

DBS – Normální formy, normalizace

Michal Valenta

Katedra softwarového inženýrství FIT
České vysoké učení technické v Praze
©Michal Valenta, 2014

BI-DBS, ZS 2014/15

<https://edux.fit.cvut.cz/courses/BI-DBS/>



Návrh relačního schématu

Existují dva přístupy:

- 1 normalizační teorie
Metoda návrhu pomocí funkčních závislostí.
- 2 z konceptuálního schématu
Metoda použití transformačních pravidel.

Poznámka: použijeme-li transformaci z konceptuálního schématu, **nemáme zaručeno**, že výsledné schéma bude normalizované! Samotné konceptuální schéma totiž nemusí být normalizované.

Kvalita schématu a normalizace

Uvažujme relaci:

PROGRAM(NAZEV_K, JMENO_F, ADRESA, DATUM)

Aktualizační anomálie (Codd)

- Změní-li se adresa kina, je nutné ji měnit víckrát,
- nehraje-li kino zrovna nic, ztrácíme jeho adresu,
- chceme-li přidat nové kino s adresou, lze to jen když se tam hraje nějaký film.

Jak to vyřešíme?

Normalizace dekompozicí

KINO(NAZEV_K, ADRESA)

MA_NA_PROGRAMU(NAZEV_K, JMENO_F, DATUM)

$MA_NA_PROGRAMU[NAZEV_K] \subseteq KINO[NAZEV_K]$

Dekompozicí jsme se zbavili všech aktualizačních anomálií.

Funkční závislost – neformálně

Hodnoty některých atributů funkčně závisí na hodnotách jiných atributů. **Například:**

- 1 Ke každému kinu existuje nejvýše jedna adresa.
- 2 Pro každou dvojici {kino, film} existuje nejvýše jedno datum, kdy dané kino má daný film na programu.

Což budeme zapisovat:

- 1 $\text{NAZEV_K} \rightarrow \text{ADRESA}$
- 2 $\{\text{NAZEV_K}, \text{JMENO_F}\} \rightarrow \text{DATUM}$

A číst:

- 1 “Atribut NAZEV_K (funkčně) určuje atribut ADRESA .”
nebo:
“Atribut ADRESA (funkčně) závisí na atributu NAZEV_K .”
- 2 “Dvojice (atributů) $\text{NAZEV_K}, \text{JMENO_F}$ (funkčně) určuje atribut DATUM .”
nebo:
“Atribut DATUM (funkčně) závisí na (dvojici atributů) $\text{NAZEV_K}, \text{JMENO_F}$.”

Funkční závislosti – integritní omezení

- funkční závislosti (FZ) vyjadřují integritní omezení
- **připomínka**: integritní omezení (obecně) jsou tvrzení, které určují jaká data v databázi být mohou a jaká ne
- **připomínka**: schéma relační databáze je $\{R(A), I\}$
- FZ uvádí do souvislosti prvky z domén příslušných atributů, je to funkce $f : A_1 \rightarrow A_2$

Kvalita schématu - příklad

Mějme databázi s rozvrhem předmětů:

Rozvrh (Přednáška, Učitel, Místnost, Hodina, Student, Znamka)

Nechť platí toto (vnitropodnikové) pravidlo (IO1):

“Každá přednáška je přednášena nejvýše jedním učitelem.”

Z pohledu DB schématu:

“K jedné hodnotě z dom(Přednáška) se přiřadí nejvýše jedna hodnota z dom(Učitel).”

Předmět \rightarrow Učitel

což budeme v dalším výkladu zkracovat:

$P \rightarrow U$

Kvalita schématu - příklad

Nechť ve schématu ROZVRH jsou zakódovány aktualizační anomálie.
Nahradme schéma množinou schémat tak, aby výsledek měl
“**rozumné vlastnosti**”.

výchozí schéma: $R(P, U, M, H, S, Z)$ (stučněji: $PUMHSZ$)
možné náhrady:

$$R_I = \{PU, HMP, HUM, PSZ, HSM\}$$

$$R_{II} = \{PU, HSP, PSZ, HSM\}$$

$$R_{III} = \{PU, HSM, PSZ, HMP\}$$

$$R_{IV} = \{PU, HMP, PSZ, HSP\}$$

$$R_V = \{HMPU, PSZ, HSM\}$$

$$R_{VI} = \{PU, HMP, PSZ\}$$

$$R_{VII} = \{PSUHM, PSZ\}$$

Které ze schémat $R_I..R_{VII}$ je nejlepší?

