



# NORMALIZACIJA



# NORMALIZACIJA

- Normalizacija je postupak **logičkog projektovanja baze podataka** kojim se **odstranjuju anomalije u njenom održavanju**.
- Redudansa (ponavljanje, dupliranje) podataka je jedan od glavnih uzroka pojave anomalija.
- Određena struktura baze podataka se smatra dobrom ako joj je logička redudansa minimalna.

## NENORMALIZOVANA RELACIJA I ANOMALIJE ODRŽAVANJA (2)

Nenormalizovana relacija je relacija koja poseduje neke vrednosti atributa koje nisu "atomske", odnosno relacija koja poseduje "grupe sa ponavljanjem".

### STUDENT

BI	IME	SEM	ŠSMER	IMERUK	ŠPRED	NAZPRED	OCENA
21	ZORAN	5	01	BATA	121	MATEMAT	7
					323	BAZEPOD	8
					056	SOCIOLO	8
77	ANA	7	01	BATA	056	MARKSIZ	10
					121	MATEMAT	5
36	PERA	4	02	MIKA	323	BAZEPOD	8
					456	ELEKTRON	9
					442	FIZIKA	6
					056 S	OCIOLO	8

Jedna n-torka ove nenormalizovane relacije bi se mogla zapisati ovako:

(21, ZORAN, 5, 01, BATA, { (121, MATEMAT, 7), (323, BAZEPOD, 8), (056, SOCIOLO, 8) })

## NENORMALIZOVANA RELACIJA I ANOMALIJE ODRŽAVANJA (3)

- **Anomalije u dodavanju:**

Ako je u novom nastavnom planu definisan novi predmet, ne mogu se ubaciti podaci o tom predmetu dok ga neki student ne položi.

Ili, ako se otvori neki novi smer, ne mogu se ubaciti podaci o tom smeru dok ga neki student ne upiše.

- **Anomalije u brisanju:**

Ako je jedan predmet (FIZIKA) položio samo jedan student (PERA) i ako se on ispiše sa fakulteta, odnosno izbaciti odgovarajuća n-torka, gube se i sve informacije o tom predmetu.

Ako je taj student bio i jedini student na nekom smeru, gube se i sve informacije o tom smeru.

- **Anomalije u ažuriranju:**

Ako se promeni naziv nekog predmeta ili rukovodioc nekog smeru, to se mora učiniti na onoliko mesta koliko je studenata položilo taj predmet, odnosno koliko je studenata upisano na dati smer.

# NENORMALIZOVANA RELACIJA I PROBLEMI IZVEŠTAVANJA

- Izveštavanje za datu strukturu
  - Data struktura relacije je veoma pogodna za izveštaj na zahtev za izdavanje uverenja o položenim ispitima.
  - "Prikaži listu predmeta, imena svih studenata koji su ga položili i prosečnu ocenu na predmetu", za datu strukturu relacije zahtevao bi znatno složeniji program.
- Postupkom normalizacije logicka struktura baze podataka se dovodi u takav oblik (ili, drugim recima, relacije se dovode u normalne forme) u kome se **izbegavaju anomalije u održavanju i problemi u izveštavanju**.

# ARMSTRONGOVE AKSIOME IZVOĐENJA

- Aksiome izvođenja su pravila koja utvrđuju važenje nekih funkcionalnih zavisnosti u relaciji R, na osnovu važenja nekih drugih funkcionalnih zavisnosti u R.
- Mogu se izvesti na osnovu osobina f-nih zavisnosti, pa u svojoj suštini nisu prave aksiome.
- Armstrongove aksiome (u modifikovanom obliku) su:
  - Aksioma refleksivnosti:  
Funkcionalna zavisnost  $X \rightarrow X$  uvek važi.
  - Aksioma proširenja:  
Ako u relaciji R važi  $X \rightarrow Y$  i ako je Z podskup skupa  $Atr(R)$ , onda važi i  $X, Z \rightarrow Y$ .
  - Aksioma aditivnosti:  
Ako u relaciji R važi  $X \rightarrow Y$  i  $X \rightarrow Z$ , onda važi i  $X \rightarrow Y, Z$ .
  - Aksioma projektivnosti:  
Ako u relaciji R važi  $X \rightarrow Y, Z$ , onda važi i  $X \rightarrow Y$  i  $X \rightarrow Z$ .

# ARMSTRONGOVE AKSIOME IZVOĐENJA

- Aksioma tranzitivnosti:  
Ako u relaciji R važi  $X \rightarrow Y$  i  $Y \rightarrow Z$ , onda važi i  $X \rightarrow Z$ .
- Aksioma pseudotranzitivnosti:  
Ako u relaciji R važi  $X \rightarrow Y$  i  $Y, Z \rightarrow W$ , onda važi i  $X, Z \rightarrow W$ .

