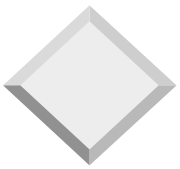


IS, Binární E-R model, transformace do RMD

Jaroslav Pokorný
MFF UK, Praha
pokorny@ksi.ms.mff.cuni.cz

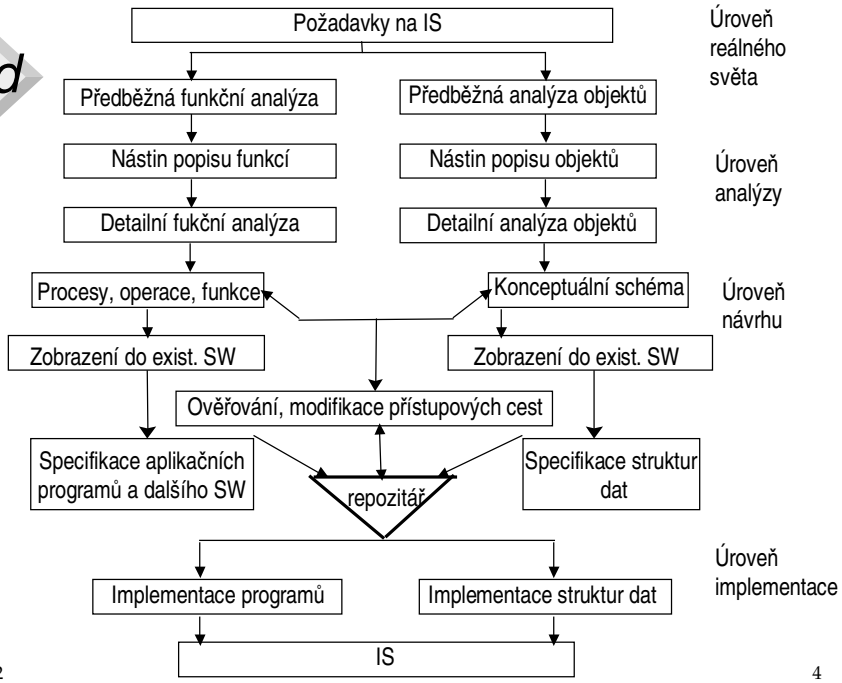
Osnova předmětu DB2-FEL

1. IS, E-R modely, konceptuální schéma v binárním E-R modelu.
2. Pravidla pro kreslení E-R diagramů, transformace schémat do RMD.
3. Metodologie návrhu E-R schématu. Návrh ve velkém (4 přístupy).
4. Funkční analýza pomocí diagramů toků dat.
5. Integrace datové a funkční analýzy.
6. OO a OR přístup k databázím, SQL:1999.
7. Vyhodnocování dotazu.

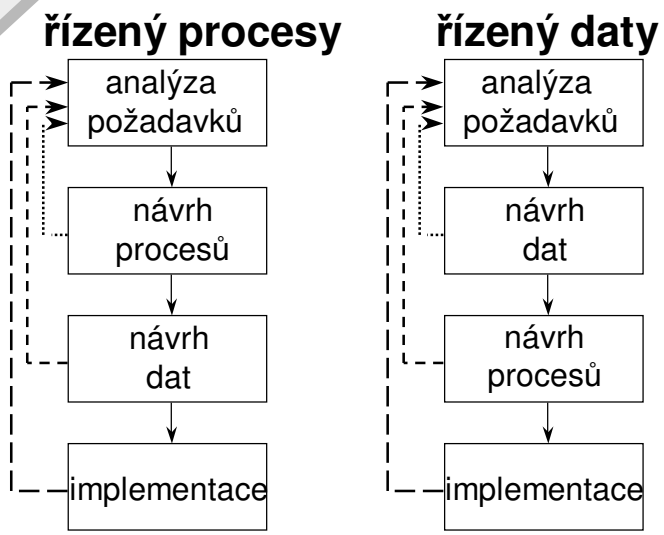
Osnova předmětu DB2-FEL

8. Optimalizace
9. Distribuované DBS – architektury, distribuované zpracování dotazu.
10. Zobecněné transakce, repliky, dvoufázový potvrzovací protokol.
11. Datové sklady – architektury, multidimenzionální modelování.
12. Textové databáze – boolský a vektorový model.
13. Relevance, zpětná vazba, možnosti implementace.