Odhalení funkčních závislostí mezi atributy

Odhalení FZ mezi atributy schématu:

P	U	M	H	S	Z
Programování	Kryl	S7	Po9	Novák	2
Programování	Kryl	S3	Út3	Novák	2
Programování	Kryl	S7	Po9	Volák	3
Programování	Kryl	S3	Út3	Volák	3
Systemy	Král	S4	Po7	Zíka	1
Systemy	Král	S4	Po7	Tupý	2
Systemy	Král	S4	Po7	Novák	2
Systemy	Král	S4	Po7	Bílý	1

? platí ? $U \rightarrow HM$! jistě neplatí !, tedy: $U \nrightarrow HM$

zřejmě platí: $P \rightarrow U$, $HM \rightarrow P$, $HU \rightarrow M$, $HS \rightarrow M$

? a co toto ? : $PS \rightarrow Z$

X-hodnota

Mějme schéma $R(A)$, uvažujme $X \subseteq A$

X-hodnota

Jsou-li atributy v $\mathbf{X} \{X_1 : \text{dom}(X_1), \dots, X_n : \text{dom}(X_n)\}$, pak **X-hodnotou** je libovolný prvek z kartézského součinu $\text{dom}(X_1) \times \text{dom}(X_2) \times \dots \times \text{dom}(X_n)$.

x1	x2	x3	x4	x5	x6	x7	x8	x9

Funkční závislost

Mějme schéma $R(A)$.

Funkční závislost

Mějme množiny atributů $B \subseteq A$, $C \subseteq A$. Říkáme, že C závisí funkčně na B (nebo B funkčně určuje C), jestliže ke každé B -hodnotě existuje nejvýše jedna C -hodnota.

značíme:

$$B \rightarrow C$$

resp.:

$$B \nrightarrow C$$

Pozorování

$PS \rightarrow S$ platí vždy

$$\left. \begin{array}{l} PS \rightarrow S \\ PS \rightarrow Z \end{array} \right\} \Leftrightarrow PS \rightarrow SZ$$

Z výchozí množiny funkčních závislostí lze pomocí “určitých” **pravidel** odvozovat další další FZ

Sada “korektních” odvozovacích pravidel

Mějme $R(A)$, nechť $X \subseteq A$, $Y \subseteq A$, $Z \subseteq A$.

Armstrongova pravidla

triviální funkční závislosti

jestliže $Y \subseteq X$, pak $X \rightarrow Y$ (FZ1)

př.: $UM \rightarrow U$

tranzitivita

jestliže $X \rightarrow Y$ a $Y \rightarrow Z$, pak $X \rightarrow Z$ (FZ2)

př.: $HS \rightarrow HM$ a $HM \rightarrow P$, pak také platí $HS \rightarrow P$

kompozice pravé strany

jestliže $X \rightarrow Y$ a $X \rightarrow Z$, pak $X \rightarrow YZ$ (FZ3)

dekompozice pravé strany

jestliže $X \rightarrow YZ$, pak $X \rightarrow Y$ a $X \rightarrow Z$ (FZ4)

Použití odvozovacích pravidel

Mějme vstupní relaci $R(M, H, U, P, S, Z)$

a sadu funkčních závislostí:

$$F = \{P \rightarrow U, HM \rightarrow P, HU \rightarrow M, PS \rightarrow Z, HS \rightarrow M\}$$

Odvodíme:

- Podle (FZ1) platí $HM \rightarrow H$ a $HU \rightarrow H$.
- Podle (FZ3) z $HU \rightarrow H$ a $HU \rightarrow M$ odvodíme $HU \rightarrow HM$.
- Podle (FZ2) z $HM \rightarrow P$ a $P \rightarrow U$ odvodíme $HM \rightarrow U$.
- Podle (FZ3) z $HM \rightarrow H$ a $HM \rightarrow U$ odvodíme $HM \rightarrow HU$.

Vidíme, že HM a HU jsou **funkčně ekvivalentní**: $HM \leftrightarrow HU$

Tranzitivní uzávěr, klíč relace

Uzávěr množiny atributů X^+ vzhledem k F

Uzávěr množiny atributů X^+ vzhledem k F je množina všech atributů funkčně závislých na X . Označujeme jej X^+ .

Klíč relace

Mějme $R(A)$, nechť $K \subseteq A$.

K je klíčem schématu $R(A)$, jestliže splňuje dvě vlastnosti:

- 1 $K \rightarrow A$
- 2 neexistuje $K' \subset K$ taková, že $K' \rightarrow A$.

Příklad – nalezení klíče relace

Mějme vstupní relaci $R(M, H, U, P, S, Z)$

a sadu funkčních závislostí:

$$F = \{P \rightarrow U, HM \rightarrow P, HU \rightarrow M, PS \rightarrow Z, HS \rightarrow M\}$$

Úkol: najděte alespoň jeden klíč relace R vzhledem k F .

$$P^+ = \{P, U\}$$

$$HM^+ = \{H, M, P, U\}$$

$$HU^+ = \{H, U, M, P\}$$

$$PS^+ = \{P, S, Z, U\}$$

$$HS^+ = \{H, S, M, P, Z, U\}$$

Protože $H^+ = \{H\}$ a $S^+ = \{S\}$, je **HS** klíčem relace R .

Normální formy – motivační příklad

PROGRAM	NÁZEV_K	JMÉNO_F	ADRESA	DATUM
	Blaník	Top gun	Václavské nám. 4	29.03.94
	Blaník	Kmotr	Václavské nám. 4	08.03.94
	Mír	Nováček	Starostrašnická 3	10.03.94
	Mír	Top gun	Starostrašnická 3	09.03.94
	Mír	Kmotr	Starostrašnická 3	08.03.94

Integritní omezení:

- IO1: Klíčem schématu je NÁZEV_K, JMÉNO_F.
- IO2: Každé kino má právě jednu adresu.

Relace obsahuje redundance a mohou nastat aktualizací anomálie.

Normální formy – motivační příklad

Intuitivním řešením je dekompozice

ADRESÁŘ(NÁZEV_K, ADRESA),

PROGRAMY(NÁZEV_K, JMÉNO_F, DATUM)

PROGRAMY

NÁZEV_K	JMÉNO_F	DATUM
Blaník	Top gun	29.03.94
Blaník	Kmotr	08.03.94
Mír	Nováček	10.03.94
Mír	Top gun	09.03.94
Mír	Kmotr	08.03.94

ADRESÁŘ

NÁZEV_K	ADRESA
Blaník	Václavské nám. 4
Mír	Starotrašnická 3

- adresa kina je pouze jednou (odstraněna redundance)
- lze evidovat i kino, kde se (právě) nic nehraje (nehrozí ztráta informace o kinu, když bude 'stát')

podstata řešení: odstraněna závislost neklíče (adresa) na pouhém podklíči(Název_k)

Normální formy – motivační příklad

FILM1	JMÉNO_F	HEREC	OBČANSTVÍ	ROK
	Černí baroni	Landovský	CZ	94
	Top gun	Cruise	USA	86
	Kmotr	Brando	USA	72
	Nováček	Brando	USA	90
	Vzorec	Brando	USA	80

Integritní omezení:

- IO1: Klíčem schématu je JMÉNO_F.
- IO2: Každý herec má právě jedno občanství

Relace obsahuje redundance a mohou nastat aktualizací anomálie.

Normální formy – motivační příklad

Intuitivním řešením je dekompozice

OSOBNÍ_ÚDAJE(HEREC, OBČANSTVÍ)

FILM2(JMÉNO_F, HEREC, ROK)

OSOBNÍ_ÚDAJE

HEREC	OBČANSTVÍ
Landovský	CZ
Cruise	USA
Brando	USA

FILM2

JMÉNO_F	HEREC	ROK
Černí baroni	Landovský	94
Top gun	Cruise	86
Kmotr	Brando	72
Nováček	Brando	90
Vzorec	Brando	80

- občanství herce je pouze jednou (odstraněna redundance)
- lze evidovat i občanství herce, jehož filmy vypadly z db (nehrozí ztráta informace o občanství herce, který “stojí”)

podstata řešení: odstraněna závislost neklíče (občanství) na jiném neklíči (herce)

Normální formy – motivační příklad – rozbor

V obou předchozích příkladech byly neklíčové atributy závislé na klíči. Některé z nich však nepřímo - tranzitivně.

V prvním případě šlo o tranzitivitu:

klíč \rightarrow podklíč \rightarrow neklíč

V druhém případě šlo o tranzitivitu:

klíč \rightarrow neklíč \rightarrow neklíč

Jsou-li všechny neklíčové atributy závislé na klíči přímo a nikoliv tranzitivně, pak je schéma ve 3NF.

Poznámka1: Má-li schéma více klíčů ($\text{klíč1} \leftrightarrow \text{klíč2}$), nebude nám vadit $\text{klíč1} \rightarrow \text{klíč2} \rightarrow \text{neklíč}$.

Poznámka2: Jsou-li všechny atributy schématu součástí nějakého klíče, je schéma ve 3NF.

Tranzitivní závislost, 3NF

Tranzitivní závislost

Mějme $R(A)$ Nechť $X \subset A$, $Y \subset A$ a $C \in A$, $C \notin X$ a $C \notin Y$.

Nechť dále $X \rightarrow Y \rightarrow C$ a neplatí, že $Y \rightarrow X$.

Pak říkáme, že C je tranzitivně závislý na X .

Třetí normální forma (3NF)

Říkáme, že schéma relace R je ve 3. normální formě (3NF), jestliže každý neklíčový atribut schématu R není tranzitivně závislý na žádném klíči schématu.

