

Normalizálás

A normalizálás az adatbázis-tervezés egyik módszere. Célja, hogy az adatok kezdeti káoszából egy áttekinthető, és tárolási szempontból is megfelelő relációs modellt kapjunk. A módszer nem más, mint egy táblázatszétbontó eljárás. A kiinduló helyzet egyetlen táblázat, amelybe minden tárolni kívánt adat össze van zsúfolva. Az eljárás során általában három, esetenként négy lépésben a kezdeti táblát több egymással kapcsolatban lévő táblázattá bontjuk. Az eljárás végén kapott táblázatok anomáliamentesek, redundanciájuk pedig, jóval kisebb, mint a kiinduló táblázatnak.

A normalizálás lépései

A normalizálás során feladatunk, hogy relációinkat úgy alakítsuk át, hogy azok megfeleljenek a normál formák kritériumainak. Először az első normál forma kritériumait kell teljesítenünk, majd a másodikét és a harmadikét. Általában a harmadik normál formánál megállhatunk, csak ritka esetben kell a negyedikig eljutni.

Az első normál forma létrehozása

A tárolási igények összegyűjtése után már tisztában vagyunk azzal, hogy milyen adatokat akarunk tárolni. Első feladatunk, hogy megpróbáljuk az adatokat táblázatos formába rendezni, mégpedig úgy, hogy a táblázat megfeleljen a következő feltételeknek:

- az oszlopok száma és sorrendje minden sorban azonos legyen

Ellenpélda:

- ne legyen két egyforma sor
- ne legyen két egyforma oszlop
- ne legyen többértékű attribútum

Ellenpélda (Az „oktatott tárgy” több értéket is felvehet):

oktató kódja	oktató neve	tanszék	oktatott tárgy
GJ1	Gipsz Jakab	Informatika	Pascal C++

A többértékű attribútum megszüntetése:

oktató kódja	oktató neve	tanszék	oktatott tárgy
GJ1	Gipsz Jakab	Informatika	Pascal
GJ1	Gipsz Jakab	Informatika	C++

- minden sort egyértelműen lehessen azonosítani egy egyedi kulcs segítségével

Példa: A statisztikai hivatal felmérést kíván készíteni a TV-nézők és az általuk nézett filmek közötti viszonyról. Feladatunk a felmérés adatainak tárolásához szükséges adatbázis megtervezése.

Tárolni kell a nézők igazolványszámát, nevét, lakhelyét (csak a település nevét), azt, hogy egy bizonyos filmet hány alkalommal láttak, a filmek címét, hosszát, gyártási évét, műfaját, a gyártó nevét és címét.

Feltételezzük, hogy azonos című filmeket nem tárolunk, illetve hogy a filmeknek egyetlen gyártója van, vagyis nincsenek koprodukcióban készült filmek.

Az első normál forma (néhány rekorddal):

FILMNÉZŐK

<u>igsz</u>	<u>név</u>	<u>lakhely</u>	<u>alkalom</u>	<u>f_cím</u>	<u>hossz</u>	<u>gy_év</u>	<u>műfaj</u>	<u>gy_név</u>	<u>gy_cím</u>
ad12345678	Gipsz Jakab	Eger	4	Star Wars	100	1977	sci-fi	Lucasfilm	Hollywood, USA
ad12345678	Gipsz Jakab	Eger	1	Gyűrűk ura	180	2001	kaland	New Line Cinema	New Zeland
ad98765432	Kukor Ica	Győr	2	Star Wars	100	1977	sci-fi	Lucasfilm	Hollywood, USA

Az elsődleges kulcsot az **igsz** és az **f_cím** mezők kombinációjából képezzük.

A nem kulcs attribútumokat az elsődleges kulcs egyértelműen meghatározza, vagyis funkcionálisan függenek az elsődleges kulcstól.

A **funkcionális függés** azt jelenti, hogy egy attribútum (vagy attribútumhalmaz) egyértelműen meghatároz más attribútumokat, vagyis azok függenek tőle.

A kapott reláció első normál formában van, mert megfelel az első normál forma követelményeinek.

A második normál forma létrehozása

Ha vannak olyan attribútumok, amelyeket az összetett elsődleges kulcs egy része is egyértelműen meghatároz, akkor ezeknek az attribútumoknak a kulcs részével együtt egy külön relációt hozunk létre. Ennek megfelelően több táblázatra bontjuk az első normál formában lévő táblázatunkat. Az így kapott táblázatok nem kulcs attribútumait a kulcs egésze határozza meg, vagyis a kulcs nem bontható úgy részekre, hogy valamely rész meghatározza akár egy attribútumot is. Ebben az esetben a nem kulcs attribútumok teljesen függenek az elsődleges kulcstól.

