

Adatbázisrendszerek

Normálformák, normalizálás

Nem hivatalos tervezési irányelvek relációs adatbázisokhoz

- Mit jelent a relációs adatbázis-tervezés?
 - Az attribútumok csoportosítását, hogy „jó” relációsémákat alkossanak
- A relációsémák két szintje:
 - A logikai „felhasználói nézet” szint
 - A tárolási (fizikai) „alap reláció” szint
- A tervezés elsősorban az alap relációkkal foglalkozik
- Mik a „jó” alap relációk követelményei?
- Először a jó relációs tervezés nem hivatalos irányelveit tárgyaljuk
- Ezután a funkcionális függések és a normálformák formális definícióit tekintjük át:
 - 1NF (első normálforma)
 - 2NF (második normálforma)
 - 3NF (harmadik normálforma)
 - BCNF (Boyce–Codd-féle normálforma)
- A függések egyéb típusaival, további normálformákkal a későbbi előadásokban foglalkozunk

A reláció attribútumainak szemantikája

1. (nem hivatalos) irányelv

Egy reláció minden egyes rekordja egy egyedet vagy kapcsolat-előfordulást reprezentáljon. (Az egyes relációkra és azok attribútumaira külön-külön vonatkozik.)

- Különböző egyedek (DOLGOZÓ-k, OSZTÁLY-ok, PROJEKT-ek) attribútumai nem keverendők egyazon relációban.
- Más egyedekre való hivatkozás csak külső kulcsok használatával történjen.
- Az egyedekre és a kapcsolatokra vonatkozó attribútumokat a lehető legjobban el kell különíteni egymástól.

Tömören

Olyan sémát kell tervezni, ami könnyen magyarázható relációról relációra. Az attribútumok szemantikájának könnyen értelmezhetőnek kell lennie.

Redundáns információk a rekordokban és a karbantartási anomáliák

Amikor az információt redundánsan tároljuk, az

- tárhelyet pazarol,
- karbantartási anomáliákat okoz, amelyek lehetnek
 - beszúrási anomáliák,
 - törlési anomáliák és
 - módosítási anomáliák.

2. (nem hivatalos) irányelv

Olyan sémát tervezzünk, amelyben nem jelennek meg beszúrási, törlési és módosítási anomáliák.

Ha mégis előfordulnak anomáliák, akkor jegyezzük fel azokat, hogy az alkalmazások számításba vehessék őket.

Példa módosítási anomáliára

Tekintsük az alábbi relációt:

DOLG_PROJ(Dszsz, Pszám, Dnév, Pnév, Órák)

Példa módosítási anomáliára

Ha megváltoztatjuk a P1 számú projekt nevét „Számlázás”-ról például „Ügyfél-számlakezelés”-re, akkor ezt a módosítást mind a 100 olyan dolgozó esetén végre kell hajtanunk, aki a P1 projekten dolgozik.

Példa beszúrási anomáliára

Tekintsük az alábbi relációt:

DOLG_PROJ(Dszsz, Pszám, Dnév, Pnév, Órák)

Példa beszúrási anomáliára

Nem tudunk új projektet beszúrni, ha nincs hozzárendelve egyetlen dolgozó sem.

Példa beszúrási anomáliára – megfordítva

Nem tudunk új dolgozót beszúrni, ha nincs hozzárendelve egyetlen projekthez sem.

Példa törlési anomáliára

Tekintsük az alábbi relációt:

DOLG_PROJ(Dszsz, Pszám, Dnév, Pnév, Órák)

Példa törlési anomáliára

Ha törlünk egy projektet, akkor az összes olyan dolgozó is törlődik, aki az adott projekten dolgozik.

Példa törlési anomáliára – a másik oldalról

Ha egy dolgozó egyedülként dolgozik egy projekten, akkor a dolgozó törlése a szóban forgó projekt törlését is maga után vonja.

NULL értékek a rekordokban

3. (nem hivatalos) irányelv

A relációkat úgy kell megtervezni, hogy a rekordjaik a lehető legkevesebb NULL értéket tartsalmazzák.

Azok az attribútumok, amelyek gyakran vesznek fel NULL értéket, külön relációkba tehetők (az elsődleges kulccsal).

A NULL értékek okai:

- Az attribútum nem értelmezhető vagy érvénytelen.
- Az attribútumérték ismeretlen (de létezhet).
- Az érték biztosan létezik, de nem elérhető.

