faptul ca din cele **n** reguli obtinute prin descompunere se poate obtine dependenta initiala, deci inlocuirea acesteia nu duce la pierderea vreunei corelatii existente.

R3. Pseudotranzitivitate: Fie R o schema de relatie si X, Y, Z, W $\subseteq$ R.

Daca $X \to Y$ si $YZ \to W$ atunci si $XZ \to W$

Exercitiu: Demonstrati cele trei reguli folosind axiomele lui Armstrong.

Exemplu de demonstratie:
Pentru regula R3: Augmentam prima dependenta cu Z. Obtinem $XZ \to YZ$. Din aceasta dependenta si din $YZ \to W$ obtinem prin tranzitivitate $XZ \to W$, qed.

4.2.4. Inchiderea unei multimi de DF

Pornind de la un set de dependente functionale F si utilizand axiomele si regulile obtinem o multitudine de alte dependente, triviale sau nu. Multimea tuturor dependentelor functionale care se pot deduce din F se numeste ***inchiderea multimii de dependente F***, notata cu F^+. Definitia formala a acestei inchideri este urmatoarea:

$$F^+ = \{X \to Y \mid F \Rightarrow X \to Y\}$$

Unde prin $\Rightarrow$ am notat faptul ca dependenta respectiva de poate deduce din F folosind axiomele si regulile.

Multimea F^+ contine foarte multe dependente, inclusiv dependente triviale ca:

$ABC \to A$, $ABC \to B$, $ABC \to C$, $ABC \to AB$, $ABC \to AC$, $ABC \to BC$ sau $ABC \to ABC$

Ea nu se calculeaza, algoritmii care au nevoie de ea ocolind intr-un fel sau altul calculul acesteia. Introducerii acestei notiuni s-a facut pentru a explica, in cazul descompunerii unei scheme de relatie, care sunt dependentele mostenite de elementele descompunerii de la relatia initiala (paragraful 4.3) si pentru a putea defini formal alte notiuni:

Acoperirea unei multimi de DF: Fie R o schema de relatie si F, G doua multimi de dependente pentru R. Se spune ca F ***acopera*** pe G daca si numai daca $G \subseteq F^+$.

Echivalenta a doua multimi de dependente: Fie R o schema de relatie si F, G doua multimi de dependente pentru R. Se spune ca F ***e echivalenta*** cu G daca si numai daca F acopera pe G si G acopera pe F (deci $G \subseteq F^+$ si $F \subseteq G^+$, deci $F^+ = G^+$)

Forma canonica a unei multimi de DF: Din definitiile de mai sus rezulta ca o multime de dependente poate fi inlocuita cu alta echivalenta continand alte dependente. In cazul in care aceasta multime indeplineste conditiile urmatoare se spune ca este in ***forma canonica***:

- Orice dependenta are in partea dreapta un singur atribut. Acest lucru se poate obtine aplicand regula descompunerii prezentata anterior.
- Multimea de dependente este minimala, nici una dintre dependente neputand sa fie dedusa din celelalte (altfel spus nu exista dependente redundante).

Exemplul 1: Fie R = ABCDE o schema de relatie si F multimea de dependente functionale asociata, cu $F = \{ AB \to CDE, C \to DE \}$:
Aplicam regula de descompunere. Obtinem:

$$F = \{ AB \to C, AB \to D, AB \to E, C \to D, C \to E \}$$

Multimea nu este insa minimala deoarece AB → D si AB → E se pot deduce prin tranzitivitate din AB → C impreuna cu C → D, C → E. Rezulta ca forma canonica a lui F este:

$$F = \{ AB \to C, C \to D, C \to E \}$$

Exemplul 2: Pentru relatia Produse(IdP, NumeP, Qty, IdF, NumeF, AdresaF, IdF) din paragraful 4.1. avand multimea de dependente functionale:

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

Forma canonica a lui F este:

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

De asemenea au fost eliminate doua dependente redundante:

IdP → NumeF si
IdP → AdresaF

4.2.5. Chei si superchei

In acest moment putem sa dam o definitie echivalenta a cheii pe baza dependentelor functionale:

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:

- F ⇒ X → R (deci X → R se poate deduce din F)

si

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

Deci o cheie determina functional toate atributele relatiei si este minimala: nici o submultime stricta a sa nu determina functional pe R. Se observa faptul ca aceasta definitie este echivalenta cu cea din capitolul 3: cunoscandu-se valorile pe atributele X sunt unic determinate valorile pentru toate atributele relatiei, deci este unic determinat tuplul din relatie.

In cazul in care doar prima conditie este indeplinita multimea X se numeste **supercheie**.

Observatie: Faptul ca o supercheie nu este constransa de minimalitate nu inseamna insa ca ea nu poate fi minimala. Rezulta ca orice cheie este in acelasi timp si supercheie, reciproca nefiind insa adevarata.

Exemplu: Fie R = ABCDE si F = { AB → C, C → D, C → E }. Atunci AB este cheie pentru R:

- Din AB → C, C → D si C → E obtinem prin tranzitivitate AB → D si AB → E
- Din AB → C, AB → D si AB → E obtinem prin reuniune AB → CDE
- Din AB → CDE obtinem (augmentare cu AB) AB → ABCDE, deci AB → R

Rezulta ca AB este supercheie pentru R. In paragraful 4.2.8. vom vedea cum se poate demonstra si faptul ca AB este minimala, deci este nu numai supercheie ci chiar cheie pentru R.