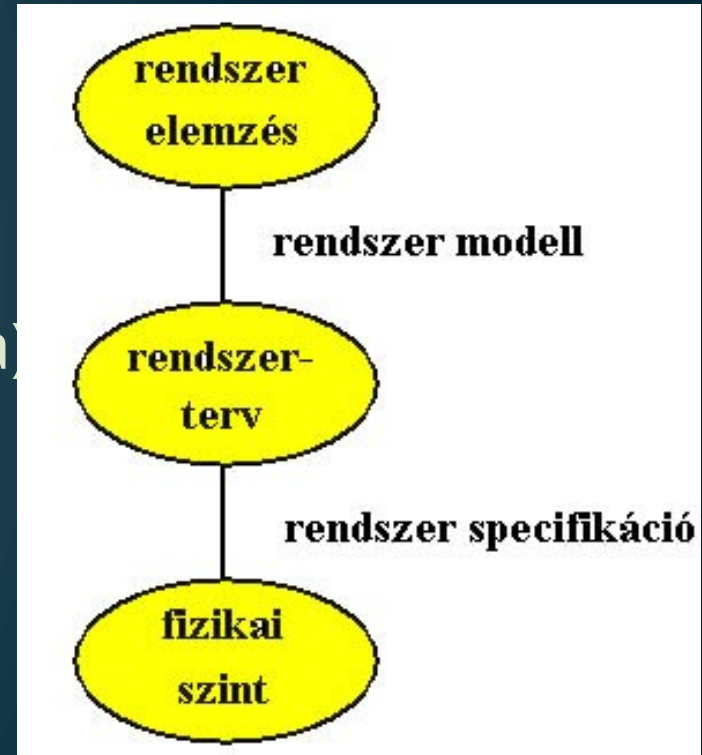


# Normalizálás

Gugolya László

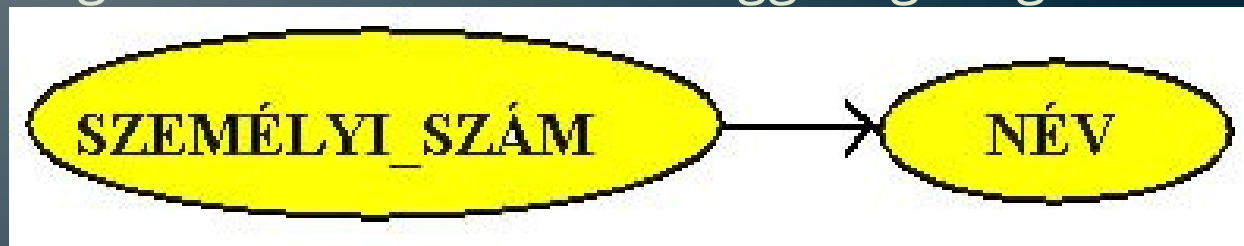


# Bevezető fogalmak

- funkcionális függőség
- reláció kulcs
- redundancia

# Adatok közötti funkcionális kapcsolat

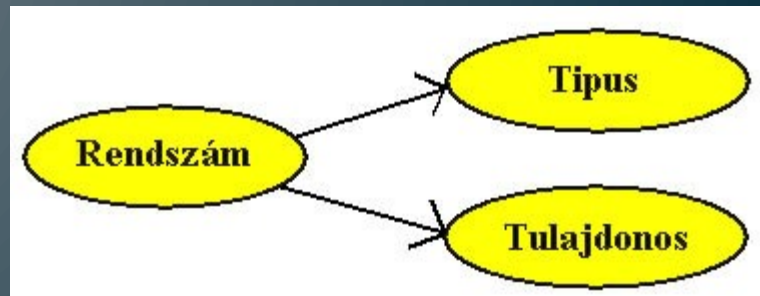
- Adatok között akkor áll fenn funkcionális kapcsolat, ha egy vagy több adat konkrét értékéből más adatok egyértelműen következnek.
- Például a személyi szám és a név között funkcionális kapcsolat áll fenn, mivel minden embernek különböző személyi száma van.
- Ezt a **SZEMÉLYI\_SZÁM** -> **NÉV** kifejezéssel jelöljük vagy pedig egy diagrammal. Funkcionális függőség diagram:



- A funkcionális függőség bal oldalát a függőség meghatározójának nevezzük

# Adatok közötti funkcionális kapcsolat

- Az adatok közötti funkcionális függőségek az adatok természetéből következnek. A tervezés során nagyon fontos.
- A funkcionális függőség jobb oldalán több attribútum is állhat: **AUTÓ\_RENDSZÁM -> TIPUS, TULAJDONOS**

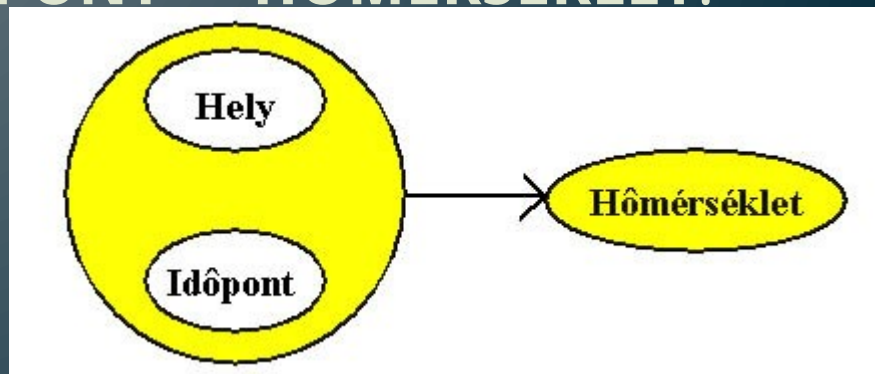


- Előfordulhat, hogy két attribútum kölcsönösen függ egymástól: **FÉRJ\_SZEM\_SZÁMA <-> FELESÉG\_SZEM\_SZÁMA**.

# Adatok közötti funkcionális kapcsolat

- A funkcionális függőség bal oldalán több attribútum is megjelenhet, melyek együttesen határozzák meg a jobb oldalon szereplő attribútum értékét.
- Például hőmérsékletet mérünk különböző helyeken és időben úgy, hogy a helyszínek között azonosak is lehetnek:

**HELY, IDŐPONT -> HŐMÉRSÉKLET.**



# Adatok közötti funkcionális kapcsolat

- A funkcionális függőségek speciális esete a teljes funkcionális függőség.
- Erről akkor beszélhetünk, ha a meghatározó oldalon nincsen felesleges attribútum. Például a **RENDSZÁM, TÍPUS -> SZÍN** funkcionális függőség nem teljes funkcionális függőség, mivel a rendszám már egyértelműen meghatározza a kocsí színét, ehhez nincs szükség a típusra is.

# Adatok közötti funkcionális kapcsolat

- A funkcionális függőség bevezetése után a relációk egy másik, matematikai jelölése:

**reláció\_név=({attribútumok},{funkcionális függőségek listája})**

- Például:

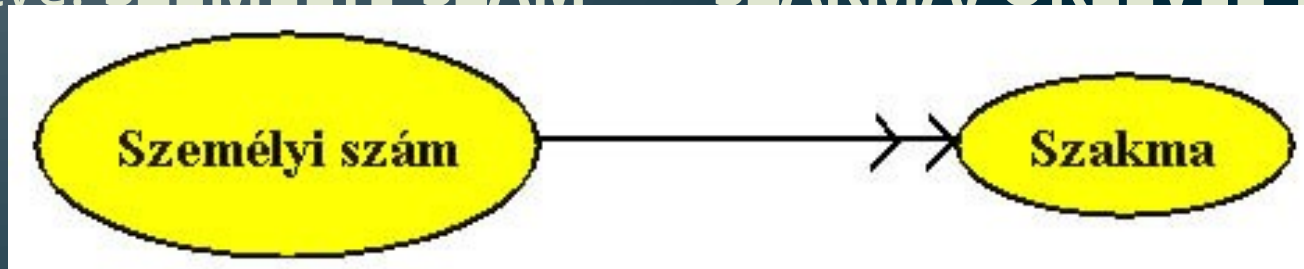
Személyi szám	Név	Munkahely

SZEMÉLYEK=({SZEMÉLYI\_SZÁM, NÉV, MUNKAHELY},  
{SZEMÉLYI\_SZÁM -> NÉV, SZEMÉLYI\_SZÁM -> MUNKAHELY})  
írható le a funkcionális függőségekkel együtt, feltételezve, hogy  
mindenkinek csak egy munkahelye van.



# Adatok közötti többértékű függőség

- Az adatok között fennálló kapcsolatok közül nem mindegyik fejezhető ki a funkcionális függőség segítségével. Például minden embernek lehet több szakmája, illetve ugyanazzal a szakmával több ember is rendelkezhet. Ebben az esetben egyik irányban sincs egyértelmű függőség.
- Ez egy többértékű függőség, az egyik attribútumhoz egy másik attribútum csoportja, halmaza kapcsolódik:  
**SZEMÉLYI\_SZÁM ->> SZAKMA**  
bővítve: **SZEMÉLYI\_SZÁM ->> SZAKMA, OKLEVÉL\_KELTE**



# Adatok közötti többértékű függőség

- A funkcionális és a többértékű függőség között kapcsolat van. Nagyon gyakran ugyanazt a függőségi kapcsolatot kifejezhetjük funkcionális és többértékű függőséggel is.
- Egy üzemben különböző termékeket gyártanak, melyek mindegyike többfajta alkatrészből tevődik össze. Szeretnénk nyilvántartani termékenként a felhasznált alkatrészek mennyiségét. Ezt leírhatjuk funkcionális függőség segítségével:  
**TERMÉK\_AZONOSÍTÓ,ALKATRÉSZ\_AZONOSÍTÓ -> MENNYISÉG,**  
többértékű függőséggel is:  
**TERMÉK\_AZONOSÍTÓ ->> ALKATRÉSZ\_AZONOSÍTÓ, MENNYISÉG.**
- A funkcionális függőségeket mindig előnyben kell részesíteni a többértékű függőséggel szemben.

# Reláció kulcs fogalma

- A reláció kulcs a reláció egy sorát azonosítja egyértelműen.
- A reláció - definíció szerint- nem tartalmazhat két azonos sort, ezért minden relációban létezik kulcs
- A reláció kulcsnak a következő feltételeket kell teljesítenie
  - az attribútumok egy olyan csoportja, melyek csak egy sort azonosítanak (egyértelműség)
  - a kulcsban szereplő attribútumok egyetlen részhalmaza sem alkot kulcsot
  - a kulcsban szereplő attribútumok értéke nem lehet definiálatlan (NULL)

# Reláció kulcs fogalma

- SZEMÉLY\_ADATOK=({ SZEMÉLYI\_SZÁM, SZÜL\_ÉV, NÉV}).

Diák		
Személyi szám	Születési év	Név

- NAPLÓ=({SZEMÉLYI\_SZÁM, TANTÁRGY, DÁTUM, OSZTÁLYZAT})

Napló			
Személyi szám	Tantárgy	Dátum	Osztályzat

# Reláció kulcs fogalma

- Léteznek olyan relációk is, melyekben nem csak egy, hanem több kulcs is található.

KONZULTÁCIÓ={TANÁR, IDŐPONT, DIÁK}

Konzultáció		
Tanár	Időpont	Diák

- Egy személy egy időben csak egy helyen tartózkodhat. Így kulcs lehet a TANÁR, IDŐPONT és a DIÁK, IDŐPONT attribútumok is
- Vegyük észre azt, hogy a kulcsok nem önkényes döntések következtében alakulnak ki, hanem az adatok természetéből következnek, mint a funkcionális vagy a többértékű függőség.

# Redundancia fogalma

- A logikai adatbázis tervezés egyik fő célja a redundanciák megszüntetése.
- Redundanciáról akkor beszélünk, ha valamely tényt vagy a többi adatból levezethető mennyiséget ismételten (többszörösen) tároljuk az adatbázisban.
- A redundancia, a szükségtelen tároló terület lefoglalása mellett, komplikált adatbázis frissítési és karbantartási műveletekhez vezet, melyek könnyen az adatbázis inkonzisztenciáját okozhatják.
- Egy adatbázis akkor inkonzisztens, ha egymásnak ellentmondó tényeket tartalmaz.
- A fizikai tervezés során az adatbázis műveletek gyorsítása érdekében esetleg redundáns attribútumokat is bevezetünk.

