Adatbázisok 1. (nappali/esti)
Relációs adatmodell

Mi is az az adatmodell?

- Az adatmodell információ vagy adatok leírására szolgáló jelölés. A leírás részei:
  - az adatok struktúrája.
  - Az adatokon végezhető műveletek. A DBMS esetében általában kevesebb műveletet hajthatunk végre, mint egy általános célú programnyelv esetében. Itt azonban a kevesebb, több. A műveletek egyedi hatékony megvalósításán túl, több művelet - egy lekérdezés - együttes optimalizációja is lehetővé válik.
  - Az adatra vonatkozó megszorítások. Pl. egy személyigazolvány-számhoz nem tartozhat két különböző személy.
  - Legfontosabb adatmodellek: relációs és féligstrukturált (XML).

Példa féligstrukturált adatra (XML)

```
<?xml version="1.0" encoding="UTF-8"?>
<bár típus="étterem">
    <név>Makk 7-es</név>
    <város>Budapest</város>
    <tulaj>Géza</tulaj>
    <telefon>+36-70-123-2345</telefon>
    <telefon>+36-70-123-2346</telefon>
</bár>
<bár típus="kocsma">
    <név>Lórúgás</név>
    <város>Eger</város>
    <telefon>+36-30-451-1894</telefon>
</bár>
</xml>
```

Relációs adatmodell I.

- A relációs adatmodellben az adatokat kétdimenziós táblákban (relációkban) tároljuk.

Példa a fogalmak illusztrálására

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

Relációs adatmodell I.

- A relációs adatmodellben az adatokat kétdimenziós táblákban (relációkban) tároljuk.
- A reláció fejrészében találhatók az attribútumok.
- Minden attribútumhoz hozzátartozik egy értékkészlet.

Példa a fogalmak illusztrálására

attribútumok

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

Relációs adatmodell I.

- A relációs adatmodellben az adatokat kétdimenziós táblákban (relációkban) tároljuk.
- A reláció fejrészében találhatók az attribútumok.
- Minden attribútumhoz hozzátartozik egy értékkészlet.
- A reláció neve és a reláció-attribútumok halmaza együtt alkotják a relációsémát.
- A reláció attribútumainak sorrendje felcserélhető (nem változtatja meg a relációt)
- A séma egy adatmodellben általánosságban azt adja meg, hogy egy-egy adatelem milyen „formájú” adatokat tárol.

Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Sör (név, ország).

attribútumok

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

Relációs adatmodell II.

- Egy-egy reláció soroknak egy halmaza.
- Halmaz: tehát nem számít a sorrend, valamint egy elem csak egyszer szerepelhet.

Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Sör (név, ország).

attribútumok

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

sorok

Relációs adatmodell II.

- Egy-egy reláció soroknak egy halmaza.
- Halmaz: tehát nem számít a sorrend, valamint egy elem csak egyszer szerepelhet.
- A reláció sorainak halmazát előfordulásnak nevezzük.
- $\rho \subset X_1 \times \ldots \times X_n$ esetén az attribútumok értékkészlete adja az X_i alaphalmazokat ($1 \leq i \leq n$), egy-egy előfordulás pedig egy-egy relációnak „feleltethető meg”.
- (A $X_1 \times \ldots \times X_n$ Descartes-szorzatnak az elemeinél az első érték az X_1 alaphalmazból jön, a második X_2 alaphalmazból stb. az összes lehetséges módon.)
- Az attribútumok sorrendje láttuk, hogy nem rögzített a relációsémában. Egy-egy előfordulás ábrázolása esetén viszont rögzítésre kerül.
- A lényeg, hogy a soroknál egy attribútum értékét az attribútum nevével azonosítjuk, és nem pedig azzal, hogy hányadik helyen szerepel a soron belül.

Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Sör (név, ország).

attribútumok

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

előfordulás

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

sorok

Relációs adatmodell III.

- Az adatbázis tulajdonképpen relációk halmaza. A megfelelő relációsémák halmaza adja az adatbázissémát, a hozzá tartozó előfordulások pedig az adatbázis-előfordulást.
- Egy sor elemeit mezőnek (komponens) nevezzük. Minden mező csak atomi értéket vehet fel. Léteznek bonyolultabb adatmodellek is, ahol egy mező értéke lehet halmaz, lista, tömb, rekord, referencia stb.
- Megjegyzés: a gyakorlatban sokszor megengedik a sorok ismétlődését, hiszen az ismétlődések megszüntetése nagyon időigényes.

Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Sör (név, ország).

- Az adatbázis sémája:

Sör (név, ország),

Bár (név, város, tulaj),

Felszolgál (sör, bár, ár).

attribútumok

Sör

név	ország
Soproni	Magyar
Kozel	Cseh
Dreher	Német

Bár

név	város	tulaj
Makk 7-es	Budapest	Géza
Lórúgás	Győr	Ica

előfordulás

Felszolgál

sör	bár	ár
Kozel	Makk 7-es	320
Dreher	Makk 7-es	400
Soproni	Lórúgás	280

sorok

Mire kell odafigyelni?

