

Relační model dat (Codd 1970)

Odkud vychází, co přináší?

Formální abstrakce nejjednodušších souborů.

Relační kalkul a relační algebra (dotazovací prostředky).

Metodika pro posuzování kvality relačního schématu.

Metodika pro návrh kvalitního relačního schématu.

- jména **atributů** A_1, A_2, \dots, A_n
- **domény** atributů, D_i nebo **dom**(A_i); 1NF
- **n-tice** (a_1, \dots, a_n) $a_i \in A_i$
- množina n-tic $\subset D_1 \times D_2 \times \dots \times D_n \sim$ **relace**
- **schéma relace** $R(A_1:D_1, \dots, A_n:D_n)$, zkráceně $R(A)$
- **jméno schématu relace** R

Inuitivně (pracovně), ale nepřesně:

relace = tabulka, schéma = záhlaví tabulky

Relační model dat

příklad

Kino*

Film*

KINO	NÁZEV_K	ADRESA
	Blaník	Václ. n. 4
	Vesna	V olšínách 6
	Mír	Strašnická 3
	Domovina	V dvorcích

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

Schémata relací:

Kino (Název_k, Adresa, Jméno_v)

Film (Jméno_f, Herec , Rok)

Program(Název_k, Jméno_f, Čas)

Chceme sledovat více herců daného filmu, co s tím?

Jaký je vztah tabulky FILM na obrázku k relaci FILM* ?

Vztah daný film je dáván v daném kině?

Relační model dat

Chceme sledovat více herců
daného filmu, co s tím?

<u>JMENO_F</u>		HEREC	HEREC2	ROK
Černí baroni		<u>JMENO_F</u>	HEREC	ROK
K	<u>JMENO_F</u>	Černí baroni	Vetchý,Brando	1994
T	Černí baroni	Kmotr	Brando,Cruise	1972
M	Černí baroni	Top Gun	Cruise,Vetchý	1986
V	Kmotr	Nováček	Brando	1990
	Nováček	Vzorec	Brando	1980
	Top Gun	Vetchý	1980	
	Top Gun	Cruise	1986	
	Vzorec	Brando	1980	

Relační model dat

Chceme sledovat více herců daného filmu, co s tím?

<u>JMENO_F</u>	<u>HEREC</u>	ROK
Černí baroni	Brando	1994
Černí baroni	Vetchý	1994
Kmotr	Cruise	1972
Kmotr	Brando	1972
Nováček	Brando	1990
Top Gun	Vetchý	1986
Top Gun	Cruise	1986
Vzorec	Brando	1980

“Klasické” relační řešení (není normalizované => budou zde update anomálie, viz normální formy !!!)

Relační model dat tabulková terminologie

Jaký je vztah tabulky FILM na obrázku k relaci FILM*

Schéma relace

jméno atributu

atribut

relace

n-tice relace

záhlaví tabulky

jméno sloupce

sloupec

tabulka

řádek tabulky

Odlišnosti:

V relaci nezáleží na pořadí n-tic.

Relace neobsahují duplicitní n-tice.

Relační model dat

integritní omezení

je nutné zajistit, aby se do relací dostala pouze "správná" data - přípustné n-tice

- úplná definice relačního schématu:

(R, I) ... **schéma relační databáze**

$R = \{ R_1, R_2, \dots, R_k \},$

I ... množina IO

- **přípustná relační databáze** se schématem (R, I) je množina relací $R_1^*, R_2^*, \dots, R_k^*$ takových, že jejich n-tice vyhovují tvrzením v **I**.

Relační model dat,

integritní omezení

Definice 3.3.1. **Klíč** K schématu $R(A)$ je **minimální** množina atributů z A , která jednoznačně určí n -tice relace R^* .

Tvrzení 3.3.1. Necht' K je klíč schématu $R(A)$. Pak pro každou přípustnou relaci R^* platí:

jsou-li u a v dvě různé n -tice z R^* , pak $u[K] \neq v[K]$.

Relační model dat *integritní omezení*

Příklad:

KINO(NÁZEV_K, ADRESA),
FILM(JMÉNO_F, HEREC, ROK)
MÁ_NA_PROGRAMU(NÁZEV_K, JMÉNO_F, DATUM)

IO1: primární Klíče

IO2: Cizí klíče

IO3: V kinech se nehraje více, než dvakrát týdně

IO4: Jeden film se nedává více, než ve třech kinech

Relační model dat,

Příklad IO

KNIHA (ISBN, AUTOR, TITUL)

EXEMPLAR (ISBN, INV_C, NAKUPU, CENA, ...)

CTENAR(C_CT, JMENO, D_ZPET)

VYPUJCKA(INV_C, C_CT, D_ZPET)

REZERV(ISBN, C_CT, D_REZ)

IO1: primární klíče

IO2: cizí klíče

IO3: knihu lze rezervovat, jsou-li všechny exempláře půjčeny

- jak definovat IO nad schématem úložiště?
- jak zajistit dodržování daného IO?

Vyjádření IO:

- deklarativní
- procedurální na straně klienta
- procedurální na straně serveru

- databázový **dotaz** nad schématem S je výraz, který vrací odpověď se schématem T
 - def. oborem jsou všechna úložiště se schématem S
 - oborem hodnot jsou všechny relace se schématem T
 - data v odpovědi pocházejí z databáze
 - odpověď nezávisí na fyzickém uložení dat
- **dotazovací jazyk** je množina všech použitelných výrazů

Relační model dat *manipulace s relacemi*

- vložení n-tice do dané relace
- zrušení/změna daných n-tic v dané relaci

algoritmus manipulačních operací zahrnuje
kontrolu dodržování IO

- zadání dotazu

relační algebra, relační kalkul, SQL

Relační model dat

souhrn

- relace (tabulky) jsou v 1NF, tj. komponenty jejich n-tic (řádků) jsou atomické (dále nedělitelné)
- přístup k prvkům relace (řádkům tabulky) dle obsahu
- jedinečné n-tice (řádky),
jde o zdůraznění množinového myšlení v RMD
- abstrakce je nezávislá na fyzickém uložení dat
- existují silné prostředky pro manipulaci s daty
- existují metody návrhu schématu úložiště v relační databázi, které vedou na schéma “dobrých vlastností”

Dotazovací jazyk

Relační algebra

Předpoklady: relace $R(A)$, $S(B)$

- *selekce* (restrikce) R podle podmínky φ
Značení: $R(\varphi) = \{ u \mid u \in R \text{ a } \varphi(u) \}$
kde φ je $(t1 \Theta t2)$ nebo $(t1 \Theta a)$
- *projekce* R na množinu atributů C , kde $C \subseteq A$
Značení: $R[C] = \{ u[C] \mid u \in R \}$
- *přirozené spojení* R a S se schématy $R(A)$ resp. $S(B)$
Značení: $T(C) = R * S = \{ u \mid u[A] \in R \text{ a } u[B] \in S \}$
kde $C = A \cup B$
- *přejmenování atributů*,
značení: $t1 \rightarrow \text{alias}$

MA_NA_PROGRAMU(Nazev_K = 'Mír')[Jmeno_F, Datum]*
FILM[Herec \rightarrow Hvezda]

Relační algebra

dotazovací jazyk

- Množinové operace:

- sjednocení \cup

- průnik \cap

- rozdíl \setminus

- kartézský součin \times

- Minimální množina operací:

- B = {kartézský součin,
selekce, projekce, přejmenování,
sjednocení, rozdíl }

Relační algebra

- Programování vs. relační algebra
 - relační algebra je jazyk velmi vysoké úrovně
- Dotazovací jazyk, který umožňuje realizovat relační algebru se nazývá *relačně úplný*.
- Komerční svět:
 - SQL,
 - jazyky formulářů,
 - obrázkové jazyky