# Redundancia fogalma

- Redundanciát tartalmazó reláció:

Tanár	Tantárgy	Össz_óraszám	Tanított_órák
Kiss Péter	Adatbázis kezelés	64	12
Nagy Andrea	Matematika	32	8
Szabó Miklós	Adatbázis kezelés	64	4
Kovács Rita	Matematika	32	5
	Angol	48	

- A fenti relációban a tantárgyak össz óraszámát annyiszor tároljuk, ahány tanár tanítja az adott tantárgyat. A példa kedvéért feltételeztük, hogy egy tantárgyat több tanár is tanít. A redundancia a következő hátrányokkal jár.

# Redundancia fogalma

Tanár	Tantárgy	Össz_óraszám	Tanított_órák
Kiss Péter	Adatbázis kezelés	64	12
Nagy Andrea	Matematika	32	8
Szabó Miklós	Adatbázis kezelés	64	4
Kovács Rita	Matematika	32	5
	Angol	48	

- Ha egy tantárgy össz óraszámja megváltozik több helyen kell módosítani a relációban.
- Valahányszor egy új tanár kerül be a relációba ugyanannak a tantárgynak az előző soraiból kell elővenni az össz óraszám adatot.
- Az utolsó sorban szereplő tantárgy (angol) esetén még nem került kitöltésre a tanár személye. Új tanárnak a listára történő felvételekor ezt az esetet másként kell kezelni. Ilyenkor csak két üres értéket (tanár, tanított órák) kell átírni.



# Redundancia fogalma

- A redundanciát meg kell különböztetni az értékek duplikált (többszörös) tárolásától.
- A duplikált adattárolásra szükségünk lehet a relációkban, míg a redundanciát el kell kerülni.

Termék	Alkatrész	Darab
Nyomtató	papír adagoló	1
Nyomtató	64Kb memória	2
Számítógép	1.2 MB floppy	1
Számítógép	1 MB memória	4

- A táblázat a termék oszlopban többször tartalmazza a nyomtató és számítógép adatokat. Ez azonban nem okoz redundanciát, mivel egy termék több alkatrészből is állhat, így nem ugyanannak a ténynek a többszörös tárolásáról van szó, hanem egy másik tényt fejezünk ki, melyhez elengedhetetlen a duplikált tárolás.
- A redundancia fordul elő akkor is, ha levezett vagy levezethető mennyiségeket tárolunk a relációkban.

# Redundancia megszüntetése, a relációk normál alakjai

- A logikai tervezés célja egy redundancia mentes reláció rendszer, relációs adatbázis.
- A reláció elmélet módszereket tartalmaz a redundancia megszüntetésére, az úgynevezett normál formák segítségével.
- Öt normál formát különböztetünk meg.
- A különböző normál formák egymásra épülnek, a második normál formában levő reláció első normál formában is van.
- A tervezés során a legmagasabb normál forma elérése a cél.
- Az első három normál forma a funkcionális függőségekben található redundanciák, míg a negyedik és ötödik a többértékű függőségekből adódó redundanciák megszüntetésére koncentrálnak.

# Redundancia megszüntetése, a relációk normál alakjai

Még két fogalom:

- Elsődleges attribútumnak nevezzük azokat az attribútumokat, melyek legalább egy reláció kulcsban szerepelnek.
- A többi attribútumot nem elsődlegesnek nevezzük.

# Első normál forma (1NF)

- Egy reláció első normál formában van, ha minden attribútuma egyszerű, nem összetett adat.

Szakkörök

Szakkör	Tanár	Diákok	
Számítástechnika	Nagy Pál	Név	Osztály
		Kiss Rita	III.b
		Álmos Éva	II.c
Video	Gál János	Név	Osztály
		Réz Ede	I.a
		Vas Ferenc	II.b

Szakkörök

Szakkör	Tanár	Diák	Osztály
Számítástechnika	Nagy Pál	Kiss Rita	III.b
Számítástechnika	Nagy Pál	Álmos Éva	II.c
Video	Gál János	Réz Ede	I.a
Video	Gál János	Vas Ferenc	II.b

# Első normál forma (1NF)

- Annak eldöntése, hogy egy attribútumot egyszerűnek vagy összetettnek tekintünk nem mindig egyértelmű, az adatok felhasználásától is függ.
- A döntéseink során, hogy egy vagy több attribútumot tervezünk az adat tárolására, tartsuk szem előtt egyszerűbb több oszlopból egyet csinálni, mint egy oszlop tartalmát több részre vágni.
- Például egy vagy két attribútumban tároljuk a személyek vezetéke és keresztnévét. Amennyiben a nevek között nem akarunk külön-külön keresztnév és vezetéknév szerint keresni, akkor elfogadható lehet az egy mezőben tárolás.

