

Normalizacija

Normalizacija baze podataka je iterativni postupak kojim se iz modela baze podataka nastoji otkloniti višestruko ponavljanje istih podataka (**redundacija**) i **povećati stabilnost baze**. Redundantni zapisi, osim bespotrebnog zauzimanja prostora, dovode i do problema sa promjenama sadržaja, tj. do tzv. **anomalija**. Potpuno eliminisanje redundacije podataka u bazi podataka je skoro nemoguće ostvariti. Realni cilj pri projektovanju baze podataka je **kontrolisana redundansa podataka**.

Pri normalizaciji je potrebno da obezbijedimo **reverzibilnost**, što znači da preuređivanjem ne smije doći do gubitka informacija sadržanih u polaznoj relaciji i da na osnovu normalizovanih relacija, **mora biti moguća rekonstrukcija polazne nenormalizovane relacije**.

Tehnike normalizacije

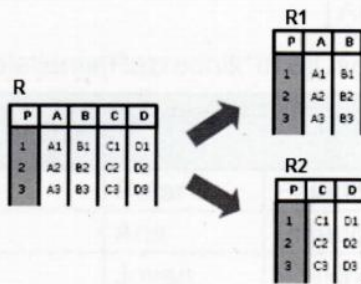
Postoje sljedeće dvije tehnike normalizacije:

1. Horizontalna normalizacija,
2. Vertikalna normalizacija.

Horizontalnom normalizacijom relacija se rastavlja na podskupove n-torki, **fragmente relacije koji zadovoljavaju određene uslove**. Horizontalna normalizacija se zasniva na operacijama **selekcije i unije**. Sama tehnika se koristi kod distribuiranih baza podataka.

Vertikalna normalizacija je postupak kojim se proizvoljna nenormalizovana šema relacije **transformiše u skup manjih i normalizovanih šema relacija**.

Vertikalna normalizacija se zasniva na operacijama **projekcije**. Primjenjujući operacije projekcije relaciju vertikalno dijelimo (razbijamo) na dvije ili više manjih relacija što je prikazano na slici.



Ilustracija vertikalne dekompozicije

Normalizacija dekompozicijom započinje od proizvoljne nenormalizovane relacione šeme i izvodi se u koracima. Vertikalna normalizacija dekompozicijom je najčešće korišćena tehnika normalizacije. Svakim korakom normalizacije relaciona šema se prevodi u višu normalnu formu, tako da se polazni skup obilježja dijeli u dva skupa i od svakog formira posebna relaciona šema.

Svaki korak normalizacije mora biti reverzibilan. Da bi rekonstruisali polaznu, nenormalizovanu relaciju, koristimo operaciju prirodnog spajanja.

Funkcionalna zavisnost

Funkcionalna zavisnost predstavlja (definiše) nezavisnost atributa odnosno polja u tabelama.

Atribut B je funkcionalno zavisn od atributa A ($A \rightarrow B$), ako za bilo koje dvije n-torke $t_1(A)$ i $t_2(A)$ za koje vrijedi $t_1(A) = t_2(A)$, istovremeno vrijedi i $t_1(B) = t_2(B)$. Tada kažemo da A funkcijski određuje B. *Ako postoji više n-torki s istom vrijednošću A, tada i atribut B mora biti isti, ili ne postoji funkcionalna zavisnost.*

Npr., ako se za poznatu vrijednost atributa `Id_Radnik` može saznati prezime radnika znači da `Id_Radnik` funkcionalno određuje prezime radnika.

Funkcionalna zavisnost je tranzitivna (indirektna) ako za $A \rightarrow B$ i $B \rightarrow C$ slijedi $A \rightarrow C$.

Anomalije

Cilj normalizacije je sprečavanje anomalija održavanja podataka. **Pod anomalijama održavanja podataka podrazumijevamo:**

- **Anomaliju dodavanja (insert)** – javlja se u onim slučajevima kada su informacije o atributima jednog entiteta zapamćene kao dio nekog drugog entiteta; što ima za posljedicu da je nemoguće izvršiti dodavanje novog entiteta ukoliko ne postoji informacija o njemu u nekom drugom entitetu.

