

POSTUPCI NORMALIZACIJE

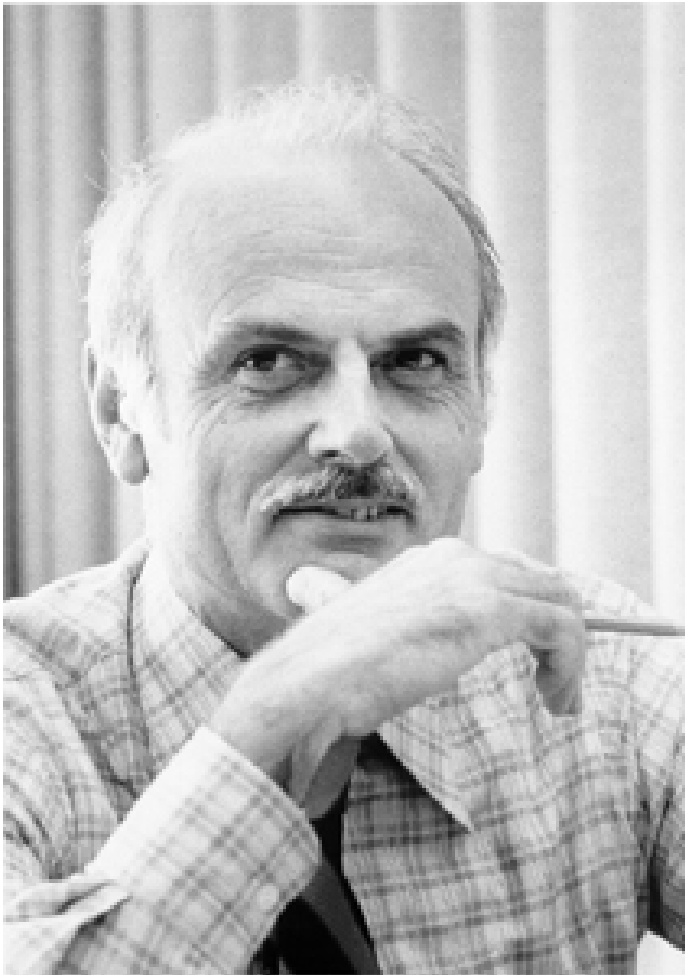
V predavanju

Dr.sc. Emir Mešković

Postupci normalizacije

- ▶ Uočena znanja o međusobnim zavisnostima atributa relacije koriste se u postupcima normalizacije
- ▶ Cilj normalizacije je:
 - ▶ Ukloniti redundanciju
 - ▶ Anomalije unosa, izmjene i brisanja
 - ▶ Neracionalno korištenje prostora za pohranu
 - ▶ Spriječiti pojavu lažnih n-torki
- ▶ Postupci normalizacije omogućavaju da se postepeno, tačno definisanom metodom, odredi dobra zamjena za loše koncipiranu shemu
- ▶ E.F. Codd
 - ▶ “Normalized database structure: A brief tutorial”, Proc. ACM SIGFIDET Workshop on Data Description, Access and Control, 1971

Postupci normalizacije



- ▶ “I called it normalization because then-President Nixon was talking a lot about normalizing relations with China. I figured that if he could normalize relations, so could I” – ‘A “Freshside” Chat’, DBMS, Dec. 1993)

Normalne forme

- ▶ **Zasnovane na funkcijskim zavisnostima**
 - ▶ Prva normalna forma – 1 NF
 - ▶ Druga normalna forma – 2 NF
 - ▶ Treća normalna forma – 3 NF
 - ▶ Boyce-Coddova normalna forma – BCNF
- ▶ **Zasnovana na višeznačnim zavisnostima**
 - ▶ Četvrta normalna forma – 4 NF
- ▶ **Zasnovana na spojnim zavisnostima**
 - ▶ Projekcijsko-spojna normalna forma - PJNF

Postupci normalizacije

- ▶ **Dekompozicija**

- ▶ Početne relacije (relacijske sheme) se dekomponuju na osnovu uočenih funkcijskih zavisnosti

- ▶ **Sinteza**

- ▶ Zadan je skup atributa i nad njima skup funkcijskih zavisnosti iz kojih se sintetiziraju relacijske sheme koje zadovoljavaju 3 NF

Postupak normalizacije dekompozicijom

- ▶ Dekompozicija – relacijska shema R zamjenjuje se shemama R_1, R_2, \dots, R_n pri čemu vrijedi $R = R_1 R_2 \dots R_n$
- ▶ $R_i \subseteq R$ $r_i(R_i) = \pi_{R_i}(R)$
- ▶ Dekompozicija relacijske sheme R na sheme R_1, R_2, \dots, R_n obaviće se bez gubitka informacija ako vrijedi:
 - ▶ $r_1(R_1) \bowtie r_2(R_2) \bowtie \dots \bowtie r_n(R_n) = r(R)$
 - ▶ $\pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \dots \bowtie \pi_{R_n}(r) = r$

Razlaganje relacija bez gubitka na dvije projekcije

<u>OSOBA(matBr,</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
12345	Pirić	Damir	75000	Tuzla
23456	Đurić	Maja	71000	Sarajevo
34567	Pejić	Dino	72000	Zenica
45678	Pirić	Damir	71000	Sarajevo

- ▶ U kojim zadacima će se relacija osoba dekomponirati bez gubitka informacija? Zašto?