Mivel attribútumok halmazáról van szó, a Példa 1 és Példa 2 relációk nevüktől eltekintve azonosak.

Példa 1

A	B	C
a	b	c
d	a	a
c	b	d

Példa 2

B	C	A
b	c	a
a	a	d
b	d	c

Mivel sorok halmazáról van szó, a Példa 1 és Példa 3 relációk nevüktől eltekintve azonosak.

Példa 3

A	B	C
c	b	d
d	a	a
a	b	c

Példa 4

A	B	C
c	b	d
c	b	d
a	b	c

Ebben a modellben a Példa 4 nem reláció.

Példa megszorításra

- Az attribútumok egy halmaza egy kulcsot alkot egy relációra nézve, ha a reláció bármely előfordulásában nincs két olyan sor, amelyek a kulcs összes attribútumának értékein megegyeznének.
- Ilyen egy attribútumú kulcs például a személyi igazolvány-szám vagy a TAJ szám.
- Megjegyzés: egy kulcs nem feltétlenül egy attribútumból áll. Például a bár táblában valószínűleg az a jó, ha a név és a város együtt alkotják a kulcsot.
- A kulcsot aláhúzás jelöli:

  Bár (<u>név</u>, <u>város</u>, tulaj).

Vigyázat!

Ennél a konkrét előfordulásnál választhatnánk a nevet kulcsnak, sok esetben viszont ez nem megfelelő, hiszen több különböző ember is él ugyanazzal a névvel.

név	**telefon**
Grasshaus Ignác	20-234-4567
Menyhért Lipót	20-564-2345
Bereg Anna	20-345-1231

Feladat

- Hány különböző módon reprezentálható egy reláció-előfordulás (az attribútumok és sorok sorrendjét figyelembe véve), ha az előfordulásnak 4 attribútuma és 5 sora van?

Mit nevezünk algebrának?

- Egy algebra általában műveleteket és atomi operandusokat tartalmaz.
- Az algebra lehetővé teszi kifejezések megfogalmazását az atomi operandusokon és az algebrai kifejezéseken végzett műveletek alkalmazásával kapott relációkon.
- Fontos tehát, hogy minden művelet végeredménye reláció, amelyen további műveletek adhatók meg.
- A relációs algebra atomi operandusai a következők:
  - a relációkhoz reprezentáló változók
  - konstansok, amelyek véges relációt fejeznek ki

Relációs algebra (műveletek) I.

- Projekció (vetítés). Adott relációt vetít le az alsó indexben szereplő attribútumokra. Példa: $\Pi_{A,\ B}(R)$

A	B	C
a	b	c
c	d	e
g	a	d

→

A	B
a	b
c	d
g	a

Relációs algebra (műveletek) II.

- Szelekció (kiválasztás). Kiválasztja az argumentumban szereplő reláció azon sorait, amelyek eleget tesznek az alsó indexben szereplő feltételnek.
- $R(A_1, \ldots, A_n)$ reláció esetén a σ_F kiválasztás F feltétele a következőképpen épül fel:
  - atomi feltétel: $A_i \, \theta \, A_j$, $A_i \, \theta$ c, ahol c konstans, $\theta \in \{=, <, >\}$,
  - ha B_1, B_2 feltételek, akkor $\neg B_1$, $B_1 \wedge B_2$, $B_1 \vee B_2$ is feltételek.
  - $\neg$: logikai NEM, $\wedge$: logikai ÉS, $\vee$: logikai VAGY
- Példa: $\sigma_{A=a \vee C=d}(R)$ (a $\neq$, $\leq$, $\geq$ műveleteket ezentúl értelemszerűen használjuk)

A	B	C
a	b	c
c	d	e
g	a	d

A	B	C
a	b	c
g	a	d

Relációs adatmodell (műveletek) III.

- Mivel sorok halmazáról van szó, így értelmezhetők a szokásos halmazműveletek: az unió, a metszet és a különbség. A műveletek alkalmazásának feltétele, hogy az operandusok attribútumai megegyezzenek és azonos sorrendben szerepeljenek. Példa: R – S:

R

A	B	C
a	b	c
c	d	e
g	a	d

S

A	B	C
a	b	c
c	d	e
g	d	f

$\Rightarrow$

A	B	C
g	a	d

Relációs algebra (műveletek) IV.

- A Descartes-szorzat is értelmezhető. Itt természetesen nem fontos az attribútumok egyenlősége. A két vagy több reláció azonos nevű attribútumait azonban meg kell különböztetni egymástól. Példa: R $\times$ S.

