

# Normálne formy na základe funkčných a spojovacích závislostí

- a) 2NF, 3NF, HBCNF
- b) Spojovacia závislosť
- c) ETNF

**Normalizácia** je proces, ktorý vedie k návrhu, v ktorom sa minimalizuje **redundancia**, **nadbytočnosť** v databázovej schéme. Nedávno bola definovaná [1] nová normálna forma **ETNF** - *essential tuple normal form*, v ktorej každá **n-tica** je **nevyhnutná** (essential tuple). Vďaka nej, na zabezpečenie n-tíc s minimálnou redundanciou nie je potrebná 4NF a 5NF: ETNF je slabšia ako 5NF ale v eliminácii nadbytočných n-tíc je rovnako efektívna.

Najprv na základe funkčných závislostí definujeme druhú, tretiu a Heath-Boyce-Codd - ovu normálnu formu, ktoré ilustrujeme príkladmi. Potom po definícii **spojovacej závislosti** sa pozrieme podrobnejšie na vlastnosti HBCNF a na záver zdefinujeme **tri typy n-tíc** a **ETNF**.

## a) 2NF, 3NF, HBCNF

Uvažujme relačnú schému  $R \equiv R(\mathcal{A}, \mathcal{F})$ ,  $\mathcal{A} = \{A_1, \dots, A_n\}$ . Vieme, že  $A_i$  je kľúčový atribút, ak je prvkom (kompozitného) kandidátneho kľúča. Zovšeobecnením KK je nadkľúč.

Pripomíname, že množina atribútov  $X$  je **podkľúč** (PK) / **nadkľúč** (NK), ak  $X$  je podmnožinou / nadmnožinou **nejakého, aspoň jedného** kandidátneho kľúča,  $X \subseteq KK_i / KK_i \subseteq X$ .

Poznamenáme, že

- každý kľúč  $K$  je NK, ale naopak to neplatí (ak  $X = \{W, K\}$ , kde  $W \notin \mathcal{A}$ , potom  $X$  je NK, ale  $X$  nie je  $K$  - prečo nie je?);
- $\mathcal{A}$  je NK;
- nulová množina je PK pre ľubovoľnú relačnú schému  $R$ .

### Definícia 1

Uvažujme relačnú schému  $R(\mathcal{A}, \mathcal{F})$ . Ak pre **ľubovoľnú** (aj implicitnú) **netriviálnu** funkčnú závislosť

$$X \rightarrow Y,$$

vyplývajúcu z  $\mathcal{F}$ , kde  $X \subset \mathcal{A}$ ,  $Y \subset \mathcal{A}$ , platí

<p><b>jedna</b> vlastnosť z troch</p> <ul style="list-style-type: none"> <li>- <math>X</math> je nadkľúč - NK,</li> <li>- <math>Y</math> je podkľúč - <math>Y \subseteq KK_i</math>,</li> <li>- <math>X</math> nie je podkľúč - <math>X \not\subseteq KK_j</math>,</li> </ul> <p>potom a len potom je <math>R</math> v <b>2NF</b>,</p>	<p><b>jedna</b> vlastnosť z dvoch</p> <ul style="list-style-type: none"> <li>- <math>X</math> je nadkľúč,</li> <li>- <math>Y</math> je podkľúč - <math>Y \subseteq KK_i</math>,</li> </ul> <p>potom a len potom je <math>R</math> v <b>3NF</b>,</p>	<p><b>Že</b></p> <ul style="list-style-type: none"> <li>- <math>X</math> je nadkľúč,</li> </ul> <p>potom a len potom je <math>R</math> v <b>HBCNF</b>.</p>
--	---	--

Môže sa v danej definícii X rovnať  $\mathcal{A}$ , teda  $X \subseteq \mathcal{A}$  ?

Poznamenáme, že z daných normálnych foriem pochopiteľne 2NF je najslabšia, lebo pre jej splnenie stačí platnosť jednej vlastnosti z troch.