- ▶ **Zadatak 1**

- ▶ $r_1 = \pi_{matBr, prezime, ime}(osoba)$
- ▶ $r_2 = \pi_{matBr, postBr, nazivGrad}(osoba)$
- ▶ $r_3 = r_1 \bowtie r_2$

- ▶ **Zadatak 2**

- ▶ $r_4 = \pi_{prezime, ime, postBr, nazivGrad}(osoba)$
- ▶ $r_5 = \pi_{matBr, prezime}(osoba)$
- ▶ $r_6 = r_4 \bowtie r_5$

- ▶ **Zadatak 3**

- ▶ $r_7 = \pi_{matBr, prezime, ime, postBr}(osoba)$
- ▶ $r_8 = \pi_{postBr, nazivGrad}(osoba)$
- ▶ $r_9 = r_7 \bowtie r_8$

Razlaganje relacija bez gubitka na dvije projekcije

► U zadatku 1

- $R_1 = \{\text{matBr}, \text{prezime}, \text{ime}\}$ $K_{R1} = \{\text{matBr}\}$
- $R_2 = \{\text{matBr}, \text{postBr}, \text{nazivGrad}\}$ $K_{R2} = \{\text{matBr}\}$

► U zadatku 2

- $R_4 = \{\text{prezime}, \text{ime}, \text{postBr}, \text{nazivGrad}\}$ $K_{R1} = \{\text{prezime}, \text{ime}, \text{postBr}\}$
- $R_5 = \{\text{matBr}, \text{prezime}\}$ $K_{R2} = \{\text{matBr}\}$

► U zadatku 3

- $R_7 = \{\text{matBr}, \text{prezime}, \text{ime}, \text{postBr}\}$ $K_{R1} = \{\text{matBr}\}$
- $R_8 = \{\text{postBr}, \text{nazivGrad}\}$ $K_{R2} = \{\text{postBr}\}$

► Relacija se bez gubitaka razlaže na dvije projekcije ako:

- projekcije imaju zajedničke atribute
- zajednički atributi su **ključ u barem jednoj od projekcija**

Prva normalna forma – 1 NF

- ▶ Definicija prve normalne forme proizilazi iz same definicije relacije
- ▶ Relacijska shema je u 1 NF ako:
 - ▶ neključni atributi funkcijski ovise o ključu
 - ▶ domene sadrže samo jednostavne (nedjeljive) vrijednosti
 - ▶ vrijednost svakog atributa je samo jedna vrijednost iz domene tog atributa
- ▶ Shema baze podataka $R = \{R_1, R_2, \dots, R_n\}$ je u prvoj normalnoj formi ako je svaka relacijska shema u 1 NF

1 NF – primjer 1

- ▶ Preduzeće evidentira podatke o zaposlenicima:

	<u>ZAP(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>datRod,</u>	<u>sifOdjel)</u>
	12345	Pirić	Damir	23.10.63	111
zap(ZAP)	23456	Đurić	Maja	16.07.61	22
	34567	Pejić	Dino	08.11.64	111

$$K_{ZAP} = \{ matBr \}$$

- ▶ Domene svih atributa sadrže jednostavne vrijednosti
- ▶ Vrijednost svakog atributa je samo jedna vrijednost iz domene tog atributa
- ▶ Postoje funkcijske zavisnosti:
 - ▶ $matBr \rightarrow prezime$ $matBr \rightarrow ime$
 - ▶ $matBr \rightarrow datRod$ $matBr \rightarrow sifOdjel$
- ▶ Relacijska shema ZAP je u 1 NF

Neključni atributi relacije funkcijski ovise o ključu relacije

1 NF – primjer 2

- ▶ Preduzeće evidentira podatke o zaposlenicima i njihovoj djeci – korsnicima zdravstvenog osiguranja (1. način):

	<u>ZAP I (matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>imenaDjece)</u>
	12345	Pirić	Damir	Deni, Jasna
<i>zap I (ZAP I)</i>	23456	Đurić	Maja	Goran
	34567	Pejić	Dino	Denis, Maja, Irma

$$K_{ZAP I} = \{ matBr \}$$

- ▶ Domena atributa *imenaDjece* ne sadrže jednostavne vrijednosti
- ▶ Relacijska shema *ZAP I* nije u 1 NF

1 NF – primjer 2

- ▶ Preduzeće evidentira podatke o zaposlenicima i njihovoj djeci – korsnicima zdravstvenog osiguranja (2. način):

