

# Adatbázisok 1.

## Relációs adatmodell

# Mi is az az adatmodell?

- Az adatmodell (*data model*) információ vagy adatok leírására szolgáló jelölés. A leírás részei:
  - az adatok struktúrája (*data structure*).
  - Az adatokon végezhető műveletek (*operations*). 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 (*constraints*). Pl. egy személyigazolvány-számhoz nem tartozhat két különböző személy.
  - Legfontosabb adatmodellek: relációs és féligstrukturált (XML).

# Relációs adatmodell I.

- A relációs adatmodellben (*relational model*) az adatokat kétdimenziós táblákban (*tables*), relációkban (*relations*) tároljuk.

# Példa a fogalmak illusztrálására

## Tea

név	ország
Yellow label	Skócia
Ice Tea	Svájc
Gyógytea	Írország

## Teázó

név	város	tulaj
Makk 7-es	Budapest	Géza
Zöld Teaház	Győr	Ica

## Felhasználó

tea	teázó	ár
Ice Tea	Makk 7-es	920
Yellow label	Makk 7-es	950
Gyógytea	Zöld Teaház	450

# Relációs adatmodell I.

- A relációs adatmodellben az adatokat kétdimenziós táblákban (relációkban) tároljuk.
- A reláció fejrésében található az **attribútumok** (*attributes*).
- Minden attribútumhoz hozzátartozik egy **értékkészlet**.

# Példa a fogalmak illusztrálására



# Relációs adatmodell I.

- A relációs adatmodellben az adatokat kétdimenziós táblákban (relációkban) tároljuk.
- A reláció fejrészében található az attribútumok.
- Minden attribútumhoz hozzátartozik egy értékkészlet.
- A reláció neve és a reláció-attribútumok halmaza együtt alkotják a **relációsémát** (*relation schema*).
- 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.:

Tea (név, ország).

attribútumok

Tea

név	ország
Yellow label	Skócia
Ice Tea	Svájc
Gyógytea	Írország

Teázó

név	város	tulaj
Makk 7-es	Budapest	Géza
Zöld Teaház	Győr	Ica

Felhasználó

tea	teázó	ár
Ice Tea	Makk 7-es	920
Yellow label	Makk 7-es	950
Gyógytea	Zöld Teaház	450



# Relációs adatmodell II.

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

# Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Tea (név, ország).

attribútumok

**Tea**

név	ország
Yellow label	Skócia
Ice Tea	Svájc
Gyógytea	Írország

**Teázó**

név	város	tulaj
Makk 7-es	Budapest	Géza
Zöld Teaház	Győr	Ica

**Felhasználó**

tea	teázó	ár
Ice Tea	Makk 7-es	920
Yellow label	Makk 7-es	950
Gyógytea	Zöld Teaház	450

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** (*relation instance*) nevezzük.
- $\rho \subset X_1 \times \dots \times X_n$  esetén az attribútumok értékkészlete adja az  $X_i$  alaphalmazokat ( $1 \leq i \leq n$ ), egy-egy előfordulás pedig egy-egy (mat.) relációnak „feleltethető meg”.
- (A  $X_1 \times \dots \times X_n$  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.:

Tea (név, ország).

attribútumok

**Tea**

név	ország
Yellow label	Skócia
Ice Tea	Svájc
Gyógytea	Írország

**Teázó**

név	város	tulaj
Makk 7-es	Budapest	Géza
Zöld Teaház	Győr	Ica

előfordulás

**Felhasználó**

tea	teázó	ár
Ice Tea	Makk 7-es	920
Yellow label	Makk 7-es	950
Gyógytea	Zöld Teaház	450

sorok

# Relációs adatmodell III.

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

# Példa a fogalmak illusztrálására

- Egy reláció sémája, pl.:

Tea (név, ország).