**a) Príklady.**

Uvažujme štyri príklady s tromi atribútmi (v DB teórii AB je množina zjednotení atribútov A, B)

- 1)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow B\}$ ;
- 2)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow B, B \rightarrow C\}$ ;
- 3)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{AC \rightarrow B, B \rightarrow C\}$ ;
- 4)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow BC\}$ .

**Dokážme, že**

- prvý príklad nie je v 2NF;
- druhý príklad je v 2NF, ale nie je v 3NF;
- tretí príklad je v 3NF, ale nie je v HBCNF;
- štvrtý príklad je v HBCNF.

jedna vlastnosť z troch - X je nadkľúč, - Y je podkľúč $- Y \subseteq KK_i$ , - X nie je podkľúč $- X \not\subseteq KK_j$ , potom R je v <b>2NF</b> .	jedna vlastnosť z dvoch - X je nadkľúč, - Y je podkľúč $- Y \subseteq KK_i$ , potom R je v <b>3NF</b> .	že - X je nadkľúč, potom R je v <b>HBCNF</b> .
---	--	--

**Dôkaz:**

**OK** - pre všetky FZi platí aspoň jedna vlastnosť; **NO** - aspoň pre jednu FZ neplatí ani jedna vl.

- 1)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow B\}$ ;  
 - 2NF **NO**: vo FZi  $A \rightarrow B$  a) atr. A **nie** je NK; b) je B PK?  $\{AC\}^+ = \mathcal{A}$ , teda AC je KK, ale **neplatí**  $B \subseteq AC$ ; c) **ani**  $A \not\subseteq AC$  (pretože  $A \subseteq AC$ , teda A je podkľúč, neplatí  $A \not\subseteq AC$ );
- 2)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow B, B \rightarrow C\}$ ;  
 - 2NF **OK**:  $\{A\}^+ = \mathcal{A}$  a pre  $A \rightarrow B$  **platí** A je NK a pre  $B \rightarrow C$  **platí**  $B \not\subseteq A$ ; 3NF **NO**: pre  $B \rightarrow C$  atr. B nie je NK **ani**  $C \subseteq A$ ;
- 3)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{AC \rightarrow B, B \rightarrow C\}$ ;  
 - 3NF **OK**:  $\{AC\}^+ = \mathcal{A}$  a pre  $AC \rightarrow B$ : AC je NK a pre  $B \rightarrow C$ :  $C \subseteq AC$ ; HBCNF **NO**: B nie je NK;
- 4)  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C\}$  a  $\mathcal{F} = \{A \rightarrow BC\}$ .  
 - HBCNF **OK**:  $\{A\}^+ = \mathcal{A}$ .

**b) Príklad.**

Nech  $R \equiv R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{A} = \{A, B, C, D\}$ ,  $\mathcal{F} = \{ACD \rightarrow B, AC \rightarrow D, AC \rightarrow B, D \rightarrow C\}$ .

2a) Skontrolujte, že  $\{AC\}^+ = \mathcal{A}$  a preto AC je KK. Existuje ďalší dvojatribútový KK?

2b) Bude R v HBCNF?

2c) Bude R v 3NF? Na cvičení.

Zdôrazňujeme, že HBCNF musí platiť pre ľubovoľnú explicitnú alebo implicitnú FZ  $X \rightarrow Y$  v  $R$  a  $\mathcal{F}$  štandardne obsahuje explicitné FZi. Nižšie uvidíme, že závislosti je možné zaviesť aj cez iné obmedzenia, konštrukcie, ako napr. SZi  $\mathcal{J}$ , z ktorých potom sa dajú vydedukovať implicitné FZi. Preto formuláciu “FZ vyplývajúca z  $\mathcal{F}$ ” treba rozšíriť na “FZ vyplývajúca z  $\mathcal{F}$  a  $\mathcal{J}$ ”.

## b) Spojovacia závislosť

V databázovej teórii okrem

- **funkčnej závislosti FZ** sa rozlišuje aj
- viachodnotová závislosť (Multy valued dependency MVD) a
- **spojovacia závislosť SZ** (Join dependency JD).