	<i>ZAP2</i> (<i>matBr</i>	<i>prezime</i> ,	<i>ime</i> ,	<i>imeDj</i> ,	<i>datRodDj</i>)
<i>zap2</i> (<i>ZAP2</i>)	12345	Pirić	Damir	Deni	16.05.90
				Jasna	02.09.93
	23456	Đurić	Maja	Goran	14.12.87
				Denis	22.04.85
	34567	Pejić	Dino	Maja	11.06.88
				Irma	06.10.90

$$K_{ZAP2} = \{ matBr \}$$

- ▶ Domene sadrže jednostavne vrijednosti, ali vrijednost atributa *imeDj* nije uvijek samo jedna vrijednost iz domene tog atributa (isto vrijedi i za atribut *datRodDj*)
- ▶ Relacijska shema *ZAP2* nije u 1 NF

1 NF – primjer 2

- ▶ Preduzeće evidentira podatke o zaposlenicima i njihovoj djeci – kornicima zdravstvenog osiguranja (3. način):

	<i>ZAP3(matBr</i>	<i>prezime,</i>	<i>ime,</i>	<i>imeDj,</i>	<i>datRodDj)</i>
	12345	Pirić	Damir	Deni	16.05.90
	12345	Pirić	Damir	Jasna	02.09.93
<i>zap3(ZAP3)</i>	23456	Đurić	Maja	Goran	14.12.87
	34567	Pejić	Dino	Denis	22.04.85
	34567	Pejić	Dino	Maja	11.06.88
	34567	Pejić	Dino	Irma	06.10.90

$$K_{ZAP3} = \{ matBr \}$$

- ▶ Domene sadrže jednostavne vrijednosti
- ▶ Vrijednost svakog atributa je samo jedna vrijednost iz domene tog atributa
- ▶ Postoje funkcijske zavisnosti:
 - ▶ $matBr \rightarrow prezime$ $matBr \rightarrow ime$
- ▶ Ne postoje funkcijske zavisnosti
 - ▶ $matBr \rightarrow imeDj$ $matBr \rightarrow datRodDj$
- ▶ Relacijska shema ZAP3 nije u 1 NF

Normalizacija na 1 NF

- ▶ Izdvajanjem atributa koji nisu funkcijski ovisni o ključu
 - ▶ U posebnu relaciju izdvaja se skup atributa koji se ponavlja sa jednakom kratnošću, zajedno sa ključem originalne relacije

	<u>ZAP4(matBr</u>	<u>prezime,</u>	<u>ime)</u>	
	12345	Pirić	Damir	
<i>zap4(ZAP4)</i>	23456	Đurić	Maja	$K_{ZAP4} = \{ matBr \}$
	34567	Pejić	Dino	

	<u>DIJETE(matBr</u>	<u>imeDj,</u>	<u>datRodDj)</u>	
	12345	Deni	16.05.90	
	12345	Jasna	02.09.93	
<i>dijete(DIJETE)</i>	23456	Goran	14.12.87	$K_{DIJETE} = \{ matBr, imeDj \}$
	34567	Denis	22.04.85	
	34567	Maja	11.06.88	
	34567	Irma	06.10.90	

Normalizacija na 1 NF

- ▶ Izdvajanjem atributa koji nisu funkcijski ovisni o ključu
- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ X \rightarrow Y, XZ \rightarrow V \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ X \}$
- ▶ R ne zadovoljava 1 NF
- ▶ Normalizacijom na 1 NF R se zamjenjuje shemama:
 - ▶ $R_1 = XY$ $K_{R1} = \{ X \}$
 - ▶ $R_2 = XZV$ $K_{R2} = \{ XZ \}$
- ▶ Relacija $r(R)$ se normalizacijom na prvu normalnu formu zamjenjuje projekcijama:
 - ▶ $r_1 = \pi_{XY}(R)$ $r_2 = \pi_{XZV}(R)$
- ▶ Operacija je izvedena bez gubitka informacija
 - ▶ Relacijske sheme imaju zajedničke attribute (X)
 - ▶ Zajednički atributi su ključ u R_1

Normalizacija na 1 NF

► Promjenom ključa

	<u>ZAP3(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>imeDj,</u>	<u>datRodDj)</u>
zap3(ZAP3)	12345	Pirić	Damir	Deni	16.05.90
	12345	Pirić	Damir	Jasna	02.09.93
	23456	Đurić	Maja	Goran	14.12.87
	34567	Pejić	Dino	Denis	22.04.85
	34567	Pejić	Dino	Maja	11.06.88
	34567	Pejić	Dino	Irma	06.10.90

$$K_{ZAP3} = \{ matBr, imeDj \}$$

Normalizacija na 1 NF

- ▶ Promjenom ključa
- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ X \rightarrow Y, XZ \rightarrow V \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ X \}$
- ▶ R ne zadovoljava 1 NF
- ▶ Normalizacijom na 1 NF u relacijskoj shemi R definiše se novi ključ:
 - ▶ $K_R = \{ XZ \}$

