

## 9. Normalizácia dátového modelu

### 9.1 Funkčné závislosti

#### 9.1.1 Definícia funkčných závislostí

##### **Definícia - Funkčná závislosť**

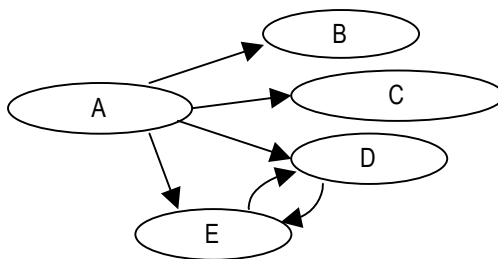
Ak existuje relácia **R** s relačnou schémou **R(A,B)**, kde **A** a **B** sú množiny atribútov, potom hovoríme, že množina atribútov **B** je funkčne závislá od množiny atribútov **A**, vtedy a len vtedy, keď pre každú kombináciu hodnôt atribútov **A** existuje práve jedna kombinácia hodnôt atribútov **B**.

◆

Z predchádzajúcej definície vyplýva, že hodnota atribútov **A** determinuje hodnotu atribútov **B**. Označenie  $A \rightarrow B$  znamená, že množina atribútov **B** je funkčne závislá od množiny atribútov **A** a označenie  $A \nrightarrow B$  znamená, že množina atribútov **B** nie je funkčne závislá od množiny atribútov **A**.

##### □ *Príklad 9.3*

Majme reláciu **R(A, B, C, D, E)**, potom graf funkčných závislostí môže vyzeráť nasledovne:



##### **Definícia - Elementárna funkčná závislosť**

Elementárna funkčná závislosť je taká funkčná závislosť  $A \rightarrow B$ , keď atribút **A** je atomický.

◆

##### **Definícia - Úplná funkčná závislosť**

Ak existuje relácia **R** s relačnou schémou **R(A,B)**, kde **A** a **B** sú množiny atribútov, potom hovoríme, že množina atribútov **B** je úplne funkčne závislá od množiny atribútov **A**, vtedy a len vtedy, keď množina atribútov **B** je funkčne závislá od celej množiny atribútov **A** a nie od ľubovoľnej kombinácie podmnožiny atribútov množiny atribútov **A**.

◆

##### *Poznámka*

Úplnú funkčnú závislosť hľadáme ak vo funkčnej závislosti  $A \rightarrow B$ , je atribút **A** kompozitný (t. j. množina atribútov). Ak atribút **A** je atomický, tak každá funkčná závislosť je úplnou funkčnou závislosťou.

**Definícia - Vzájomná funkčná závislosť**

Ak existuje relácia **R** s relačnou schémou **R(A,B)** hovoríme, že množiny atribútov **A** a **B** sú vzájomne funkčne závislé, ak platí, že **B** je funkčne závislé od **A** ( $A \rightarrow B$ ) a zároveň **A** je funkčne závislé od **B** ( $B \rightarrow A$ ).

◆

Označenie  $A \leftrightarrow B$  znamená, že množiny atribútov **A** a **B** sú vzájomne funkčne závislé.

**Definícia - Tranzitívna funkčná závislosť**

Ak existuje relácia **R** s relačnou schémou **R(A,B,C)**, ak platí že **B** je funkčne závislé od **A** ( $A \rightarrow B$ ), a zároveň **C** je funkčne závislé od **B** ( $B \rightarrow C$ ), potom **C** je funkčne závislé od **A** ( $A \rightarrow C$ ).

◆

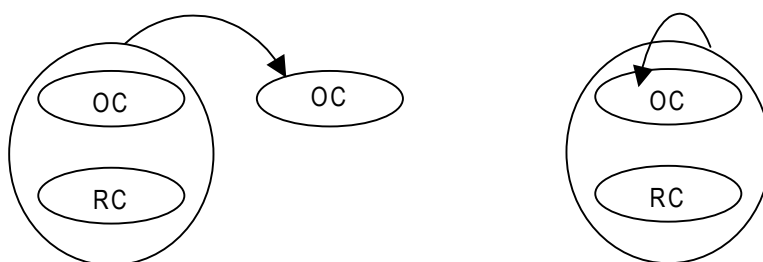
**Definícia - Triviálna funkčná závislosť**

Ak existuje relácia **R** s relačnou schémou **R(A,B)**, kde **A** a **B** sú množiny atribútov, triviálna funkčná závislosť je keď platí, že **A** je funkčne závislé od dvojice **A,B**.

