

**Príklad 1** (8 b)

Daná je relačná databázová schéma:

Čapujú(Krčma, Pivo)  
 Ľúbi(Pijan, Pivo)  
 Navštívil(Idn, Pijan, Krčma, od, do)  
 Vypil(Idn, Pivo, Množstvo),

kde *Idn* je identifikátor jednotlivých návštev. Slúži na spájanie tabuliek *Vypil* a *Navštívil*. *Od* a *do* sú typu date/time.

Napište v datalógu, kalkule a algebre na:

1. Pijanov, ktorí navštívili menej ako 5 krčiem.
2. Pijana, ktorý navštívil všetky krčmy, kde čapujú aspoň jednu značku piva, ktorú ľúbi.

**Riešenia 1.1:**

Sú možné dva spôsoby riešenia: Sú to pijani, ktorí nenavštívili žiadnu krčmu  $\{Pijan: \exists(P) \text{Ľúbi}(Pijan, P) \wedge \neg \exists(I, K, o, d) \text{Navštívil}(I, Pijan, K, o, d)\}$ , alebo  $\cup$  pijani, ktorí navštívili menej ako 5 krčiem

$$\{Pijan: \cup(k=\text{count}(Krčma))(\exists(I,o,d)\text{Navštívil}(I, Pijan, Krčma, o, d)) \wedge k < 5 \}$$

Datalóg:

NavštívilKrčmu(Pijan, K)  $\leftarrow$  Navštívil(I, Pijan, K, o, d).  
 Neskúsení(Pijan)  $\leftarrow$  Ľúbi(Pijan, P),  $\neg$  NavštívilKrčmu(Pijan, K);  
 $\leftarrow$  subtotal(NavštívilKrčmu;Pijan;k=count(K)),  $k < 5$ .

Kalkul:

$$\{Pijan: \exists(P) \text{Ľúbi}(Pijan, P) \wedge \neg \exists(I, K, o, d) \text{Navštívil}(I, Pijan, K, o, d) \vee \cup(k=\text{count}(Krčma))(\exists(I,o,d) \text{Navštívil}(I, Pijan, Krčma, o, d)) \wedge k < 5 \}$$

Algebra:

$$\prod_{Pijan} \text{Ľúbi} - \prod_{Pijan} \text{Navštívil} \cup \prod_{Pijan} \sigma_{k < 5} \Gamma_{Pijan, k \leftarrow \text{count}(Krčma)} \prod_{Pijan, Krčma} \text{Navštívil}$$

**Iné riešenie:** Sú to všetci pijani okrem tých, čo navštívili 5 a viac krčiem.

Datalóg:

VšetciPijani(Pijan)  $\leftarrow$  Ľúbi(Pijan, P);  
 $\leftarrow$  Navštívil(I, Pijan, K, o, d).  
 NavštívilKrčmu(Pijan, K)  $\leftarrow$  Navštívil(I, Pijan, K, o, d).  
 Skúsení(Pijan)  $\leftarrow$  subtotal(NavštívilKrčmu;Pijan;k=count(K)),  $k \geq 5$ .  
 Neskúsení(Pijan)  $\leftarrow$  VšetciPijani(Pijan),  $\neg$  Skúsení(Pijan).

Kalkul:

$$\{Pijan: \exists(P) (\text{Ľúbi}(Pijan, P) \vee \exists(I, K, o, d) \text{Navštívil}(I, Pijan, K, o, d)) \wedge \neg(\cup(k=\text{count}(Krčma))(\exists(I,o,d) \text{Navštívil}(I, Pijan, Krčma, o, d)) \wedge k \geq 5) \}$$

Algebra:

$$(\prod_{Pijan} \text{Ľúbi} \cup \prod_{Pijan} \text{Navštívil}) - \prod_{Pijan} \sigma_{k \geq 5} \Gamma_{Pijan, k \leftarrow \text{count}(Krčma)} \prod_{Pijan, Krčma} \text{Navštívil}$$

**Riešenia 1.2:**

Vyhovuje pijan, ktorý splňuje podmienku: ak v nejakej krčme čapujú pivo, ktoré ľúbi, potom tú krčmu navštívil.

Kalkul:

$$\{Pijan: \forall(Krčma, Pivo) \exists(I, o, d) \\ \text{Ľúbi}(Pijan, Pivo) \wedge \text{Čapujú}(Krčma, Pivo) \Rightarrow \text{Navštívil}(I, Pijan, Krčma, o, d)\}$$

V datalógu a algebre nevieme túto formulu realizovať priamo, použijeme preto jej negáciu t.j. najdeme pijanov, čo nevyhovujú. Aby sme dostali vyhovujúcich, vynecháme nevyhovujúcich z pomedzi všetkých pijanov.