Druga normalna forma – 2 NF

- ▶ Relacijska shema R je u 2 NF ako je u 1 NF i ako je:
 - ▶ Svaki atribut iz zavisnog dijela **potpuno funkcijski ovisan o ključu**
- ▶ Zadana je relacijska shema R i skupovi atributa X i Y iz R ($X, Y \subseteq R$)
- ▶ Neka u R vrijedi funkcijska zavisnost $X \rightarrow Y$
- ▶ Funkcijska zavisnost $X \rightarrow Y$ je **nepotpuna** ako postoji atribut ili skup atributa Z koji je podskup od X , za koji vrijedi $Z \rightarrow Y$
- ▶ Ili
- ▶ FZ $X \rightarrow Y$ je **nepotpuna** ako
- ▶ $(\exists Z)(Z \subset X) : Z \rightarrow Y$

2 NF – primjer 3

	<u>ZAP3(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>imeDj,</u>	<u>datRodDj)</u>
	12345	Pirić	Damir	Deni	16.05.90
	12345	Pirić	Damir	Jasna	02.09.93
<i>zap3</i> (ZAP3)	23456	Đurić	Maja	Goran	14.12.87
	34567	Pejić	Dino	Denis	22.04.85
	34567	Pejić	Dino	Maja	11.06.88
	34567	Pejić	Dino	Irma	06.10.90

$$K_{ZAP3} = \{ matBr, imeDj \}$$

- ▶ Relacijska shema ZAP3 zadovoljava 1 NF
- ▶ Postoje funkcijske zavisnosti:
 - ▶ $matBr, imeDj \rightarrow prezime$ $matBr, imeDj \rightarrow ime$
- ▶ Ali postoje i funkcijske zavisnosti
 - ▶ $matBr \rightarrow prezime$ $matBr \rightarrow ime$
- ▶ $matBr, imeDj \rightarrow prezime, ime$ **je nepotpuna funkcijska zavisnost!**
- ▶ Relacijska shema ZAP3 nije u 2 NF

2 NF – primjer 3

- ▶ Normalizacijom na 2 NF nastaju:
 - ▶ relacijska shema koja sadrži skup atributa koji su bili nepotpuno funkcijski ovisni o ključu i dio ključa o kojem su potpuno funkcijski ovisni
 - ▶ relacijska shema koja sadrži ključ originalne relacije i skup atributa koji su potpuno funkcijski ovisni o ključu

	<u>ZAP4(matBr</u>	<u>prezime,</u>	<u>ime)</u>
	12345	Pirić	Damir
<i>zap4(ZAP4)</i>	23456	Đurić	Maja
	34567	Pejić	Dino

$$K_{ZAP4} = \{ matBr \}$$

	<u>DIJETE(matBr</u>	<u>imeDj,</u>	<u>datRodDj)</u>
	12345	Deni	16.05.90
	12345	Jasna	02.09.93
<i>dijete(DIJETE)</i>	23456	Goran	14.12.87
	34567	Denis	22.04.85
	34567	Maja	11.06.88
	34567	Irma	06.10.90

$$K_{DIJETE} = \{ matBr, imeDj \}$$

Normalizacija na 2 NF

- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ X \rightarrow Y, XZ \rightarrow V \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ XZ \}$
- ▶ R zadovoljava 1 NF
- ▶ Funkcijska zavisnost $XZ \rightarrow Y$ je nepotpuna $\Rightarrow R$ ne zadovoljava 2 NF
- ▶ Normalizacijom na 2 NF R se zamjenjuje shemama:
 - ▶ $R_1 = XY$ $K_{R_1} = \{ X \}$
 - ▶ $R_2 = XZV$ $K_{R_2} = \{ XZ \}$
- ▶ Relacija $r(R)$ se normalizacijom na 2 NF zamjenjuje projekcijama:
 - ▶ $r_1 = \pi_{XY}(R)$ $r_2 = \pi_{XZV}(R)$
- ▶ Operacija je izvedena bez gubitka informacija
 - ▶ Relacijske sheme imaju zajedničke attribute (X)
 - ▶ Zajednički atributi su ključ u R_1

Treća normalna forma

- ▶ Relacija (relacijska shema) je u 3 NF ako je u 1 NF i ako:
 - ▶ niti jedan atribut iz zavisnog dijela nije tranzitivno funkcijski ovisan o ključu
- ▶ Posredna funkcijska zavisnost
- ▶ Zadano je:
 - ▶ Relacijska shema R
 - ▶ Skupovi atributa $X, Y, Z \in R$
 - ▶ Skup funkcijskih zavisnosti F
- ▶ Atribut Z je tranzitivno ovisan o X ako vrijedi:
 - ▶ $X \rightarrow Y, Y \not\rightarrow X$ i $Y \rightarrow Z$ (proizilazi iz F)
 - ▶ $Z \not\subset XY$