Definícia 4NF a 5NF sa zakladajú práve na MVD a SZ [1]. Nižšie po definícii SZi uvedieme iba definíciu HBCNF a ETNF, z ktorých posledná sa explicitne neopiera o SZ, definíciu 4NM a 5NM nie a ukážeme, ako sú HBCNF a ETNF prepojené.

Tu iba poznamenáme, že z ETNF vyplýva 4NF. Preto definíciu MVD, potrebnú pre 4NF, neuvádzame (navyše MVD je možné vyjadriť aj pomocou SZi).

**Relačná schéma**  $R$  je trojica, pozostávajúca z konečnej množiny atribútov a množiny obmedzení, ktoré špecifikujú  $R$  pomocou množiny FZi  $\mathcal{F}$  a SZi  $\mathcal{J}$

$$R(\mathcal{A}, \mathcal{F}, \mathcal{J}).$$

**Relácia**  $r$  je inštanciou relačnej schémy  $R$  ak

- $\mathcal{A}$  je množina atribútov  $r$  a
- $r$  vyhovuje obmedzeniam, zadaných pomocou závislostí  $\mathcal{F}$  a  $\mathcal{J}$ .

Je dobre známe, že množina FZi môže mať za následok (imply) inú FZ. To isté platí aj pre SZi. Hovoríme, že **implicitné** FZi a SZi **logicky** vyplývajú z **explicitne** zadaných FZi a SZi.

**Príklad 1:** Uvažujme  $R = R(\mathcal{A}, \mathcal{F})$ , kde  $\mathcal{F} = \{X \rightarrow Y, Y \rightarrow Z\}$ . Potom  $X \rightarrow Y$ ,  $Y \rightarrow Z$  sú explicitné FZi a  $X \rightarrow Z$  je implicitná FZ.

**Definícia 2.** Nech  $S_1 \cup \dots \cup S_k = \mathcal{A}$ . Hovoríme, že relácia  $r$  ako inštancia  $R$  *splňa spojovaciú závislosť* (SZ)

$$\bowtie\{S_1, \dots, S_k\},$$

ak vždy, keď  $t_1, \dots, t_k$  sú ľubovoľné  $n$ -tice relácie  $r$ , a existuje  $n$ -tica  $t$  (nie nutne z  $r$ ) taká, že

$$t[S_i] = t_i[S_i], \quad i=1, 2, \dots, k$$

pre každý komponent  $S_i$ , potom  $t$  je  $n$ -tica  $r$ .

Poznamenáme, že namiesto  $\bowtie\{S_1, \dots, S_k\}$  sa používa aj označenie  $\ast\{S_1, \dots, S_k\}$ .

Názov ( $k$ -komponentná) *spojovacia závislosť* vyjadruje skutočnosť, že

relácia  $r$  vyhovuje, podlieha **SZi**  $\bowtie\{S_1, \dots, S_k\}$  vtedy a len vtedy, keď  $r$  je **bezstratové spojenie** (lossless join) **jej projekcií**  $r[S_1], \dots, r[S_k]$ ,

$$r \text{ vyhovuje SZi } \bowtie\{S_1, \dots, S_k\} \Leftrightarrow r = \bowtie_{i=1}^k r[S_i]$$

kde  $r = \bowtie_{i=1}^k r[S_i]$  znamená, že  $r$  možno znovu vytvoriť spojením viacerých (menších) relácií,

z ktorých každá má podmnožinu atribútov  $r$ , teda spojením projekcií  $r[S_1], \dots, r[S_k]$ .

Poznamenáme, že medzi komponentami SZi musí existovať aspoň jeden spoločný atribút.

**SZ** sa vyjadruje aj slovami:  **$r$  sa rozkladá bezstratovo na jej projekcie** (decomposes losslessly, nonloss decomposition).

