

Řízení databázových systémů

Přednáška 2

Roman Špánek

První normální forma – 1NF

- Def:
 - **Neredukovatelná** je taková množina, kterou nelze rozložit na systém jednodušších množin bez ztráty informace.
- Def:
 - Relace je v první normální formě, pokud všechny domény jejích atributů jsou neredukovatelné (**atomické**).
- Značíme: R je v 1NF.

Normalizace a normální formy

- Závislosti mezi daty
- Funkční závislost

Funkční závislosti

- Požadavky na databázi jsou speciálním druhem IO
- Budeme jim říkat **funkční závislosti**
 - Předmět_název → předmět_zkratka
 - osobní_číslo → jméno_studenta

Funkční závislosti

- Funkční závislost je definována mezi dvěma množinami atributů v rámci jednoho schématu relace.
- Jedná se tedy o vztahy mezi **daty**

X-hodnoty

- Mějme schéma relace $R(A)$, dále mějme $X \subseteq A$
- jsou-li atributy v X , $X_1:\text{dom}(X_1)$, ... , $X_n:\text{dom}(X_n)$, pak X-hodnotou budeme nazývat libovolný prvek z kartézského součinu $\text{dom}(X_1) \times \dots \times \text{dom}(X_n)$.

Funkční závislosti pak lze definovat jako:

- Mějme množiny atributů $B, C \subseteq A$.
- C funkčně závisí na B , pokud ke každé B -hodnotě lze přiřadit (existuje) nejvýše jedna C -hodnota.
- Funkční závislost označíme jako
- $B \rightarrow C$.
- Pokud naopak C funkčně nezávisí na B , značíme:
 $B \nrightarrow C$

Pozor.

- Nelze funkční závislosti odvozovat od samotných dat v databázi!!!
- Tak bychom mohli odhalit a snažit se i odstranit funkční závislosti, které ve skutečnosti závislostmi nejsou.

Klíč

Mějme:

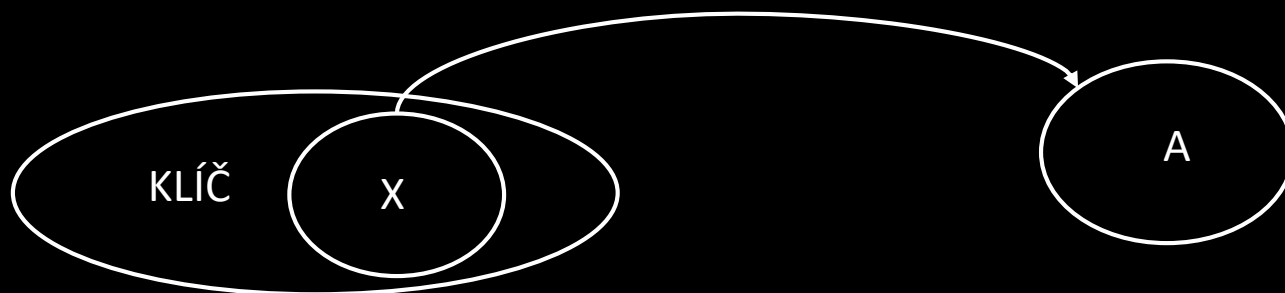
- Relační schéma $R(A)$
- $K \subseteq A$
- Pak K je klíčem schématu R pokud platí:
 1. $K \rightarrow A$
 2. Neexistuje K' , která je vlastní podmnožinou K , taková že $K' \rightarrow A$.