- **Anomaliju brisanja (delete)** – to je inverzija anomalije dodavanja; kad je nemoguće izbrisati jedan entitet, ukoliko prethodno ne izbrišemo drugi.
- **Anomaliju promjene (ažuriranja – update)** – javlja se u slučaju kada promjenu podataka o jednom entitetu treba izvršiti na više od jedne kopije podataka.

Raspored

Predmet	Aktiv	Nastavnik
Multimedija	Informatika	Marko Marković
Računarski sistemi	Informatika	Jovan Jovanović
Windows	Informatika	Marko Marković
Engleski	Strani jezici	Ana Petrović
Njemački	Strani jezici	Vito Lakić

Anomalija dodavanja: Ne može se dodati nastavnik koji ne predaje nijedan predmet.

Anomalija brisanja: Ako se izbriše Predmet Računarski sistemi briše se i Nastavnik.

Anomalija ažuriranja: Ako se promijeni Aktiv nastavniku Marku Markoviću, moraju se brisati dva reda.

Da bi se izbjegle anomalije, potrebno je provesti normalizaciju tabele.

Forme normalizacije

O normalizaciji se obično govori u obliku formi. **Normalne forme definišu način kako podaci treba da budu strukturisani i uređeni.** Normalizacija je prema tome proces definisanja strukture baze, odnosno relacija koje ju čine.

Forme su hijerarhijski poredane, tako da je prva najjednostavnija. Definirano je šest normalnih formi. U praksi se smatra da je relaciona šema (tabela) normalizovana, ako je u trećoj normalnoj formi.

Standardne normalne forme **su aditivne**, dakle, ako je neki model sveden na treću normalnu formu, automatski je u drugoj i prvoj normalnoj formi.

Prva normalna forma

Relaciona šema R nalazi se u 1NF ako su svi neključni atributi funkcijski zavisni o ključu od R. *Baza podataka nalazi se u prvoj normalnoj formi ako se svaka relaciona šema u njoj nalazi u prvoj normalnoj formi.*

Princip 1NF je da **svaki atribut torke može sadržati samo prostu (atomsku) vrijednost.** To znači da u tabeli treba unositi pojedinačne vrijednosti, ne grupe ili kompozitne objekte.

Za svođenje relacije na 1NF, treba razložiti svaku neatomsku vrijednost.

Pretplatnik

IdPretplatnik	Ime	Prezime	Telefon
101	Pero	Perić	051/123-987 057/234-555
102	Marko	Marković	055/321-456

Relacija (tabela) Pretplatnik nije u 1NF jer kolona Telefon sadrži dva telefonska broja. Tabelu možemo prevesti u 1NF tako da dodamo atribut Telefon2 ili napravimo novu tabelu koja sadrži ključ ove tabele i atribut Telefon.

Pretplatnik 1

IdPretplatnik	Ime	Prezime	Telefon1	Telefon2
101	Pero	Perić	051/123-987	057/234-555
102	Marko	Marković	055/321-456	

Ponekad se umjesto proste vrijednosti kaže: Relacija je u prvoj normalnoj formi ako su skalarni svi domeni u kojima su definisani njeni atributi. Pojam skalar se koristi kao sinonim za prost podatak.

Druga normalna forma

Relacija je u drugoj normalnoj formi (2NF), ako je u 1NF i svi neključni atributi relacije u potpunosti zavise od primarnog ključa.

Zahtjevi 2NF su:

1. Uklanjanje podskupova podataka koji se nalaze u više redova i njihovo smještanje u posebne tabele.
2. Kreiranje veza između novih tabela i tabela sa kojima su spojene korišćenjem spoljnih ključeva.