# Második normál forma (2NF)

- Az első normál forma nem elegendő feltétel a redundanciák megszüntetésére.
- Egy reláció második normál alakjában nem tartalmazhat tényeket a reláció kulcs egy részére vonatkozóan.
- A második normál forma definíciója két feltétellel írható le.
  1. A reláció első normál formában van
  2. A reláció minden nem elsődleges attribútuma teljes funkcionális függőségben van az összes reláció kulccsal

# Második normál forma (2NF)

- KONFERENCIA = ({TEREM, IDŐPONT, ELŐADÁS, FÉRŐHELY}, {TEREM, IDŐPONT-> ELŐADÁS, TEREM->FÉRŐ

**Konferencia**

Terem	Időpont	Előadás	Férőhely
B	10:00	Mitológia	250
A	8:30	Irodalom	130
B	11:30	Színház	250
A	11:00	Festészet	130
A	13:15	Régészet	130

**Konferencia**

Terem	Időpont	Előadás
B	10:00	Mitológia
A	8:30	Irodalom
B	11:30	Színház
A	11:00	Festészet
A	13:15	Régészet

**Termek**

Terem	Férőhely
A	130
B	250

# Második normál forma (2NF)

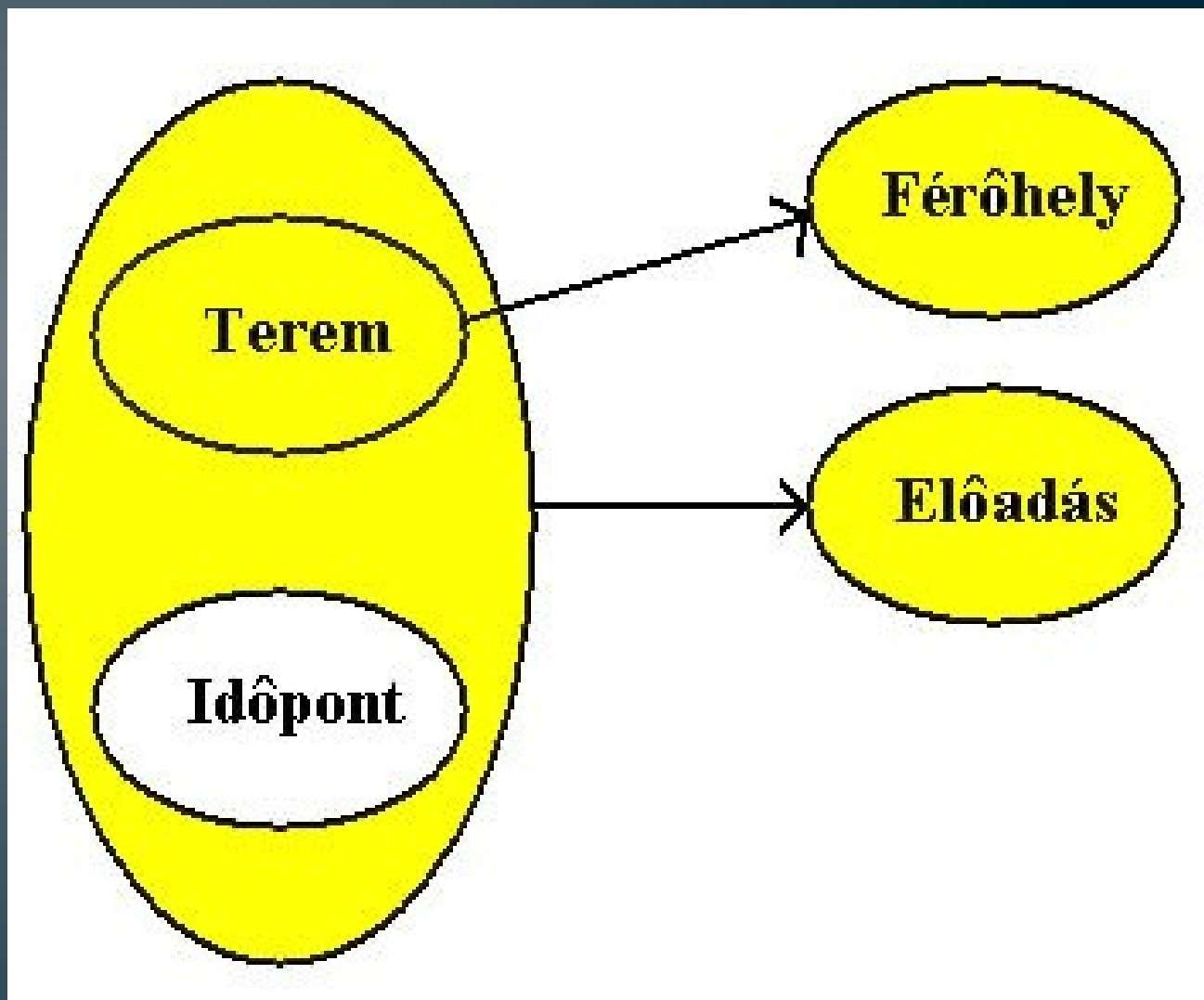
- A reláció kulcsa a TEREM, IDŐPONT attribútumok, ezek a reláció elsődleges attribútumai.
- A másodlagos attribútumok (ELŐADÁS, FÉRŐHELY) közül a FÉRŐHELY csak a reláció kulcs egy részétől függ, ezért nincs második normál formában.
- A felbontás után keletkezett két reláció már második normál formában van.