## PRVA NORMALNA FORMA (1NF)

- Relacija R je u Prvoj normalnoj formi (1NF) ako su sve vrednosti njenih atributa atomske.
- Termin normalizovana relacija se koristi za relacije koje su u 1NF.
- Svi upitni jezici zasnovani na relacionoj algebr zahtevaju normalizovane relacije.



## PRVA NORMALNA FORMA (1NF)

- **Relacija R je u Prvoj normalnoj formi (1NF) ako su sve vrednosti njenih atributa atomske.**
- Termin **normalizovana** relacija se koristi za relacije koje su u 1NF.
- Svi upitni jezici zasnovani na relacionoj algebr zahtevaju normalizovane relacije.

### STUDENT

BI	IME	SEM	ŠSMER	IMERUK	ŠPRED	NAZPRED	OCENA
21	ZORAN	5	01	BATA	121	MATEMAT	7
21	ZORAN	5	01	BATA	323	BAZEPOD	8
21	ZORAN	5	01	BATA	056	SOCIOL	8
77	ANA	7	01	BATA	056	MARKSIZ	10
77	ANA	7	01	BATA	121	MATEMAT	5
36	PERA	4	02	MIKA	323	BAZEPOD	8
36	PERA	4	02	MIKA	456	ELEKTRON	9
36	PERA	4	02	MIKA	442	FIZIKA	6
36	PERA	4	02	MIKA	056	SOCIOL	8

# FUNKCIONALNE ZAVISNOSTI

Data je relacija  $R$  sa atributima  $X$  i  $Y$ , moguće složenim. Atribut  $Y$  je funkcionalno zavisen od atributa  $X$  (ili  $X$  funkcionalno određuje  $Y$ ),

$$R.X \rightarrow R.Y,$$

ako i samo ako svakoj vrednosti  $X$  odgovara jedna i samo jedna vrednost  $Y$ .

Primer.

STUDENT I (BI, IME, SEM, ŠSMER, IMERUK) (I)

BI  $\rightarrow$  IME

BI  $\rightarrow$  SEM

BI  $\rightarrow$  ŠSMER

BI  $\rightarrow$  IMERUK

Atribut  $Y$  relacije  $R$  je funkcionalno zavisen od atributa  $X$  relacije  $R$  ako i samo ako kad god dve  $n$ -torke relacije  $R$  imaju istu  $x$ -vrednost one moraju imati istu  $y$ -vrednost.

# POTPUNE FUNKCIONALNE ZAVISNOSTI

Atribut  $Y$  relacije  $R$  je **potpuno funkcionalno zavisan** od kolekcije atributa  $X$  relacije  $R$  ako je funkcionalno zavisan od kolekcije atributa  $X$ , a nije funkcionalno zavisan ni od jednog pravog podskupa kolekcije atributa  $X$ .

Primere potpune i nepotpune funkcionalne zavisnosti možemo pokazati na relaciji PRIJAVA:

(a)  $BI, \check{S}PRED \twoheadrightarrow OCENA$

$BI \not\rightarrow OCENA$

$\check{S}PRED \not\rightarrow OCENA$

atribut  $OCENA$  je potpuno funkcionalno zavisan od složenog atributa  $BI, \check{S}PRED$ .