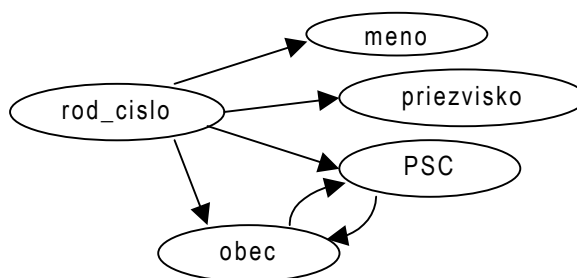
◆

□ **Príklad 9.4 – Triviálna funkčná závislosť:**

$A, B \rightarrow A$

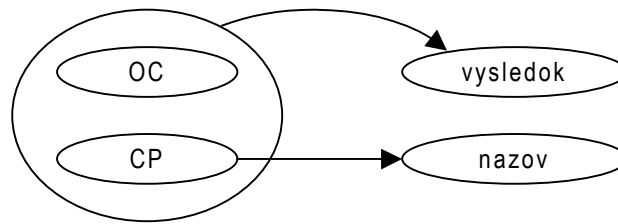
□ **Príklad 9.5 – Funkčné závislosti:**

Osobne\_udaje(rod\_cislo, meno, priezvisko, PSC, Obec)



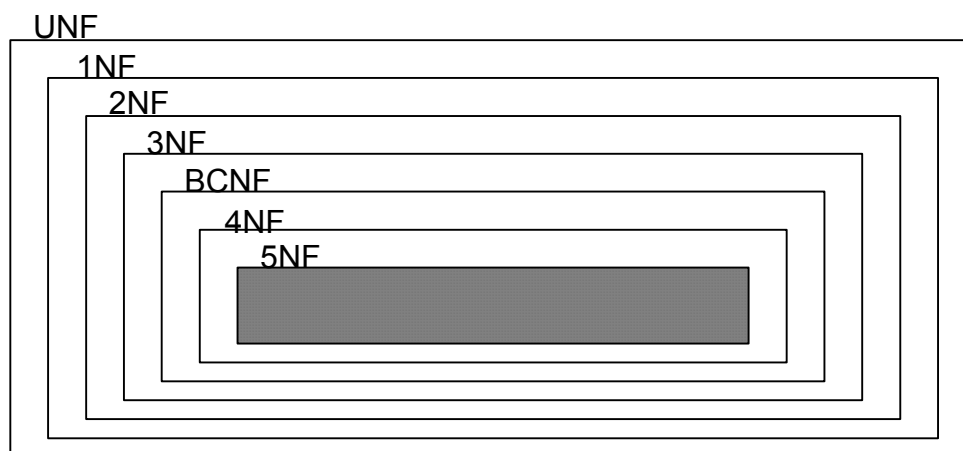
Funkčná závislosť	Elementárna	Úplná	Vzájomná	Tranzitívna
$RC \rightarrow \text{meno}$	Y	Y	N	N
$RC \rightarrow \text{priezvisko}$	Y	Y	N	N
$RC \rightarrow \text{PSC}$	Y	Y	N	Y
$RC \rightarrow \text{obec}$	Y	Y	N	Y
$\text{PSC} \rightarrow \text{obec}$	Y	Y	Y	N

---

Zap\_predmety(OC, CP, vysledok, nazov)

FZ	Elementárna	Úplná	Vzájomná	Tranzitívna
<b>(OC, CP)→vysledok</b>	N	Y	N	N
<b>CP→nazov</b>	Y	Y	N	N

## 9.2 Normalizácia



Relácie, ktoré nesplňujú podmienky príslušnej normálnej formy sú dekomponované na menšie relácie, ktoré musia spĺňať podmienky:

- bezstratového spojenia v prípade potreby tvorby odvodených relácií
- zachovania funkčných závislostí, čo znamená, že všetky funkčné závislosti v dátovom modeli budú zachované aj po dekompozícii.