Klíč

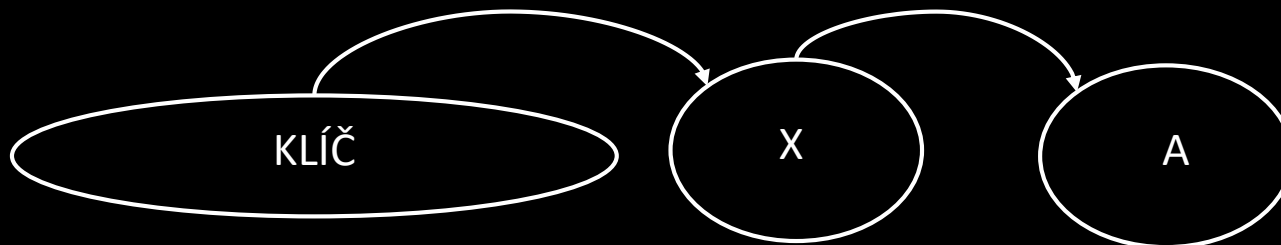
- Relace může mít více klíčů (množin splňujících předchozí požadavky:
 - *Kandidátní klíče,*
 - *Primární klíč – pouze jeden vybraný z množiny kandidátních klíčů*

3NF

- Pokud existuje FZ $X \rightarrow A$ pak 3NF je porušena pokud



X je vlastní podmnožina nějakého klíče K



X není vlastní podmnožina nějakého klíče K,
pak se jedná o tzv. tranzitivní závislost

3NF

- 3NF zajišťuje nezávislost neklíčových atributů
- Relace je ve třetí normální formě, pokud žádný neklíčový atribut není závislý na jiném neklíčovém atributu.

Boyce-Coddova normální forma BCNF

- Relace R je v BCNF právě tehdy, když pro každou netriviální závislost $X \rightarrow Y$, kde X a Y jsou množiny atributů a zároveň Y není podmnožinou X , platí, že X je nadmnožinou nějakého klíče, nebo X je klíčem relace R .
 - Jinak řečeno relace R je v BCNF tehdy a jen tehdy, když každý determinant funkční závislosti v relaci R je zároveň kandidátním klíčem relace R .
- Zajišťuje, že schéma nebude obsahovat redundanci detekovatelnou pomocí funkčních závislostí

BCNF příklad

- V této relaci platí dvě netriviální funkční závislosti:
 $\{\text{Město}, \text{Ulice}\} \rightarrow \text{PSČ}$ a $\text{PSČ} \rightarrow \text{Město}$
Protože neplatí $\text{Ulice} \rightarrow \text{PSČ}$ ani $\text{Město} \rightarrow \text{PSČ}$, tvoří dvojice $\{\text{Město}, \text{Ulice}\}$ klíč schématu.
- Klíčem je ale i $\{\text{Ulice}, \text{PSČ}\}$ platí totiž $\text{PSČ} \rightarrow \text{Město}$, nikoliv však $\text{PSČ} \rightarrow \text{Ulice}$.
- $\{\text{PSČ}, \text{Ulice}\}$ je kandidátním klíčem schématu.
- Schéma má všechny atributy atomické a nemá žádný neklíčový atribut a tudíž je v 3.NF, ale není v BCNF.
- Tento fakt vede k tomu, že nelze evidovat města s PSČ bez znalosti Ulice a krom toho jsou v relaci redundantní data, pokud by se evidovalo velké množství ulic v jednom městě, začal by to být problém.
- Klasické řešení, rozpad na dvě tabulky. Vzhledem k tomu, že neplatí $\text{PSČ} \rightarrow \text{Ulice}$, musíme spojit PSČ a Ulice. Výsledkem tudíž budou relace Města(PSČ, Město) a Ulice(PSČ, Ulice)

Adresa		
Město	Ulice	PSČ
Praha 10	Černokostelecká	100 00
Jihlava	Žižkova	586 01
Praha 10	Vrátkovská	100 00
Brno	Dvořákova	589 74
Praha 6	Chaloupeckého	160 00

Jak navrhovat schémata v odpovídající normální formy?