Álrekordok

- A relációs adatbázisok rossz tervezése bizonyos összekapcsolási műveletek esetén hibás eredményhez vezethet.
- A „veszteségmentes összekapcsolás” tulajdonsággal garantáljuk, hogy az összekapcsolási műveletek értelmes eredményt adnak.

4. (nem hivatalos) irányelv

- A relációkat úgy kell megtervezni, hogy kielégítsék a veszteségmentes összekapcsolás feltételét.
- Egy tetszőleges relációkon végrehajtott természetes összekapcsolás nem állíthat elő álrekordokat.

Álrekordok

A dekompozíciók két fontos tulajdonsággal rendelkeznek:

- (a) A megfelelő összekapcsolás nemadditív vagy veszteségmentes.
- (b) Megőrzi a funkcionális függéseket.

Jegyezzük meg, hogy

- az (a) tulajdonság különösen fontos, és nem áldozható fel,
- a (b) tulajdonság kevésbé szigorú és feláldozható.

Relációsémák normalizációja

Definíció

A **normalizáció** az a folyamat, amelynek során szétbontjuk a normálformák definíciójának nem eleget tevő relációsémákat úgy, hogy az attribútumaikat dekompozícióval több kisebb relációsémába helyezzük át.

Definíció

A normálforma a relációsémák kulcsai és a bennük fennálló funkcionális függések segítségével megfogalmazott feltétel, amellyel megállapítható, hogy a relációséma egy adott normálformában van-e.

Relációsémák normálformái

- 2NF, 3NF, BCNF
a relációsémák **kulcsai** és a bennük fennálló **funkcionális függések** alapján
- 4NF
kulcsok és **többértékű függések** alapján
- 5NF
kulcsok és **join függések** alapján
- További tulajdonságok lehetnek szükségesek a jó relációs tervezés biztosításához (**veszteségmentes összekapcsolás, függésmegőrzés**)

A normálformák használata a gyakorlatban

- A normalizációt úgy hajtják végre a gyakorlatban, hogy a kapott tervek magas színvonalúak legyenek, és rendelkezzenek az elvárt tulajdonságokkal.
- A normálformák gyakorlati alkalmazhatósága megkérdőjelezhetővé válik, ha az alapjukat képező megszorítások nehezen értelmezhetők vagy nehezen ismerhetők fel.
- Az adatbázis-tervezőknek **nem szükséges** a lehető legmagasabb normálformáig normalizálniuk (általában csak 3NF-ig, BCNF-ig vagy 4NF-ig).
- A **denormalizáció** az a folyamat, amelynek során magasabb normálformájú relációk összekapcsolását letároljuk alap relációként – alacsonyabb normálformában.

Kulcsok és a kulcsokat alkotó attribútumok

Definíció

Egy $R = \{A_1, A_2, \dots, A_n\}$ relációséma **szuperkulcsa** minden olyan $S \subseteq R$ attribútumhalmaz, amelyre igaz, hogy bármely R feletti legális r relációban nincs két olyan t_1 és t_2 rekord, amelyekre $t_1[S] = t_2[S]$ teljesül.

Definíció

A K **kulcs** egy olyan szuperkulcs, amelyből **bármely** attribútum eltávolítása azt eredményezi, hogy K már nem lesz szuperkulcs többé.

Kulcsok és kulcsokat alkotó attribútumok

Definíció

- Ha egy relációsémának egynél több attribútumhalmaza van ami alkalmas kulcsnak, akkor ezeket **kulcsjelölteknek** nevezzük. A kulcsjelöltek közül egy **tetszőlegesen** kiválasztott lesz az **elsődleges kulcs**, a többit másodlagos kulcsoknak nevezzük. Minden relációsémának kell, hogy legyen elsődleges kulcsa.

Definíció

- Az R relációséma egy attribútumát R egy **elsődleges attribútumának** nevezzük, ha eleme R **valamely kulcsjelöltjének**. Egy attribútumot **másodlagos (leíró)** attribútumnak hívunk, ha nem elsődleges attribútum, azaz nem eleme egyetlen kulcsjelöltnek sem.

Első normálforma

Tiltja

- az összetett attribútumokat,
- a többértékű attribútumokat,
- a beágyazott relációkat: az olyan attribútumokat, amelyek értékei a különálló rekordokban nem atomiak.

A reláció definíciójának részét képezi.