- Az adatbázis sémája:

Tea (név, ország),

Teázó (név, város, tulaj),

Felhasználó (tea, teázó, ár).

attribútumok

Tea

név	ország
Yellow label	Skócia
Ice Tea	Svájc
Gyógytea	Írország

Teázó

név	város	tulaj
Makk 7-es	Budapest	Géza
Zöld Teaház	Győr	Ica

előfordulás

Felhasználó

tea	teázó	ár
Ice Tea	Makk 7-es	920
Yellow label	Makk 7-es	950
Gyógytea	Zöld Teaház	450

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

**Példa 3**

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

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

**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** (*key*) 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 teázó táblában valószínűleg az a jó, ha a név és a város együtt alkotják a kulcsot.
- A kulcsot aláhúzás jelöli:  
Teázó (név, város, 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** (*operators*) és **atomi operandusokat** (*operands*) tartalmaz.
- A **relációs algebra** (*relational algebra*) lehetővé teszi **kifejezések** (*expressions*) 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** (*variables*)
  - **konstansok** (*constants*), amelyek véges relációt fejeznek ki

# Relációs algebra (műveletek) I.

- **Projekció** (vetítés) (*projection*). 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) (*selection*). Kiválasztja az argumentumban szereplő reláció azon sorait, amelyek eleget tesznek az alsó indexben szereplő feltételnek.
- $R(A_1, \dots, A_n)$  reláció esetén a  $\sigma_F$  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

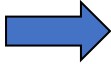


A	B	C
a	b	c
g	a	d

## Relációs adatmodell (műveletek) III.

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

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



A	B	C
g	a	d

## Relációs algebra (műveletek) IV.

- A **Descartes-szorzat** (*product*) 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			S						
A	B	C	B	D					
a	b	c	b	r					
c	d	e	q	s					
g	a	d							
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** (*renaming*). 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.



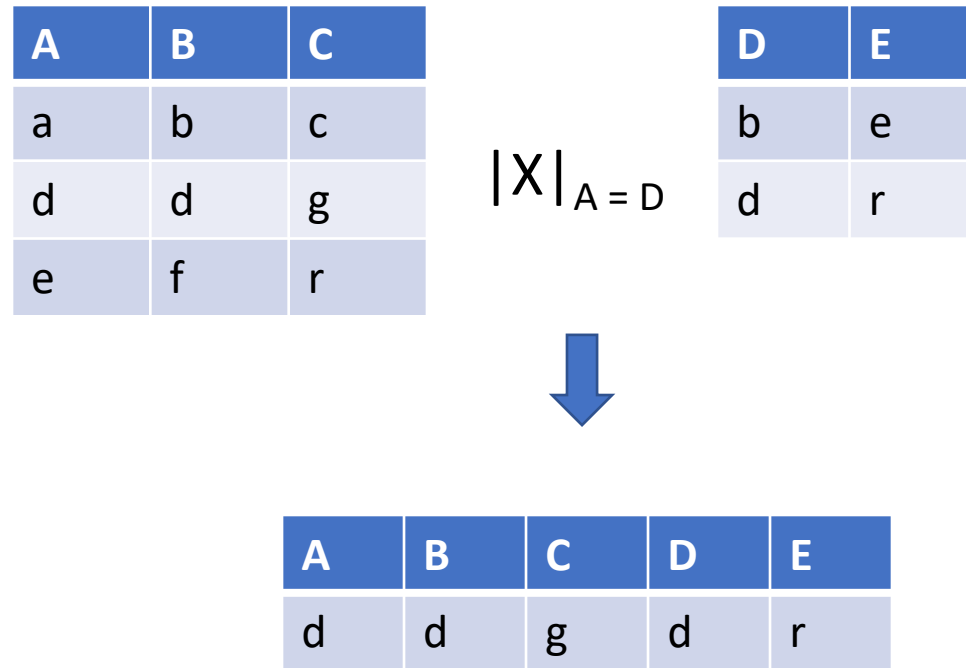
# Théta-összekapcsolás I.

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

$R \bowtie_F S = \sigma_F (R \times S)$  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.



- Egyen-összekapcsolás (*equi join*): ha a théta-összekapcsolásban a  $\Theta$  helyén = szerepel.