KONFERENCIA = ({TEREM, IDŐPONT, ELŐADÁS}, {TEREM, IDŐPONT- $\rightarrow$  ELŐADÁS})

TERMEK = (TEREM, FÉRŐHELY}, {TEREM- $\rightarrow$  FÉRŐHELY}).



# Második normál forma (2NF)



# Második normál forma (2NF)

Másik példa

**Hőmérsékletek**

Terem	Időpont	Hőmérséklet	Radiátor
213	98.11.18	23	2
213	98.11.24	22	2
213	98.12.05	21	2
214	98.12.05	21	3
214	98.12.15	20	3

**Konferencia**

Terem	Időpont	Hőmérséklet
213	98.11.18	23
213	98.11.24	22
213	98.12.05	21
214	98.12.05	21
214	98.12.15	20

**Termek**

Terem	Radiátor
213	2
214	3

# Harmadik normál forma (3NF)

- A második normál formájú relációkban nem lehetnek olyan tények, amelyek a reláció kulcs részeihez kapcsolódnak.
- Azonban ennek ellenére is lehet bennük redundancia, ha olyan tényeket tartalmaznak, amelyek a nem elsődleges attribútumokkal állnak kapcsolatban.
- Ezt a lehetőséget szünteti meg a harmadik normál forma.
- Egy reláció harmadik normál formában van, ha
  1. A reláció második normál formában van.
  2. A reláció nem tartalmaz funkcionális függőséget a nem elsődleges attribútumok között.

# Harmadik normál forma (3NF)

**Szakkörök**

Szakkör	Tanár	Születési év
Képzőművész	Sár Izodor	1943
Iparművész	Sár Izodor	1943
Karate	Erős János	1972

**Szakkörök**

Szakkör	Tanár
Képzőművész	Sár Izodor
Iparművész	Sár Izodor
Karate	Erős János

**Tanárok**

Tanár	Születési év
Erős János	1972
Sár Izodor	1943



# Harmadik normál forma (3NF)

- Eredeti:  
SZAKKÖRÖK = ({SZAKKÖR, TANÁR, SZÜLETÉSI ÉV},  
{SZAKKÖR- $\rightarrow$  TANÁR, SZÜLETÉSI ÉV, TANÁR- $\rightarrow$  SZÜLETÉSI ÉV}).
- Felbontás után a nem elsődleges attribútumok közötti függőséget kivettük az eredeti relációból:  
SZAKKÖRÖK = ({SZAKKÖR, TANÁR}, {SZAKKÖR- $\rightarrow$  TANÁR})  
TANÁROK = ({TANÁR, SZÜLETÉSI ÉV}, {TANÁR- $\rightarrow$  SZÜLETÉSI ÉV}).