Első normálformára hozás

A legegyszerűbb módszerrel

(c)

OSZTÁLY

Onév	<u>Oszám</u>	Ovez_szs	<u>Ohelyszín</u>
Kutatás	5	2 551208 2219	Vác
Kutatás	5	2 551208 2219	Tiszafüred
Kutatás	5	2 551208 2219	Budapest
Humán erőforrás	4	2 690329 1099	Kecskemét
Központ	1	1 371110 4518	Budapest

(a)

OSZTÁLY

Onév	<u>Oszám</u>	Ovez_szs	Ohelyszín

(b)

OSZTÁLY

Onév	<u>Oszám</u>	Ovez_szs	Ohelyszín
Kutatás	5	2 551208 2219	{ Vác, Tiszafüred, Budapest }
Humán erőforrás	4	2 690329 1099	{ Kecskemét }
Központ	1	1 371110 4518	{ Budapest }

Első normálformára hozás

Dekompozícióval

(a)

OSZTÁLY

Onév	<u>Oszám</u>	Ovez_szzs	Ohelyszínek

(b)

OSZTÁLY

Onév	<u>Oszám</u>	Ovez_szzs	Ohelyszínek
Kutatás	5	2 551208 2219	{ Vác, Tiszafüred, Budapest }
Humán erőforrás	4	2 690329 1099	{ Kecskemét }
Központ	1	1 371110 4518	{ Budapest }

<u>Oszám</u>	<u>Ohelyszín</u>
5	Vác
5	Tiszafüred
5	Budapest
4	Kecskemét
1	Budapest

Onév	<u>Oszám</u>	Ovez_szzs
Kutatás	5	2 551208 1129
Humán erőforrás	4	2 690329 1099
Központ	1	1 371110 4518

Második normálforma

Definíció

- Egy $X \rightarrow Y$ funkcionális függés **teljes funkcionális függés**, ha X-ből **bármely** A attribútumot eltávolítva a függés a továbbiakban már **nem** áll fenn, azaz bármely $A \in X$ attribútum esetén $(X - \{A\})$ már nem határozza meg funkcionálisan Y-t.

Definíció

- Egy $X \rightarrow Y$ funkcionális függés **részleges függés**, ha valamely $A \in X$ attribútum eltávolítható X-ből úgy, hogy a függés továbbra is fennáll, azaz **valamely** $A \in X$ esetén $(X - \{A\}) \rightarrow Y$.

Definíció

- Egy R relációséma **második normálformában** (2NF-ben) van, ha R **minden** másodlagos (leíró) attribútuma **teljesen** funkcionálisan függ R elsődleges kulcsától.

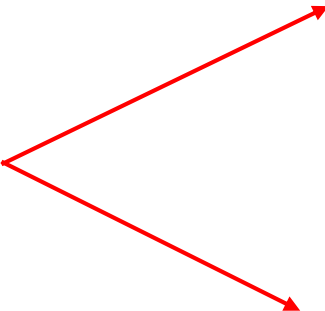
Második normálformára hozás

Konferencia

<u>időpont</u>	<u>terem</u>	előadó	férőhely
10:00	F0	Nagypál Kadosa	195
10:30	F0	Mekk Elek	195
10:30	F01	Erdei Virág	145
11:00	F02	Gyere Pál	40
11:00	F01	Rózsa Kata	145

$\{\text{időpont, terem}\} \rightarrow \{\text{férőhely}\}$

DE $\{\text{terem}\} \rightarrow \{\text{férőhely}\}$



<u>terem</u>	férőhely
F0	195
F01	145
F02	40

<u>időpont</u>	<u>terem</u>	előadó
10:00	F0	Nagypál Kadosa
10:30	F0	Mekk Elek
10:30	F01	Erdei Virág
11:00	F02	Gyere Pál
11:00	F01	Rózsa Kata

Második normálforma

A definícióból következik két egyszerű kritérium, mely megkönnyítheti a 2NF típusú relációk felismerését (amennyiben csak egy kulcsjelölt volt):

1. Ha a kulcs **egyetlen attribútumból** áll, (vagyis egyszerű) akkor a reláció 2NF típusú.
2. Ha a relációban **nincsenek másodlagos attribútumok** (tehát minden attribútum része a kulcsnak) akkor a reláció 2NF típusú.

Harmadik normálforma

A harmadik normálforma a tranzitív függés fogalmán alapul.

Definíció

- Egy R relációséma $X \rightarrow Y$ funkcionális függése **tranzitív függés**, ha létezik egy olyan Z attribútumhalmaz, amely **nem kulcsjelölt és nem része R egyetlen kulcsának sem**, és fennáll $X \rightarrow Z$, illetve $Z \rightarrow Y$.