3 NF – primjer 4

	<u>OSOBA(matBr,</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
	12345	Pirić	Damir	75000	Tuzla
osoba(OSOBA)	23456	Đurić	Maja	71000	Sarajevo
	34567	Pejić	Dino	75000	Tuzla

$$K_{OSOBA} = \{ matBr \}$$

- ▶ Relacijska shema *OSOBA* zadovoljava 1 NF
- ▶ Postoji funkcijska zavisnost:
 - ▶ $matBr \rightarrow prezime, ime, postBr, nazivGrad$
- ▶ Ali postoje i funkcijska zavisnost
 - ▶ $postBr \rightarrow nazivGrad$ (ne postoji $postBr \rightarrow matBr$)
- ▶ $matBr \rightarrow nazivGrad$ je tranzitivna zavisnost!
- ▶ Relacijska shema *OSOBA* ne zadovoljava 3 NF

Normalizacija na 3 NF

- ▶ Normalizacijom na 3 NF nastaju:
 - ▶ relacijska shema koja sadrži skup atributa relacijske sheme OSOBA koji su tranzitivno ovisni o ključu, te srednji skup atributa uočene tranzitivne zavisnosti
 - ▶ relacijska shema koja sadrži ključ relacijske sheme OSOBA i neključne attribute relacijske sheme OSOBA koji nisu tranzitivno ovisni o ključu

<u>GRAD(<i>postBr</i>, <i>nazivGrad</i>)</u>			$K_{GRAD} = \{ postBr \}$
grad(GRAD)	75000	Tuzla	
	71000	Sarajevo	

<u>OSOBA I (<i>matBr</i>, <i>prezime</i>, <i>ime</i>, <i>postBr</i>)</u>					$K_{OSOBA I} = \{ matBr \}$
osoba I (OSOBA I)	12345	Pirić	Damir	75000	
	23456	Đurić	Maja	71000	
	34567	Pejić	Dino	75000	

Normalizacija na 3 NF

- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ X \rightarrow YZV, Z \rightarrow V \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ X \}$
- ▶ R zadovoljava 1 NF
- ▶ Funkcijska zavisnost $X \rightarrow V$ je tranzitivna $\Rightarrow R$ ne zadovoljava 3 NF
- ▶ Normalizacijom na 3 NF R se zamjenjuje shemama:
 - ▶ $R_1 = XYZ$ $K_{R1} = \{ X \}$
 - ▶ $R_2 = ZV$ $K_{R2} = \{ Z \}$
- ▶ Relacija $r(R)$ se normalizacijom na 2 NF zamjenjuje projekcijama:
 - ▶ $r_1 = \pi_{XYZ}(R)$ $r_2 = \pi_{ZV}(R)$
- ▶ Operacija je izvedena bez gubitka informacija
 - ▶ Relacijske sheme imaju zajedničke attribute (Z)
 - ▶ Zajednički atributi su ključ u R_2

Normalizacija na 3NF - komentar

- ▶ Normalizacija na 2 NF nije nužan preduslov za provođenje normalizacije na 3 NF jer se nepotpune FZ mogu posmatrati kao tranzitivne FZ
- ▶ **Primjer:** zadana je shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{XY \rightarrow ZV, X \rightarrow Z\}$. Ključ relacije $K_R = XY$. R je u 1 NF, ali nije u 2 NF jer postoji nepotpuna FZ $XY \rightarrow Z$. Međutim, postoji i tranzitivna funkcijska zavisnost $XY \rightarrow Z$ ($XY \rightarrow X \wedge X \rightarrow Z$).
- ▶ Normalizacijom na 3NF shema R se zamjenjuje shemama:
 - ▶ $R_1 = XZ$ $R_2 = XYV$
 - ▶ $K_{R_1} = X$ $K_{R_2} = XY$
- ▶ R_1 i R_2 su u 2NF i 3NF
- ▶ Preporuka: normalizaciju ipak obavljati postepeno:
 - ▶ $1NF \Rightarrow 2NF \Rightarrow 3NF$

3 NF – primjer 5

	<i>OSOBA2</i> (<i>matBr</i> ,	<i>prezime</i> ,	<i>ime</i> ,	<i>JMBG</i>)
	12345	Pirić	Damir	1710977180025
<i>osoba2</i> (<i>OSOBA2</i>)	23456	Đurić	Maja	1812982185011
	34567	Pejić	Dino	0209979180016

- ▶ Postoji funkcijska zavisnost:
 - ▶ $matBr \rightarrow prezime, ime, JMBG$
- ▶ Ali postoje i funkcijska zavisnost
 - ▶ $JMBG \rightarrow prezime, ime$ i $JMBG \rightarrow matBr$
- ▶ *matBr* i *JMBG* su mogući ključevi
- ▶ Relacijska shema *OSOBA2* zadovoljava 3 NF
 - ▶ $K1_{OSOBA2} = \{ matBr \}$
 - ▶ $K2_{OSOBA2} = \{ JMBG \}$

Normalizacija na 3 NF – dodatna razmatranja

- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ X \rightarrow YZV, Z \rightarrow V, Z \rightarrow X \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ X \}$
- ▶ R zadovoljava 1 NF
- ▶ Zbog $X \rightarrow Z$ i $Z \rightarrow X$
 - ▶ Funkcijsku zavisnost $Z \rightarrow V$ nije potrebno ukloniti jer u tom slučaju nema redundancije
- ▶ Z je također ključ u R
 - ▶ $K1_R = \{ X \}$
 - ▶ $K2_R = \{ Z \}$
- ▶ X i Z su mogući ključevi $\Rightarrow R$ zadovoljava 3 NF

Primjer normalizacije

- ▶ Zadana je relacijska shema *STUD* i trenutna vrijednost relacije *stud*

<i>(matBr,</i>	<i>prezime,</i>	<i>ime,</i>	<i>sifPred,</i>	<i>nazPred,</i>	<i>datIspr,</i>	<i>ocj,</i>	<i>sifNas,</i>	<i>prezNas)</i>
12345	Pirić	Damir	1111	Baze pod.	23.01.12	5	1001	Mešković
12345	Pirić	Damir	1111	Baze pod.	14.02.12	5	1002	Pjanić
12345	Pirić	Damir	1111	Baze pod.	26.06.12	7	1001	Mešković
12345	Pirić	Damir	2222	Razvoj softv.	05.02.12	8	1003	Hasanović
12345	Pirić	Damir	3333	Oper. sist.	03.07.12	9	1002	Pjanić
23456	Đurić	Maja	1111	Baze pod.	23.01.12	8	1001	Mešković

- ▶ Pretpostavljeni ključ je $K_{STUD} = \{ matBr \}$
- ▶ Na osnovu značenja podataka odrediti funkcijske zavisnosti
- ▶ Normalizirati relacijsku shemu *STUD* na 1 NF, 2 NF, 3 NF

Normalizacija na 1 NF

- ▶ Da li ključ funkcijski određuje zavisni dio?
- ▶ Postoji FZ:
 - ▶ $matBr \rightarrow prezime, ime$
- ▶ Ne postoje FZ:
 - ▶ $matBr \rightarrow sifPred$ $matBr \rightarrow nazPred$
 - ▶ $matBr \rightarrow datIsplit$ $matBr \rightarrow ocj$
 - ▶ $matBr \rightarrow sifNas$ $matBr \rightarrow prezNas$
- ▶ $student = \pi_{matBr, prezime, ime}(stud)$
- ▶ $ispit = \pi_{matBr, sifPred, nazPred, datIsplit, ocj, sifNas, prezNas}(stud)$

Normalizacija na 1 NF

<u>student(matBr,</u>	<u>prezime,</u>	<u>ime</u>
12345	Pirić	Damir
23456	Đurić	Maja

2 NF?

student ima jednostavan ključ
ispit ne zadovoljava 2 NF

$$K_{STUDENT} = \{ matBr \}$$

<u>ispit(matBr,</u>	<u>sifPred,</u>	<u>datIsplit,</u>	<u>nazPred</u>	<u>ocj,</u>	<u>sifNas,</u>	<u>prezNas)</u>
12345	1111	23.01.12	Baze pod.	5	1001	Mešković
12345	1111	14.02.12	Baze pod.	5	1002	Pjanić
12345	1111	26.06.12	Baze pod.	7	1001	Mešković
12345	2222	05.02.12	Razvoj softv.	8	1003	Hasanović
12345	3333	03.07.12	Oper. sist.	9	1002	Pjanić
23456	1111	23.01.12	Baze pod.	8	1001	Mešković

$$K_{ISPIT} = \{ matBr, sifPred, datIsplit \}$$

Normalizacija na 2 NF

- ▶ $ispitI = \pi_{matBr, sifPred, datIsplit, ocj, sifNas, prezNas}(ispit)$
- ▶ $predmet = \pi_{sifPred, nazPred}(ispit)$

$predmet(sifPred, \underline{nazPred})$	
1111	Baze pod.
2222	Razvoj softv.
3333	Oper. sist.

3 NF?

student zadovoljava 3 NF

predmet zadovoljava 3 NF

ispitI ne zadovoljava 3 NF

$$K_{PREDMET} = \{ sifPred \}$$

$ispitI(matBr, sifPred, datIsplit, ocj, \overset{\curvearrowright}{sifNas, prezNas})$					
12345	1111	23.01.12	5	1001	Mešković
12345	1111	14.02.12	5	1002	Pjanić
12345	1111	26.06.12	7	1001	Mešković
12345	2222	05.02.12	8	1003	Hasanović
12345	3333	03.07.12	9	1002	Pjanić
23456	1111	23.01.12	8	1001	Mešković

$$K_{ISPITI} = \{ matBr, sifPred, datIsplit \}$$

Normalizacija na 3 NF

- ▶ $ispit2 = \pi_{matBr, sifPred, datIsplit, ocj, sifNas}(ispit1)$
- ▶ $nastavnik = \pi_{sifNas, prezNas}(ispit1)$