1. Úvod



1. Úvod



Obsah

- Obecná definice E-R modelu
- Binární E-R model
- Transformace E-R $\rightarrow$ RMD

E-R schéma - definice

E-R schéma je čtveřice <***E***, ***R***, ***H***, ***CC***>, kde ***E*** je množina typů entit E_i s atributy A_i, i=1,...,n, $n \geq 1$. ***R*** je množina typů vztahů, ***H*** množina hierarchií typů entit a ***CC*** je množina kardinalit. Dále:

- Každý typ vztahu R je dán n-ticí (E_{i1},…,E_{im}), m≥2, a množinou atributů A_R.
- Existuje $\boldsymbol{E}_S \subset \boldsymbol{E}$ (eventuálně prázdná) obsahující *slabé typy entit*. Pro každý $E \in \boldsymbol{E}_S$ existuje alespoň jedna posloupnost E_1,…,E_s, taková že $E_1 = E$, E_{i-1} je identifikačně závislá na E_i, i = 2,…,s-1, a E_s je silný typ entity. Identifikační klíč E je tvořen sjednocením všech částečných a úplných identifikačních klíčů ze všech těchto posloupností.

E-R schéma - definice

- Typy entit jsou strukturovány do jednoduchých hierarchií. Každá *hierarchie typů entit* H_E je podmnožina $\boldsymbol{E} \times \boldsymbol{E}$ (eventuálně prázdná), kde
  - typ entity E je zdroj hierarchie s identifikačním klíčem K_E,
  - graf odpovídající H_E je stromem s kořenem E, který není vlastním podstromem žádné jiné hierarchie
- pro každý typ vztahu $R \in \boldsymbol{R}$ existuje množina min-max IO $\boldsymbol{CC}_R \subseteq \boldsymbol{CC}$.

E-R informační báze - definice

Informační báze pro dané E-R schéma je množina *entitních množin* typů z ***E*** a *vztahových množin* popsaných v ***R***, které vyhovují příslušným min-max IO.

Je-li $(E_i, E_j) \in H_E$ a E_i^* a E_j^* jsou odpovídající entitní množiny, pak $E_i^* \subseteq E_j^*$.

Pz.: informační báze se implementuje pomocí databáze.

Pz.: neuvažujeme jiná IO než kardinality.

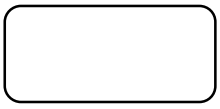
Binární E-R model

V *binárním E-R modelu* jsme omezeni v ***R*** pouze na binární typy vztahů, dále pak na kardinality 1:1 nebo 1:N.

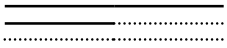
$\Rightarrow$ pro min-max IO v $<E_1:(a,b), E_2:(c,d)>$ vždy buď $a, b \in \{0,1\}$ nebo $c, d \in \{0,1\}$.

Př.: binární E-R model ORACLE (viz Designer 2000 a Developer 2000). Obsahuje ještě některé další úpravy a rozšíření.

Značení: typy entit



typy vztahů



povinné a nepovinné členství

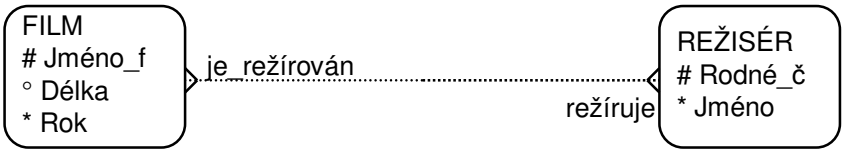
Binární E-R model

- atributy
  - *klíčové* jsou označeny znakem #,
  - *povinné* (jde o totální funkce) jsou označeny znakem *,
  - *nepovinné* (jde o parciální funkce) jsou označeny znakem °.
  - V E-R diagramu se informace o atributech zapisují dovnitř obdélníka.
- vztahy pomocí dvou „inverzních“ jmen
- identifikační závislost

Binární E-R model

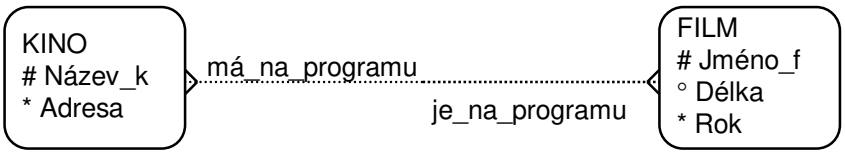
Binární E-R model

Co se vztahy s kardinalitou M:N?