# Természetes összekapcsolás

- Természetes összekapcsolás (*natural join*):  $R(A_1, \dots, A_n)$ ,  $S(B_1, \dots, B_m)$  sémájú táblák esetén  $R \bowtie S$  azon sorpárokat tartalmazza R-ből illetve S-ből, amelyek R és S azonos attribútumain megegyeznek.
- A természetes összekapcsolás *asszociatív*, azaz:  
 $(R_1 \bowtie R_2) \bowtie R_3 = R_1 \bowtie (R_2 \bowtie R_3)$ , és *kommutatív*, azaz :  
 $R_1 \bowtie R_2 = R_2 \bowtie R_1$ .

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

 $\bowtie$ 

B	D
b	e
d	r



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

# Miért olyan gyakori?

Felszolgál

teázó	tea
Makk 7-es	Yellow label
Zöld Teaház	Ice Tea
Zöld Teaház	Gyógytea

Látogat

név	teázó
Péter	Makk 7-es
Feri	Zöld Teaház

|X|



teázó	tea	név
Makk 7-es	Yellow label	Péter
Zöld Teaház	Ice Tea	Feri
Zöld Teaház	Gyógytea	Feri

A természetes összekapcsolás kifejezhető a többi alpművelettel:

$$R \bowtie S \equiv \pi_L(\sigma_C(R \times S)),$$

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

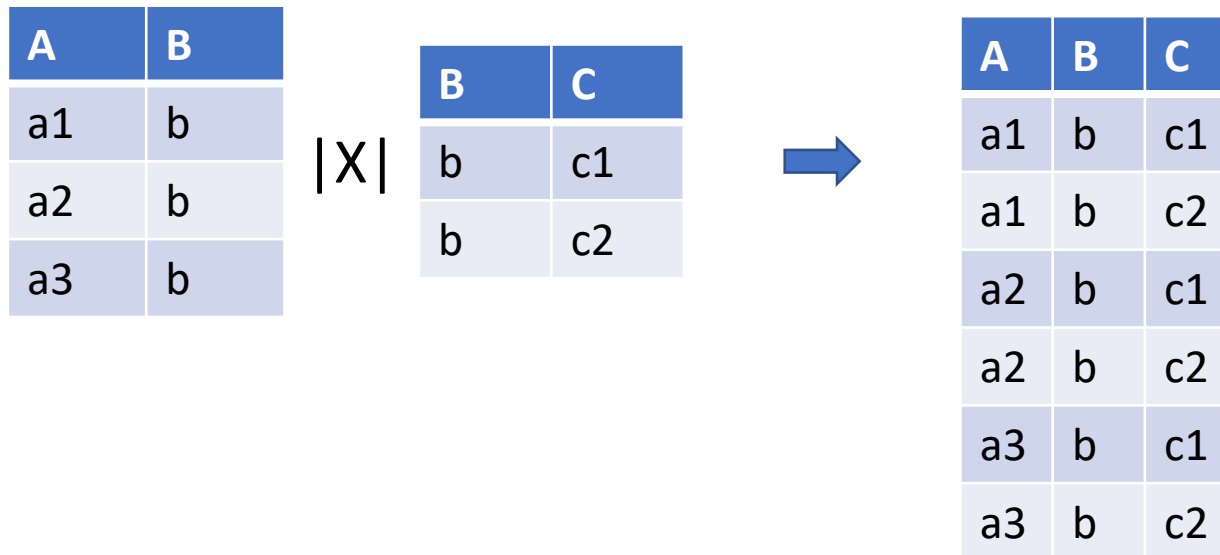
# Relációs adatmodell - példák

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

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

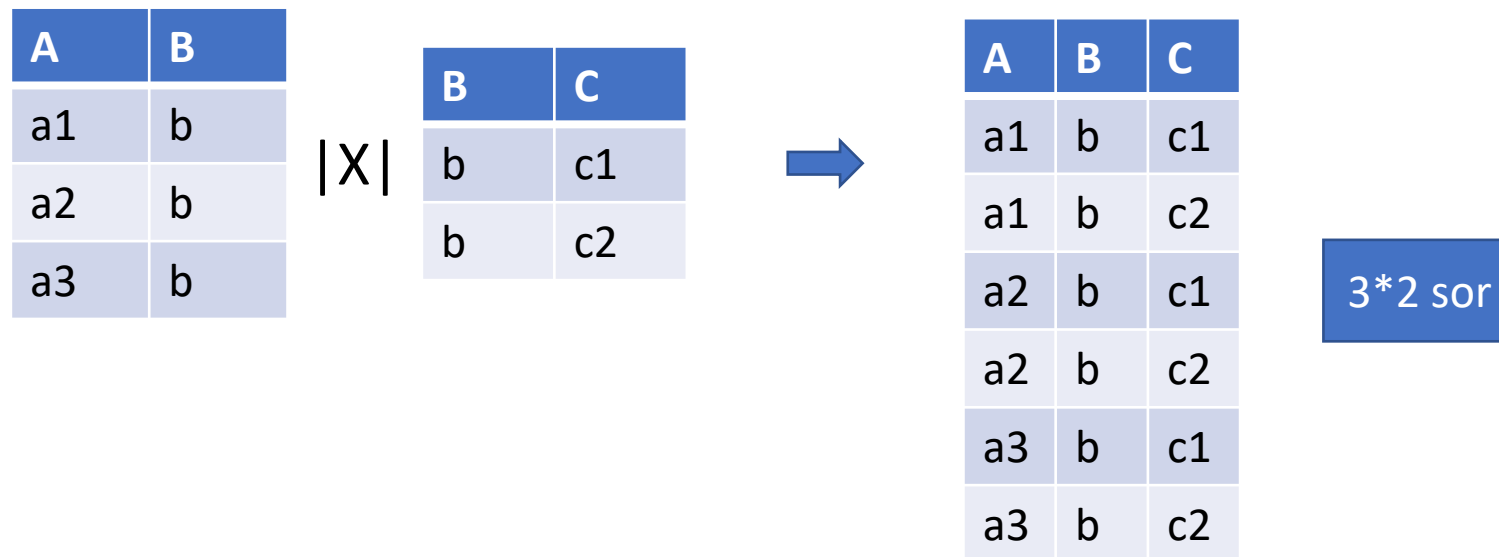
# Relációs adatmodell - példák

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