A szétbontással létrejött táblázatok között kapcsolatot kell létrehozni, annak érdekében, hogy az eredetileg összetartozó adatokat, melyek most már nem egy, hanem több táblázatban vannak, össze lehessen párosítani. A kapcsolatot egyszerűbb esetben létrehozhatjuk úgy, hogy az egyik tábla elsődleges kulcsát a másik táblában is szerepeltetjük, mint idegen kulcsot. Ennek feltétele, hogy az így létrehozott idegen kulcs minden egyes rekordnál kitöltött legyen, illetve egyértékű attribútum legyen. Ha ezek a feltételek nem teljesülnek (még úgy sem, hogy a két táblázatot felcseréljük), akkor kizárólag kapcsoló táblázattal lehet a kapcsolatot megvalósítani. A kapcsoló tábla a kapcsolni kívánt táblák elsődleges kulcsait tartalmazza, amelyek egyenként idegen kulcsok.

A példánk:

Az összetett kulcs **igsz** része egyértelműen meghatározza a néző nevét és lakcímét, így ezekkel az attribútumokkal új relációt hozunk létre.

A kulcs **f_cím** része a *hossz*, *gy_év*, *műfaj*, *gy_név*, *gy_cím* attribútumokat határozza meg, ezekkel is új relációt hozunk létre.

Az alkalom attribútum az egyetlen, amelyet az összetett kulcs határoz meg.

Esetünkben három reláció jön létre. Valójában a NÉZŐ táblát kell a FILM táblával összekapcsolni. A néző igazolványszámát nem hozhatjuk létre a FILM relációban is, mert az ott többértékű attribútum lenne (egy filmet többen is látnak). A dolog fordítva sem működik, mert a film címe sem szerepelhet a NÉZŐ relációban, hiszen ott többértékű attribútum lenne (egy néző több filmet is láthat). Nem marad más választás, mint a kapcsolótábla. Olyan szerencsénk van, hogy az *alkalom* attribútumnak köszönhetően a kapcsolótábla már létre is jött, illetve megmaradt az eredeti relációból.

A kapcsolótábla tartalmazza a két tábla elsődleges kulcsát, amelyek együtt képzik a kapcsolótábla összetett elsődleges kulcsát, illetve egyenként idegen kulcsok is, melyek visszamutatnak a NÉZŐ és a FILM tábla elsődleges kulcsára. Ezen kívül tartalmazza azokat az attribútumokat, amelyeket az összetett kulcs határozott meg (az *alkalom* volt egyedül ilyen).

NÉZŐ (igsz, név, lakcím)

LÁTOTT (igsz, f_cím, alkalom)

FILM (f_cím, hossz, gy_év, műfaj, gy_név, gy_cím)

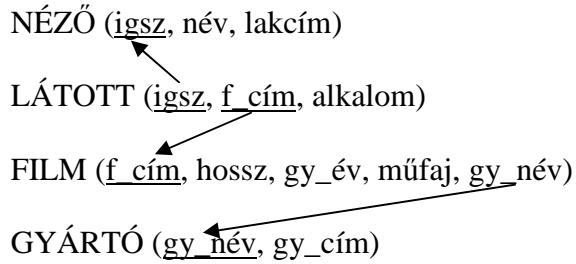
Az így kapott relációk második normál formában vannak, mert első normál formában vannak és a nem kulcs attribútumok teljesen függenek az elsődleges kulcstól.

A **teljes függés** azt jelenti, hogy egy attribútumhalmaz úgy határoz meg egyértelműen más attribútumokat, hogy a meghatározó attribútumhalmaz minimális, vagyis nem bontható úgy részekre, hogy valamely rész az eredetileg meghatározott attribútumokból akár egyet is egyértelműen meghatározzon, így a meghatározott attribútumok teljesen függenek az őket meghatározó attribútumhalmaztól.

A harmadik normál forma létrehozása

Azt az esetet, amikor egy nem kulcs attribútum egyértelműen meghatároz egy másik attribútumot **tranzitív függésnek** nevezzük. Ha ilyen függést találunk valamelyik relációban, akkor ezekkel az attribútumokkal új relációt hozunk létre, melynek elsődleges kulcsa lesz az az attribútum, amelyből a tranzitív függés kiindult. Az eredeti relációval pedig kapcsolatot kell létesíteni a második normál formánál ismertetett módon.

A FILM relációban láthatunk tranzitív függést, hiszen a gyártó neve egyértelműen meghatározza a gyártó címét.



Az így kapott relációk mindegyike harmadik normál formában van, mert második normál formában vannak, és nem tartalmazznak tranzitív függést.

Feladatok

Normalizálás segítségével tervezzük meg az adatbázisokat.

1. