**PROJEKTOVANJE BAZE PODATAKA**

**Logičko projektovanje/fizičko projketovanje baze**

Logičko projektovanje baze:

**Normalizacija** – korišćenje ideja o normalizaciji radi razbijanja „velikih“ u male relacije

**Semantičko modeliranje** – upotreba modela entiteta i odnosa radi formiranja „velikih“ relacija

**Normalizacija – uvod**

**Normalizacija** je proces zamene relacija skupom relacija koje su u pogodnijem obliku. Svrha normalizacije je **izbegavanje redundantnosti i pojedinih anomalija ažuriranja**.

U procesu normalizacije operator projekcije se više puta primenjuje na datu relaciju na takav način da spajanjem relacija može da se dođe do početne relacije. Na taj način, proces normalizacije je **reverzibilan** i čuva informacije, tj. uvek je moguće da se uzme proces iz izlaza i preslika unatrag do ulaza.

**Normalne** **forme**

* Prva normalna forma (1NF)
* Druga normalna forma (2NF)
* Treća normalna forma (3NF)
* Bojs-Kodova normalna forma (BCNF)
* Četvrta normalna forma
* Peta normalna forma

**Anomalije**

**Primer**: relacija PredmetIspit koja je dobijena „mešanjem relacija Predmet i Ispit. Relacija sadrži atribute {Id\_predmeta, Sifra, Naziv, Bodovi, Indeks, Ocena}. Primarni ključ je {Id\_predmeta, Indeks} a važi i dodatna FZ Naziv 🡪 Bodovi.

Postoji **redundantnost**: svaki Id\_predmeta 1001 kao naziv pokazuje Analiza 1. Takođe, svaka Analiza 1 za broj bodova ima vrednost 5, itd.

**Anomalije ažuriranja**:

* **Unošenje**: ne može jednostavno da se unese podatak da neki Id odgovara pojedinom nazivu predmetu dok neko nije polagao taj predmet. (jer ne postoji odgovarajući broj indeksa koji je deo p.k.)
* **Brisanje:** Ako se izbrišu podaci za Id\_predmeta, takođe se brišu i podaci o nazivu tog predmeta, njegovoj šifri, broju bodova.
* **Ažuriranje:** Pošto se isti naziv javlja na više mesta, to može da dovede do problema pri promeni naziva.

Zbog toga se vrši dekompozicija (preko projekcija) ovakvih relacija.

**Prva normalna forma**

Relvar je u **1NF** ako i samo ako u svakoj važećoj vrednosti tog relvara svaka torka sadrži tačno jednu vrednost za svaki atribut.

**Druga normalna forma**

Definicija: **Nereducibilna funkcionalna zavisnost**: Ako A🡪B I uklanjanje bilo kog atributa iz A povlači da A🡪B postaje netačno, tada je B nereducibilno zavisno od A.

Naredna definicija podrazumeva postojanje samo **jednog kandidata za ključ** koji je ujedno i primarni ključ.

Relvar je u **2NF** ako i samo ako je svaki neključni atribut nereducibilno zavisan od primarnog ključa.

**Dodatna definicija (Date**): Relvar je u **2NF** ako i samo ako za svaku netrivijalnu FZ X 🡪 Y važi barem jedan od sledećih uslova:

* X je superključ
* Y je podključ (podskupovi kandidata za ključ)
* X nije podključ

**Treća normalna forma**

Naredna definicija podrazumeva postojanje samo **jednog kandidata** za ključ koji je ujedno i primarni ključ.

Relvar je u **3NF** ako i samo ako je u 2NF i svaki neključni atribut je netranzitivno zavisan od primarnog ključa.

**Posledica**: Neključni atributi su uzajamno nezavisni

**Dodatna definicija (Date**): Relvar je u **3NF** ako i samo ako za svaku netrivijalnu FZ X 🡪 Y važi barem jedan od sledećih uslova:

* X je superključ
* Y je podključ (podskupovi kandidata za ključ)

**Definicija preko FZ:** Relvar je u 3NF ako i samo ako za svaku FZ A1A2...An 🡪 B1B2….Bm koja nije trivijalna važi: (ovo je isto kao Date, samo gluplje)

* {A1A2…An} je superključ, ili
* svaki atribut B1B2...Bm !∈ {A1A2...An} je element nekog (ne obavezno istog) kandidata za ključ

**Nedostaci treće normalne forme**

Kodova originalna definicija 3NF nije uzimala u obzir slučajeve kada relacija :

1. ima više od jednog kandidata za ključ
2. kandidat za ključ je kompozitan
3. kompozitni kandidati za ključeve se preklapaju

Ovi slučajevi se obuhvataju Bojs-Kodovom normalnom formom.