### 9.2.1 Normalizácia podľa primárneho kľúča

#### Príklad 9.13 – Normalizácia podľa normálnych foriem

Nenormalizovaná relácia (UNR)

R(#OC, RC, Meno, Priezvisko, PSC, Obec, CP<sub>1</sub>, Nazov<sub>1</sub>, Vysl<sub>1</sub>, CP<sub>2</sub>, Nazov<sub>2</sub>, Vysl<sub>2</sub>,

...

CP<sub>n</sub>, Nazov<sub>n</sub>, Vysl<sub>n</sub>)

s predpokladom, že pripúšťame duplicitné rodné čísla.

#### Poznámka

OC je atribút reprezentujúci osobné číslo študenta,

RC reprezentuje rodné číslo študenta,

PSC reprezentuje poštové smerové číslo obce, v ktorej študent má trvalé bydlisko,

CP je číslo predmetu,

a Vysl je atribút pre uloženie výsledku hodnotenia študenta z daného predmetu.

#### 9.2.1.1 Prvá normálna forma (1NF)

##### Definícia - 1. normálna forma

Relácia R je v 1. normálnej forme vtedy a len vtedy, keď neobsahuje opakujúce sa podskupiny dát (t. j. keď každý atribút je atomický).



**Riešenie** - dekompozícia – relácia je rozdelená na dve relácie tak, že nová relácia bude obsahovať opakujúcu sa podskupinu dát (CP, Nazov, Vysl) a atribúty pôvodnej relácie, ktoré zabezpečia bezstratovosť spojenia (OC).

R1(#OC, #CP, Nazov, Vysl)

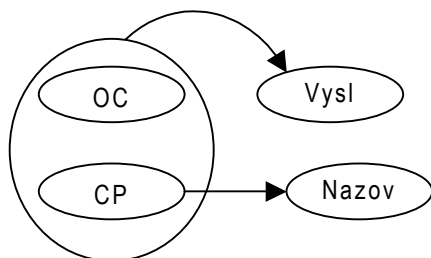
## 9.2.1.2 Druhá normálna forma (2NF)

**Definícia - 2. normálna forma**

Relácia R je v 2. normálnej forme vtedy a len vtedy, keď je v 1NF a každý neklúčový atribút je úplne funkčne závislý od primárneho kľúča.

R2 ... automaticky v 2NF

R1:



*Riešenie* - dekompozícia – relácia je rozdelená na dve relácie tak, že nová relácia obsahuje funkčnú závislosť, ktorá porušila pravidlo 2NF, pričom musí byť zachovaná podmienka spätnej rekonštrukcie.

R11(#CP, Nazov)  
R12(#OC, #CP, Vysl)

**Poznámka**

Z uvedeného je zrejmé, že každá relácia, ktorej PK je jednoatribútový spĺňa automaticky podmienku 2NF. Čiže pri analýze schém relácií má zmysel pri 2NF sa zaoberať len reláciami, ktoré majú zložený PK.

## 9.2.1.3 Tretia normálna forma (3NF)

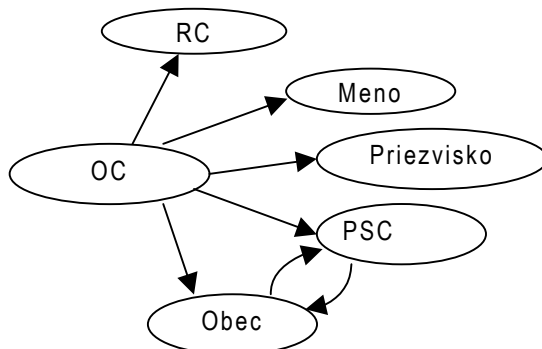
**Definícia - 3. normálna forma**

Relácia R je v 3. normálnej forme, ak je v 2NF a každý neklúčový atribút nie je tranzitívne funkčne závislý na primárnom kľúči.

Často sa stáva, že súčasťou tranzitívnej závislosti sú aj atribúty vzájomne funkčne závislé a v tom prípade je nutné navrhnúť primárny kľúč novej relácie s ohľadom na požadované kritériá (pamäťové miesto, požiadavky aplikácie, ...).