BCNF

Mějme *ROZVRH*(*MHUP*) a platí

$HU \rightarrow M$, $HM \rightarrow P$, $P \rightarrow U$

Lze odvodit klíče: *HU*, *HM*, *HP*.

$P \rightarrow U$ je závislost mezi dvěma podklíči.

ROZVRH vyhovuje kritériu pro 3NF. **Proč?**

... a přeci je v datech redundance!

ROZVRH	PREDNASKA	UCITEL	MISTNOST	HODINA
	Systémy	Král	S4	Po7
	Programování	Kryl	S7	Po9
	Programování	Kryl	S3	Ut11

BCNF

Existuje zde závislost část_klíče1 \rightarrow část_klíče2.

V našem případě: $P \rightarrow U$.

Dekompozice:

OBS(P,U)

ROZVRH1(HMP)

Opět platí, že:

- zmizela redundance v atributu U
- neztratí se informace, že Kryl přednáší Programování, když toto vypadne z rozvrhu

Řešení spočívá v odstranění závislosti části jednoho klíče na části druhého klíče.

BCNF

Říkáme, že schéma relace R je v Boyce - Coddově normální formě (BCNF), jestliže pro každou netriviální závislost $X \rightarrow Y$ platí, že X obsahuje klíč schématu R .

Poznámky:

Každé schéma, které je v BCNF, je také ve 3NF. Obrácené tvrzení obecně neplatí.

Má-li ale schéma jediný klíč, nebo jednoduché klíče, potom je-li ve 3NF je také v BCNF.

BCNF – příklad

Uvažujme schéma relace:

ADRESÁŘ(MĚSTO, ULICE, DUM, PSČ)

F: {MĚSTO, ULICE} \rightarrow PSČ, PSČ \rightarrow MĚSTO

{MĚSTO, ULICE, DUM} je klíčem (\rightarrow {PSČ, MĚSTO, ULICE, DUM})

{PSČ, ULICE, DUM} je klíčem (\rightarrow {PSČ, MĚSTO, ULICE, DUM})

Schéma nemá žádný neklíčový atribut a je tedy ve 3NF. Nikoliv však v BCNF.

ADRESÁŘ lze nahradit dekompozicí.

dekompozice1:

A1(PSČ, MĚSTO)

B1(PSČ, ULICE, DUM)

dekompozice2:

A2(MĚSTO, ULICE, PSČ)

B2(MĚSTO, ULICE, DUM)

Normalizace

Eliminaci aktualizačních anomálií zajišťujeme převedením relačního schématu do 3NF, resp. BCNF.

Normalizovat lze pomocí DEKOMPOZICE

Původní schéma:

$$R(U, F)$$

Dekomponované schéma:

$$\{R_i(U_i, F_i)\}_{i=1}^n, kde \cup_{i=1}^n U_i = U$$

Kvalita dekompozice (požadavky):

P1: Výsledná schémata by měla mít "stejnou" sémantiku.

P2: Nové relace by měly obsahovat "stejná" data, jaká by obsahovala původní relace.

P1: pokrytí původní množiny funkčních závislostí

Cílem bude, aby původní schéma a schémata získaná dekompozicí nějak odrážela stejné závislosti.

$$F^+ = (\cup_{i=1}^n F_i)^+$$

zpět k příkladu:

ADRESÁŘ(MĚSTO, ULICE, DUM, PSČ).

F: {MĚSTO, ULICE} → PSČ, PSČ → MĚSTO

Dekompozice:

SEZNAM_POŠT(PSČ, MĚSTO)

POŠTOVNÍ_RAJON(PSČ, ULICE, DUM)

Ve schématu SEZNAM_POŠT lze kontrolovat původní funkční závislost PSČ → MĚSTO.

Původní závislost {MĚSTO, ULICE} → PSČ pokryta není.

P1: pokrytí původní množiny funkčních závislostí

FILM1(JMÉNO_F, ROK, HEREC, PŘÍSLUŠNOST)

F: HEREC \rightarrow PŘÍSLUŠNOST, JMÉNO_F \rightarrow HEREC,
JMÉNO_F \rightarrow PŘÍSLUŠNOST

Dekompozice podle HEREC \rightarrow PŘÍSLUŠNOST:

OSOBNÍ_ÚDAJE(HEREC, PŘÍSLUŠNOST),
HEREC \rightarrow PŘÍSLUŠNOST

FILM2(JMÉNO_F, ROK, HEREC), JMÉNO_F \rightarrow HEREC.

Závislost JMÉNO_F \rightarrow PŘÍSLUŠNOST je pokryta, protože je odvoditelná ze závislostí, které platí na schématech OSOBNÍ_ÚDAJE a FILM2.