Primer: Relacija SPN

S P N

Lazić matematika Petrović

Lazić računarstvo Marković

Perić matematika Petrović

Perić računartsvo Marković

Značenje torke: Student S sluša predmet P kod Nastavnika N.

Pretpostavke:

1. Za svaki predmet, svaki student sluša samo nastavu kod jednog nastavnika, tj {S, P} 🡪 {N}
2. Kandidati za ključ su {S, P} I {S, N}
3. FZ {S, P} 🡪 N
4. Svaki nastavnik predaje samo jedan predmet N 🡪 P

**Bojs-Kodova normalna forma**

Relvar je u **BCNF** ako i samo ako svaka netrivijalna levo-nereducibilna FZ ima kandidat za ključ (superključ) kao svoju levu stranu.

**Manje formalno**: Relvar je u BCNF ako i samo ako su jedini kandidati za ključ leve strane FZ.

**Primer redukcije u BCNF:**

Neka su nazivi predmeta jedinstveni. Relacija PredmetDosije{Id\_predmeta, Naziv, Indeks, Ocena}

Kandidati za ključ su:

{Id\_predmeta, Indeks} 🡪 Ocena

{Naziv, Indeks} 🡪 Ocena

Naziv predmeta je jedinstven:

Id\_predmeta 🡪 Naziv

Naziv 🡪 Id\_predmeta

Id\_predmeta I Naziv su na levoj strani FZ, ali nisu kandidati za ključ => PredmetDosije nije u BCNF.

Rešenje: razbijanje PredmetDosije na dve relacije:

1. IN {Id\_predmeta, Naziv} I PD{Id\_predmeta, Indeks, Ocena} ili
2. IN {Id\_predmeta, Naziv} i PD {Naziv, Indeks, Ocena}

**Moguće redukcije**

U relaciji

Dosije {Indeks, Jmbg, Datum\_rodjenja…}

kandidati za ključ su Indeks i Jmbg. Neka važe sledeće FZ:

Indeks 🡪 Jmbg

Jmbg 🡪 Datum\_rodjenja

….

Relacija nije u 3NF jer postoji tranzitivna zavisnost: Indeks 🡪 Datum\_rodjenja

Razbijanje relacije Dosije:

Dosije\_1a{Indeks, Jmbg…}

Dosije\_1b{Jmbg, Datum\_rodjenja...}

Dosije\_2a{Indeks, Jmbg…}

Dosije\_2b{Indeks, Datum\_rodjenja…}

Dosije\_3a{Jmbg, Datum\_rodjenja…}

Dosije\_3b{Indeks, Datum\_rodjenja…}

Sva tri para jesu u 3NF I BCNF. Dekompozicija [1] i [2] ne dovode do gubitka informacija. Pri dekompoziciji [2] se javlja anomalija pri unosu. => korektno razbijanje je [1].

**Poželjne osobine pri redukciji**

1. Eliminacija anomalija
2. Mogućnost rekonstrukcije informacija – da li početni oblik relacije može da se rekonstruiše iz novih relacija koje su dobijene dekompozicijom
3. Očuvanje FZ. Ako se FZ projektuju na dekompovane relacije, da li će rekonstruisana originalna relacija (dobijena spajanjem relacija koje su rezultat dekompozicije) zadovoljavati originalne FZ?

**Pravila:**

1. Sve FZ polaznog skupa moraju da budu očuvane (direktno ili mogućim izvođenjem iz skupa relacija dobijenih dekompozicijom)
2. Ako u novodobijenim projekcijama nastalim razbijanjem osnvovne relacije postoji zajednički atribut, on mora da bude ključ u bar jednoj od novodobijenih relacija

**Algoritam za dekompoziciju relacije R u BCNF:**

1. Proveriti da li je R u BCNF. Ako jeste, proces je završen i {R} je rešenje.
2. Ako FZ X 🡪 Y narušava BCNF tada za nove relacije uzeti R1 = {X}+ i R2 = {X} U {R – {X}+}
3. Odrediti skup FZ za relacije R1 i R2
4. Primeniti rekurzivno algoritam na relacije R1 i R2, uzimajući uniju svih dekompozicija kao rešenje.

**Višeznačne zavisnosti**

Relacija PNU (Predmet Nastavnik Udžbenik)

P N U

matematika {Petrovic, Markovic} {Analiza, Linearna algebra}

racunarstvo Petrovic {Linearna algebra, Uvod u programiranje, Strukture podataka}

Značenje torke: predmet P može da predaje bilo koji nastavnik N i da koristi bilo koji udžbenik U

Pretpostavke:

1. Za dati kurs postoji proizvoljan broj nastavnika i udžbenika
2. Nastavnici i udzbenici su nezavisni
3. Nastavnik ili udzbenik može da se pridruži bilo kom kursu

U ovoj relaciji ne postoje FZ.

