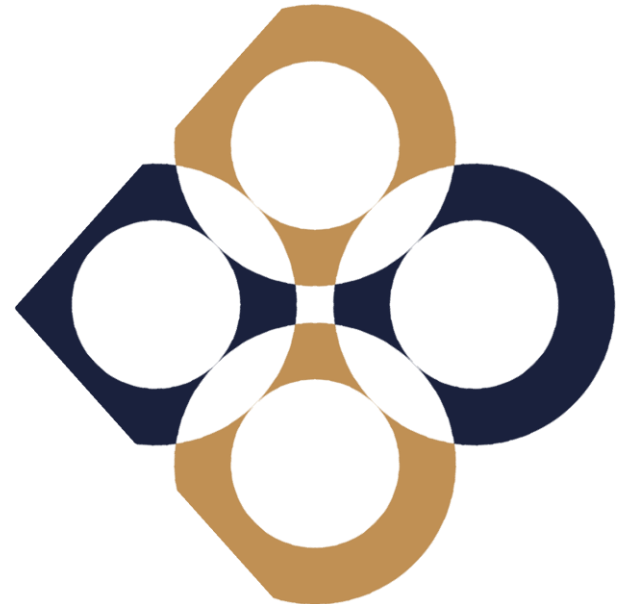


Adatbázisok előadás 03

Adatbázisok tervezése



A tervezés lépései



Egyed – kapcsolat
modell (ER-modell)



Adatbázisterv

Implementálás terve

Adatbázis-terv – hol tartunk?

☐ Adatmodell:



☐ Technikai architektúra (MS SQL):



☐ Megszorítások (kényszerek)



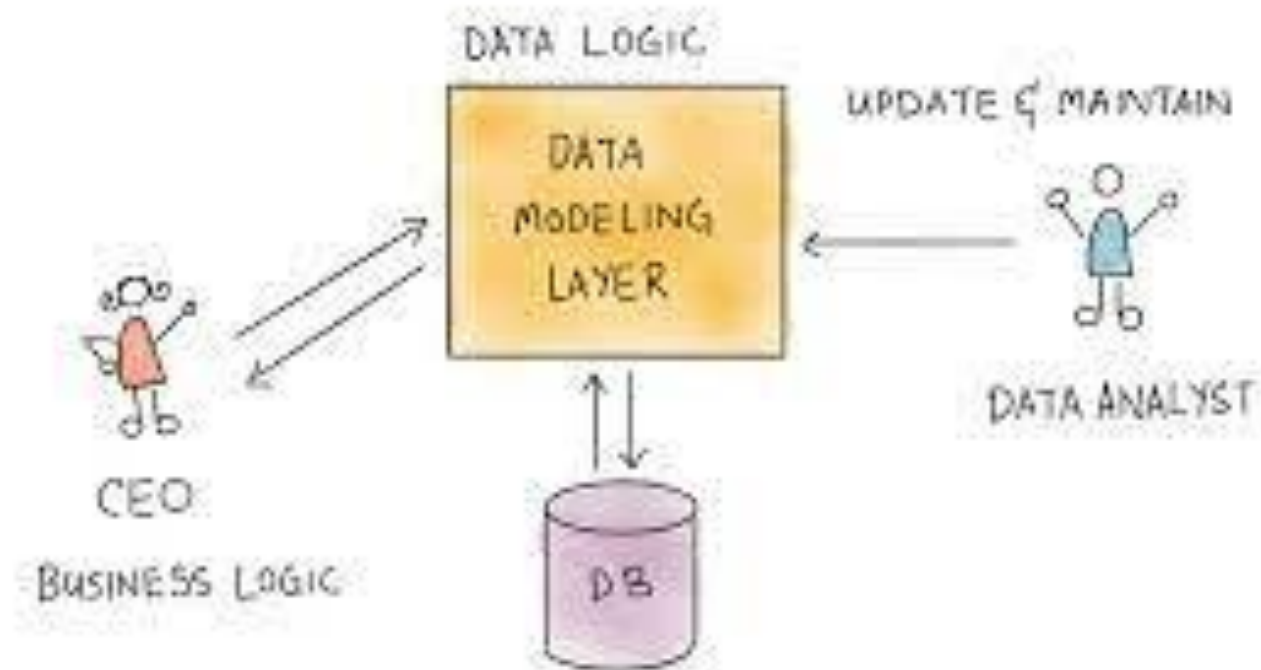
☐ **Hatékonysági megfontolások**

Mit jelent a hatékonyság?

Egy hatékony adatbázisterv ...

- ☐ Segít az adatminőségi problémák csökkentésében
- ☐ Biztosítja a nagyobb adatmennyiség hatékony kezelését
- ☐ Támogatja a változó üzleti igényeket
- ☐ Védi az adatok integritását
- ☐ Megkönnyíti az adatbázisra épülő alkalmazásfejlesztést

Hogyan lehet hatékony a tervezett adatbázis?



<https://www.linkedin.com/pulse/mastering-data-modeling-unlocking-key-efficient-albilasi-pmp/>

A hatékonyság elérésének eszközei

- ☐ Egyszerűsége törekvés
- ☐ Iteratív tervezés fejlesztési folyamatot ciklikusan ismétlődő lépésekre bontják.
- ☐ Folyamatos dokumentálás
- ☐ Együttműködés az érintettekkel
- ☐ Normalizálás vagy Denormalizálás

NORMALIZÁLÁS

A normalizálás egy olyan adatbázis tervezési technika, amely csökkenti az adattárolásban lévő redundanciát, és növeli az adatok integritását.