**Triviálna SZ** (nad  $\mathcal{A}$ ) je taká závislosť, ktorá platí pre každú reláciu s atribútmi  $\mathcal{A}$ . Ľahko nahliadneme, že SZ je triviálna vtedy a len vtedy, keď niektorý jej komponent je množinou všetkých atribútov.

**Príklad 2:** Uvažujme  $R\{\mathcal{A}, \mathcal{F}\}$  s

$$\mathcal{A} = \{D, S, P\}, \quad \mathcal{F} = \bowtie\{DS, SP, PD\},$$

kde D-dodávateľ, S-súčiastka, P-projekt.

Spojovacia závislosť  $\mathcal{F}$  znamená:

- ak dodávateľ  $d$  dodáva súčiastku  $s$  – DS (pre niektorý projekt  $p$ ),
- a súčiastka  $s$  je dodávaná do projektu  $p$  – SP (niektorým dodávateľom  $d$ ),
- a projekt  $p$  je zásobovaný dodávateľom  $d$  – PD (niektorou súčiastkou  $s$ ),

potom dodávateľ  $d$  zásobuje súčiastkou  $s$  projekt  $p$

$\Leftrightarrow$  dodávateľ  $d$  dodáva súčiastku  $s$  do projektu  $p$ .

**Príklad 3.** Uvažujme relačnú schému  $R(\mathcal{A}, \mathcal{F}, \mathcal{J})$ , kde

$$\mathcal{A} = \{A, B, C\}, \quad \mathcal{F} = \{A \rightarrow B\} \text{ a } \mathcal{J} = \bowtie\{AB, AC\}.$$

Dokážme, že  $A \rightarrow C$  je (implicitná) FZ.

Nech  $r$  je inštancia  $R$ . Predpokladajme, že dve  $n$ -tice

$$\begin{cases} t_1 = [a, b_1, c_1], \\ t_2 = [a, b_2, c_2], \end{cases}$$

sú z  $r$ . Nech  $AB \rightarrow C$  je FZ. Pochopiteľne iba z nej nevyplýva  $A \rightarrow C$ , ako to dokazujú dané dve  $n$ -tice. Uvažujme tretiu  $n$ -ticu

$$t = [a, b_1, c_2],$$

ktorá samozrejme nemusí byť z  $r$  (prečo?). Ale zo SZ-ti  $\bowtie\{AB, AC\}$  dostaneme, ako to čoskoro ukážeme, že  $t$  musí byť z  $r$ , čo sa kvôli  $AB \rightarrow C$  môže stať iba ak  $c_1 = c_2$  a preto  $A \rightarrow C$ .

Teda ostalo nám ukázať, že  $t \in r$ . Ak  $\bowtie\{AB, AC\}$  je SZ, kde  $AB \cup AC = \{A, B, C\} = \mathcal{A}$ , potom z definície SZi, z dvoch rovností

$$t[a, b_1] = t_1[a, b_1] \text{ a}$$

$$t[a, c_2] = t_2[a, c_2]$$

obdržíme, že  $t \in r$ .

**Veta 1 (Heath)**. Nech  $X$  a  $Y$  sú podmnožiny  $\mathcal{A}$  a nech  $Z = \mathcal{A} \setminus (X \cup Y)$ . Potom nasledujúce tvrdenia sú ekvivalentné

- $X \rightarrow Y$ ,
- platí SZ  $\bowtie\{XY, XZ\}$  a zároveň  $XZ$  je nadkľúč, tzn.  $XZ \rightarrow Y$ .

Poznámka:

- pôvodne Heath dokázal iba to, že z FZi  $X \rightarrow Y$  vyplýva SZ  $\bowtie\{XY, XZ\}$ ;
- ako sme to vyššie poznamenali, SZ  $\bowtie\{XY, XZ\}$  (s  $XY \cup XZ = \mathcal{A}$ ) znamená bezstratové spojenie

$$r[XY] \bowtie r[XZ] = r.$$

**Definícia 3.** Relačná schéma  $R$  je v Heath-Boyce-Codd normálnej forme (**HBCNF**), ak každá explicitná alebo implicitná FZ v  $R$  logicky vyplýva z kľúčov  $R$ .

**Veta 2.** Relačná schéma  $R$  je v HBCNF vtedy a len vtedy, ak pre každú explicitnú alebo implicitnú netriviálnu FZ  $X \rightarrow Y$  v  $R$  determinant  $X$  je nutne nadkľúč.

Pozri definíciu 1, teda: Def.1  $\Leftrightarrow$  Def.3

**Príklad 4** (pokračovanie). Ukážme, že v príklade 3  $R$  nie je v HBCNF.

Z poslednej vety (či z Def.1) vidíme, že k určeniu, či  $R$  je v HBCNF, musíme uvažovať aj implicitné FZi. V príklade 3 sme ukázali, že  $A \rightarrow C$  je netriviálna implicitná FZ. Otázka znie, bude  $A$  (nad)kľúč? Nebude, lebo  $A \rightarrow B$  neplatí. Je  $R$  v 3NF?.

Poznamenáme, že v tomto príklade implicitná FZ  $A \rightarrow C$  zohráva dôležitú úlohu. Keby sme nekorektne uvažovali iba explicitné FZi, teda iba  $AB \rightarrow C$ , potom by sme došli nesprávnemu záveru, že  $R$  je v HBCNF.

**Príklad 5.** Uvažujme relačnú schému  $R(\mathcal{A}, \mathcal{F}, \mathcal{J})$ , kde

$$\mathcal{A} = \{A, B, C, D\}, \mathcal{F} = \{A \rightarrow BCD, BC \rightarrow AD\} \text{ a } \mathcal{J} = \{\bowtie\{ABC, CD, BD\}\}.$$

Ukážeme pomocou poslednej vety, že  $R$  je v HBCNF. Nech  $X \rightarrow Y$  je ľubovoľná netriviálna FZ, kde  $X$  je ľubovoľná podmnožina  $\mathcal{A}$ . Máme ukázať, že determinant  $X$  je nadkľúč. Pretože determinanty  $A$  a  $BC$  dvoch FZí  $\mathcal{F}$  sú (nad)kľúče, stačí ukázať, že  **$X$  obsahuje  $A$  alebo  $BC$**  (prečo?). Stačí rozlišovať **štyri** prípady:  $X$  obsahuje **jeden, dva** alebo **tri** atribúty (4 prečo nie?), alebo  $X$  je prázdna množina. Ukážeme iba prípad, keď  $X$  má **tri** atribúty. V tomto prípade ľahko nahliadneme (!), že  $X$  buď obsahuje  $A$ , alebo je  $BCD$ . Potom ale  $X$  obsahuje buď  $A$  alebo  $BC$ , ktoré sú kľúče (v ďalších troch prípadoch dôjdeme k rovnakému záveru). Preto  $X$  je nadkľúč.

**Príklad 6.** Dokážte pr. 5 pre prípad, keď  $X$  má dva atribúty. **Na cvičení.**

### c) ETNF

Nielen **FZ** môže logicky vyplývať z iných závislostí, ale aj **n-tica** z iných n-tíc.

**Definícia.** Nech  $R$  je relačná schéma,  $S$  je množina n-tíc a  $t$  je n-tica. Hovoríme, že  **$t$  logicky vyplýva z  $S$**  (vzhľadom na  $R$ ), ak každá inštancia  $R$ , ktorá obsahuje všetky n-tice  $S$ , obsahuje aj  $t$ .

Teraz uvedieme tri typy n-tíc z hľadiska nadbytočnosti, z ktorých sa v definícii ETNF používa posledný.

**Definícia.** Nech je  $R(\mathcal{A})$  relačná schéma,  $r$  relácia ako inštancia  $R$  a  $t$  n-tica z  $r$ . Potom

1. n-tica  $t$  z  $r$  je **čiasťočne redundantná/nadbytočná** v  $r$  (vzhľadom na  $R$ ) ak
  - a) existuje n-tica  $t^*$  z  $r$  taká, že  $t \neq t^*$ ,
  - b) existuje  $X \subseteq \mathcal{A}$ , že  $t[X] = t^*[X]$  a zároveň
  - c) existuje netriviálna FZ  $X \rightarrow A$ , kde  $A \in \mathcal{A}$  a  $A \notin X$ .