*Riešenie* - dekompozícia - relácia je rozdelená na dve relácie tak, že z tranzitívnej funkčnej závislosti odstránime jednu funkčnú závislosť tak, že vytvoríme novú reláciu, ktorá bude navyše obsahovať atribúty potrebné pre zabezpečenie bezstratového spojenia.

R2:



R21(#PSC, Obec)

R22(#OC, RC, Meno, Priezvisko, PSC)

**9.2.2 Boyce – Coddova normálna forma (BCNF)****Definícia - Determinant**

Determinant je množina atribútov, na ktorej je iná množina atribútov funkčne závislá.

*Poznámka*

$A \rightarrow B \Rightarrow A$  je determinant

**Definícia - BCNF**

Relácia R je v BCNF vtedy a len vtedy, keď každý determinant relácie je kandidátom primárneho kľúča.

**Algoritmus normalizácie BCNF:**

1. vytvor nenormalizovanú reláciu (UNR)
2. detekuj funkčné závislosti
3. test BCNF - je v BCNF  $\Rightarrow$  KONIEC  
     nie je v BCNF  $\Rightarrow$  dekompozícia
4. opakuj kroky 2,3 pre všetky relácie, ktoré vznikli

□ *Príklad 9.14 – Normalizácia podľa BCNF:*

**Nenormalizovaná (UNF)**

R(#OC, RC, Meno, Priezvisko, PSC, Obec, #CP, Nazov, Vysl)

KPK: OC, CP  
       RC, CP

**Determinant:** OC  
                   RC  
                   PSC  
                   Obec  
                   CP  
                   OC, CP  
                   RC, CP

1. R1(#CP, Nazov)  
    KPK: CP

**Determinant:** CP

R2(#OC, RC, #CP, Meno, Priezvisko, PSC, Obec, Vysl)

KPK: OC, CP  
       RC, CP

**Determinant:** OC  
                   RC  
                   PSC  
                   OC, CP  
                   RC, CP  
                   Obec

2. R1: R3(#OC, #CP, Vysl)  
    KPK: OC, CP

**Determinant:** OC, CP

R4(#OC, RC, Meno, Priezvisko, PSC, Obec)

KPK: OC  
       RC

**Determinant:** OC

RC

PSC

Obec

3. R4: R5(#PSC, Obec)

KPK: PSC  
ObecDeterminant: PSC  
Obec

R6(#OC, RC, Meno, Priezvisko, PSC)

KPK: OC  
RCDeterminant: OC  
RC

### 9.2.3 Anomálie medzi 3NF a BCNF

V praktických aplikáciách je väčšina relácií súčasne v 3NF a aj v BCNF. Ale napriek tomu existujú situácie, v ktorých sa stretávame s anomáliami, keď relácia je v 3NF, ale nie je v BCNF, alebo naopak. Najčastejšie sa s týmto typom anomálií stretávame v takých reláciách, ktoré obsahujú:

- viac kandidátov PK
- kandidát PK je kompozitný
- kompozitní kandidáti PK sa prekrývajú

Anomálie si budeme prezentovať na príkladoch.

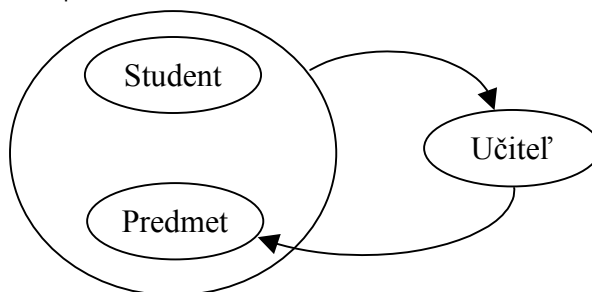
#### 9.2.3.1 Relácia je v 3NF a nie je v BCNF

□ *Príklad 9.15 – je v 3NF*

Majme reláciu VYUKA(#Student, #Predmet, #Ucitel) s nasledovnými výskytmi:

S1	P1	U1
S2	P1	U2
S2	P2	U3
S2	P3	U4
S3	P1	U1
S3	P2	U3
S4	P1	U5
S5	P1	U2

Funkčné závislosti v danom príklade môžeme identifikovať nasledovne:



Pre dekompozíciu sa nám ponúkajú 3 varianty:

1. R<sub>1</sub>(#Student, #Ucitel), R<sub>2</sub>(#Student, #Predmet)
2. R<sub>1</sub>(#Ucitel, Predmet), R<sub>2</sub>(#Student, #Predmet)
3. R<sub>1</sub>(#Ucitel, Predmet), R<sub>2</sub>(#Student, #Ucitel)

Správny výsledok zodpovedá variante číslo 3, pretože prvé dva varianty nám pri spojení generujú nadbytočné riadky. (Čitateľ si to môže samostatne overiť.)





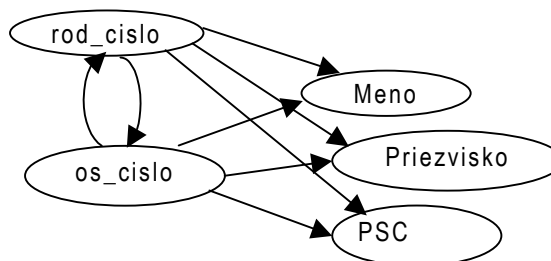
## 9.2.3.2 Relácia je v BCNF a nie je v 3NF

□ *Príklad 9.16 – je v BCNF*

Majme reláciu  $R(\#os\_cislo, rod\_cislo, Meno, Priezvisko, PSC)$ .

Kandidátmi v tejto relácii môžu byť atribúty  $os\_cislo$ , alebo  $rod\_cislo$ .

Funkčné závislosti v relácii  $R$  sú znázornené na nasledujúcom obrázku.



Z uvedeného príkladu vyplýva, že medzi atribútmi  $os\_cislo$  a  $rod\_cislo$  existuje vzájomná funkčná závislosť a taktiež existujú tranzitívne závislosti. Determinanty v tejto relácii sú atribúty  $os\_cislo$  a  $rod\_cislo$ , z čoho vyplýva, že je relácia v BCNF, pretože každý kandidát primárneho kľúča je zároveň determinantom relácie.

Avšak relácia nie je v 3NF, pretože v relácii existujú tranzitívne závislosti.

Riešením by mohla byť dekompozícia:

$R_1(\#os\_cislo, rod\_cislo)$

$R_2(\#os\_cislo, Meno, Priezvisko, PSC)$ ,

avšak väčšine riešiteľov by bolo zrejmé, že došlo k zbytočnej dekompozícii, ktorá navyše môže v aplikáciách vyvolať nadbytočnú operáciu spojenia.

## 9.2.4 Štvrtá normálna forma (4NF)

**Definícia - Viachodnotová závislosť**

Ak existuje relácia  $R$  s relačnou schémou  $R(A,B,C)$ , kde  $A, B$  a  $C$  sú množiny atribútov, potom hovoríme, že množina atribútov  $B$  je viachodnotovo závislá od množiny atribútov  $A$ , vtedy a len vtedy, keď existujú duplicitné kombinácie hodnôt atribútov  $A, B$ .

□ *Označenie:  $A \twoheadrightarrow B$* *Poznámka:*

Ak relácia  $R$  obsahuje len dva atribúty  $A$  a  $B$ , potom hovoríme o triviálnej závislosti.

□ *Príklad 8.17*

$R(OC, Zručnosť, Šport)$

OC	Zručnosť	Šport
1	Programovanie	Futbal
1	Programovanie	Hokej
1	Analýza	Futbal
1	Analýza	Hokej
2	Analýza	Futbal
2	Analýza	Atletika
2	Manažment	Futbal
2	Manažment	Atletika

V uvedenom príklade vidíme, že hoci je relácia v BCNF nevyhneme sa anomáliám pri deštruktívnych operáciách. V prípade, že by sme chceli opraviť názov zručnosti alebo športu, musíme túto opravu vykonať vo viacerých riadkoch relácie. Podobne keby sme chceli pridať novú zručnosť, tak ju musíme pridať do viacerých riadkov relácie – podľa počtu športov daného študenta.