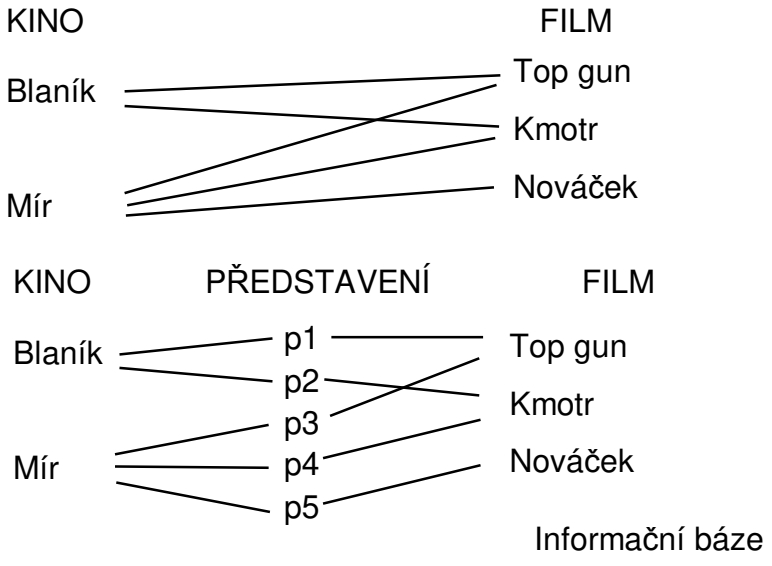
Binární E-R model

Další problém: kam umístit atribut Datum?