# Boyce/Codd normál forma (BCNF)

- A normál formák tárgyalása során eddig olyan relációkra mutattunk példákat, melyeknek csak egy reláció kulcsa van.
  - A normál formák definíciója alkalmazható a több kulccsal rendelkező relációkra is.
  - Ebben az esetben minden attribútum, mely valamely kulcsnak a része, elsődleges attribútum, de ez az attribútum függhet egy másik, ezt nem tartalmazó kulcs részétől.
  - Ha ez a helyzet fennáll, redundanciát tartalmaz a reláció. Ennek a felismerése vezetett a harmadik normál forma egy szigorúbb definíciójához, a Boyce/Codd normál formához:
- 1. A reláció harmadik normál formában van**
  - 2. Minden elsődleges attribútum teljes funkcionális függőségben van azokkal a kulcsokkal, melyeknek nem része**

# Boyce/Codd normál forma (BCNF)

- Funkcionális függőségek  
Tanár, Félév -> Tantárgy, Tantárgy, Félév -> Tanár
- A relációnak két kulcsa van:  
(Tanár, Időpont, Félév) és a (Tantárgy, Időpont, Félév)

Tantárgyak

Tanár	Időpont	Tantárgy	Félév	Diák_száma
Kiss Pál	93/1	Adatbázis	1	17
Jó Péter	93/1	Unix	1	21
Kiss Pál	93/2	Adatbázis	2	32
Jó Péter	93/1	Unix	2	19
KissPál	93/1	Adatbázis	3	25

Tantárgyak

Időpont	Tantárgy	Félév	Diák_száma
93/1	Adatbázis	1	17
93/1	Unix	1	21
93/2	Adatbázis	2	32
93/2	Unix	2	19
93/1	Adatbázis	3	25

Tanárok

Tanár	Időpont	Tantárgy
Kiss Pál	93/1	Adatbázis
Jó Péter	93/1	Unix
Kiss Pál	93/2	Adatbázis

# Negyedik normál forma (4NF)

- Sajnos még a Boyce/Codd normál forma is tartalmazhat redundanciát.
- Mindeddig csak a funkcionális függőségeket vizsgáltuk, a többértékű függőségeket nem.
- A további két normál forma a többértékű függőségekből adódó redundancia kiszűrését szolgálja.
- Egy reláció negyedik normál formában van:
  1. Harmadik normál formában van és
  2. egy  $X \twoheadrightarrow Y$  többértékű függőséget tartalmazó relációban csak az X és Y-ban megtalálható attribútumokat tartalmazza.



# Negyedik normál forma (4NF)

Egy relációban tároljuk a személyek, és barátaik nevét valamint hobbiját. Minden személynek több barátja és több hobbija is lehet.

Barátok-hobbik

Személy	Barát	Hobbi
Nagy József	Elek Attila	foci
Nagy József	Varga Attila	foci
Kiss Péter	Kiss Pál	sakk
Kiss Péter	Kiss Pál	video

Barát

Személy	Barát
Nagy József	Elek Attila
Nagy József	Varga Attila
Kiss Péter	Kiss Pál

Hobbi

Személy	Hobbi
Nagy József	foci
Kiss Péter	sakk
Kiss Péter	video

# Negyedik normál forma (4NF)

- Csak egy kulcs van és nincsenek nem elsődleges attribútumok.
- Ezek alapján a reláció harmadik, sőt Boyce/Codd normál formában van.
- Mégis tartalmaz redundanciát, ugyanaz a személy-barát illetve személy-hobby kapcsolat többször is szerepelhet.
- A reláció két többértékű függőséget tartalmaz: Személy->>Barát és Személy->>Hobbi.
- A negyedik normál forma szabályait kielégítő két relációra felbontva a redundancia megszüntethető.