#### Definícia - 4. normálna forma

Relácia **R** s relačnou schémou **R(A,B,C)** je v 4. normálnej forme, ak v každej multihodnotovej závislosti **A → B**, platí jedno z nasledujúcich pravidiel:

- je atribút **A** je superkľúčom
- atribút **B**  $\subseteq$  **A**, alebo **R = (A,B)**



Odstránenie vyššie uvedených anomálií relácie s viachodnotovými závislosťami je možné zabezpečiť dekompozíciou na dve relácie, kde každá z nich bude mať rovnaký PK.

#### □ Príklad 9.18

**R(#OC, #Zručnosť, #Šport)**

Relácia je v BCNF, pretože medzi atribútmi neexistujú funkčné závislosti, ale relácia nie je dobre navrhnutá, pretože keby sme chceli pridať nový šport niektorému z pracovníkov, je nutné priradiť ho vo väzbe na nejakú zručnosť.

Problém je možné riešiť identifikáciou viachodnotových závislostí **OC → Zručnosť** a **OC → Šport** a následnou dekompozíciou na dve relácie, ktorou odstránime zistenú anomáliu.

**R<sub>1</sub>(#OC, #Zručnosť)**

**R<sub>2</sub>(#OC, #Šport)**

#### 9.2.5 Piata normálna forma (5NF)

Napriek tomu, že sme odstránili viachodnotovú funkčnú závislosť a s ňou súvisiace anomálie, nevyhneme sa ďalším problémom. Pri denormalizácii sa môže stať, že výsledkom spojenia dekomponovaných relácií získame nadbytočné riadky relácie.

Uvedený problém bude zrejmý z nasledujúceho príkladu:

#### □ Príklad 9.19 [Date96]

Majme reláciu **SPJ(S,P,J)** a nasledovným výskytom n-tíc

SPJ	S#	P#	J#
	S1	P1	J2
	S1	P2	J1
	S2	P1	J1
	S1	P1	J1

Ak uvažujeme viachodnotovú závislosť a následnú dekompozíciu do 4NF, dostaneme relácie **SP** a **SJ**:

SP	S#	P#
	S1	P1
	S1	P2
	S2	P1

SJ	S#	J#
	S1	J2
	S1	J1
	S2	J1

Po ich spojení v snahe získať pôvodnú reláciu získame reláciu SPJ<sub>2</sub> s nadbytočným riadkom:

SPJ <sub>2</sub>	S#	P#	J#
	S1	P1	J2
	S1	P1	J1
	S1	P2	J1
nadbytočný →	S1	P2	J2
	S2	P1	J1

### Definícia – Spojovacia závislosť (Join dependency)

Majme reláciu **R** a nech **A, B, Z** sú ľubovoľné podmnožiny atribútov relácie **R**, potom hovoríme, že relácia spĺňa spojovaciu závislosť