(b)  $BI, \check{S}PRED \dashrightarrow NAZPRED$

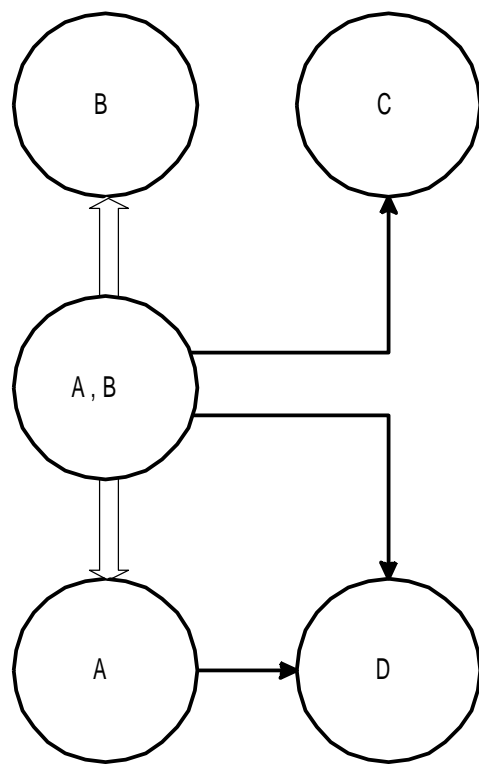
$BI \not\rightarrow NAZPRED$

$\check{S}PRED \dashrightarrow NAZPRED$

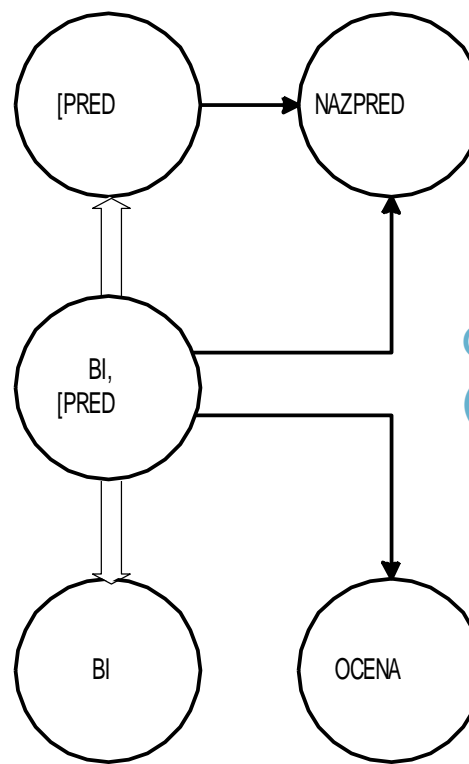
atribut  $NAZPRED$  je nepotpuno funkcionalno zavisan od složenog atributa  $BI, \check{S}PRED$ , jer je funkcionalno zavisan i od njega i od jednog njegovog dela, od  $\check{S}PRED$ .

# POTPUNE FUNKCIONALNE ZAVISNOSTI

C je potpuno funkcionalno zavisno od složenog atributa (A,B), a D nije.



sa  $\Rightarrow$  su označene trivijalne funkcije



OCENA je potpuno zavisna od (BI, ŠPRED), a NAZPRED nije.

## DEFINICIJA KLJUČA I NADKLJUČA

- Atribut  $X$  moguće složeni, je nadključ neke relacije  $R$  ako i samo ako funkcionalno određuje sve ostale attribute relacije  $R$ .
- Atribut  $X$ , moguće složeni, je ključ relacije  $R$  ako je nadključ relacije  $R$ , a nijedan njegov pravi podskup nema tu osobinu.

## DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

- Relacija R se dekomponuje u svoje projekcije bez gubljenja informacija ako prirodno spajanje tako dobijenih projekcija dovodi do polazne relacije.

PRIJAVA	BI	ŠPRED	NAZPRED	OCENA
	21	121	MATEMAT	7
	21	323	BAZEPOD	8
	21	056	SOCIOL	8
	77	056	SOCIOL	10
	77	121	MATEMAT	5
	36	323	BAZEPOD	8
	36	456	ELEKTRON	9
	36	442	FIZIKA	6
	36	056	SOCIOL	8

RI	BI	ŠPRED	OCENA
	21	121	7
	21	323	8
	21	056	8
	77	056	10
	77	121	5
	36	323	8
	36	456	9
	36	442	6
	36	056	8

R2	NAZPRED	OCENA
	MATEMAT	7
	BAZEPOD	8
	SOCIOL	8
	SOCIOL	10
	MATEMAT	5
	ELEKTRON	9
	FIZIKA	6

# DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

**R3 = R1 [OCENA \* OCENA] R2**

R3	BI	ŠPRED	NAZPRED	OCENA
----	----	-------	---------	-------

	21	121	MATEMAT	7
	21	323	BAZEPOD	8
	21	056	SOCIOI	6
	77	056	SOCIOI	10
	77	121	MATEMAT	5
	36	323	BAZEPOD	9
	36	456	ELEKTRON	9
	36	442	FIZIKA	6
	36	056	SOCIOI	8

---

	21	323	SOCIOI	8
	36	056	BAZEPOD	8
	21	056	FIZIKA	6
	36	442	SOCIOI	6
	36	323	ELEKTRON	9
	36	465	BAZEPOD	9

# DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

- Heath-ova teorema

Relacija  $R(A,B,C)$ , gde su  $A$ ,  $B$  i  $C$  podskupovi atributa, u kojoj važi  $R.A \twoheadrightarrow R.B$  može se uvek dekomponovati u svoje projekcije  $R_1(A,B)$  i  $R_2(A,C)$  bez gubljenja informacija.

- Za navedeni primer relacije PRIJAVA dekompozicija bez gubljenja informacija bi bila dekompozicija u projekcije  
 $R_1(BI, \text{ŠPRED}, OCENA)$  i  $R_2(\text{ŠPRED}, NAZPRED)$



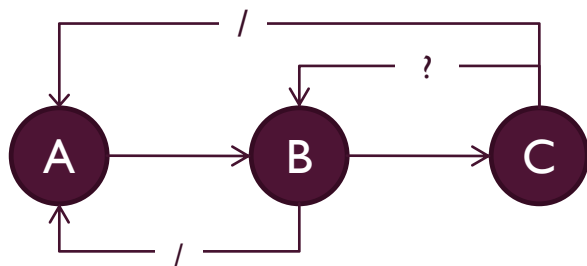
## DRUGA NORMALNA FORMA (2NF)

Relacija R je u Drugoj normalnoj formi (2NF) ako i samo ako je u 1NF i svi njeni neključni atributi potpuno i funkcionalno zavise od primarnog ključa.

- Svodenje na 2NF vrši se dekompozicijom na taj način što u jednoj projekciji ostavlja primarni ključ i svi atributi koji su potpuno funkcionalno zavisni od njega, a u drugim projekcijama se realizuju one funkcionalne zavisnosti koje su prouzrokovale nepotpune funkcionalne zavisnosti.

# TRANZITIVNE FUNKCIONALNE ZAVISNOSTI

Atribut C je **tranzitivno funkcionalno zavisan od atributa A** ako je funkcionalno zavisan od A i ako je funkcionalno zavisan od nekog atributa B koji je i sam funkcionalno zavisan od A.

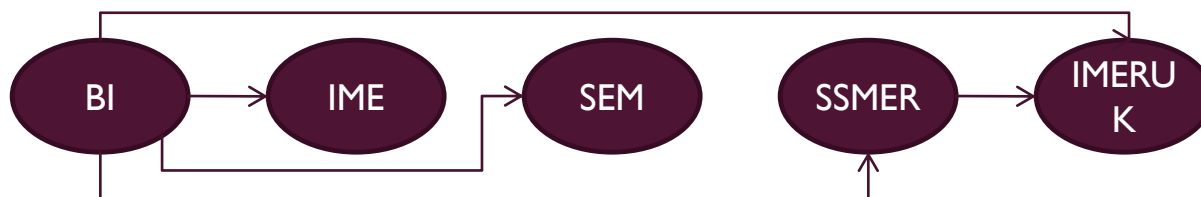


Tranzitivna funkcionalna zavisnost je redundantna, pa je nije neophodno pamtit u bazi podataka.

$f_1: BI \dashrightarrow \check{S}SMER$

$f_2: \check{S}SMER \dashrightarrow IMERUK$

$f_3 = f_1 \circ f_2 : BI \dashrightarrow IMERUK$



## TREĆA NORMALNA FORMA (3NF)

Relacija R je u Trecoj normalnoj formi (3NF) ako i samo ako je u 2NF i ako svi njeni neključni atributi **netranzitivno funkcionalno** zavise od primarnog ključa.

Svodenje na 2NF vrši se dekompozicijom na taj način što u jednoj projekciji ostavlja primarni ključ i svi atributi netranzitivno zavisni atributi, a u drugim projekcijama se realizuju funkcionalne zavisnosti koje su dovele do tranzitivnih zavisnosti.

## BOYCE-CODD-OVA NORMALNA FORMA (BCNF)

- Relacija sa tzv "preklapajucim" kandidatima za ključ (dva ili više složenih kandidata za ključ koji imaju barem jedan zajednicki atribut).

### **PRIJAVA(BI, ŠPRED, NAZPRED, OCENA)**

BI, ŠPRED  $\rightarrow$  NAZPRED

BI, ŠPRED  $\rightarrow$  OCENA

ŠPRED  $\rightarrow$  NAZPRED

NAZPRED  $\rightarrow$  ŠPRED

KK(BI, ŠPRED) i KK(BI, NAZPRED)

- **Determinanta** relacije R je bilo koji atribut, prost ili složen, od koga neki drugi atribut u relaciji potpuno funkcionalno zavisi.
- Relacija R je u Boyce-Codd-ovoj normalnoj formi (BCNF) ako i samo ako su **sve determinante u relaciji i kandidati za ključ**.

## BOYCE-CODD-OVA NORMALNA FORMA (BCNF)

PRIJAVA(BI, ŠPRED, NAZPRED, OCENA)

BI, ŠPRED  $\rightarrow$  NAZPRED, OCENA (D) (KK)

BI, NAZPRED  $\rightarrow$  ŠPRED, OCENA (D) (KK)

ŠPRED  $\rightarrow$  NAZPRED (D)

NAZPRED  $\rightarrow$  ŠPRED (D)

KK(BI, ŠPRED) i KK(BI, NAZPRED)

- Dekompozicijom, pri kojoj se iz relacije "izvlace" projekcije sa onim determinantama koje nisu kandidati za ključ, relacija se svodi na BCNF.