- A normalizálás során a reláció (tábla) rendszerint több részre bomlik szét
- A normalizálás egyes lépései során előállított állapotot normálformáknak is nevezzük

FONTOS: a normalizálást a működési célú (OLTP) adatbázisok tervezésénél alkalmazzuk (az elemzési célú adatbázisoknál egy ellentétes folyamat, a denormalizálás jellemző)

Milyen a tipikus nem hatékony (rossz) adatbázis?



Példa: oktatók nyilvántartása

Egy felsőoktatási intézmény az oktatóiról a következő adatokat tárolja:

- ☐ Oktató azonosítója, neve, fokozata
- ☐ Tanszékek azonosítója, neve
- ☐ Tantárgyak azonosítója, neve, szintje

Feltételezzük, hogy egy oktató több tárgyat is oktathat, viszont pontosan egy tanszéken tanít

Készítsünk relációs adatmodellen alapuló adatbázistervet!*

*ER-diagramot most nem készítünk

Oktatói adatbázis – kezdeti terv

Tegyünk be minden adatot egy táblába!

OKTATÓ							
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név	Tantárgy kód	Tantárgy név	Tantárgy szintje
001	Kiss Péter	Docens	T01	Matematika	Tan01, Tan02	Algebra, Analízis	BSC, BSC
002	Nagy László	Adjunktus	T02	Informatika	Tan03, Tan04	Programozás, Adatbázisok	BSC, MSC

Probléma: a relációs adatbáziskezelők megkövetelik az atomi (nem összetett) attribútumértékeket

NORMALIZÁLÁS – A lényeg

A normalizálás lényege az adatok közötti összefüggések (un. Függőségi kapcsolatok) feltárása és kezelése

Példák:

- ☐ Funkcionális függőség
- ☐ Teljes függőség
- ☐ Tranzitív függőség
- ☐ Többértékű függőség (ld. Később)
- ☐ Kapcsolásfüggőség (ld. Később)

Funkcionális függőség

Az attribútumok A halmaza funkcionálisan meghatározza az attribútumok B halmazát, ha teljesül, hogy amennyiben két rekord megegyezik az A halmazon, akkor a B-n is.

- Jele: $A \rightarrow B$
- PI: DOLGOZÓ(AZON, NÉV, SZÜLIDŐ) esetén $\{AZON\} \rightarrow \{NÉV, SZÜLIDŐ\}$
- Megj: Ha A funkcionálisan meghatározza B-t, akkor azt is mondhatjuk, hogy B funkcionálisan függ A-tól

Triviális funkcionális függőség

Triviális funkcionális függőségek: egy adott attribútumhalmaz bármely részhalmaza funkcionálisan függ az eredeti halmaztól.

pl: $\{A, B, C\} \rightarrow \{A, B\}$

Mindig igaz!

Példák:

ELADÁS (Dátum, Termékkód, Darabszám) esetén

$\{\text{Dátum}, \text{Termékkód}\} \rightarrow \{\text{Dátum}\}$

$\{\text{Dátum}, \text{Termékkód}\} \rightarrow \{\text{Termékkód}\}$

Teljes függőség

B attribútumhalmaz teljesen függ A-tól, ha nem függ külön az A egyik részalmazától sem

PI: ELADÁS (DÁTUM, TERMÉKKÓD, DARABSZÁM) esetén

{DÁTUM, TERMÉKKÓD} \rightarrow {DARABSZÁM}, de

{DÁTUM} \nrightarrow {DARABSZÁM} és {TERMÉKKÓD} \nrightarrow {DARABSZÁM}

Eladás tábla

Dátum	Termékkód	Darabszám
2020.01.05	2	5
2020.01.05	4	9
2020.01.06	4	10

Az A attribútumhalmaz részalmazától való függőséget részleges függőségnek nevezzük.

Tranzitív függőség

C attribútumhalmaz tranzitíven függ A-tól, ha van olyan B attribútumhalmaz, amelyre $A \rightarrow B$ és $B \rightarrow C$, de visszafele nem igazak a függőségek

- PI: SZAKKÖR(Szakkörnév, Tanár, Végzettség) esetén
- $\{\text{Szakkörnév}\} \rightarrow \{\text{Tanár}\}$ és $\{\text{Tanár}\} \rightarrow \{\text{Végzettség}\}$, de
- $\{\text{Végzettség}\} \not\rightarrow \{\text{Tanár}\}$ és $\{\text{Tanár}\} \not\rightarrow \{\text{Szakkörnév}\}$

Első normálforma

A reláció első normálformában van, ha minden egyes attribútumhoz maximum egy érték tartozik.

Pl: a DOLGOZÓ (AZON, NÉV, VÉGZETTSÉG)

DOLGOZÓ		
AZON	NÉV	VÉGZETTSÉG
01	Kiss Béla	Asztalos, Kőműves
02	Nagy Ervin	Lakatos, Gépész

Nem teljesül az első normálforma

DOLGOZÓ		
AZON	NÉV	VÉGZETTSÉG
05	Horváth István	Cipész
08	Balogh Tamás	Tanító

Teljesül az első normálforma

Első normálformára hozás módja

Ha egy attribútumhoz több érték is tartozik, akkor -
többnyire egy új reláció sémában - minden értéket egy-egy
külön rekordban rögzítünk.

Pl: Kis Béla végzettsége asztalos és kőműves

DOLGOZÓ	
AZON	NÉV
01	Kiss Béla
02	Nagy Ervin

VEGZETTSEG	
AZON	NÉV
05	Asztalos
08	Kőműves

DOLG_VEGZETTSEG	
DAZON	VAZON
01	05
01	08

Első normálformára hozás módja II.

Az összetett attribútumokat részattribútumokkal helyettesíthetjük

Pl: Kis Béla végzettsége asztalos és kőműves

DOLGOZÓ			
AZON	NÉV	VÉGZETTSÉG1	VÉGZETTSÉG2
01	Kiss Béla	Asztalos	Kőműves
02	Nagy Ervin	Lakatos	Gépész

Oktatói adatbázis – Első normálforma

OKTATÓ							
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név	Tantárgy kód	Tantárgy név	Tantárgy szintje
001	Kiss Péter	Docens	T01	Matematika	Tan01	Algebra	BSC
001	Kiss Péter	Docens	T01	Matematika	Tan02	Analízis	BSC
002	Nagy László	Adjunktus	T02	Informatika	Tan03	Programozás	BSC
002	Nagy László	Adjunktus	T02	Informatika	Tan04	Adatbázisok	MSC

Problémák (ld. később: Anomáliák)

- ☐ Az Algebra tárgyat csak Kiss Péter tanítja. Ha őt töröljük az adatbázisból, akkor elvesznek a tárgy adatai
- ☐ Nagy Lászlót előléptetik, docens lesz. Ezt a táblában több helyen is módosítani kell
- ☐ Új oktató érkezik, de még nincs meg, hogy melyik tárgyat fogja tanítani. Addig nem lehet őt rögzíteni.

Megoldás: a második normálforma

Második normálforma

A reláció második normálformában (2NF) van, ha teljesülnek az első normálforma feltételei és a kulcs valódi részhalmazaitól egyetlen másodlagos attribútum sem függ.

Pl: Az ELADÁS(dátum, termékkód, terméknév, darab) reláció nincs 2NF-ben, mert a kulcs: {dátum, termékkód}, {termékkód} a kulcs valódi részhalmaza, {terméknév} másodlagos attribútum, és {termékkód} --> {terméknév}

Második normálformára hozás módja

Ha az első normálformában lévő reláció minden kulcsa egy attribútumból áll, vagy nincs másodlagos attribútum, akkor automatikusan teljesül a 2NF.

Ellenkező esetben a relációt felbontjuk a normálformát sértő funkcionális függőségek mentén.

Pl: Az ELADÁS(dátum, termékkód, terméknév, darab) esetén a normálformát sértő függőség:

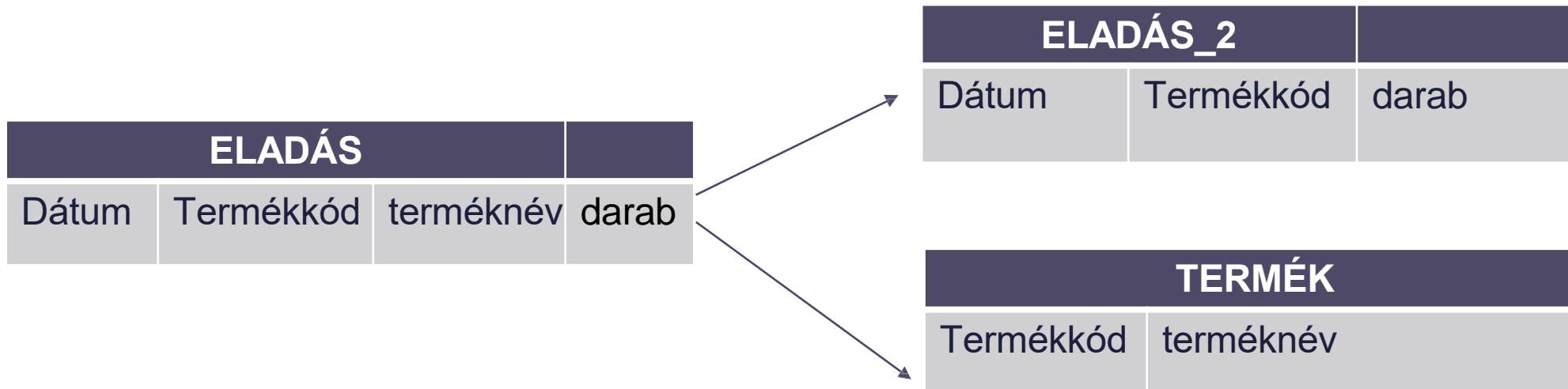
{termékkód} --> {terméknév}

Második normálformára hozás módja - Példa

PI: ELADÁS(dátum, termékkód, terméknév, darab) esetén két új reláció jön létre:

TERMÉK(Termékkód, terméknév) és

ELADÁS_2(Dátum, Termékkód, darab)



Anomáliák – Ezek megszűnnek 2NF esetén

Bővítési anomália:

Az új rekord felvitele a szükséges adatok hiánya miatt nem lehetséges.

Módosítási anomália:

Egy adott módosítást annyi helyen kell végrehajtani, ahányszor az előfordul.

Törlési anomália:

Egy adott rekord törlésekor információvesztés történhet.

Oktatói adatbázis – hozzuk második normálformára

OKTATÓ							
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név	Tantárgy kód	Tantárgy név	Tantárgy szintje
001	Kiss Péter	Docens	T01	Matematika	Tan01	Algebra	BSC
001	Kiss Péter	Docens	T01	Matematika	Tan02	Analízis	BSC
002	Nagy László	Adjunktus	T02	Informatika	Tan03	Programozás	BSC
002	Nagy László	Adjunktus	T02	Informatika	Tan04	Adatbázisok	MSC

Funkcionális függőségek:

- {Oktató azonosító , Tantárgy kód } \rightarrow { }
- {Oktató azonosító} \rightarrow {Oktató név, Oktató tudományos fokozata, Tanszék kód, Tanszék név}
- {Tantárgy kód} \rightarrow {Tantárgy név, Tantárgy szintje}

Az Oktató tábla 3 részre esik szét

OKTATÓ							
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név	Tantárgy kód	Tantárgy név	Tantárgy szintje

OKTATÓ				
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név

TANTÁRGY		
Tantárgy kód	Tantárgy név	Tantárgy szintje

OKTATÁS	
Oktató azonosító	Tantárgy kód

Ez már második normálforma (2NF)

- ❑ Az anomáliák megszűntek, de van még probléma, pl: Tanszék kód és Tanszék név?
- ❑ Ezek egy függőségi kapcsolat részei

Harmadik normálforma

A 2NF-ben lévő reláció harmadik normálformában van, ha nincsenek benne tranzitív függőségek.

PL: ELŐADÁS(Előadás, Terem, Férőhelyek) nem 3NF, mert
 $\{\text{Előadás}\} \twoheadrightarrow \{\text{Terem}\},$
 $\{\text{Terem}\} \twoheadrightarrow \{\text{Férőhelyek}\},$
és ezek fordítva nem teljesülnek, ezért a Férőhelyek attribútum tranzitívan függ az Előadástól.

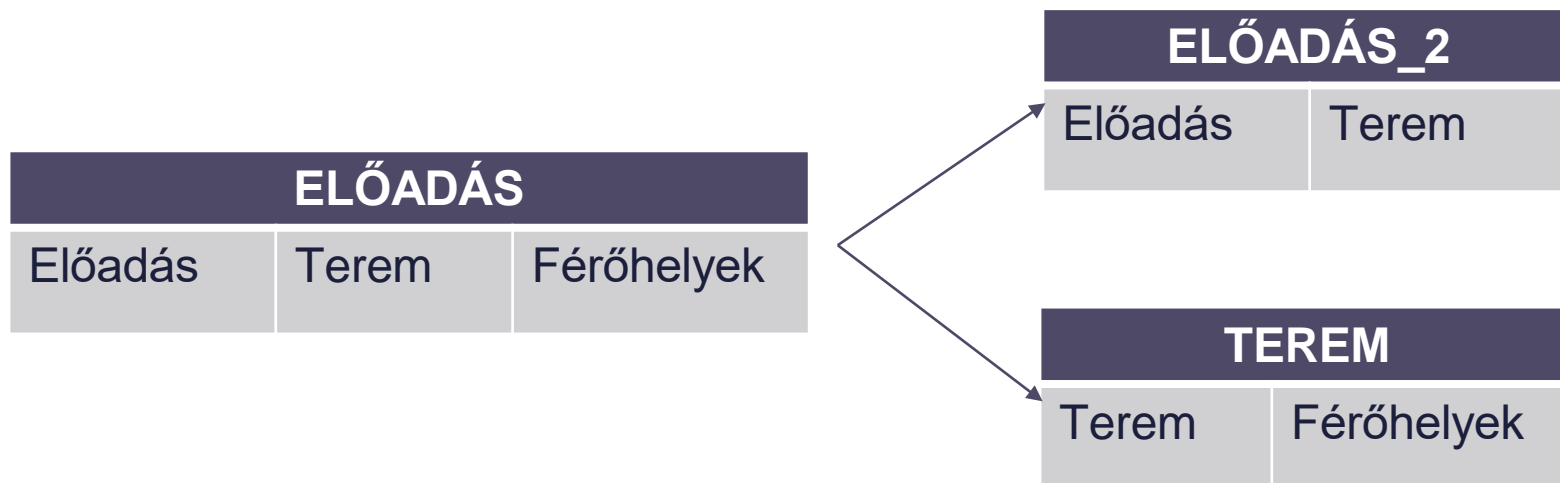
Harmadik normálformára hozás módja

A relációt részekre bontjuk a tranzitív függőségek mentén.

PL: ELŐADÁS(Előadás, Terem, Férőhelyek) reláció esetén a tranzitív függőség:
 $\{\text{Előadás}\} \rightarrow \{\text{Terem}\} \rightarrow \{\text{Férőhely}\}$

Ebből két reláció lesz:

ELŐADÁS_2(Előadás, Terem) és TEREM(Terem, Férőhelyek)



Oktatói adatbázis: hozzuk harmadik normálformára!

OKTATÓ				
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név

TANTÁRGY		
Tantárgy kód	Tantárgy név	Tantárgy szintje

OKTATÁS	
Oktató azonosító	Tantárgy kód

Tranzitív függőségek

- {Oktató azonosító} → {Tanszék kód}
- {Tanszék kód} → {Tanszék név},

de fordítva nem igaz, ezért a Tanszék név tranzitíven függ az Oktató azonosító mezőtől

A tranzitív függés miatt az Oktató táblát két részre bontjuk:

- Oktató(Oktató azonosító, Oktató név, Oktató tudományos fokozata, Tanszék kód)
- Tanszék(Tanszék kód, Tanszék név)

Tranzitív függőségek az Oktató táblában

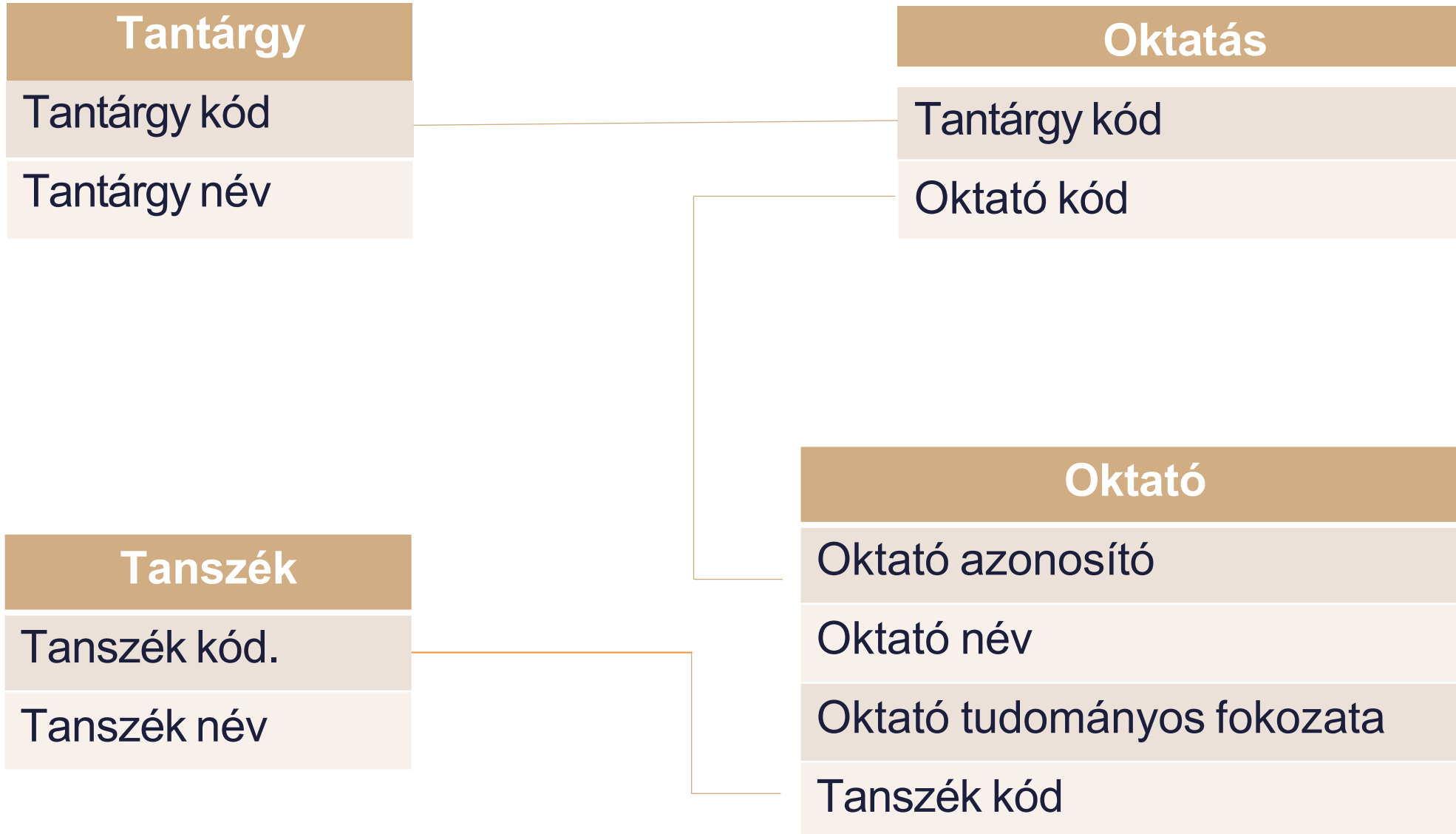
OKTATÓ				
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód	Tanszék név

OKTATÓ			
Oktató azonosító	Oktató név	Oktató tudományos fokozat	Tanszék kód

TANSZÉK	
Tanszék kód	Tanszék név

Ez már harmadik normálforma (3NF)

A végső adatbázis 3NF-ben



Normalizálás – előnyök és hátrányok

- ☐ Csökkenti a redundanciát
- ☐ Növeli az adatintegritást
- ☐ Csökkenti az adatok tárolásához szükséges helyet
- ☐ Megszünteti bizonyos anomáliákat
- ☐ Az egytáblás lekérdezések futása gyorsabb lesz

- ☐ A lekérdezések bonyolultabbak lesznek
- ☐ Az adatbázisra épülő alkalmazások lassabban futhatnak

Magasabb normálformák

Normalizálás – mindig elég a 3NF?

PI: TANÁCSADÁS(ügyfél, téma, tanácsadó, munkaidő)

- Minden tanácsadó csak egy témakörrel foglalkozik, de ugyanabból a témából más és más ügyfeleknek különböző tanácsadók adhatnak tanácsokat.
- Egy ügyfélnek több témából lehet tanácsadója, de egy adott témában mindig ugyanaz a tanácsadója van.
- Nyilvántartjuk még minden ügyfélnél a tanácsadóknak az egyes témakörökben történt tanácsadásra fordított munkaidejét is.

Itt mi lesz a kulcs? Van-e redundancia? Ha igen, hogyan szüntethető meg?

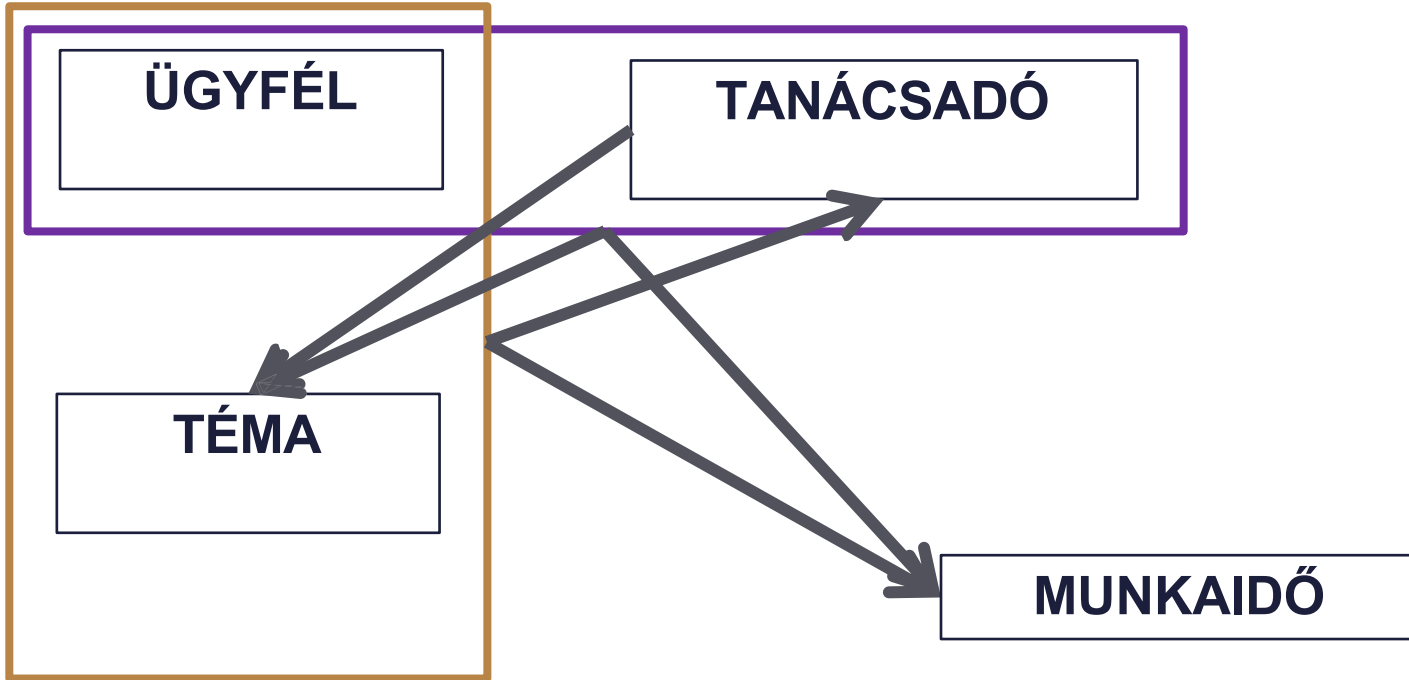
Normalizálás – mindig elég a 3NF?

TANÁCSADÁS			
ÜGYFÉL	TÉMA	TANÁCSADÓ	MUNKAIDŐ
Kiss Péter	Adózás	Nagy Ilona	10
Kiss Péter	Befektetés	Nagy András	8
Bíró Evelin	Adózás	Nagy Ivett	12
Bíró Evelin	Befektetés	Nagy András	9

Kulcs: ÜGYFÉL + TANÁCSADÓ vagy ÜGYFÉL + TÉMA

Másodlagos attribútum: MUNKAIDŐ

Függőségi diagram



- {Ügyfél, téma} → {tanácsadó, munkaidő}
- {tanácsadó} → {téma}
- {ügyfél, tanácsadó} → {téma, munkaidő}

Boyce-Codd normálforma (BCNF)

Egy 3NF-ben lévő reláció BCNF-ben van, ha bármely nem triviális funkcionális függőség baloldalán egy superkulcs áll.

Pl: TANÁCSADÁS(ügyfél, téma, tanácsadó, munkaidő) reláció esetén

- {Ügyfél, téma} \rightarrow {tanácsadó, munkaidő} baloldal: (szuper)kulcs
- {tanácsadó} \rightarrow {téma} baloldal: nem superkulcs
- {ügyfél, tanácsadó} \rightarrow {téma, munkaidő} baloldal: (szuper)kulcs

\rightarrow Ez a reláció nincs BCNF-ben

Azok a 3NF relációk, melyeknek nincs alternatív kulcsuk automatikusan BCNF-ben is vannak