# Ötödik normál forma (5NF)

- Hosszú ideig a negyedik normál formát tartották a normalizálás utolsó lépésének.
- A többértékű függőségek külön relációkban tárolásával azonban információt veszthetünk.
- Ennek bemutatására nézzünk egy példát. Egy számítógépes ismeretek oktatására szakosodott Kft. több jól képzett tanárral rendelkezik. A tanárok többfajta tanfolyam oktatására is alkalmasak. A tanfolyamok az ország különböző részeiben kerülnek megtartásra. Ezek alapján állítsuk össze a Tanár-Tanfolyam-Helyszín relációt. Ebben a relációban csupán azt kívánjuk tárolni, hogy hol és milyen tanfolyamokat tartottak a tanárok, feltételezzük, hogy ugyanazon a helyszínen egyfajta tanfolyam csak egyszer kerül megtartásra.

# Ötödik normál forma (5NF)

- A következő függőségeket írhatjuk fel:  
Tanár->>Tanfolyam Tanár->>Helyszín Tanfolyam-

Tanár-Tanfolyam-Helyszín

Tanár	Tanfolyam	Helyszín
Nagy Éva	Adatbázis I.	Szeged
Kiss Pál	Adatbázis I.	Győr
Nagy Éva	Adatbázis II.	Pécs
Kiss Pál	Adatbázis I.	Pécs

Tanár\_Tanfolyam

Tanár	Tanfolyam
Nagy Éva	Adatbázis I.
Kiss Pál	Adatbázis I.
Nagy Éva	Adatbázis II.

Tanfolyam-Helyszín

Tanfolyam	Helyszín
Adatbázis I.	Szeged
Adatbázis I.	Győr
Adatbázis II.	Pécs
Adatbázis I.	Pécs

Tanár-Helyszín

Tanár	Helyszín
Nagy Éva	Szeged
Kiss Pál	Győr
Nagy Éva	Pécs
Kiss Pál	Pécs

# Ötödik normál forma (5NF)

- Az egyetlen reláció kulcs tartalmazza az összes attribútumot (Tanár, Tanfolyam, Helyszín), ebből következik, hogy Boyce/Codd normál formában van a reláció
- Mégis tartalmaz redundanciát: két sorban is megtalálható, hogy Kiss Pál Adatbázis I. tanfolyamot tanít.
- A relációt felbontva két - csak egy többértékű függőséget tartalmazó - relációra, (Tanár, Tanfolyam) és (Tanár, Helyszín), a redundancia megszüntetésével információt is veszünk.
- A felbontás után már nem tudjuk, hogy a tanár melyik tantárgyát oktatja az adott helyszínen.
- Az eredeti relációt három relációra felbontva kapjuk meg az ötödik normál formát.

# Ötödik normál forma (5NF)

- Az ötödik normál formának megfelelő felbontás eredményeképpen a tárolandó adatmennyiség megnövekszik, a reláció három táblára bontásával.
- Általában célszerűbb egy újabb oszlop bevezetésével csak két táblára bontani a relációt.

Tanfolyamok			Tanfolyam_tanár	
ID	Tanfolyam	Helyszín	Tanár	ID
1	Adatbázis I.	Szeged	Nagy Éva	1
2	Adatbázis I.	Győr	Kiss Pál	2
3	Adatbázis II.	Pécs	Nagy Éva	3
4	Adatbázis I.	Pécs	Kiss Pál	4

# Adatbázis tervezés összefoglalása

- A normál formák tárgyalása végén megjegyezzük, hogy a harmadik normál formáig mindenféleképpen érdemes normalizálni a relációkat.
- Ez a redundanciák nagy részét kiszűri. Azok az esetek, melyekben a negyedik illetve az ötödik normál formák alkalmazására van szükség, ritkábban fordulnak elő.
- Az ötödik normál forma esetén a redundancia megszüntetése nagyobb tároló terület felhasználásával lehetséges csak. Így általában az adatbázis tervezője döntheti el, hogy az ötödik normál formát és a nagyobb adatbázist vagy a redundanciát és a komplikáltabb frissítési, módosítási algoritmusokat választja.

# Adatbázis tervezés összefoglalása

1. Egyed-kapcsolat modell.
2. Relációs adatbázis séma (1NF).
3. Relációsémák normalizálása (2NF ... 5NF).
4. Szükség esetén az egyed-kapcsolat modell módosítása a normalizálás szerint.