

OSNOVE BAZA PODATAKA

(6)

Projektovanje baza podataka
(Normalizacija)

NORMALIZACIJA

Osnovni cilj relacionih baza podataka je da odgovarajuća baza podataka:

- Ne sadrži redundansu,
- Da se može jednostavno koristiti i menjati

Formalne kriterijume prema kojima se utvrđuje da li model podataka ispunjava prethodne zahteve daju ***normalne forme***

- Procesom normalizacije želi se razviti dobar model podataka, tako da se iz zadatog početnog modela otklone slabosti:
 - Redudansa i
 - Problemi u ažuriranju
- Pod redundansom podrazumevamo višestruko memorisanje iste informacije u bazi podataka.

NORMALIZACIJA

- Cilj pri projektovanju baze podataka je eliminisati redundansu zbog negativnih posledica koje redundans donosi.
- Potpuno eliminisanje redundanse podataka u bazi podataka je gotovo nemoguće ostvariti.
- Realni cilj pri projektovanju baze podataka je kontrolisana redundansa podataka.
- Jednostavno korišćenje i menjanje podataka podrazumeva pre svega sprečavanje ***anomalija održavanja podataka.***
- Pod anomalijama održavanja podataka podrazumevamo
 - ***Anomaliju dodavanja***
 - ***Anomaliju brisanja***
 - ***Anomaliju izmene***

NORMALIZACIJA – ANOMALIJE ODRŽAVANJA PODATAKA

Zajednički uzrok anomalija jeste povezivanje opisa svojstava različitih objekata u jedan zapis u bazi podataka.

- **Anomalija dodavanja** javlja se kada su informacije o svojstvima jednog objekta memorisane u bazi podataka kao deo opisa nekog drugog objekta.
- **Anomalija brisanja** je inverzija anomalije dodavanja
- **Anomalija izmene** javlja se u slučaju kada promene podataka o jednom objektu treba izvršiti na više od jedne kopije podataka

U relacionom modelu podataka anomalije održavanja podataka se prevazilaze normalizacijom šema relacija.

NORMALIZACIJA – METODE NORMALIZACIJE

U najopštijem smislu normalizacija je postupak kojim se proizvoljna nenormalizovana šema relacije transformiše u skup manjih normalizovanih šema relacija.

- Bitna osobina koja se očekuje od normalizacije je ***reverzibilnost***, odnosno tokom normalizacije ne sme doći do gubitka informacija sadržanih u polaznoj šemi relacije.

Postoje dve tehnike normalizacije:

- Vertikalna normalizacija
- Horizontalna normalizacija

NORMALIZACIJA – TEHNIKE NORMALIZACIJE

- **VERTIKALNA NORMALIZACIJA** – iz relacije šeme se izdvajaju obeležja koja stoje u nedozvoljenim odnosima sa ostalim obeležjima u šemi.
 - Zasnovana na operacijama **projekcija i prirodni spoj**.
 - Operacijom projekcije relaciju razbijamo na dve ili više relacija.
 - Operacija prirodni spoj se koristi da bi se dokazala reverzibilnost, odnosno da je moguće rekonstruisati polaznu nenormalizovanu relaciju
- **HORIZONTALNA NORMALIZACIJA** – relacija se rastavlja na podskupove n -torki, odnosno fragmente relacije koji zadovoljavaju određene uslove.
 - Zasniva se na operacijama **selekcije i unije**.
 - Koristi se distribuiranih baza podataka

NORMALIZACIJA – VERTIKALNA

Postoje dve varijante vertikalne normalizacije:

- Normalizacija dekompozicijom
- Normalizacija sintezom
- **Normalizacija dekompozicijom** – Započinje od proizvoljne nenormalizovane relacije šeme i izvodi se u koracima.
 - Svakim korakom normalizacije relaciona šema prevodi se u višu normalnu formu,
 - Tako da se polazni skup obeležja deli u dva podskupa i od svakog formira posebna šema relacije.
 - Svaki korak normalizacije mora biti reverzibilan.
- **Normalizacija sintezom** – polazi od skupa obeležja i funkcionalnih zavisnosti zadatih u tom skupu obeležja.

NORMALIZACIJA – DEKOMPOZICIJA BEZ GUBITKA INFORMACIJA

- Dekompozicija relacije šeme $R(A_1, A_2, \dots, A_n)$ je zamena relacije šeme R sa skupom relacionih šema R_1, R_2, \dots, R_k za koje važi da je R_i podskup od R ($1 < i < k$) i $R_1 R_2 \dots R_k = R$ (unija obeležja relacionih šema R_i jednaka je skupu obeležja polazne relacije šeme R).
- Da bi dekompozicija relacije $p(P)$ na $r(R)$ i $s(S)$ bila reverzibilna presek projekcija $r(R)$ i $s(S)$ ne sme biti prazan skup i obeležja koja čine presek moraju biti kandidati za ključ u bar jednoj od dve šeme relacija (šemi relacije R ili šemi relacije S).