$nastavnik(sifNas, prezNas)$	$ispit2(matBr, sifPred, datIsplit, ocj, sifNas)$
1001 Mešković	12345 1111 23.01.12 5 1001
1002 Pjanić	12345 1111 14.02.12 5 1002
1003 Hasanović	12345 1111 26.06.12 7 1001
	12345 2222 05.02.12 8 1003
$K_{NASTAVNIK} = \{ sifNas \}$	12345 3333 03.07.12 9 1002
	23456 1111 23.01.12 8 1002

$$K_{ISPIT2} = \{ matBr, sifPred, datIsplit \}$$

- ▶ Shema baze podataka *STUSLU* sastoji se od relacijskih shema:
 - ▶ $STUSLU = \{ STUDENT, PREDMET, NASTAVNIK, ISPIT2 \}$
 - ▶ Shema baze podataka *STUSLU* zadovoljava 3 NF

Primjer normalizacije – opšti atributi

- ▶ Zadana je relacijska shema $R = ABCDEFGH$ i na njoj skup funkcijskih zavisnosti $F = \{ AB \rightarrow E, C \rightarrow F, ABCD \rightarrow GH, G \rightarrow H \}$.
- ▶ Domene atributa sadrže samo jednostavne vrijednosti, vrijednost svakog atributa je samo jedna vrijednost iz domene tog atributa.
- ▶ Odrediti primarni ključ relacije (tako da bude zadovoljen uslov 1NF kojem neključni atributi funkcijski zavise od ključa) te shemu postepeno normalizirati na 2NF i 3NF.

Primjer normalizacije – 1NF

- ▶ $R = ABCDEFGH$
- ▶ $F = \{ AB \rightarrow E, C \rightarrow F, ABCD \rightarrow GH, G \rightarrow H \}$
- ▶ Odrediti primarni ključ relacije:
 - ▶ $AB \rightarrow E \Rightarrow ABCD \rightarrow E$ (iz A2 - uvećanje)
 - ▶ $C \rightarrow F \Rightarrow ABCD \rightarrow F$ (iz A2 - uvećanje) $\Rightarrow ABCD \rightarrow EFGH$
 - ▶ $ABCD \rightarrow GH$ (iz F)
- ▶ Postoji li skup $X \subset ABCD$ za kojeg vrijedi $X \rightarrow R$
 - ▶ NE

$\Rightarrow ABCD$ je mogući ključ i može se odabrati kao primarni ključ sheme R

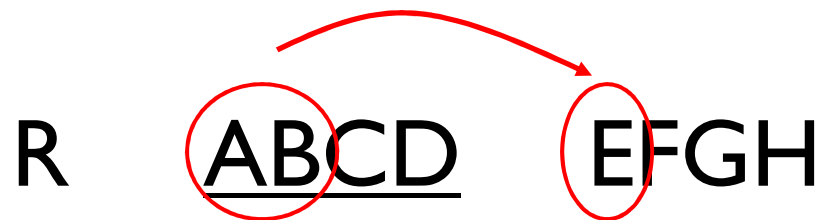
$R = ABCDEFGH$

$K_R = ABCD$

R je u 1NF

Primjer normalizacije – 2NF

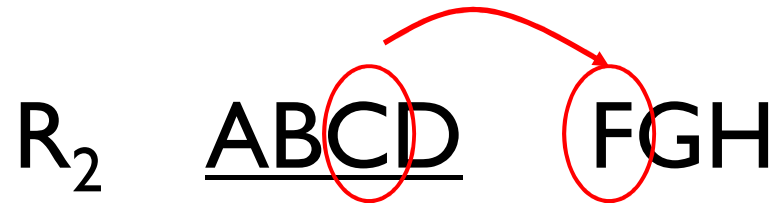
- ▶ $R = ABCDEFGH$ $K_R = ABCD$
- ▶ $F = \{ AB \rightarrow E, C \rightarrow F, ABCD \rightarrow GH, G \rightarrow H \}$
- ▶ Normalizacija na 2 NF
 - ▶ Svi atributi iz zavisnog dijela moraju biti potpuno funkcijski ovisni o ključu



- ▶ $ABCD \rightarrow E$ je nepotpuna FZ jer vrijedi $AB \rightarrow E$
 $\Rightarrow R$ nije u 2NF
- ▶ Normalizacijom na 2 NF R se zamjenjuje shemama:
 - ▶ $R_1 = ABE$ $K_{R_1} = AB$ R_1 je u 2NF
 - ▶ $R_2 = ABCDFGH$ $K_{R_2} = ABCD$ R_2 nije u 2NF

Primjer normalizacije – 2NF (nastavak)

- ▶ $R_1 = ABE$ $K_{R_1} = AB$ R_1 je u 2NF
- ▶ $R_2 = ABCDFGH$ $K_{R_2} = ABCD$ R_2 nije u 2NF
- ▶ $F = \{ AB \rightarrow E, C \rightarrow F, ABCD \rightarrow GH, G \rightarrow H \}$
- ▶ Normalizacija na 2 NF
 - ▶ Svi atributi iz zavisnog dijela moraju biti potpuno funkcijski ovisni o ključu