Datalóg:

$$\begin{aligned} \text{Nevyhovuje}(Pijan) &\leftarrow \text{Ľúbi}(Pijan, P), \text{Čapujú}(K, P), \neg \text{Navštívil}(I, Pijan, K, o, d). \\ \text{Vyhovuje}(Pijan) &\leftarrow \text{VšetciPijani}(Pijan), \neg \text{Nevyhovuje}(Pijan). \end{aligned}$$

Algebra:

$$\begin{aligned} &(\prod_{Pijan} \text{Ľúbi} \cup \prod_{Pijan} \text{Navštívil}) - \\ &\prod_{Pijan} (\prod_{Pijan, Krčma} \text{Ľúbi} \times \text{Čapujú} - \prod_{Pijan, Krčma} \text{Navštívil}) \end{aligned}$$

**Príklad 2 (6 b)**

Daná je schéma  $R = ABCDEFGH$  a množina závislostí

$$F = \{A \rightarrow GE, B \rightarrow CE, AD \rightarrow E, CD \rightarrow F, BD \rightarrow A, CDE \rightarrow ABD\}.$$

- Nájdite všetky kľúče schémy  $R$ .
- Zistite, či rozklad  $S_1 = ABCDE$  a  $S_2 = CDFGH$  sa spája bezstrátovo.
- Nájdite 3NF nelámajúcu závislosti, pozostávajúcu z najmenšieho počtu tabuliek.

**Riešenia:**

Minimálne pokrytie:

$$\{A \rightarrow G, A \rightarrow E, B \rightarrow C, B \rightarrow E, CD \rightarrow F, CDE \rightarrow A, CDE \rightarrow B\}$$

- Postupným vynechávaním odvoditeľných atribútov zo schémy  $R$  získame prvý kľúč **CDEH**.  $C$  nahradíme  $B$  a pretože  $B$  určuje aj  $E$ , ďalší kľúč je **BDH**. Nahradením  $E$  za  $A$  v kľúči **CDEH** dostaneme kľúč **ACDH**. Viac kľúčov neexistuje.

- Nespája sa bezstrátovo.** Dôvod ani jedna relácia neobsahuje nadkľúč, alebo test z „kuchárky“.

Schéma	A	B	C	D	E	F	G	H
ABCDE	*	*	*	*	*	*	*	*
CDFGH			*	*		*	*	*

Okrem červenej hviezdičky sa nič nedá doplniť.

- Neprimárne atribúty sú jedine  $F$  a  $G$ . Stačí závislosti ich určujúce vydeliť do samostatných relácií. Dekompozícia  $R_1 = \mathbf{AG}$ ,  $R_2 = \mathbf{CDF}$  a  $R_3 = \mathbf{ABCDEH}$  je teda v 3NF a neláme žiadnu závislosť. Schémy  $R_1$  a  $R_2$  nemôžeme spojiť, lebo jediný kľúč spojenej relácie **ACDFG** je **ACD** a obe v nej platné závislosti porušujú definíciu 3NF.

Podľa kucharky bez optimalizácie dostaneme rozklad **AG**, **CDF**, **ACDE**, **BCDE** a jeden kľúč napr. **BDH**.

**Úloha 3** (5b, bod za každú otázku )

1. Definujte pojem dátového modelu.
2. Popíšte trojschémovú architektúru.
3. Čo je to referenčná integrita.
4. Popíšte algoritmus systematického generovania všetkých kľúčov.
5. Popíšte spôsob, akým sa prekladá SQL do relačnej algebry. Uvedte preklad dotazu:

```
select T.a, T.b
from T(a,b,c), S1(a,b), S2(a,c)
where T.a = S1.a or T.a = S2.a;
```

Aký je výsledok, keď tabuľka S2 je prázdna ?

**Riešenia:**

1. Matematická notácia pre dáta a operácie s nimi. (Úvod slide 16)
2. Úvod slide 15
3. Inklúzna závislosť vyjadrujúca podmienku konzistencie dát podporovaná v SQL. (Normálne formy slide 24)
4. Viď kuchárka str. 2
5. Definícia sémantiky select príkazu v SQL,  $\Pi\sigma$ . Ak S2 je prázdna výsledok je prázdny, lebo kartézsky súčin, z ktorého sa vyberá je prázdny.

**Úloha 4** (4b)

Popíšte spôsob vytvárania a využitie indexovo sekvenčných súborov. Vysvetlite využitie viac úrovni indexov. Čo je to riedký a hustý index. Definujte B strom. Popíšte algoritmus vloženia a vynechania prvku v B-strome. Porovnajme rôzne varianty B stromov.

**Riešenie:** Viď prednáška Základné dátové štruktúry slidy 14 – 24. Stačil stručný výťah.

**Príklad 5** (4b)

Je potrebné kódovať texty pozostávajúce zo štyroch znakov {a,b,c,d}, ktorých pravdepodobnosti výskytu nepoznáme navrhnete vhodnú metódu:

1. Kompresujte slovo *aabbcd aabbcd*
2. Porovnajme s inými možnosťami kompresie.

**Riešenie:** Vhodné sú rôzne varianty LZW metódy. Kompresia slidy 11 – 14. Ak ste použili inú variantu ako slide 13, bolo potrebné uviesť aj spôsob dekódovania. Techniky Bzip2 (slidy 20 – 23) sú vhodné, ale pre ručný výpočet na písomke príliš pracné.