To znači da se postojeća relacija *raščlani* na nove relacije (tabele) koje zavise samo od primarnog ključa (**u 2NF nije dozvoljena parcijalna zavisnost**), kao što je prikazano na slici.



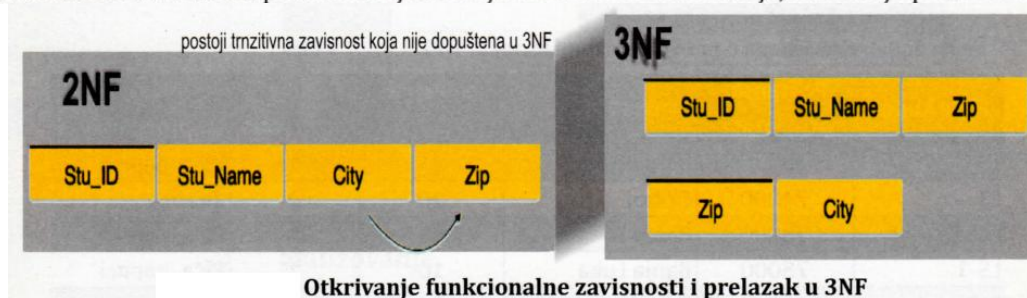
Treća normalna forma

Relacija je u trećoj normalnoj formi, ako je u drugoj normalnoj formi i **ako su svi atributi koji nisu dio nijednog ključa međusobno nezavisni** (ako svi njeni neključni atributi netranzitivno zavise od primarnog ključa).

Osnovni zahtjevi 3NF:

1. Tabela je u 1NF i 2NF.
2. Uklonjene su kolone koje nisu potpuno zavisne od primarnog ključa.

Svođenje na 3NF se provodi uklanjanjem svih podataka u tabelama koji ne zavise jedino od primarnog ključa. *Praktično, postavlja se pitanje: Da li postoje kolone koje ne zavise u potpunosti od primarnog ključa?* Treba zadržati samo podatke koji su zavisni od primarnog ključa, a one koji nisu treba premjestiti u nove tabele i formirati primarni ključ za njih i veze između relacija, kao što je prikazano na slici



Primjeri normalizacije

Koje su mane baze realizovane pomoću relacije (tabele) Pretplatnik1?

Šta kad Pero nabavi treći telefon?

Razmotri bazu Telefonac na slici sa dvije tabele: Pretplatnik2 i Tel_Broj.

Pretplatnik2		
IdPretplatnik	Ime	Prezime
101	Pero	Perić
102	Marko	Marković

Tel_Broj		
Id_Telefon	IdPretplatnik	Broj_Telefona
1000	101	051/123-987
1001	101	057/234-555
1002	102	055/321-456

Baza Telefonac

Kako bismo na osnovu baze Telefonac riješili problem studenata (Ane i Miloša) koji se bave sa više sportova, predstavljen tabelom Sport?

BrojIndeksa	JMBG	Ime	Prezime	DatumRođenja	MjestoRođenja	Sport
2014-22	221	Milica	Marković	5.5.1994.	Milići	Biciklizam
2015-1	123	Pero	Panić	12.2.1995.	Doboj	-
2015-75	1234	Marko	Marković	2.12.1995.	Prijedor	Obojka
2014-3	567	Ana	Janić	3.3.1994.	Teslić	Odbojka
2014-3	567	Ana	Janić	3.3.1994.	Teslić	Košarka
2014-2	221	Miloš	Marković	5.5.1994.	Milići	Plivanje
2014-2	221	Miloš	Marković	5.5.1994.	Milići	Vaterpolo
2014-2	221	Miloš	Marković	5.5.1994.	Milići	Biciklizam
2015-75	91234	Pajo	Pašić	2.12.1995.	Prijedor	Fudbal

Normalizuj bazu da omogući evidenciju više sportova.