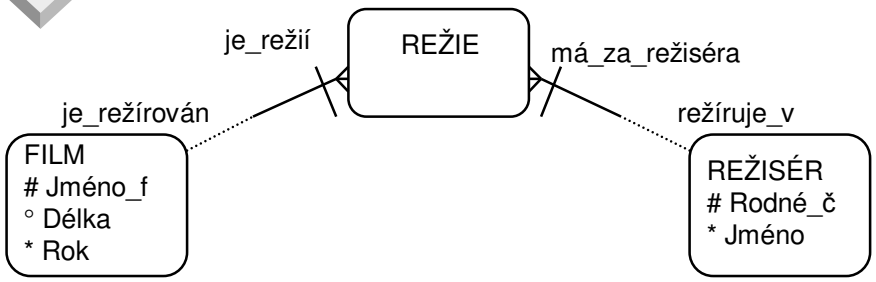


Řešení: dekompozice do dvou typů vztahů s kardinalitou 1:N

Binární E-R model

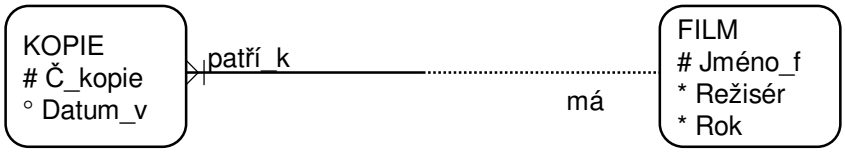


Binární E-R model



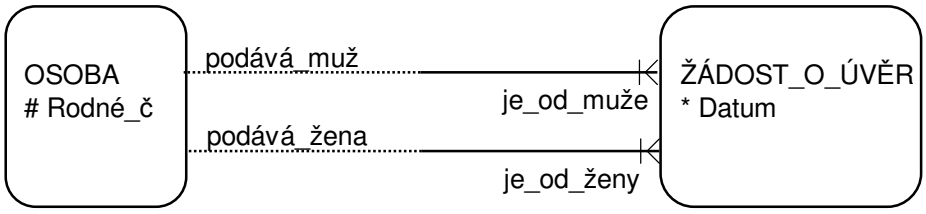
Binární E-R model

Slabý entitní typ a jeho identifikační závislost



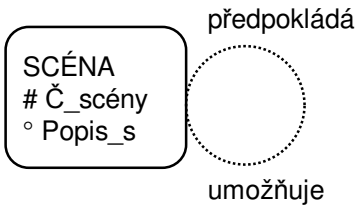
Binární E-R model

Týž identifikační vlastních ve dvou rolích

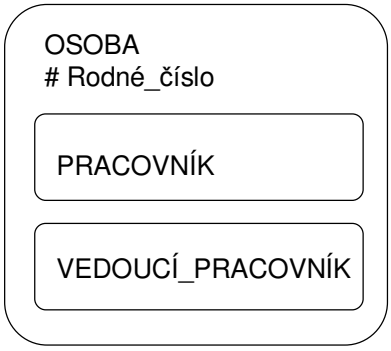


Binární E-R model

Typ rekurzivního vztahu



Hierarchie typů entit



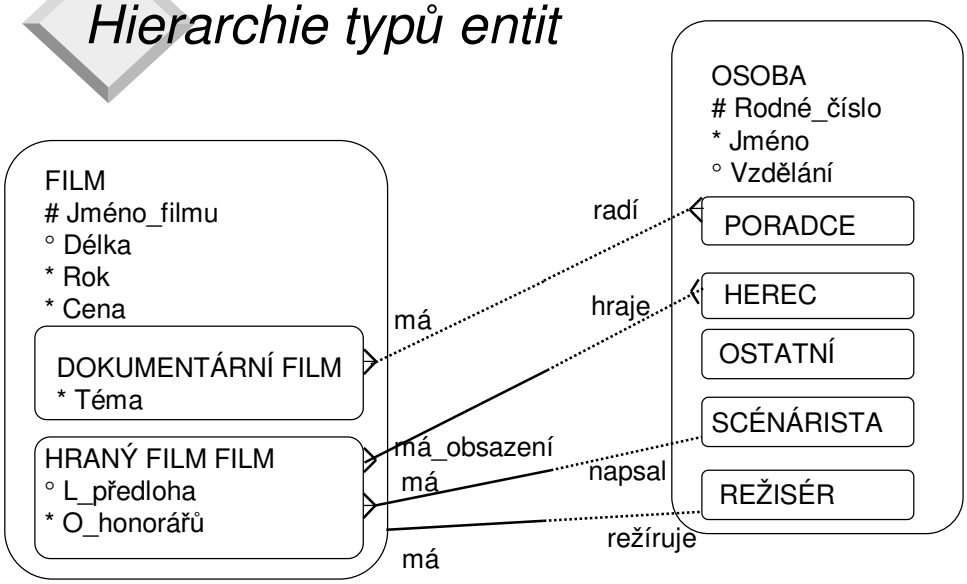
Účel:

- podtypy mají další atributy
- jsou členy dalších typů vztahů

IO v ORACLE:

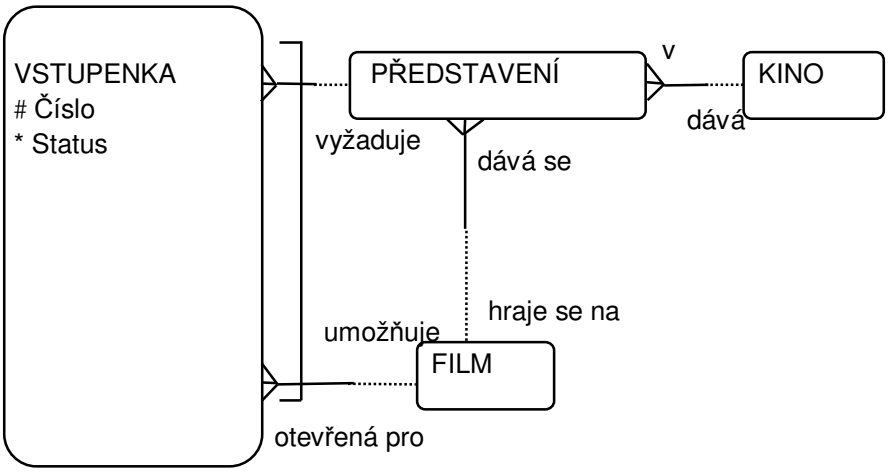
- tvoří disjunktní pokrytí

Hierarchie typů entit



Binární E-R model

Výlučný typ vztahu



Binární E-R model

Definice: Výlučný typ vztahu

Je-li X typ entity spojen dvěma výlučnými typy vztahů s typy entit Y_1 a Y_2, pak lze entity typu X rozdělit do množin X_1 a X_2 tak, že

$$X_1 \cup X_2 = X$$

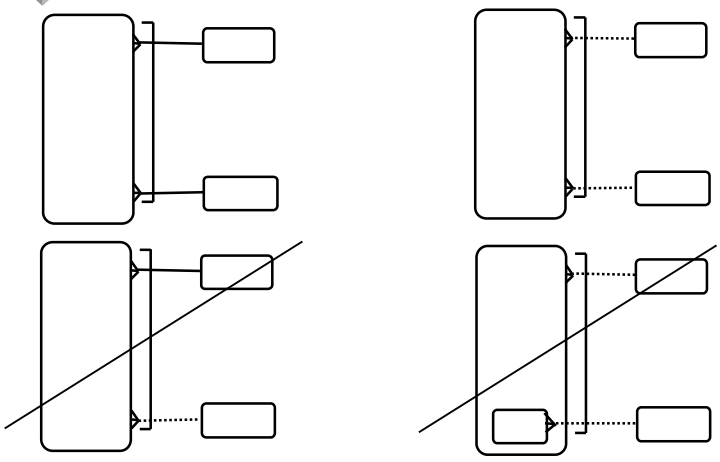
$$X_1 \cap X_2 = \emptyset$$

přičemž vztahy prvního typu jsou z $X_1 \times Y_1$

a vztahy druhého typu jsou z $X_2 \times Y_2$

Binární E-R model

IO v ORACLE:

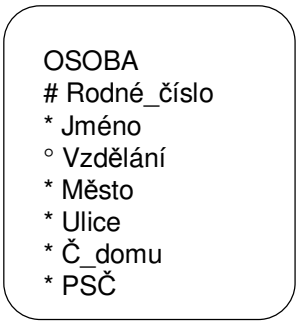


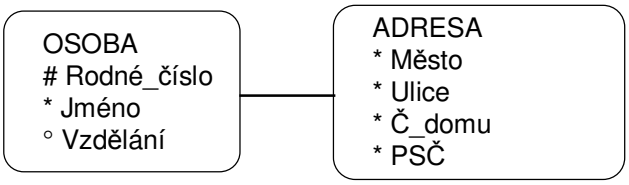
E-R schéma - definice

Další omezení:

Jsou zakázány:

- složené atributy
- vícehodnotové atributy





Příklad

Provoz půjčoven a kin.

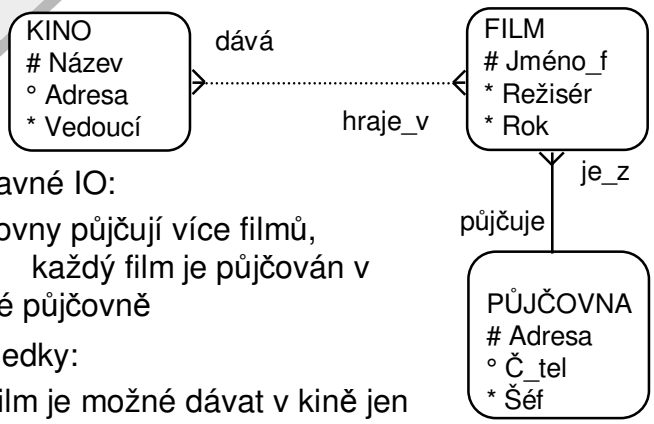
Kina jsou identifikována názvem, sleduje se adresa, vedoucí kina.

Filmy existují nezávisle na půjčovnách, jsou dány jménem filmu, známe jejich režiséra, rok natočení.

Půjčovny jsou dány adresou, znám je telefon a šéf půjčovny.

Filmy se dávají v kinech v daný den, půjčovny půjčují filmy. Každé kino dává více filmů, každý film je dáván ve více kinech.

Příklad - 1. verze



Přídavné IO:

půjčovny půjčují více filmů, každý film je půjčován v jedné půjčovně

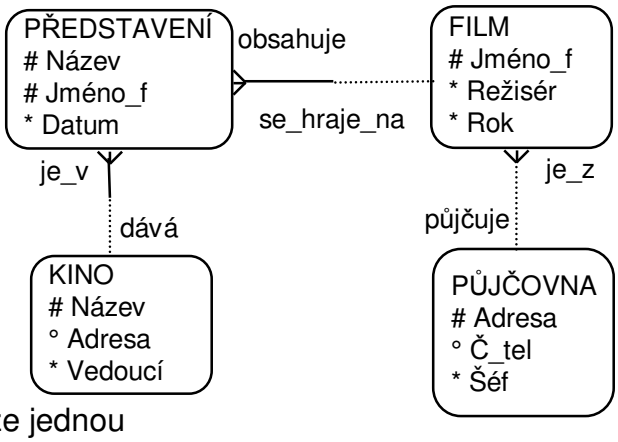
Důsledky:

IO: film je možné dávat v kině jen jednou! (neexistují kopie filmu!)

Problém: M:N mezi KINO a FILM

Chyby?: filmy musí být půjčeny, půjčovny musí půjčovat

Příklad - 2. verze



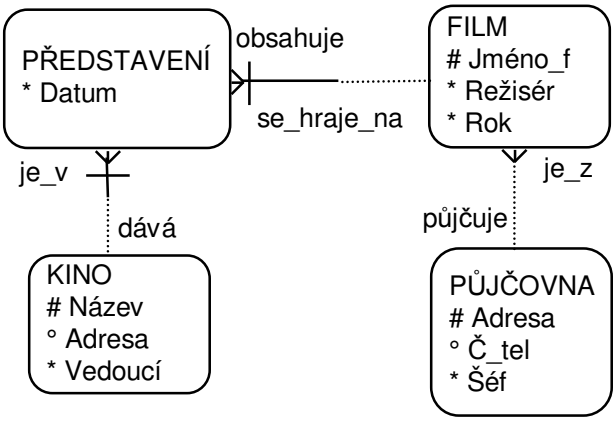
Přídavné IO:

film je stále možné promítat v kině pouze jednou

Řešení:

KOPIE filmu (slabý typ entity - identifikační závislost na FILM)

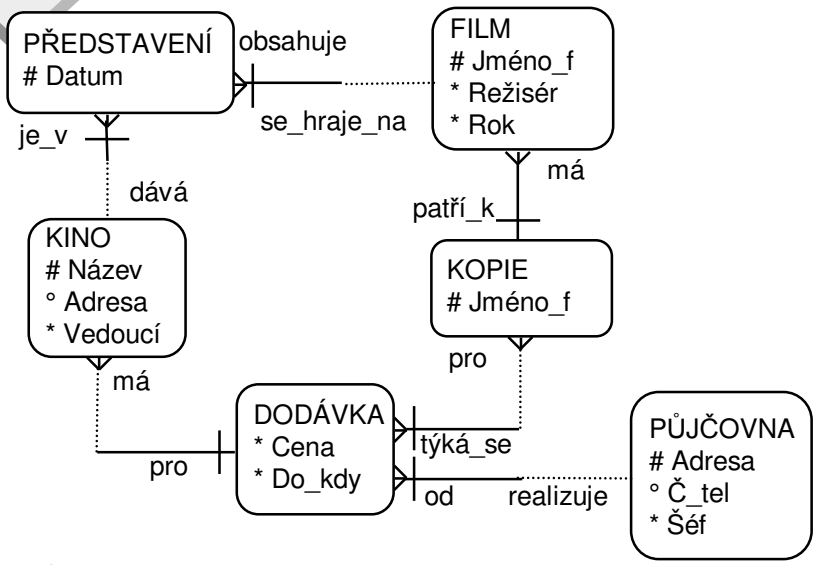
Příklad - 2. verze



Oprava klíčů:

PŘEDSTAVENÍ ke slabý entitní typ

Příklad - 3. verze



Příklad - 3. verze

Důsledky:

- jeden film může jít ve dvou kinech ve stejný den
- jeden film může jít v jednom kině ve více dnech

Chyby?: film je na programu a nemusí být z vypůjčené kopie

Transformace E-R → RMD

Strategie: konceptuální konstrukty → schémata relací + IO (referenční integrita, …)

- Reprezentace typu entity

E → schéma relace téhož jména

zásady: jméno typu entity - jednotné číslo

jméno relace - množné číslo

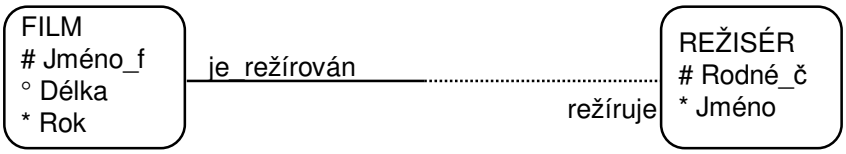
- primární klíč odpovídá identifikačnímu klíči
- typům atributů se přiřadí domény

Pz.: pro reprezentaci IZ typů entit je třeba znát schémata odpovídajících identifikačním vlastníkům

Transformace E-R → RMD

- Reprezentace typu vztahu s kardinalitou 1:1

R → vazba pomocí existujících relací



REŽISÉŘI(RODNÉ_Č, JMÉNO)

FILMY(JMÉNO_F, RODNÉ_Č, DÉLKA, ROK)

NOT NULL

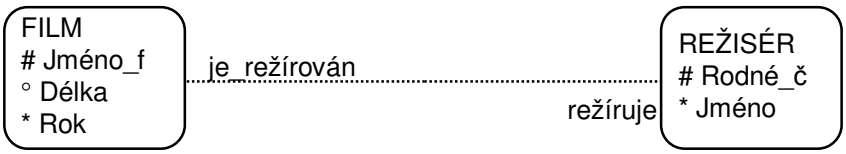
Problémy: co má být primárním klíčem?

kam s vazbou?