Relační algebra

selekce

KINO(NÁZEV_K, ADRESA)

FILM(JMÉNO_F, HEREC, ROK)

MÁ_NA_PROGRAMU(NÁZEV_K, JMÉNO_F, DATUM)

R1 := MÁ_NA_PROGRAMU(NÁZEV_K = 'Mír')

MÁ_NA_PROGRAMU	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

R1	NÁZEV_K	JMÉNO_F	DATUM
	Mír	Nováček	10.03.94
	Mír	Top gun	09.03.94
	Mír	Kmotr	08.03.94

Relační algebra

projekce

$R2 := R1[JMEÑO_F, DATUM]$

R1	NÁZEV_K	JMÉNO_F	DATUM
	Mír	Nováček	10.03.94
	Mír	Top gun	09.03.94
	Mír	Kmotr	08.03.94

R2	JMÉNO_F	DATUM
♣	Nováček	10.03.94
♠	Top gun	09.03.94
♦	Kmotr	08.03.94

Relační algebra

přirozené spojení

$R2 := R1[JMÉNO_F, DATUM]$

R2	JMÉNO_F	DATUM
	Nováček	10.03.94
	Top gun	09.03.94
	Kmotr	08.03.94

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

$R3 := FILM * R2$

R3	JMÉNO_F	HEREC	ROK	DATUM
	Nováček	Brando	90	10.03.94
	Top gun	Cruise	86	09.03.94
	Top gun	McGillis	86	09.03.94
	Kmotr	Brando	72	08.03.94

Relační algebra *filtrace stejných n-tic*

$R4 := R3[HEREC \rightarrow Hvezda]$

R3	JMĚNO_F	HEREC	ROK	DATUM	R4	Hvezda
	Top gun	Cruise	86	09.03.94		Cruise
	Kmotr	Brando	72	08.03.94		Brando
	Nováček	Brando	90	10.03.94		McGillis
	Top gun	McGillis	86	09.03.94		

bez mezikroků:

{ MÁ_NA_PROGRAMU(NÁZEV_K='Mír')[JMĚNO_F,DATUM]
* FILM } [HEREC → Hvezda]

Česky: „Seznam herců (hvězd), kteří hrají ve filmech dávaných v kině Mír“.

Relační algebra

Θ -spojení

kde $\Theta \in \{<, >, =, \geq, \leq, \neq\}$

Mějme relace se schématy $R(A)$ a $S(B)$

$R^* [t1 \Theta t2] S^* = \{ u \mid u[A] \in R^*, u[B] \in S^*, u.t1 \Theta u.t2 \}$

Příklad: Mějme $R(A, B, C)$ a $S(B, C, D, E)$, schéma T značí výsl. operace $R[A < B] S$.

R	A	B	C
	8	2	3
	1	2	3
	2	1	4
	3	6	7
	3	8	9

S	B	C	D	E
	3	4	2	3
	3	3	2	3
	1	4	5	6
	2	3	4	7

T	A	R.B	R.C	S.B	S.C	D	E
	1	2	3	3	4	2	3
	1	2	3	3	3	2	3
	1	2	3	2	3	4	7
	2	1	4	3	4	2	3
	2	1	4	3	3	2	3

Relační algebra

polospojení

- *levé Θ -polospojení*, kde $\Theta \in \{<, >, =, \geq, \leq, \neq\}$,
mějme relace se schématy $R(A)$ a $S(B)$
 $R^* < t1 \Theta t2] S^* = (R^* [t1 \Theta t2] S^*)[A]$
- *pravé Θ -polospojení*, kde $\Theta \in \{<, >, =, \geq, \leq, \neq\}$,
mějme relace se schématy $R(A)$ a $S(B)$
 $R^* [t1 \Theta t2 > S^* = (R^* [t1 \Theta t2] S^*)[B]$
- *levé přirozené polospojení* $R(A), S(B)$
Značení: $R^* <^* S^* = (R^* * S^*) [A]$
- *pravé přirozené polospojení* $R(A), S(B)$
Značení: $R^* * > S^* = (R^* * S^*) [B]$

Operace antijoin $A \bar{\bowtie} B$

Motivace: tato operace se používá jako elementární v Oracle query executoru. Můžeme jí vidět v prováděcích plánech SQL dotazů při vyhodnocení klauzule NOT EXISTS.