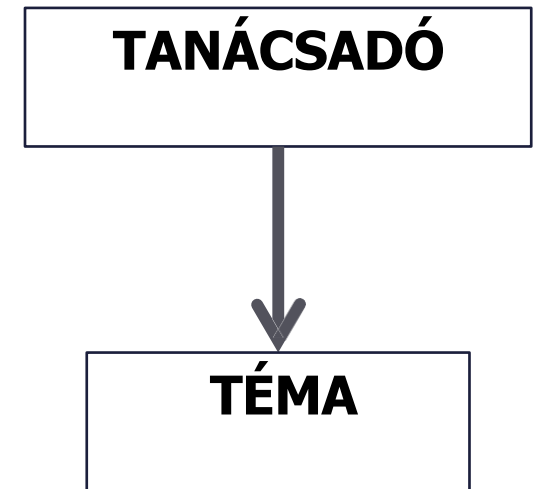
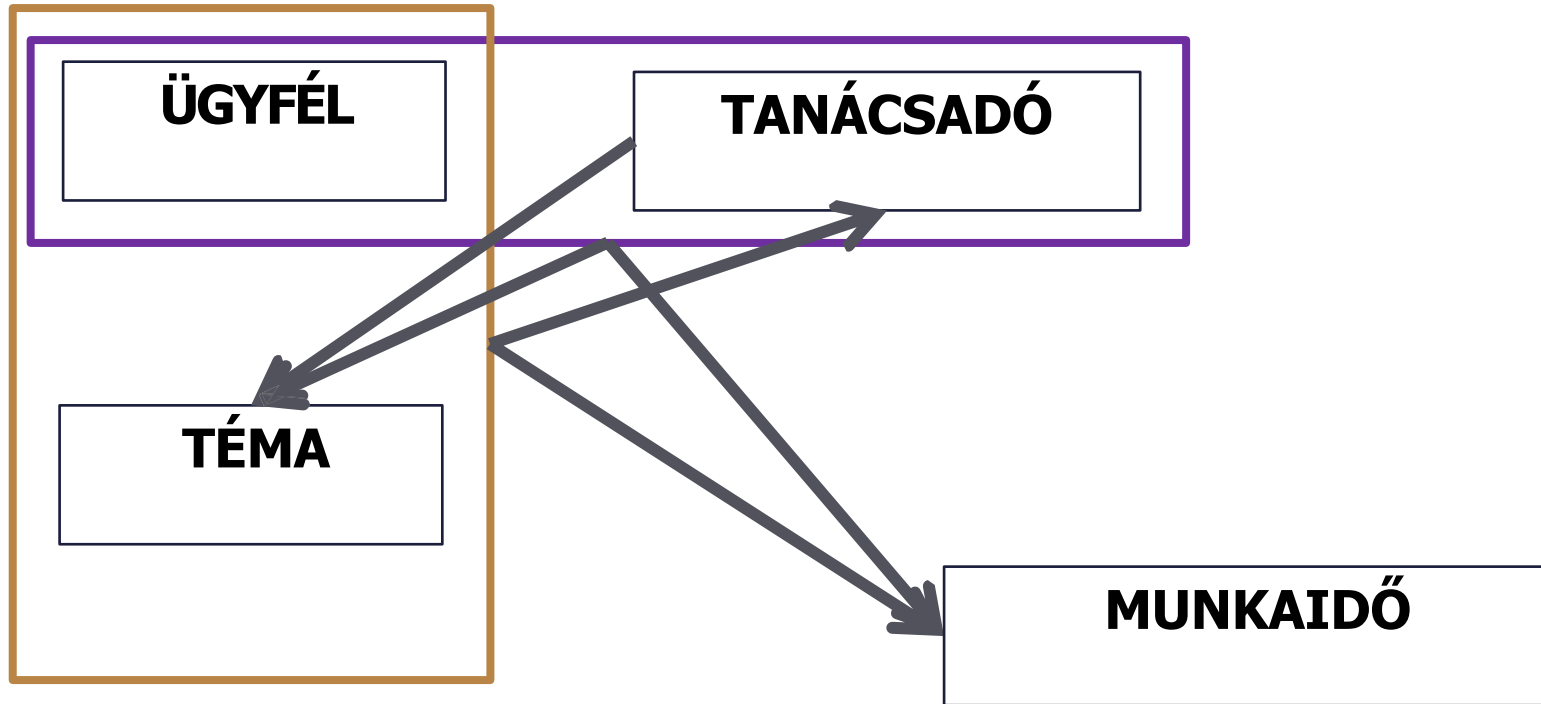
Boyce-Codd normálformára hozás módja

A relációt felbontjuk a normálformát sértő funkcionális függőségek mentén.

TANÁCSADÁS			
ÜGYFÉL	TÉMA	TANÁCSADÓ	MUNKAIIDŐ
Kiss Péter	Adózás	Nagy Ilona	10
Kiss Péter	Befektetés	Nagy András	8
Bíró Evelin	Adózás	Nagy Ivett	12
Bíró Evelin	Befektetés	Nagy András	9

Kulcs: ÜGYFÉL + TANÁCSADÓ vagy ÜGYFÉL + TÉMA

Szétbontás



Új relációk:

TANÁCSADÓ (TANÁCSADÓ, TÉMA)

TANÁCSADÁS_2(ÜGYFÉL, TANÁCSADÓ, MUNKAI DŐ)

Normalizálás – mindig elég a 3NF?

FELMÉRÉS

Személy	Hobbi	Barát
Kiss Béla	Foci	Nagy Róbert
Kiss Béla	Foci	Nagy Ivó
Nagy László	Sakk	Kiss Béla
Nagy László	Foci	Kiss Béla

Itt mi lesz a kulcs? Van-e redundancia? Ha igen, hogyan szüntethető meg?

Negyedik normálforma (4NF)

Egy 3NF-ben lévő reláció negyedik normálformában van, ha bármely nem triviális többértékű függőség* baloldalán szuperkulcs áll.