Transformace E-R → RMD

- Reprezentace typu vztahu s kardinalitou 1:1

R → vazba pomocí existujících relací



REŽISÉŘI (RODNÉ_Č, JMÉNO)

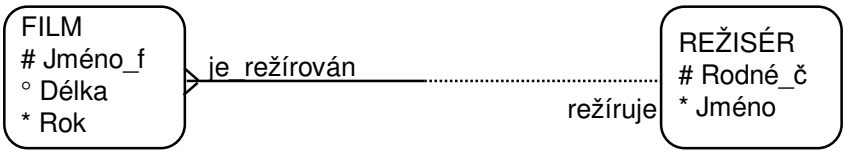
FILMY(JMÉNO_F, RODNÉ_Č, DÉLKA, ROK)

nemusí být NOT NULL

Transformace E-R → RMD

- Reprezentace typu vztahu s kardinalitou N:1

R → vazba pomocí existujících relací



REŽISÉŘI(RODNÉ_Č, JMÉNO)

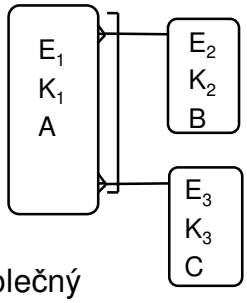
FILMY(JMÉNO_F, RODNÉ_Č, DÉLKA, ROK)

Problémy: co má být primárním klíčem?
kam s vazbou?

odpadají

Transformace E-R → RMD

- Reprezentace výlučných typu vztahů



a) jestliže dom(K_2)=dom(K_3)

$R_1(\underline{K_1}, ID, K, A)$

$R_2(\underline{K_2}, B)$

$R_3(\underline{K_3}, C)$

kde ID je atribut rozlišující typ vztahu, K je společný atribut pro klíče z K_2 a $K_3$3

b) pomocí cizích klíčů

R_1 se nahradí $R_1(\underline{K_1}, K_2, K_3, A)$

+ IO: nesmí existovat řádek s K_2 = NULL ∧ K_3 = NULL

Transformace E-R → RMD

- Reprezentace podtypů entit

a) 1 tabulka pro entity zdroje

FILMY(JMÉNO_F, DÉLKA, ROK, CENA, TYP_FILMU, TÉMA, L_PŘEDLOHA, O_HONORÁŘŮ, RODNÉ_Č)

b) tabulky podtypů

DOKUMNTÁRNÍ_FILMY(JMÉNO_F, DÉLKA, ROK, CENA, TÉMA)

HRANÝ_FILMY(JMÉNO_F, DÉLKA, ROK, CENA, L_PŘEDLOHA, O_HONORÁŘŮ, RODNÉ_Č)

Transformace E-R → RMD

- Reprezentace podtypů entit

c) tabulky pro zdrojový typ a podtypy

FILMY(JMÉNO_F, DÉLKA, ROK, CENA)

DOKUMENTÁRNÍ_FILMY(JMÉNO_F, TÉMA)

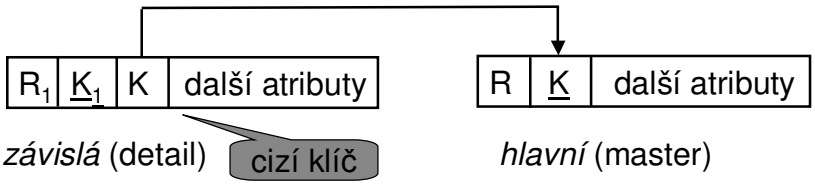
HRANÝ_FILMY(JMÉNO_F, L_PŘEDLOHA, O_HONORÁŘŮ, RODNÉ_Č)

Pz.: v SQL lze možnosti a) a b) realizovat pomocí pohledů. Je třeba typ filmu.

Problém: transformace jsou zatím neúplné

Referenční integrita

- popisuje vztah mezi daty dvou relací pomocí pojmů primární a cizí klíč



Def. podle SQL92: hodnota cizího klíče je v n-tici relace R_1 prázdná nebo je obsažena jako hodnota primárního klíče (případně UNIQUE atributu) v R.

\+ chování aktualizačních operací

Referenční integrita

Pz.: v SQL

KINA

PŘEDSTAVENÍ

FOREIGN KEY (NÁZEV_K) REFERENCES KINO

- INSERT

omezené vložení n-tice *u* do závislé relace

není dovoleno, neexistuje-li pro cizí klíč *k* n-tice *u* v hlavní relaci n-tice, která má *k* jako primárním klíč.

Referenční integrita

- DELETE

omezené odstranění n-tice *u* z hlavní relace

není dovoleno, existuje-li pro primární klíč *k* n-tice *u* v závislé relaci n-tice s cizím klíčem *k*.

kaskádové odstranění n-tice *u* z hlavní relace

odstranění n-tice *u* s primárním klíčem *k* z hlavní relace je propagováno do závislé relace, kde se odstraní všechny n-tice mající cizí klíč roven *k*.

Referenční integrita

nahrazení cizího klíče prázdnou hodnotou

odstranění n-tice *u* s primárním klíčem *k* z hlavní relace je propagováno do závislé relace, kde se ve všech n-ticích majících cizí klíč roven *k* nahradí *k* hodnotou NULL.

nahrazení cizího klíče implicitní hodnotou

odstranění n-tice *u* s primárním klíčem *k* z hlavní relace je propagováno do závislé relace, kde se ve všech n-ticích majících cizí klíč roven *k* nahradí *k* implicitní hodnotou.