R

A	B	C
a	b	c
c	d	e
g	a	d

S

B	D
b	r
q	s

$\Rightarrow$

A	R.B	C	S.B	D
a	b	c	b	r
a	b	c	q	s
c	d	e	b	r
c	d	e	q	s
g	a	d	b	r
g	a	d	q	s

Relációs algebra (műveletek) V.

- Egyes esetekben szükség lehet egy adott reláció attribútumainak átnevezésére. A $\rho_{S(C,\ D,\ E)}(R)$ az R(A, B, C) reláció helyett veszi az S relációt, melynek sorai megegyeznek R soraival, az attribútumai pedig rendre C, D, E.
- Ha az attribútumokat nem szeretnénk átnevezni, csak a relációt, ezt $\rho_S(R)$-rel jelöljük.

Théta-összekapcsolás I.

- A gyakorlatban szinte kizárólag valamilyen összekapcsolásra visszavezethető műveletet használnak abban az esetben, amikor a lekérdezés megválaszolásához több táblából kell kigyűjteni az adatokat.
- Théta-összekapcsolás: $R(A_1,\ldots,A_n)$, $S(B_1,\ldots,B_m)$ sémájú táblák esetén:

  $R \;|X|_F\, S = \sigma_F\,(R \times S)$ teljesül, itt F
  - elemi feltétel $A_i\ \Theta\ B_j$, $A_i\ \Theta\ c$, ahol $\Theta \in \{=, <, >\}$ és c konstans,
  - vagy összetett feltétel, azaz: ha B_1, B_2 feltétel, akkor $\neg B_1$, $B_1 \wedge B_2$, $B_1 \vee B_2$ is feltétel.

Théta-összekapcsolás II.

A	B	C
a	b	c
d	d	g
e	f	r

$|X|_{A = D}$

D	E
b	e
d	r

A	B	C	D	E
d	d	g	d	r

- Egyen-összekapcsolás (equi join): ha a théta-összekapcsolásban a Θ helyén = szerepel.

Természetes összekapcsolás

- Természetes összekapcsolás: $R(A_1,\ldots,A_n)$, $S(B_1,\ldots,B_m)$ sémájú táblák esetén R |X| S azon sorpárokat tartalmazza R-ből illetve S-ből, amelyek R és S azonos attribútumain megegyeznek.
- A természetes összekapcsolás asszociatív, azaz:

  $(R_1 \ |X|\ R_2)\ |X|\ R_3 = R_1\ |X|\ (R_2\ |X|\ R_3)$, és kommutatív, azaz :

  $R_1\ |X|\ R_2 = R_2\ |X|\ R_1$.

A	B	C
a	b	c
d	d	g
e	f	r

|X|

B	D
b	e
d	r

A	B	C	D
a	b	c	e
d	d	g	r

Miért olyan gyakori?

Felszolgál

kocsma	sör
Makk 7-es	Dreher
Lórúgás	Kozel
Lórúgás	Gösser

|X|

Látogat

név	kocsma
Péter	Makk 7-es
Feri	Lórúgás

kocsma	sör	név
Makk 7-es	Dreher	Péter
Lórúgás	Kozel	Feri
Lórúgás	Gösser	Feri

A természetes összekapcsolás kifejezhető a többi alapművelettel:

$R \; |X| \; S \equiv \pi_L(\sigma_C \; (R \times S))$,

itt: C a közös attribútumok egyenlőségét írja elő, L pedig csak egyszer veszi a közös attribútumokat.