Definíció

- Egy R relációséma **harmadik normálformában** (3NF-ben) van, ha 2NF-ben van, és **nincs** R-nek olyan **másodlagos** (leíró) attribútuma, amely **tranzitívan függne** az elsődleges kulcstól.

Harmadik normálformára hozás

KÖNYV(isbn, szerzo1, cím, kiadó_az, kiadó_név)

Csak az első szerzőt tároljuk.

Kulcs: isbn

Egyszerű kulcs → 2NF OK.

3NF ?

Nem, mert {isbn} → {kiadó_név} tranzitív függés

hiszen {isbn} → {kiadó_az} és {kiadó_az} → {kiadó_név} és kiadó_az nem kulcsjelölt

KÖNYV(isbn, szerzo1, cím, *kiadó_az*)

KIADÓ(kiadó_az, kiadó_név)

A 2NF és 3NF általános definíciója

Definíció

- Egy R relációséma második normálformában (2NF-ben) van, ha R-nek nincs olyan másodlagos (leíró) attribútuma, amely részlegesen függne R **bármely** kulcsától.

Definíció

- Egy R relációséma harmadik normálformában (3NF-ben) van, ha valahányszor egy $X \rightarrow A$ **nemtriviális** funkcionális függés fennáll R-en, akkor vagy
 - (a) X egy superkulcsa R-nek, vagy
 - (b) A egy elsődleges attribútuma R-nek.