$$*(A, B, \dots Z),$$

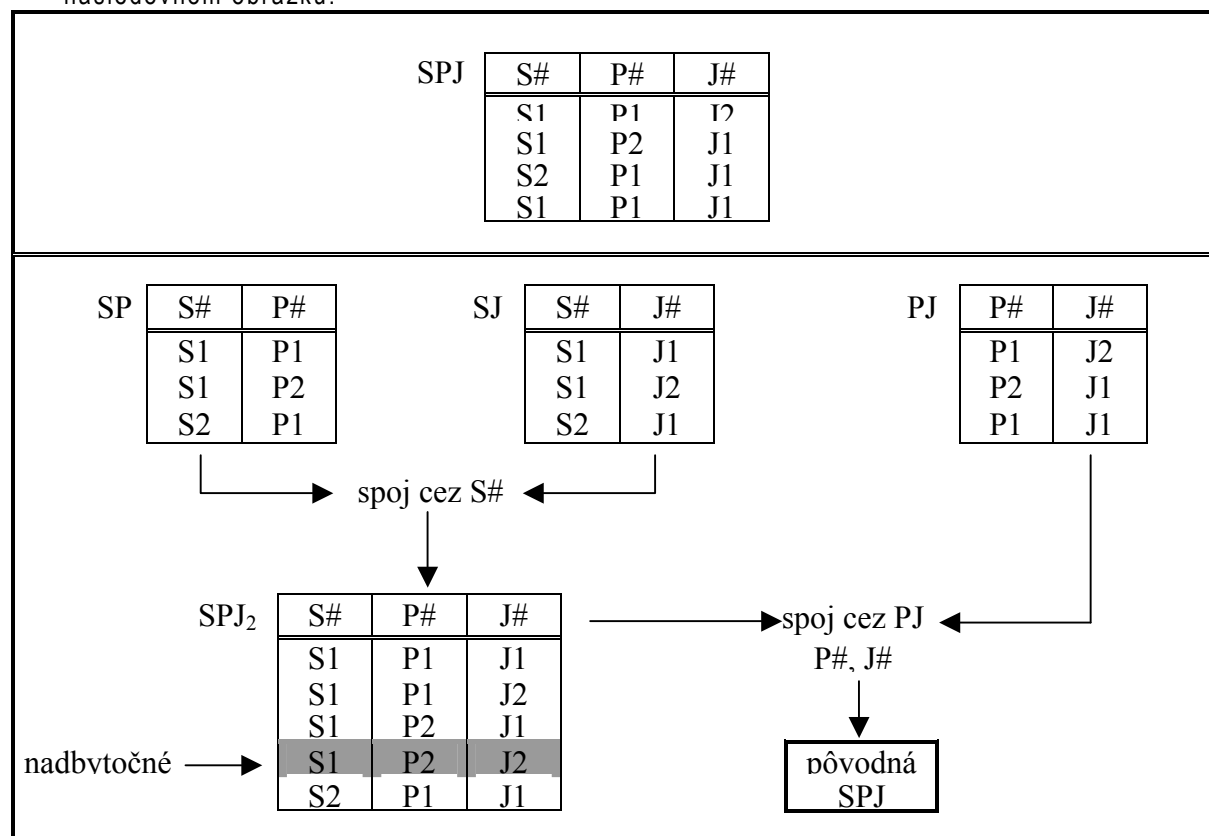
vtedy a len vtedy, ak pôvodnú reláciu získame spojením projekcií nad atribútmi **A, B, Z**.

### Definícia – Piata normálna forma (5NF)

Relácia **R** je v 5NF vtedy a len vtedy, ak každá spojovacia závislosť v relácii **R** zahŕňa kandidáta primárneho kľúča relácie **R**.

#### □ Příklad 9.20

Celkovo si proces skvalitňovania dátového modelu pomocou 5NF môžeme prezentovať na nasledovnom obrázku:



□ *Príklad 9.21*

Majme reláciu Student(#os\_cislo, rod\_cislo, meno, priezvisko, adresa). V tejto relácii máme nasledovných KPK: os\_cislo, rod\_cislo. Potom táto relácia spĺňa spojovaciu závislosť:

$((\#os\_cislo, rod\_cislo, meno, priezvisko), (\#os\_cislo, adresa))$

Pôvodnú reláciu študent získame spojením jej projekcií

$(\#os\_cislo, rod\_cislo, meno, priezvisko)$

a  $(\#os\_cislo, adresa)$

Uvedené relácie sú v 5NF, pretože kandidáti primárneho kľúča sa vyskytujú vo všetkých spojovacích závislostiach relácie R.

□ *Príklad 9.22*

Majme reláciu Student(#os\_cislo, rod\_cislo, meno, priezvisko, adresa). V tejto relácii máme nasledovných KPK: os\_cislo, rod\_cislo. Potom táto relácia spĺňa spojovaciu závislosť:

$((\#os\_cislo, meno, priezvisko), (\#os\_cislo, rod\_cislo), (\#rod\_cislo, adresa))$

Pôvodnú reláciu študent získame spojením jej projekcií

$(\#os\_cislo, meno, priezvisko),$

$(\#os\_cislo, rod\_cislo)$

a  $(\#rod\_cislo, adresa)$

Uvedené relácie sú v 5NF, pretože kandidáti primárneho kľúča sa vyskytujú vo všetkých spojovacích závislostiach relácie R.