Referenční integrita

- UPDATE

omezená změna n-tice *u* v závislé relaci

není dovolena, neexistuje-li pro klíč *k* cizího klíče změněné n-tice *u* v závislé relaci n-tice v hlavní relaci s primární klíčem *k*.

omezená změna n-tice *u* v hlavní relaci

změna primárního klíče *k* n-tice *u* není dovolena, existuje-li v závislé relaci n-tice s cizím klíčem *k*.

kaskádová změna n-tice *u* v hlavní relaci

změna primárního klíče *k* n-tice *u* je propagována do závislé relace na všechny n-tice s cizím klíčem *k*.

Referenční integrita

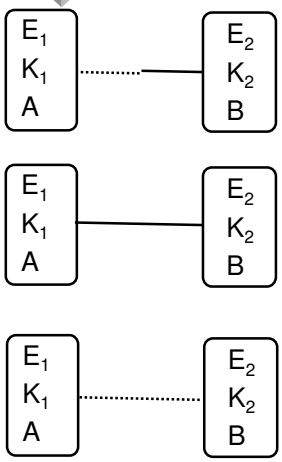
- nahrazení cizího klíče prázdnou hodnotou

změna primárního klíčem *k* n-tice *u* z hlavní relace je propagováno do závislé relace, kde se ve všech n-ticích majících cizí klíč roven *k* nahradí hodnotou *k* NULL.

- nahrazení cizího klíče implicitní hodnotou

změna primárního klíčem *k* n-tice *u* z hlavní relace je propagováno do závislé relace, kde se ve všech n-ticích majících cizí klíč roven *k* nahradí *k* hodnotou implicitní hodnotou.

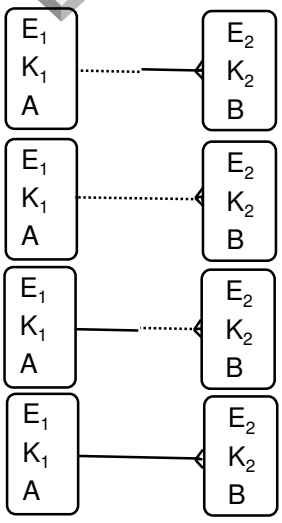
Transformace E-R → RMD: upřesnění



$R_1(\underline{K_1}, A)$
$R_2(\underline{K_2}, K_1, B)$
K_1 UNIQUE NOT NULL
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$

$R_1(\underline{K_1}, A)$
$R_2(\underline{K_2}, K_1, B)$
K_1 UNIQUE
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$

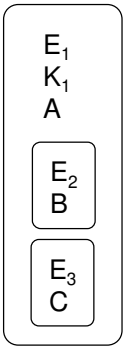
Transformace E-R → RMD: upřesnění



$R_1(\underline{K_1}, A)$
$R_2(\underline{K_2}, K_1, B)$
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$

$R_1(\underline{K_1}, A)$
$R_2(\underline{K_2}, K_1, B)$
K_1 NOT NULL
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$

Transformace E-R → RMD: upřesnění

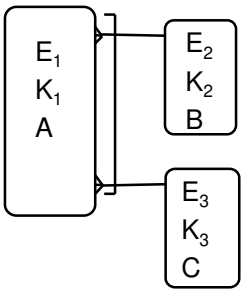


a) $R_1(\underline{K_1}, T, A, B, C)$
kde T je atribut rozlišující podtypy

b) $R_1(\underline{K_1}, A, B)$
$R_2(\underline{K_1}, A, C)$

c) $R_1(\underline{K_1}, A)$
$R_2(\underline{K_1}, B)$
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$
$R_3(\underline{K_1}, C)$
FOREIGN KEY K_1 REFERENCES $R_1(\underline{K_1})$
Pz.: schází vlastnost disjunktního pokrytí

Transformace E-R → RMD: upřesnění



a) $R_1(\underline{K_1}, ID, K, A)$
kde T je atribut rozlišující vztahy
$R_2(\underline{K_2}, B)$
$R_3(\underline{K_3}, C)$

b) $R_1(\underline{K_1}, K_2, K_3, A)$
FOREIGN KEY K_2 REFERENCES $R_2(\underline{K_2})$
FOREIGN KEY K_3 REFERENCES $R_3(\underline{K_3})$
$R_2(\underline{K_2}, B)$
$R_3(\underline{K_3}, C)$
Pz.: schází vlastnost výlučnosti