Šema 1. Relaciona šema u prvoj normalnoj formi (1 NF)

Šifra kupca	Poštanski broj kupca	Grad kupca	Količina narudžbe	Šifra proizvoda	Naziv proizvoda
LS-1	78000	Banja Luka	21	1	Pića, napitci
S1	74000	Doboj	20	2	Žitarice
LM-2	78000	Banja Luka	45	1	Pića, napitci
LS-1	78000	Banja Luka	10	1	Pića, napitci
S2	79000	Prijedor	50	2	Žitarice
T3	76300	Bijeljina	20	3	Grickalice
T4	76300	Bijeljina	10	4	Kafa
M1	78430	Prnjavor	60	1	Pića, napitci
M2	71123	Istočno Sarajevo	25	2	Žitarice
LS-1	78000	Banja Luka	10	5	Meso

Prevedi šemu 1 u 2NF.

Šema 2. Relaciona šema u drugoj normalnoj formi (2 NF)

Šifra kupca	Šifra proizvoda	Količina narudžbe	Šifra kupca	Poštanski broj kupca	Grad kupca	Šifra proizvoda	Naziv proizvoda
LS-1	1	21	LS-1	78000	Banja Luka	1	Pića, napitci
S1	2	20	S1	74000	Doboj	2	Žitarice
LM-2	1	45	LM-2	78000	Banja Luka	1	Pića, napitci
LS-1	1	10	LS-1	78000	Banja Luka	1	Pića, napitci
S2	2	50	S2	79000	Prijedor	2	Žitarice
T3	3	20	T3	76300	Bijeljina	3	Grickalice
T4	4	10	T4	76300	Bijeljina	4	Kafa
M1	1	60	M1	78430	Prnjavor	1	Pića, napitci
M2	2	25	M2	71123	Istočno Sarajevo	2	Žitarice
LS-1	5	10	LS-1	78000	Banja Luka	5	Meso

Prevedi šemu 2 u 3NF.

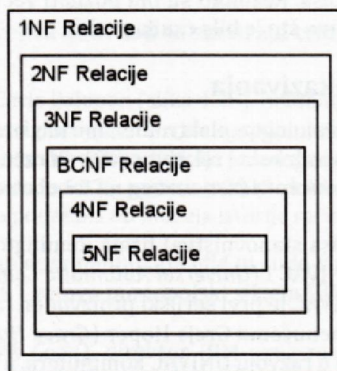
Šema 3. Relaciona šema u trećoj normalnoj formi (3 NF)

Šifra kupca	Poštanski broj kupca		Poštanski broj kupca	Grad kupca
LS-1	78000		78000	Banja Luka
S1	74000		74000	Doboj
LM-2	78000		78000	Banja Luka
LS-1	78000		78000	Banja Luka
S2	79000		79000	Prijedor
T3	76300		76300	Bijeljina
T4	76300		76300	Bijeljina
M1	78430		78430	Prnjavor
M2	71123		71123	Istočno Sarajevo
LS-1	78000		78000	Banja Luka

Ovdje date šeme 1–3 su samo podloga za razmišljanje. Treba ih proširiti s datumom (datumima fakturisanja i prispjeca), jediničnom cijenom, načinom plaćanja i kreirati veze da bi postale upotrebljive u praksi.

Dalje normalizovanje

Prve tri normalne forme su dovoljne u velikoj većini slučajeva. Više normalne forme su bazirane na drugoj vrsti zavisnosti.



Prikaz pozicije i odnosa normalnih formi

(hijerarhijska organizacija: enkapsulacija–aditivnost kao osobina NF)

Poboljšana 3NF se zove i Bojs-Kodova normalna forma (BCNF). Relacija je u Bojs-Kodovoj normalnoj formi (BCNF), ako je svaka njena determinanta i kandidat za ključ.

Četvrta normalna forma uklanja više vrijednosne zavisnosti.

Peta normalna forma uklanja zavisnosti udruživanja.

Pitanja:

- Šta je redundancija?
- Koje su tehnike normalizacije?
- Šta podrazumijevamo pod anomalijom održavanja podataka?
- Koje su forme normalizacije?