PRIJAVA1(BI, ŠPRED, OCENA)

PREDMET(ŠPRED, NAZPRED)

- Svaka relacija koja je u BCNF je sigurno i u 2NF i 3NF. Obrnuto ne važi.

## DEKOMPOZICIJA NA ZAVISNE I NEZAVISNE PROJEKCIJE

- Projekcije R1 i R2 relacije R su nezavisne ako i samo ako važi sledeće:
  1. Svaka funkcionalna zavisnost u R se može logicki dedukovati iz funkcionalnih zavisnosti u R1 i R2 i
  2. Zajednicki atribut relacija R1 i R2 je kandidat za ključ barem u jednoj od njih.

## VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

### **PREDMET NASTAVNIK KNJIGA**

INF-SIST	BRANKO	MARTIN
	KRCA	DATE
	BAJA	
SIST-ANAL	VLADAN	DEMARCO
		SARSON

Jedan predmet predaje više nastavnika. Za jedan predmet se koristi više knjiga. Ne postoji nikakva veza između nastavnika i knjiga.

### ■ Normalizovana

#### **PREDMET NASTAVNIK KNJIGA**

INF-SIST	BRANKO	MARTIN
INF-SIST	BRANKO	DATE
INF-SIST	KRCA	MARTIN
INF-SIST	KRCA	DATE
INF-SIST	BAJA	MARTIN
INF-SIST	BAJA	DATE
SIST-ANAL	VLADAN	DEMARCO
SIST-ANAL	VLADAN	SARSON

# VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

RASP(PREDMET, NASTAVNIK)      i      UDŽBENIK(PREDMET, KNJIGA)

<b>RASP</b>	<b>PREDMET</b>	<b>NASTAVNIK</b>	<b>UDŽBENIK</b>	<b>PREDMET</b>	<b>KNJIGA</b>
	INF-SIST	BRANKO		INF-SIST	MARTIN
	INF-SIST	KRCA		INF-SIST	DATE
	INF-SIST	BAJA		SIST-ANAL	DEMARCO
	SIST-ANAL	VLADAN		SIST-ANAL	SARSON

- Dekompozicija je bez gubljenja informacija.
- Veze koje postoje između atributa ove relacije nazivaju se višeznacnim vezama.



## VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

- U relaciji  $R(A, B, C)$  postoji višeznacna zavisnost  $A \twoheadrightarrow B$  ako za datu vrednost  $A$ , postoji skup od nula, jedne ili više vrednosti  $B$ , a taj skup vrednosti ni na koji nacin ne zavisi od vrednosti atributa  $C$ . Atributi  $A$ ,  $B$  i  $C$  mogu biti složeni.
- U relaciji  $R(A, B, C)$  postoji višeznacna zavisnost  $A \twoheadrightarrow B$  ako i samo ako kad god u njoj postoje  $n$ -torke

$\langle a, b, c \rangle$  i  $\langle a, b', c' \rangle$ ,

postoje takode i  $n$ -torke

$\langle a, b, c' \rangle$  i  $\langle a, b', c \rangle$ .

Atributi  $A$ ,  $B$  i  $C$  mogu biti složeni.

- U prethodnom primeru važile su

$\text{PREDMET} \twoheadrightarrow \text{NASTAVNIK}$  i  $\text{PREDMET} \twoheadrightarrow \text{KNJIGA}$ .

## VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

- Relacija R je u Četvrtoj normalnoj formi (4NF) ako i samo ako kad god postoji višeznacna funkcionalna zavisnost, na primer  $A \twoheadrightarrow B$ , tada svi atributi relacije moraju takode biti funkcionalno zavisni od A.
- Relacija R je u 4NF ako u njoj nisu date dve (ili više) nezavisne višeznacne cinjenice.

## ZAVISNOSTI SPAJANJA I PETA NORMALNA FORMA (5NF)

- U relaciji  $R(X, Y, \dots, Z)$  postoji zavisnost spajanja ako i samo ako relacija  $R$  rezultuje iz prirodnog spajanja njenih projekcija po  $X, Y, \dots, Z$ , gde su  $X, Y, \dots, Z$  podskupovi atributa relacije  $R$ .
- Relacija  $R$  je u Petoj normalnoj formi ako i samo ako se svaka zavisnost spajanja može pripisati kandidatu za ključ.