(Dôvod, prečo sa hovorí, že n-tica  $t$  je **čiasťočne** redundantná, je nasledujúci. Pretože  $t[X] = t^*[X]$  a v relácii  $r$  platí FZ  $X \rightarrow A$ , dostaneme  $t[A] = t^*[A]$ . Intuitívne teda informácia o tom, že ktoré hodnoty  $A$  sú spojené s ktorými hodnotami  $X$ , je daná aj s  $t$  a  $t^*$  a nielen s FZ  $X \rightarrow A$ .)

**Veta 3.** Nech  $R$  je ľubovoľná schéma. Potom  $R$  je v HBCNF vtedy a len vtedy, ak žiadna inštancia  $R$  nemá čiasťočne redundantnú n-ticu.

2. n-tica  $t$  z  $r$  je **plne redundantná** v  $r$  (vzhľadom na  $R$ ), ak existuje množina  $S$  n-tíc z  $r$  taká, že
  - $t \notin S$  a
  - **$t$  logicky vyplýva z  $S$**  vzhľadom na  $R$ .

(**Plne** odkazuje na skutočnosť, že **celá** n-tica je **nadbytočná**.)

Intuitívne n-tica  $t$  je plne nadbytočná, ak jej prítomnosť v  $r$  logicky vyplýva z ďalších existujúcich n-tíc z  $r$ .)

Poznamenáme, že kým čiastočná redundancia súvisí s HBCNF, plná redundancia súvisí s anomáliami vymazávania, pozri [1].

3. n-tica je **nevyhnutná** (essential), ak **nie je** ani čiastočne ani plne redundantná.

**Príklad 7.** Uvažujme príklad 2, teda schému  $R\{\mathcal{A}, \mathcal{F}\}$  s

$$\mathcal{A} = \{D, S, P\}, \quad \mathcal{F} = \bowtie\{DS, SP, PD\}$$

a inštanciu  $r$  relačnej schémy  $R$ . Ukážme, že n-tica  $(d, s, p)$  z  $r$  je **plne redundantná**, teda informácia o  $(d, s, p)$  je v  $r$  dvakrát.

Skutočne. Nech relácia  $r$  je inštanciou triedy  $R$  (takže  $r$  má atribúty  $D, S$  a  $P$  a  $r$  spĺňa SZ  $\mathcal{F}$ ). Predpokladajme, že  $r$  má n-tice  $(d, s, p')$ ,  $(d', s, p)$  a  $(d, s', p)$  a možno aj iné n-tice a že  $d \neq d'$ ,  $s \neq s'$  a  $p \neq p'$ . Zo SZi  $\mathcal{F} = \bowtie\{DS, SP, PD\}$  vyplýva  $(t[DS]=t_1[DS], \dots)$  a že  $r$ , ako inštancia  $R$ , musí **obsahovať aj n-ticu**  $(d, s, p)$ , ktorá nie je prvkom  $S = \{(d, s, p'), (d', s, p), (d, s', p)\}$ . Preto  $(d, s, p)$ , v súlade s definíciou, je plne redundantná. Takže v relácii  $r$  informácia o n-tici  $(d, s, p)$  je zastúpená dvakrát:

- po prvé, **explicitne**, pomocou samotnej n-tice  $(d, s, p)$  z  $r$  a
- po druhé, **implicitne**, pomocou n-tíc  $(d, s, p')$ ,  $(d', s, p)$  a  $(d, s', p)$  a spojovacej závislosti  $\mathcal{F}$ .