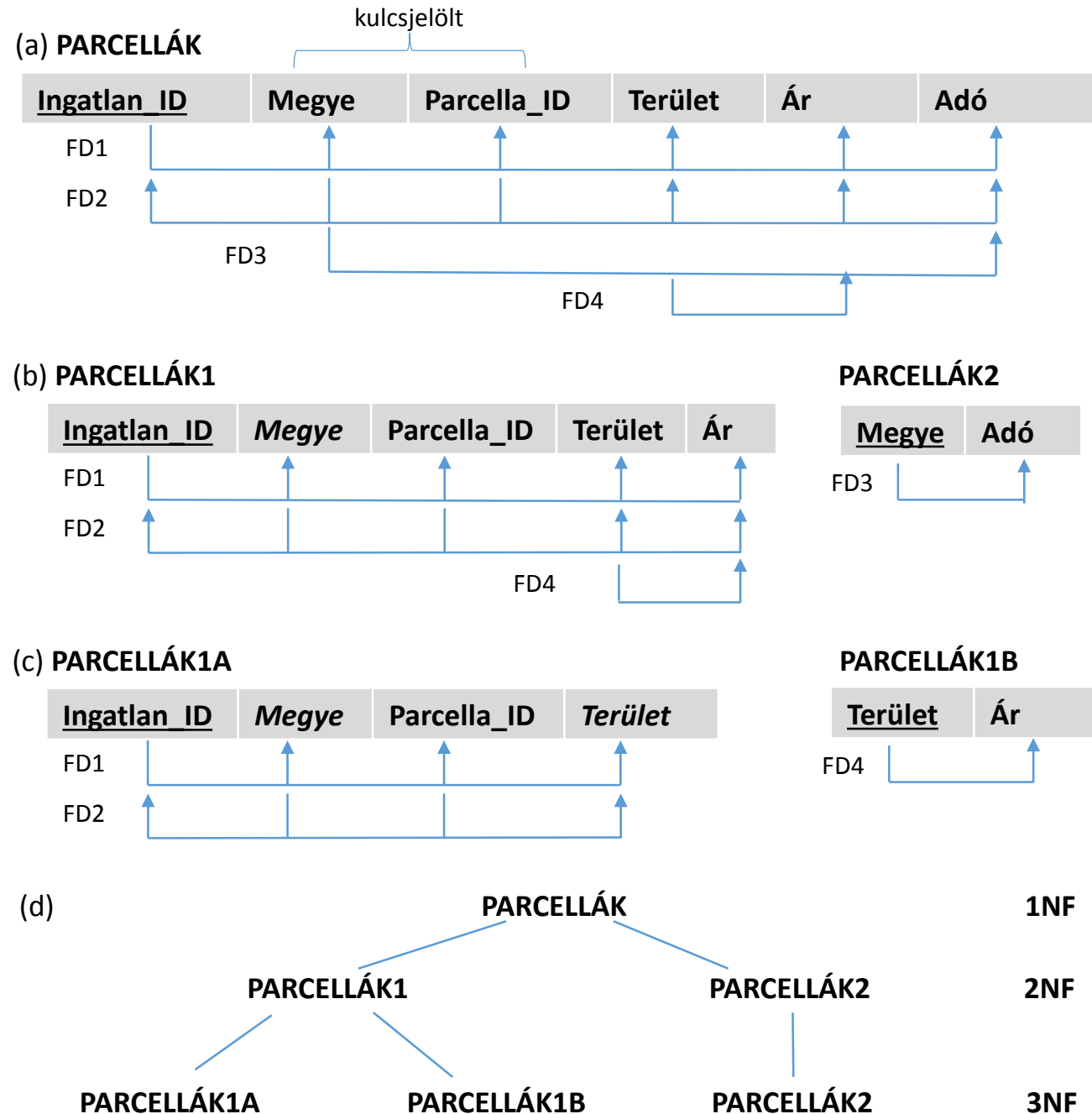
Boyce–Codd-féle normálforma

Definíció

- Egy R relációséma **Boyce–Codd-féle normálformában** (BCNF-ben) van, ha valahányszor egy $X \rightarrow A$ **nemtriviális** funkcionális függés fennáll R-en, akkor X egy superkulcsa R-nek.

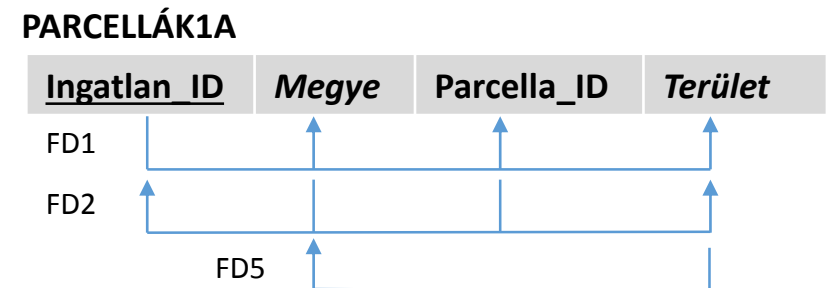
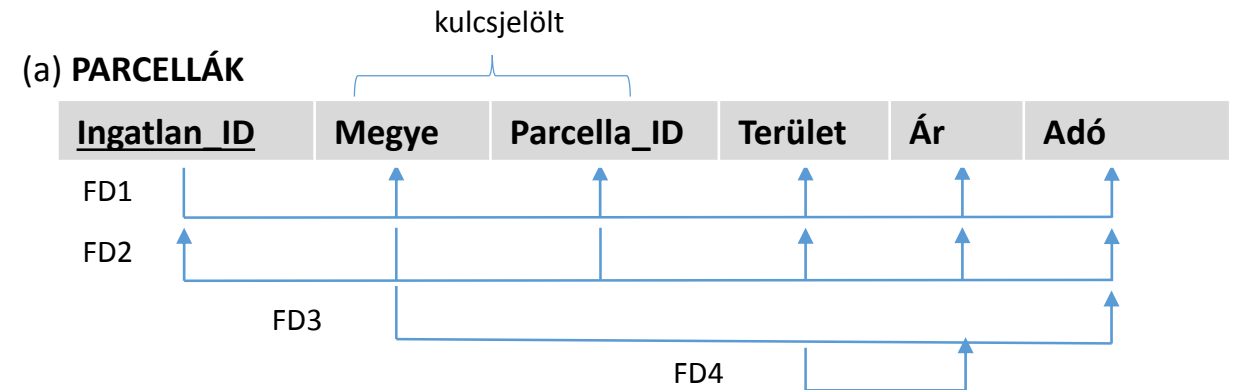
Példa normalizálásra

- Adott funkcionális függések FD1-FD4
- PARCELLÁK nincs 2NF-ben (FD3 miatt)
- PARCELLÁK1 nincs 3NF-ben (FD4 miatt)



Boyce-Codd-féle normálforma

- Tételezzük fel, hogy
 - $\text{Domain}\{\text{Megye}\} = (\text{Baranya}, \text{Csongrád})$
 - $\text{Terület}\{0,5; 0,6; 0,7; 0,8; 0,9; 1\}$ Baranya
 - $\text{Terület}\{1,1; 1,2; 1,3; \dots 2,0\}$ Csongrád
- Azaz FD5: $\text{Terület} \rightarrow \text{Megye}$
- PARCELLÁK1A
 - 3NF, mert a Megye elsődleges attribútum
 - Nem BCNF, mert a Terület nem superkulcsa a PARCELLÁK1A-nek