Primarni ključ: {Predmet, Nastavnik, Udzbenik}

Relvar je u 1NF, 2NF, 3NF I BCNF.

Relvar PNU poseduje redundantnost: ako obe torke (p, n1, u1) i (p, n2, u2) postoje tada moraju da postoje i torke (p, n1, u2) i (p, n2, u1)

Posledica je anomalija pri ažuriranju: da bi se uneo da novi nastavnik predaje matematiku moraju da se unesu dve torke, po jedna za svaki udžbenik

Intiuitivno, može da se izvrši dekompozicija:

{Predmet, Nastavnik } i

{Predmet, Udzbenik}

Dekompozicija ne sledi iz FZ (kojih i nema) već iz postojanja višeznačnih zavisnosti.

**Četvrta normalna forma**

**Definicija**: Neka je R relvar i neka su A, B i C podskupovi atributa od R. Kaže se da je B **višeznačno zavisno (VZ)** od A, u oznaci A 🡪🡪 B, ako i samo ako u svakoj mogućoj važećoj vrednosti od R, skup vrednosti B koji se uparuje sa parom (vrednost A, vrednost C) zavisi jedino od vrednosti A i nezavisan je od vrednosti C.

Relvar je u **4NF** ako i samo ako je u BCNF i svaki put kada postoje podskupovi A i B atributa od R takvi da je zadovoljena netrivijlana višeznačna zavisnost A 🡪🡪 B, tada su svi atributi od R takođe funkcionalno zavisni od A.

Primedba: VZ A 🡪🡪 B je trivijalna ako je ili A nadskup od B ili A∩B sadrži sve atribute od R.

**Peta normalna forma:**

Definicija: Neka je R relvar i neka su A,B,....Z podskupovi atributa od R. Tada R zadovoljava **zavisnost spajanja** **(ZS)** \*{A,B,….Z} ako I samo ako je R u 4NF i svaka moguća važeća vrednost u R je jednaka spajanju njenih projekcija na A, B,...Z.

Relvar je u **5NF** (projekcija-spajanje-NF) ako i samo ako je R u 4NF i svaka netrivijalna zavisnost spajanja koja važi u R je posledica kandidata za ključ u R, gde

* **Zavisnost spajanja** \*{A,B,…Z} u R je **trivijalna** akko je najmanje jedan od A,B,..Z skup svih atributa R
* **Zavisnost spajanja** \*{A,B,..Z} je **posledica kandidata za ključ** relvara R akko je svaki od A,B,...Z nadključ za R.

**Redefinicije normalnih formi (preko FZ)**

Relaciona promenljiva R je

1. normalizovana akko je u 1NF.
2. u 2NF akko za svaki ključ K u R i svaki neključni atribut A iz R

FZ K 🡪 {A} koja važi u R je nereducibilna (tj. K je minimalni ključ)

1. u 3NF akko svaka netrivijalna FZ X 🡪 Y koja važi u R ili je je X superključ ili je Y podskup ključa
2. u BCNF akko u svakoj netrivijalnoj FZ X 🡪 Y koja važi u R, X je superključ

**Proces normalizacije**

1. Uzeti projekcije originalnog relvara u 1NF radi eliminisanja FZ koje nisu nereducibilne. Dobijeni skup relvara je u 2NF.
2. Uzeti projekcije relvara u 2NF radi eliminisanja tranzitivnih zavisnosti. Dobijeni skup relvara je u 3NF
3. Uzeti projekcije relvara u 3NF radi eliminisanja preostalih FZ u kojima na levoj strani nije ključ. Dobijeni skup je u BCNF.
4. Uzeti projekcije relvara u BCNF radi eliminisanja VZ koje nisu i FZ. Dobijeni skup relvara je u 4NF.
5. Uzeti projekcije relvara koji su u 4NF i eliminisati ZS koje ne slede iz kandidata za ključ(eve). Dobijeni skup relvara je u 5NF.

**Denormalizacija**

U praksi se često ne sprovodi puna normalizacija zbog dobrih performansi.

* Puna normalizacija dovodi do velikog broja logički razdvojenih relvara.
* Veliki broj razdvojenih relvara znači veliki broj razdvojenih datoteka u kojima se čuvaju.
* Veliki broj datoteka znači veliki broj U/I operacija.

U praksi se normalizacija najčešće sprovodi do **3NF.**

Primeri: videti slajdove