- Dekompozice
- Syntéza

Dekompozice schématu R

- Dekompozici schématu relace R myslíme proces dělení původní relace na více schémat, tak že platí:
 - Pokud A je množina atributů původního schématu a A_i pro $i=1,2,\dots,n$, $n>1$ je množina i -tého schématu relace, pak platí že sjednocení všech A_i je rovno A

Bezztrátová dekompozice

- Do jaké míry jsme při rozdělování na menší relace netratili **sémantiku**, která byla v původní relaci obsažena?
- Mělo by platit:
 - Výsledná schémata by měla mít **stejnou sémantiku**
 - Výsledné relace by měly **obsahovat stejná data**, jaká obsahovala původní relace

Zachování sémantiky

- Sémantika schématu je dána pomocí integritních omezení (IO), která jsou vyjádřeny (nejen) pomocí funkčních závislostí (FZ)

Příklad

- Mějme schéma relace

FILM(NAZEV, HEREC, NARODNOST,ROK)

- A FZ:

- NAZEV \rightarrow HEREC
- HEREC \rightarrow NARODNOST
- NAZEV \rightarrow ROK

- TFZ

NAZEV \rightarrow HEREC \rightarrow NARODNOST

Provedeme dekompozici do 3NF

- PERSONAL(HEREC,NARODNOST)
- FILM2(NAZEV,HEREC,ROK)
- Platí FZ jako u původního schématu?
 - NAZEV→HEREC
 - HEREC →NARODNOST
 - NAZEV →ROK
- TFZ
NAZEV→HEREC→NARODNOST

ANO

- PERSONAL(HEREC,NARODNOST)
- FILM2(NAZEV,HEREC,ROK)
 - NAZEV→HEREC
 - HEREC →NARODNOST
 - NAZEV →ROK

Jiný příklad

- Schéma relace
- ADRESAR(MĚSTO,ULICE,PSC)
- A FZ
 - $\{MESTO,ULICE\} \rightarrow PSC$
 - $PSC \rightarrow MESTO$
- TFZ
 - $\{MESTO,ULICE\} \rightarrow PSC \rightarrow MESTO$
- Provedeme dekompozici
 - $ULICE(\underline{ULICE},PSC)$
 - $MESTA(\underline{PSC},MESTO)$

Platí ještě všechny FZ?

- $\{MESTO, ULICE\} \rightarrow PSC$
- $PSC \rightarrow MESTO$
- Dekompozice
 - $ULICE(\underline{ULICE}, PSC)$
 - $MESTA(\underline{PSC}, MESTO)$

Pokrytí funkčních závislostí

- Tato vlastnost zaručuje zachování sémantiky
- Mějme relační schéma databáze $R=\{S(A,F)\}$ a dekompozici $R1=\{Ri(Ai, Fi), 1\leq i\leq n, n\geq 1\}$. Říkáme, že $R1$ má **vlastnost pokrytí závislostí**, pokud platí:

$$F^+ = \left(\bigcup_{i=1}^n F_i \right)^+$$

Jinými slovy

- Vezmeme-li funkční závislosti v jednotlivých R_i a vytvoříme z nich uzávěr, výsledkem by měl být shodný s uzávěrem z F
- Uzávěr: je množina všech FZ, která plynou z množiny FZ v F .

Jak tento uzávěr spočítat?

- Označme X, Y, Z : množiny atributů z relace R :
 1. Reflexivita: pokud $X \subseteq Y$, pak $X \rightarrow Y$
 2. Rozšíření: pokud $X \rightarrow Y$, pak $XZ \rightarrow YZ$ pro všechny Z
 3. Tranzitivita: pokud $X \rightarrow Y$ a $Y \rightarrow Z$, pak $X \rightarrow Z$
- Armstrongova pravidla

Bezztrátové spojení

- Požadavky na dekompozici byly:
 - Výsledná schémata by měla mít stejnou sémantiku
 - Výsledné relace by měly obsahovat stejná data, jaká obsahovalo původní relace

- Pro každou přípustnou relaci S^* by mělo platit:

$$S^* = \bigast_{i=1}^n S^*[A_i]$$

- Kde S^* je označení relace
- A^* je operace spojení