**Príklad 8 (rozšírenie príkladu 2 o  $\mathcal{F}$ ):** Uvažujte  $R\{\mathcal{A}, \mathcal{F}, \mathcal{J}\}$  s

$$\mathcal{A} = \{D, S, P\}, \quad \mathcal{F} = \{DS \rightarrow P\} \text{ a } \mathcal{J} = \bowtie\{DS, SP, PD\},$$

kde  $D$ -dodávateľ,  $S$ -súčiastka,  $P$ -projekt. Ukážte, že

- a)  $R$  je v HBCNF,
- b) žiadna inštancia  $R$  nemá **čiasťočne redundantnú** n-ticu.

Na cvičení: a) vyplýva z vety 3, b) je to na poslednej stránke článku.

**Definícia.** Relačná schéma  $R$  je v normálnej forme nevyhnutných n-tíc (essential tuple normal form - **ETNF**), ak každá n-tica ľubovoľnej inštancie  $R$  je nevyhnutná.

Z nasledujúcich dvoch viet, ktoré spájajú HBCNF a ETNF, sa prvá opiera o komponent SZi a druhá o kľúč.

**Veta 4**  $\Leftrightarrow$ . Nech relačná schéma  $R$  je špecifikovaná iba s FZ-mi a SZ-mi. Potom  $R$  je v ETNF vtedy a len vtedy, ak je v HBCNF a niektorý **komponent** každej explicitnej SZ  $R$  je **nadkľúč**.

**Veta 5**  $\Rightarrow$ . Nech relačná schéma  $R$  je špecifikovaná iba s FZ-mi a SZ-mi a je v HBCNF. Ak niektorý **kľúč** sa skladá iba z **jedného** atribútu, potom  $R$  je v ETNF.

**Príklad 9** (pokračovanie príkladu 5).

$$R = R(\mathcal{A}, \mathcal{F}, \mathcal{J}): \mathcal{A} = \{A, B, C, D\}, \quad \mathcal{F} = \{A \rightarrow BCD, BC \rightarrow AD\} \text{ a } \mathcal{J} = \{\bowtie\{ABC, CD, BD\}\}.$$

Ukážme, že R je v ETNF.

Pretože na základe príkladu 5, R je v HBCNF, vďaka poslednej vety 5, R bude aj v ETNF, lebo z dvoch nadkľúčov A a BC, prvý sa skladá iba z jedného atribútu.

**Príklad 10** Uvažujme pr. 8. Ak vieme dodatočne, že žiadna inštancia R nemá plne redundantnú n-ticu, ukážte, že R je v ETNF. Na cvičení. Vyplýva to z pr. 8 (čiastocna redundantnosť), z dodatocnej znalosti tohto príkladu, plna redundantnosť, a z definície ETNF.

Koľko viet obsahuje tvrdenie o HBCNF a koľko o ETNF?

[1] Darwen, Date, Fagin, A normal form for preventing redundant tuples in relational databases, 2012 \_\_\_\_\_

[2] Chriss Date, Database design and relational theory (1st and 2nd edition), O'Reilly Media, 2012, Apress 2019

---

Chriss Date \_\_\_\_\_

### A Relational Model - set theory and logic.

"Learn the relational model first, SQL second."

- 1a. Learn the relational model, by reading the right books and/or attending the right courses.
2. Go out and get your hands dirty working on a real project for a year or three.
- 1b. Come back and read those books and/or attend those courses again.

Finally, read either or both of Ted Codd's first two papers every year.

- "Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks", IBM Research Report RJ599, August 19th, 1969, reprinted in ACM SIGMOD Record 38, No. 1 (March 2009) \_\_\_\_\_

- "A Relational Model of Data for Large Shared Data Banks", CACM 13, No. 6, June 1970, \_\_\_\_\_

Poznámka. Zo SZi  $\bowtie$  môže vyplývať aj FZ aj redundantnosť n-tice

SZ  $\bowtie$  sa požíva v príkladoch 3. a 7. na logický dôkaz, že

- a) teoretická n-tica  $t = [a, b1, c2]$
- b) konkrétna n-tica  $t = [d, s, p]$  z r

musí byť v r.

Preto ...

- a)  $c1 = c2$  a  $A \rightarrow C$
- b) t je plne redundantná.