Zavedení: $A \bar{\bowtie} B \equiv A - A <^* B$ (není standard!)

Najděte kina, která nic nehrají:

$KINO - KINO <^* MA_NA_PROGRAMU$

$KINO \bar{\bowtie} MA_NA_PROGRAMU$

Relační algebra

příklad

- D1. Dotaz, který vybere seznam kin, která něco hrají.
R := MA_NA_PROGRAMU [NÁZEV_K]
- D2. Dotaz, který vybere seznam kin, která nic nehrají.
KINO [NÁZEV_K] \ MA_NA_PROGRAMU [NÁZEV_K]
- D3. Dotaz, který vybere seznam kin, která hrají film Top Gun.
MA_NA_PROGRAMU (JMENO_F = 'Top Gun') [NÁZEV_K]
- D4. Dotaz, který vybere seznam filmů, které hraje kino s adresou 'Zvonková'
**{ KINO (ADRESA='Zvonková') [NAZEV_K] *
MA_NA_PROGRAMU } [JMENO_F]**

Relační algebra

množinový rozdíl

D5. Seznam kin, která hrají něco s hercem 'M.Brando'.

**D5 := { FILM (HEREC='M.Brando') [JMENO_F] *
MA_NA_PROGRAMU } [NAZEV_K]**

D6. Seznam kin, kde **nedávají žádný** film s M. Brando".

Přepis D6: Dotaz, který vybere všechna kina **s výjimkou** těch,
která dávají některý film s M.Brando

KINO[NÁZEV_K] \ D5

Relační algebra negace a existenční kvantifikace

D7. Seznam kin, ve kterých se nehraje některý film s M. Brando".

1. krok :

získejme univerzum pro množiny dvojic {kino, film s M.Brando}

2. krok:

získejme reálnou množinu dvojic {kino, film s M.Brando}

3. krok:

odečtením získaných množin získáme množinu v programu nezrealizovaných dvojic {kino, film s M.Brando}

4. krok:

projekcí na první atribut získáme požadovaný seznam kin

Relační algebra

dotaz D7

{kino | v kině nehrají některý film s M. Brendo}

1. krok: získáme univerzum pro množiny dvojic {kino, film s M.Brando}

R:=KINO[NÁZEV_K] × FILM(HEREC='Brando')[JMÉNO_F]

2. krok: získáme reálnou množinu dvojic {kino, film s M.Brando}

S:= MA_NA_PROGRAMU[NÁZEV_K, JMÉNO_F]

3. krok: odečtením získaných množin získáme množinu v programu nezrealizovaných dvojic {kino, film s M.Brando}

T:=R \ S

4. krok: projekcí na první atribut získáme požadovaný seznam kin

U := T[NÁZEV_K]

Poznámka: v S opravdu není nutné dělat selekci na filmy s M. Brando.

Relační algebra všeobecná kvantifikace

D8. Zadejte dotaz, který vybere z databáze seznam kin, ve kterých se dávají **všechny** filmy s M. Brando".

$$\forall x.P(x) = \neg \exists x.(\neg P(x))$$

Přepis D8: Seznam kin, ve kterých něco hrají, a zároveň **není** pravdou, že **nehrají** některý z filmů s M. Brando.

2. přepis: Seznam kin, ve kterých něco hrají **s výjimkou** kin, kde **nehrají** některý film s M. Brendo}

3. přepis: {kino | kino něco hraje} \ {kino | v kině **nehrají** některý film s M. Brendo}

Relační algebra

dotaz D8

$\{\text{kino} \mid \text{kino něco hraje}\} \setminus$
 $\{\text{kino} \mid \text{v kinu nehrají některý film s M. Brendo}\}$

$K := \text{MA_NA_PROGRAMU}[\text{NÁZEV_K}] \dots$ kina, kde něco hrají

$R := \text{MA_NA_PROGRAMU}[\text{NÁZEV_K}] \times$
 $\text{FILM}(\text{HEREC} = \text{'Brando'})[\text{JMÉNO_F}]$

R obsahuje všechny možné dvojice (kino, film), kde kino je v provozu a film je s M.Brando

$S := R \setminus \text{MA_NA_PROGRAMU}[\text{NÁZEV_K}, \text{JMÉNO_F}]$

S obsahuje dvojice Kino-Film ve významu „daný film (ve kterém hraje Brando) není na programu daného (aktivního) kina

$T := S[\text{NÁZEV_K}]$

T obsahuje aktivní kina, kde nedávají kterýkoliv z filmů s M.Brando

$U := K \setminus T,$

U obsahuje kina, ve kterých hrají všechny filmy s M. Brando.

Operace dělení - $A \div B$

Pro dotazy se všeobecnou kvantifikací můžeme zavést operaci databázové dělení.

Předpokládejme $A(x,y) B(y)$ pak:

$$A \div B = A[x] - ((A[x] \times B) - A)[x]$$

Výsledkem jsou všechny hodnoty **x** z A, které v A tvoří dvojici s **každým** prvkem **y** z B.

Dotaz D8, pak lze zapsat:

**MA_NA_PROGRAMU[NAZEV_K, JMENO_F] \div
FILM (HEREC = 'Brando') [JMENO_F]**

Relační algebra

všeob. kvantif. příklad 2

D9. Seznam kin, ve kterých se dávají **všechny** filmy, které jsou na programu.

Přepis D9: Seznam kin, ve kterých něco hrají, a zároveň **není** pravdou, že **nehrají** některý z filmů programu.

2. přepis: Seznam kin, ve kterých něco hrají **s výjimkou** kin, ve kterých **nehrají** některý film z programu}

3. přepis: {kino | kino něco hraje} \
{kino | v kině **nehrají** některý film z programu}

Relační algebra

Dotaz D9

$\{\text{kino} \mid \text{kino něco hraje}\} \setminus$

$\{\text{kino} \mid \text{v kinu nehrají některý film z programu}\}$

K := **MA_NA_PROGRAMU**[NÁZEV_K] ... kina, kde něco hrají

R := **MA_NA_PROGRAMU**[NÁZEV_K] \times
MA_NA_PROGRAMU[JMÉNO_F]

R obsahuje všechny možné dvojice (kino, film), kde kino je v provozu a film je na programu

S := **R** \setminus **MA_NA_PROGRAMU**[NÁZEV_K, JMÉNO_F]

S obsahuje dvojice Kino-Film ve významu „daný film (který je na programu) není na programu daného (aktivního) kina

T := **S** [NÁZEV_K]

T obsahuje aktivní kina, kde nedávají některý film z programu

U := **K** \setminus **T**,

U obsahuje kina, ve kterých hrají všechny filmy z programu

Relační algebra

všeob. kvant. příklad 3

D10. Zadejte dotaz, který vybere z databáze kino, ve kterém jsou konána **všechna** představení programu.

Poznámka:

Odpověď bude prázdná, nebo bude obsahovat jediné kino.

Přepis D10: Zadejte dotaz, jehož odpověď bude obsahovat seznam kin, u nichž **není** pravdou, že **nehraje** některé představení programu.