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

U kontekstu vertikalne normalizacije definisano je šest normalnih formi:

- Prva normalna forma (1NF)
- Druga normalna forma (2NF)
- Treća normalna forma (3NF)
- Boyce/Codd-ova normalna forma (BCNF)
- Četvrta normalna forma
- Peta normalna forma

Zadataka postupka normalizacije je da relacionu šemu prvo transformiše u 1NF, zatim u 2NF, 3NF i tako redom. Što je redni broj normalne forme veći to je i uslov koji se postavlja strožiji

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

- **Prva normalna forma** – Šema relacije je u prvoj normalnoj formi ako i samo ako je domen svakog od njenih obeležja skup atomarnih vrednosti.
- Obzirom da je u kontekstu relacionog modela sama relacija definisana kao neprazan podskup Dekartovog proizvoda atomarnih vrednosti, sledi da je svaka šema relacije u 1NF.
- **Druga normalna forma** – Relaciona šema R nalazi se u 2NF ako je svako neključno obeležje od R potpuno funkcionalno zavisno od kandidata ključa.
 - Ako relaciona šema R nije u 2NF, postoji takva dekompozicija relacione šeme R u skup relacionih šema koje su u 2NF.

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

- Za relaciju šemu $R(A_1, A_2, \dots, A_n)$ koja nije u 2NF postoje podskupovi obeležja X i Y skupa obeležja $\{A_1, A_2, \dots, A_n\}$ takvi da:
 - Y nisu ključna obeležja
 - X je kandidat za ključ
 - X funkcionalno određuje Y je parcijalna FZ
 - X se može predstaviti kao $X = X'X''$ (kao unija skupova obeležja X' i X'' gde je X' funkcionalno određuje Y potpuna FZ
 - Z je skup svih obeležja šeme relacije R koja nisu ni u X ni u Y

Šemu relacije $R(A_1, A_2, \dots, A_n) = R(XYZ)$ dekomponujemo na šeme relacija $R_1(\underline{X'}Y)$ i $R_2(\underline{X}Z)$.

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

- **Treća normalna forma** – Relaciona šema R nalazi se u 3NF ako je u 1NF i ako ni jedno neključno obeležje u R nije tranzitivno zavisno od ključa od R .
 - Ako relaciona šema $R(A_1, A_2, \dots, A_n)$ nije u 3NF, postoji takva dekompozicija relacije šeme R u skup relacionih šema koje su u 3NF.
 - Za relaciju šemu $R(A_1, A_2, \dots, A_n)$ koja nije u 3NF postoje podskupovi X , Y i Z skupa obeležja $\{A_1, A_2, \dots, A_n\}$ takvi da:
 - Y i Z nisu ključna obeležja,
 - X je kandidat za ključ i važi X funkcionalno određuje Y i Y funkcionalno ne određuje X
 - Y funkcionalno određuje Z je potpuna FZ
- Šemu relacije $R(A_1, A_2, \dots, A_n) = R(XYZW)$

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

- Za relaciju šemu $R(A_1, A_2, \dots, A_n)$ koja nije u 3NF postoje podskupovi X , Y i Z skupa obeležja $\{A_1, A_2, \dots, A_n\}$ takvi da:
 - Y i Z nisu ključna obeležja,
 - X je kandidat za ključ i važi X funkcionalno određuje Y i Y funkcionalno ne određuje X
 - Y funkcionalno određuje Z je potpuna FZ
 - W je skup svih obeležja šeme relacije R koja nisu ni u X ni u Y ni u Z

Šemu relacije $R(A_1, A_2, \dots, A_n) = R(\underline{X}YZW)$ dekomponujemo (zamenjujemo) šemama relacija $R_1(\underline{Y}Z)$ i $R_2(\underline{X}YW)$.

VERTIKALNA NORMALIZACIJA DEKOMPOZICIJOM

Da bi pokazali šta znači minimalna logička redudansa, koje su njene posledice i kako se one mogu eliminisati analizirajmo jedan primer loše projektovanog zapisa.

Prikazana je struktura zapisa koji sadrži grupu sa ponavljanjem, što je čest slučaj u klasičnoj i nekorektno projektovanoj baza podataka.

RADNIK

(Šifra, Ime, Prezime, Adresa, S_Odeljenja, Naziv_Odeljenja, S_Kursa, Naziv_Kursa, Ocena)