# Relációs adatmodell - példák

- $R \bowtie S$  ugyanazt jelöli, mint  $R \times S$



# Relációs adatmodell - példák

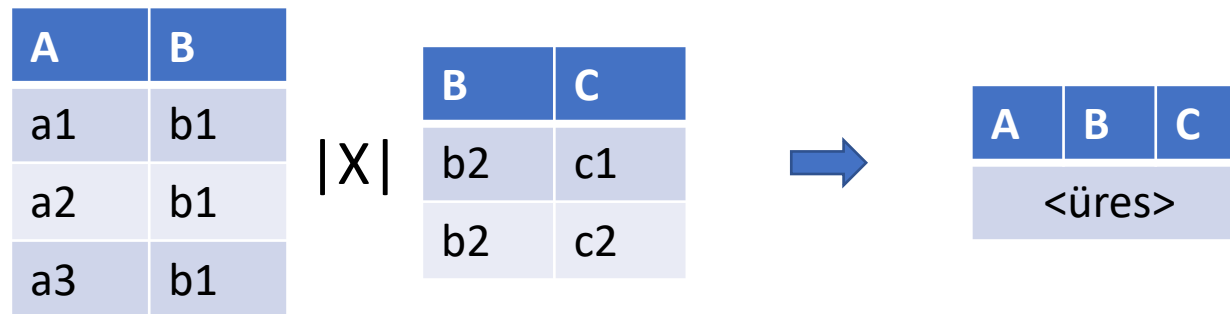
A	B
a1	b1
a2	b1
a3	b1

|X|

B	C
b2	c1
b2	c2



# Relációs adatmodell - példák



# Relációs adatmodell - példák

$R$

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

# Relációs adatmodell - példa

$R$				$\sigma_{A=a \vee C=d}(R)$		
A	B	C		A	B	C
a	b	c		a	b	c
c	d	e		g	a	d
g	a	d				



# Relációs adatmodell - példák

$R$

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



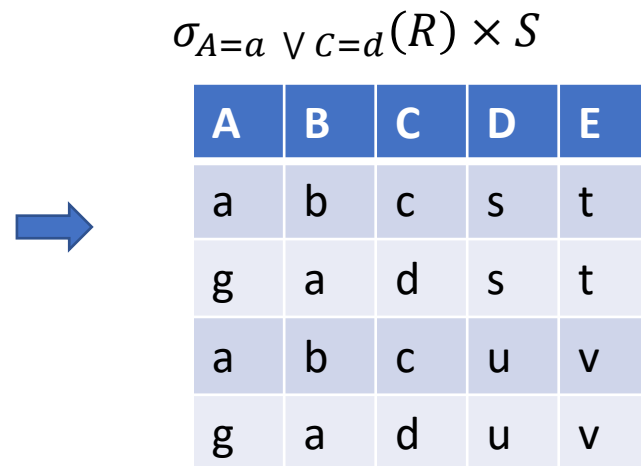
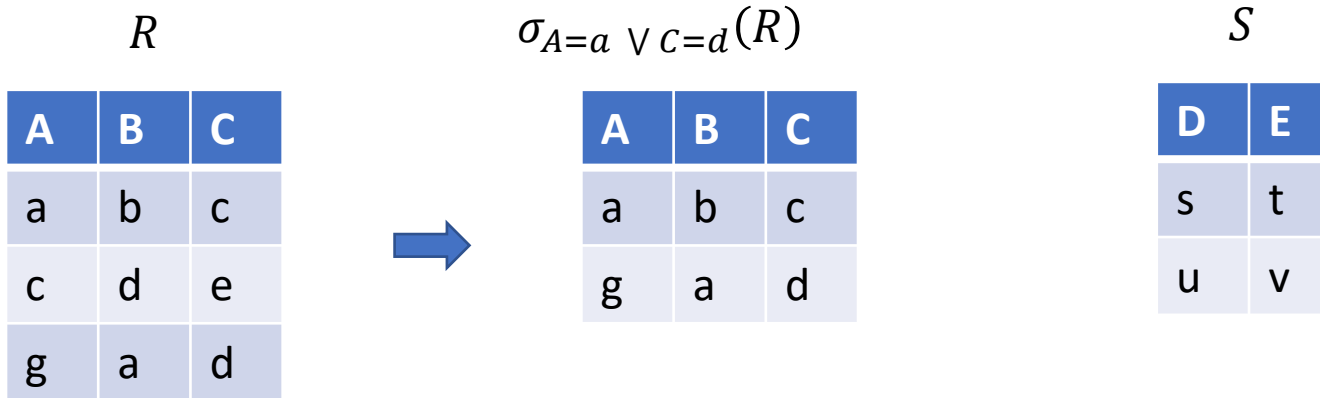
$$\sigma_{A=a \vee C=d}(R)$$