- ▶ $ABCD \rightarrow F$ je nepotpuna FZ jer vrijedi $C \rightarrow F \Rightarrow R_2$ nije u 2NF
- ▶ Normalizacijom na 2 NF R_2 se zamjenjuje shemama:
 - ▶ $R_{21} = CF$ $K_{R_{21}} = C$ R_{21} je u 2NF
 - ▶ $R_{22} = ABCDGH$ $K_{R_{22}} = ABCD$ R_{22} je u 2NF

Primjer normalizacije – 3NF

- ▶ $R_1 = ABE$ $K_{R_1} = AB$ R_1 je u 3NF
- ▶ $R_{21} = CF$ $K_{R_{21}} = C$ R_{21} je u 3NF
- ▶ $R_{22} = ABCDGH$ $K_{R_{22}} = ABCD$ R_{22} nije u 3NF
- ▶ $F = \{ AB \rightarrow E, C \rightarrow F, ABCD \rightarrow GH, G \rightarrow H \}$
- ▶ Normalizacija na 3 NF

- ▶ Niti jedan atribut iz zavisnog dijela ne smije biti tranzitivno ovisan o ključu

R_{22} ABCD \Rightarrow GH $\Rightarrow ABCD \rightarrow H$ je tranzitivna FZ

- ▶ Normalizacijom na 3 NF R_{22} se zamjenjuje shemama:
 - ▶ $R_{221} = GH$ $K_{R_{221}} = G$ R_{221} je u 3NF
 - ▶ $R_{222} = ABCDG$ $K_{R_{222}} = ABCD$ R_{222} je u 3NF
- ▶ Shema baze podataka u 3NF se sastoji od relacijskih shema:

▶ R_1, R_{21}, R_{221} i R_{222}

Boyce-Coddova normalna forma - BCNF

- ▶ Relacijska shema je u BCNF ako je u 1 NF i ako:
 - ▶ niti jedan atribut iz R nije tranzitivno ovisan o bilo kojem ključu od R , ili
 - ▶ sve funkcijske zavisnosti u R proizlaze iz ključeva od R

$DZ(\underline{sifPacijent}, \textit{ambulanta}, sifLjekar)$				$K_{DZ} = \{ sifPacijent, ambulanta \}$
$dz(DZ)$	12345	ORL	123	
	12345	KARD	234	
	23456	ORL	345	
	23456	KARD	234	

- ▶ Relacija dz , odnosno relacijska shema DZ zadovoljava 1 NF
- ▶ Postoji FZ $sifLjekar \rightarrow ambulanta$
- ▶ Relacija dz , odnosno relacijska shema DZ zadovoljava 3 NF
 - ▶ Atribut *ambulanta* je tranzitivno ovisan o ključu
- ▶ Relacija dz , odnosno relacijska shema DZ ne zadovoljava BCNF

Normalizacija na BCNF

► Nastaju relacijske sheme

<u>DZI (sifLjekar, ambulanta)</u>		<u>LIJECI (sifPacijent, sifLjekar)</u>	
123	ORL	12345	123
dzl (DZI) 234	KARD	12345	234
345	ORL	lijeci(LIJECI) 23456	345
		23456	234

$$K_{DZI} = \{ \text{sifLjekar} \}$$

$$K_{LIJECI} = \{ \text{sifPacijent, sifLjekar} \}$$

► Problemi sa BCNF

- Izvođenjem transformacija gubi se početni skup funkcijskih zavisnosti
- Ne smanjuje se redundancija u svakom slučaju
- Zaključak: normalizaciju na BCNF nije obavezno provoditi

Normalizacija na BCNF

- ▶ Neka su X, Y, Z i V atributi ili skupovi atributa
- ▶ Zadana je relacijska shema $R = XYZV$ i na njoj skup funkcijskih zavisnosti $F = \{ XY \rightarrow ZV, Z \rightarrow X \}$
- ▶ Neka je pretpostavljeni ključ $K_R = \{ XY \}$
- ▶ R zadovoljava 3 NF
- ▶ Zbog funkcijske zavisnosi $Z \rightarrow X$
 - ▶ $\Rightarrow R$ ne zadovoljava BCNF
- ▶ Normalizacijom na BCNF R se zamjenjuje shemama:
 - ▶ $R_1 = ZYV$ $K_{R_1} = \{ ZY \}$
 - ▶ $R_2 = ZX$ $K_{R_2} = \{ Z \}$
- ▶ Relacija $r(R)$ se normalizacijom na BCNF zamjenjuje projekcijama:
 - ▶ $r_1 = \pi_{ZYV}(R)$ $r_2 = \pi_{ZX}(R)$
- ▶ Operacija je izvedena bez gubitka informacija
 - ▶ Relacijske sheme imaju zajedničke attribute (Z)
 - ▶ Zajednički atributi su ključ u R_2

Normalizacija - zaključak

- ▶ Shema baze podataka ima dobra svojstva ako zadovoljava 3 NF
- ▶ Shema baze podataka zadovoljava 3 NF ako svaka njezina relacijska shema zadovoljava 3 NF