2. přepis: Seznam kin, ve kterých něco hrají **s výjimkou** kin, ve kterých se **nehraje** některé představení z programu

3. přepis: $\{\text{kino} \mid \text{kino něco hraje}\} \setminus \{\text{kino} \mid \text{v kinu nehrají některé představení z programu}\}$

Relační algebra

Dotaz D10

$\{\text{kino} \mid \text{kino něco hraje}\} \setminus$

$\{\text{kino} \mid \text{v kinu nehrájí některé představení z programu}\}$

K := **MA_NA_PROGRAMU**[**NÁZEV_K**] ... kina, kde něco hrají

R := **MA_NA_PROGRAMU**[**NÁZEV_K**] \times
MA_NA_PROGRAMU[**JMÉNO_F**, **DATUM**]

R obsahuje všechny možné dvojice (kino, představení), kde kino je v provozu a představení je dvojice (film, datum) ve významu daný film je na programu v dané datum (den a čas).

S := **R** \setminus **MA_NA_PROGRAMU**[**NÁZEV_K**, **JMÉNO_F**, **DATUM**]

S obsahuje dvojice Kino-představení ve významu „dané představení (které je na programu) není dáváno v daném aktivním kině“

T := **S** [**NÁZEV_K**]

T obsahuje aktivní kina, ve kterých se nekoná kterékoliv představení z programu

U := **K** \setminus **T**,

U obsahuje kino, ve kterém jsou dávána všechna představení z programu

Operace “za” relačním modelem

1. Prázdná hodnota: NULL, UNKNOWN

2. (levé, pravé, plné) vnější spojení :

motivace: seznam filmů ve všech kinech, v seznamu
necht' jsou i kina, kde se zrovna nic nehraje

Chceme získat:

R	NÁZEV_K	ADRESA	JMÉNO_F	DATUM
	Blaník	Václ.n. 4	Top gun	29.03.92
	Blaník	Václ.n. 4	Kmotr	08.03.92
	Mír	Starostrašnická 3	Nováček	10.03.92
	Mír	Starostrašnická 3	Top gun	09.03.92
	Mír	Starostrašnická 3	Top gun	08.03.92
	Mír	Starostrašnická 3	Kmotr	08.03.92
	Vesna	V olšínách 6	NULL	NULL
	Domovina	V dvorcích	NULL	NULL

dotaz **KINO * MÁ_NA_PROGRAMU** problém neřeší

Operace “za” relačním modelem

- *levé vnější přirozené spojení*

$$\mathbf{R} *_L \mathbf{S} = (\mathbf{R} * \mathbf{S}) \cup (\mathbf{R} - \mathbf{R}^{<*} \mathbf{S}) \times \{\text{NULL}, \dots, \text{NULL}\}$$

poznámky: $(\mathbf{R} *_L \mathbf{S})[\mathbf{A}] = \mathbf{R}$

$$(\mathbf{R} *_L \mathbf{S})[\mathbf{B}] = (\mathbf{R}^{*>} \mathbf{S}) \cup \{\text{NULL}, \dots, \text{NULL}\}$$

- *pravé vnější přirozené spojení*

$$\mathbf{R} *_R \mathbf{S} = (\mathbf{R} * \mathbf{S}) \cup (\{\text{NULL}, \dots, \text{NULL}\} \times (\mathbf{S} - \mathbf{R}^{*>} \mathbf{S}))$$

- *plné vnější přirozené spojení*

$$\mathbf{R} *_F \mathbf{S} = (\mathbf{R} *_L \mathbf{S}) \cup (\mathbf{R} *_R \mathbf{S})$$

Operace “za” relačním modelem

- Program kin s doplněnými podrobnosti o daném kinu. V seznamu nechť jsou i kina, kde se zrovna nic nehraje

KINO *_L MÁ_NA_PROGRAMU

- Dvojice kino-film Jedna dvojice má význam „v kině je na programu film“. V seznamu nechť jsou i kina, kde se zrovna nic nehraje a filmy, které nejsou nikde hrány.

KINO *_L MÁ_NA_PROGRAMU *_R FILM