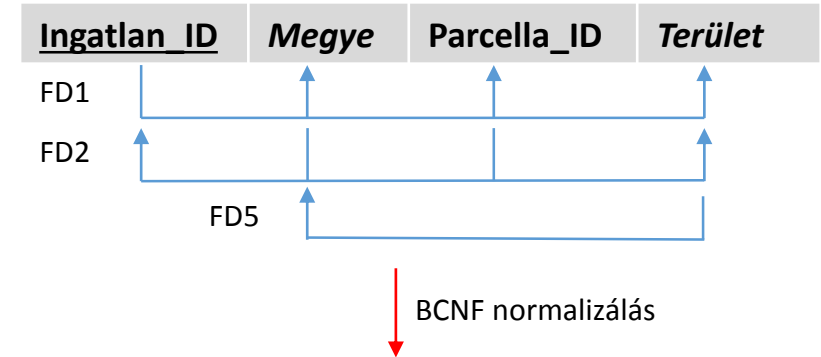


Egy R relációséma Boyce–Codd-féle normálformában (BCNF-ben) van, ha valahányszor egy $X \rightarrow A$ nemtriviális funkcionális függés fennáll R-en, akkor X egy superkulcsa R-nek.

Boyce-Codd-féle normálforma

- Egy R relációséma harmadik normálformában van, ha valahányszor egy $X \rightarrow A$ nemtriviális funkcionális függés fennáll R-en, akkor **vagy**
 - (a) X egy szuperkulcsa R-nek, vagy
 - (b) **A egy elsődleges attribútuma R-nek.**
- Egy R relációséma Boyce–Codd-féle normálformában (BCNF-ben) van, ha valahányszor egy $X \rightarrow A$ nemtriviális funkcionális függés fennáll R-en, akkor X egy szuperkulcsa R-nek.
- A (b) ábra egy tipikus séma mely 3NF-ben van, de nincs BCNF-ben.

(a) PARCELLÁK1A



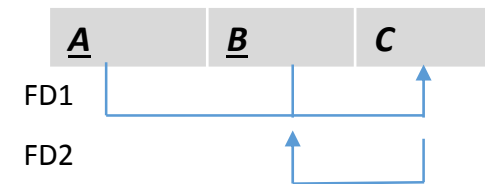
PARCELLÁK1AX

<u>Ingalan_ID</u>	Parcella_ID	Terület
-------------------	-------------	---------

PARCELLÁK1AY

<u>Terület</u>	Megye
----------------	-------

(b) R



Példa normalizálásra -1

TANFOLYAM (szsz, d.név, beosztás, fizetés, t.kód, t.név, díj, óraszám, eredmény)

Tegyük fel, hogy a t.kód, t.név, díj, óraszám, eredmény többértékű attribútumok.

DOLGOZÓ (szsz, d.név, beosztás, fizetés)

JÁR (szsz, t.kód, t.név, díj, óraszám, eredmény)

2NF

$t.kód \rightarrow \{t.név, díj, óraszám\}$

TANFOLYAM (t.kód, t.név, díj, óraszám)

JÁR (szsz, t.kód, eredmény)

3NF

Igen, amennyiben beosztástól nem függ a fizetés, illetve az óraszámától a tanfolyam díja.

Példa normalizálásra -2

Szobanövények(megnevezés, helye, dátum, víz, fény, alkalom, kié, állapot, gondozó, gtel, gmail)

Elsődleges kulcs??

Szobanövények(azon, megnevezés, helye, dátum, víz, fény, alkalom, kié, állapot, gondozó, gtel, gmail)

2NF ?

3NF ?

{gondozó} → {gtel, gmail}, {megnevezés} → {víz, fény}

Gondozó_személy(gondozó, gtel, gmail)

Gondozás(megnevezés, víz, fény)

Szobanövények(azon, *megnevezés*, helye, dátum, alkalom, kié, állapot, *gondozó*)

Példa normalizálásra -3

Cégautó(rendszám, típus, évjárat, vhorog, ajtószám, férőhely, cstér, szín, fogyasztás)

2NF ?

3NF ?

{típus} → {ajtószám, férőhely, cstér}

Típusok(típus, ajtószám, férőhely, cstér)

Cégautó(rendszám, *típus*, évjárat, vhorog, szín, fogyasztás)