Relációs adatmodell - példák

- R |X| S ugyanazt jelöli, mint $R \bowtie S$

A	B
a1	b
a2	b
a3	b

|X|

B	C
b	c1
b	c2

Relációs adatmodell - példák

- R |X| S ugyanazt jelöli, mint $R \bowtie S$

A	B
a1	b
a2	b
a3	b

|X|

B	C
b	c1
b	c2

➡

A	B	C
a1	b	c1
a1	b	c2
a2	b	c1
a2	b	c2
a3	b	c1
a3	b	c2

Relációs adatmodell - példák

- R |X| S ugyanazt jelöli, mint $R \bowtie S$

A	B
a1	b
a2	b
a3	b

|X|

B	C
b	c1
b	c2

→

A	B	C
a1	b	c1
a1	b	c2
a2	b	c1
a2	b	c2
a3	b	c1
a3	b	c2

3*2 sor

Relációs adatmodell - példák

A	B
a1	b1
a2	b1
a3	b1

|X|

B	C
b2	c1
b2	c2

Relációs adatmodell - példák

A	B
a1	b1
a2	b1
a3	b1

|X|

B	C
b2	c1
b2	c2

→

A	B	C
<üres>		

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d

$\sigma_{A=a\ \vee\ C=d}(R)$

A	B	C
a	b	c
g	a	d

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d

$\sigma_{A=a \ \vee \ C=d}(R)$

A	B	C
a	b	c
g	a	d

S

D	E
s	t
u	v

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d

$\sigma_{A=a\ \vee\ C=d}(R)$

A	B	C
a	b	c
g	a	d

S

D	E
s	t
u	v

$\sigma_{A=a\ \vee\ C=d}(R) \times S$

A	B	C	D	E
a	b	c	s	t
g	a	d	s	t
a	b	c	u	v
g	a	d	u	v

Relációs adatmodell - példák

- A műveletek kifejezhetők a többi alapművelettel
- Például:
- Természetes összekapcsolás: R |X| S $\equiv \pi_L(\sigma_C$ (R $\times$ S)), itt: C a közös attribútumok egyenlőségét írja elő, L pedig csak egyszer veszi a közös attribútumokat.
- Théta-összekapcsolás: R $|X|_F$ S = σ_F (R $\times$ S) teljesül, itt F valamilyen feltétel

A	B	C
a	b	c
d	d	g
e	f	r

$|X|_{A = D}$

D	E
b	e
d	r

A	B	C	D	E
d	d	g	d	r

Relációs adatmodell - példák

- Mivel sorok halmazáról van szó, így értelmezhetők a szokásos halmazműveletek: az unió, a metszet és a különbség. A műveletek alkalmazásának feltétele, hogy az operandusok attribútumai megegyezzenek és azonos sorrendben szerepeljenek.

R

A	B	C	D
a	b	c	t
c	d	e	v
g	a	d	u

$\pi_{A,B,C}(R)$

A	B	C
a	b	c
c	d	e
g	a	d

S

A	B	C
a	b	c
t	u	v
c	d	e
g	a	d

$\pi_{A,B,C}(R) - S$

A	B	C
<üres>		

Relációs adatmodell - példák

Egy relációkon értelmezett operátor akkor **monoton**, ha bármelyik argumentumrelációhoz egy újabb sort hozzávéve az eredmény tartalmazza az összes olyan sort, amelyet addig tartalmazott és esetleg újabb sorokat is.

$\pi_{A,B,C}(R)$

A	B	C
a	b	c
c	d	e
g	a	d

S

A	B	C
t	u	v

$\pi_{A,B,C}(R) - S$

A	B	C
a	b	c
c	d	e
g	a	d

Relációs adatmodell - példák

$\pi_{A,B,C}(R)$

A	B	C
a	b	c
c	d	e
g	a	d

S

A	B	C
t	u	v
g	a	d

$\pi_{A,B,C}(R) - S$

A	B	C
a	b	c
c	d	e

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d

S

A	B	C
t	u	v
g	a	d

$R \cap S$

A	B	C
g	a	d

Relációs adatmodell - példák

R

A	B	C
a	b	c
c	d	e
g	a	d
t	u	v

S

A	B	C
t	u	v
g	a	d

$R \cap S$

A	B	C
g	a	d
t	u	v

Relációkra vonatkozó megszorítások

- A megszorításokat kétféleképpen fejezhetjük ki (legyenek R és S relációs algebrai kifejezések):
  - $R = \emptyset$, azaz R-nek üresnek kell lennie,
  - $R \subseteq S$, azaz R eredményének minden sorának benne kell lennie S eredményében.
- A két megszorítás kifejezőerő szempontjából azonos:
  - $R \subseteq S$ így is kifejezhető: $R - S \subseteq \emptyset$,
  - míg $R = \emptyset$, $R \subseteq \emptyset$ alakban is írható.

Hivatkozási épség megszorítás

- Hivatkozási épség megszorítás: ha egy érték megjelenik valahol egy környezetben, akkor ugyanez az érték egy másik, az előzővel összefüggő környezetben is meg kell, hogy jelenjen.
- Példa: a Sör(név, ország), Felszolgál(sör, bár, ár) táblák esetén megköveteljük, hogy csak olyan sörök szerepeljenek a Felszolgál táblában, amelyek a Sör táblában is szerepelnek.
- A megszorítás: $\Pi_{\text{sör}}(\text{Felszolgál}) \subseteq \Pi_{\text{név}}(\text{Sör})$.
- Általában: $\Pi_A(R) \subseteq \Pi_B(S)$.

Kulcs és egyéb megszorítások

- Példa: a Bár(név, város, tulaj) relációban a (név, város) attribútumhalmaz kulcs.

  $\sigma_{B1.név=B2.név \wedge B1.város=B2.város \wedge B1.tulaj \neq B2.tulaj}(B_1 \times B_2) = \emptyset$

- Tegyük fel, hogy csak a budapesti vagy madridi bárokkal szeretnénk foglalkozni. Ennek kifejezése:

  $\sigma_{(város \neq 'Budapest') \wedge (város \neq 'Madrid')}(B) = \emptyset$.