Pl: FELMÉRÉS(Személy, Hobbi, Barát) reláció esetén

- Kulcs: Személy + Hobbi + Barát
- Többértékű függőségek:
 - {Személy} $\rightarrow\rightarrow$ {Hobbi} baloldal: nem szuperkulcs
 - {Személy} $\rightarrow\rightarrow$ {Barát} baloldal: nem szuperkulcs

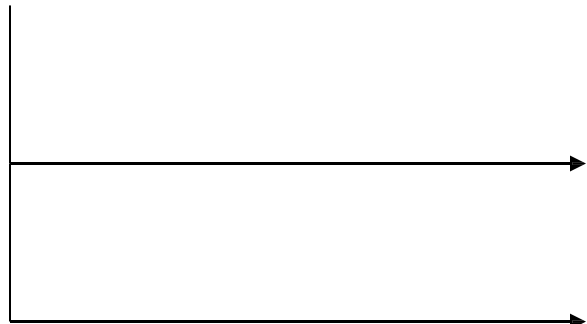
→ Ez a reláció nincs 4NF-ben

A funkcionális függőség általánosítása. $A \rightarrow\rightarrow B$ esetén egy adott A attribútum(halmaz) értékéből több B attribútum(halmaz) érték is következhet. Pl: Személyi szám $\rightarrow\rightarrow$ Végzettség

Negyedik normálformára hozás módja

A relációt felbontjuk a normálformát sértő többértékű függőségek mentén.

FELMÉRÉS(Személy, Hobbi, Barát)



HOBBI(Személy, Hobbi)

BARÁTSÁG(Személy, Barát)

Normalizálás – mindig elég a 4NF?

OKTATÁS		
Tanár	Tanfolyam	Helyszín
Kiss Béla	Adatbázisok	Szeged
Kiss Béla	Programozás	Pécs
Nagy László	Adatbázisok	Győr
Nagy László	Adatbázisok	Pécs

Elég, ha két részre bontjuk?

TANÁR(Tanár, Tanfolyam) és TANFOLYAM(Tanfolyam, Helyszín) → inf. veszteség

Ötödik normálforma (5NF)

Egy 4NF-ben lévő reláció ötödik normálformában van, ha bármely nem triviális kapcsolásfüggés* superkulcsokból áll

Pl: OKTATÁS(Tanár, Tanfolyam, Helyszín) reláció esetén

- Kulcs: Tanár + Tanfolyam + Helyszín
- Kapcsolás függés:
 - {Tanár, Tanfolyam}, {Tanfolyam, Helyszín}, {Tanár, Helyszín} összekapcsolva visszaadja az OKTATÁS relációt, de egyik sem superkulcs

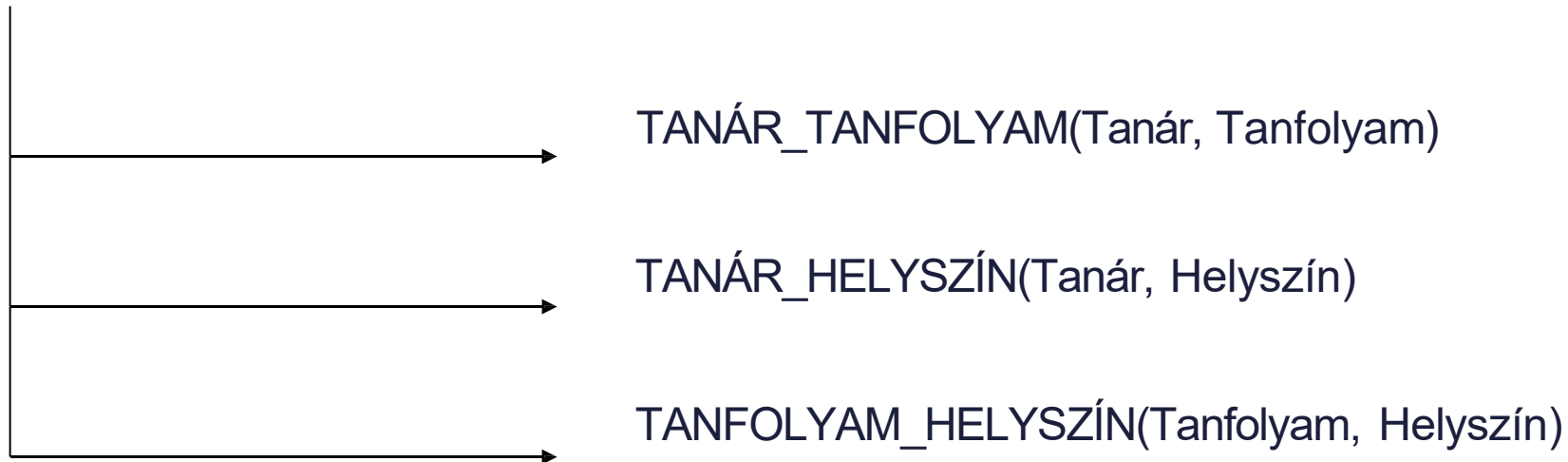
→ Ez a reláció nincs 5NF-ben

A többértékű függőség általánosítása (Join Dependency). Ilyenkor a reláció veszteségmentesen felbontható több relációra. A kapott relációk összekapcsolása (Join) visszaadja az eredeti relációt.

Ötödik normálformára hozás módja

A relációt felbontjuk a normálformát sértő kapcsolás függőség mentén.

OKTATÁS(Tanár, Tanfolyam, Helyszín)



A feladatot másképpen is megoldhatjuk, pl. új oszlop (mesterséges kulcs) bevezetésével

Normálformák és függőségek (összefoglalás)

	1NF	2NF	3NF	4NF	5NF
Decomposition of Relation	R	R ₁₁ R ₁₂	R ₂₁ R ₂₂ R ₂₃	R ₃₁ R ₃₂ R ₃₃ R ₃₄	R ₄₁ R ₄₂ R ₄₃ R ₄₄ R ₄₅
Conditions	Eliminate Repeating Groups	Eliminate Partial Functional Dependency	Eliminate Transitive Dependency	Eliminate Multi-values Dependency	Eliminate Join Dependency

<https://www.javatpoint.com/dbms-normalization>



**Köszönöm
a figyelmet!**