A	B	C
a	b	c
g	a	d

$S$

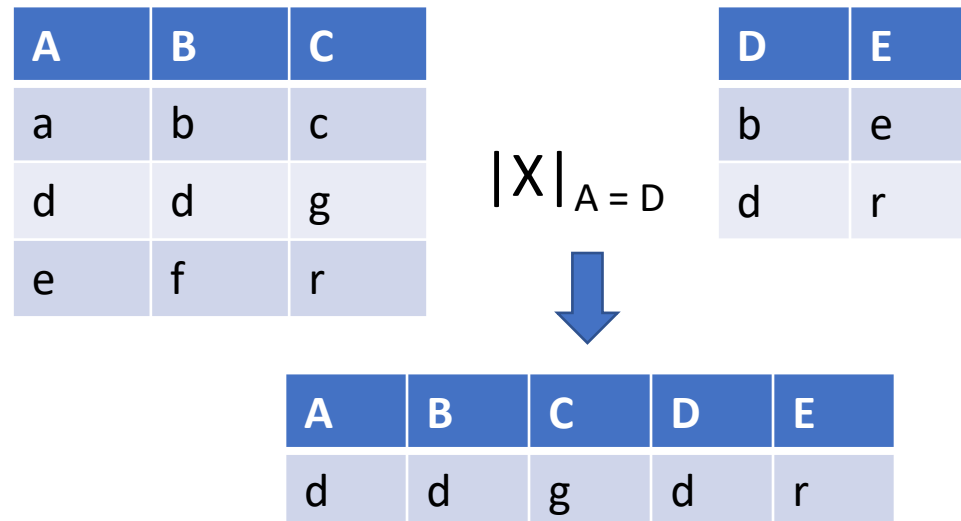
D	E
s	t
u	v

# Relációs adatmodell - példák



# Relációs adatmodell - példák

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



# Relációs adatmodell - példák

- Mivel sorok halmazáról van szó, így értelmezhetők a szokásos halmazműveletek: az **unió**, a **metset** 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** (*monotone*), 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

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

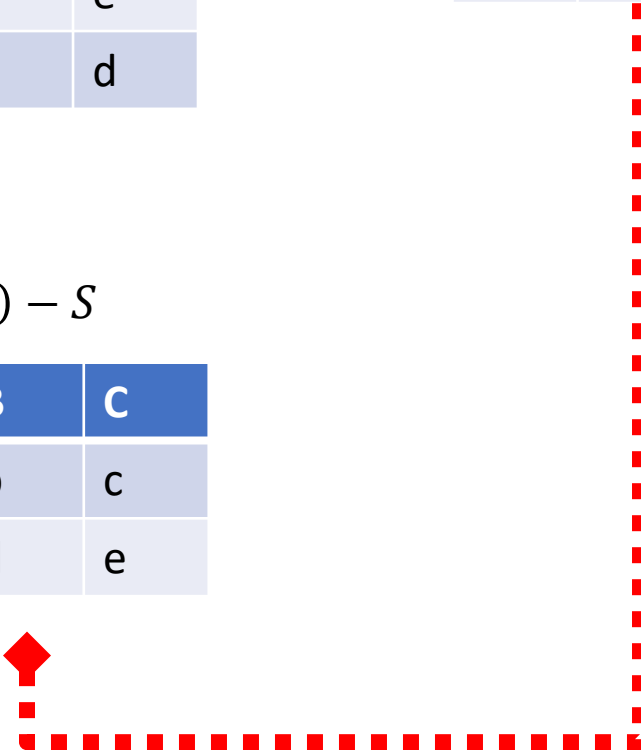
A	B	C
a	b	c
c	d	e
g	a	d

$S$

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 (*referential integrity constraint*): ha egy érték megjelenik valahol egy környezetben, akkor ugyanez az érték egy másik, az előzővel összefüggő környezetben is meg kell, hogy jelenjen.
- Példa: a Tea(név, ország), Felszolgál(tea, teázó, ár) táblák esetén megköveteljük, hogy csak olyan teák szerepeljenek a Felszolgál táblában, amelyek a Tea táblában is szerepelnek.
- A megszorítás:  $\Pi_{\text{tea}}(\text{Felszolgál}) \subseteq \Pi_{\text{név}}(\text{Tea})$ .
- Általában:  $\Pi_A(R) \subseteq \Pi_B(S)$ .

# Kulcs és egyéb megszorítások

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

$$\sigma_{T1.név=T2.név \wedge T1.város=T2.város \wedge T1.tulaj \neq T2.tulaj}(T_1 \times T_2) = \emptyset$$

(ahol  $T_1$  a  $\rho_{T1}(\text{Teázó})$ ,  $T_2$  a  $\rho_{T2}(\text{Teázó})$  rövidítése)

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

$$\sigma_{(város \neq 'Budapest') \wedge (város \neq 'Madrid')}(T) = \emptyset.$$

(ahol  $T$  a  $\rho_T(\text{Teázó})$  rövidítése)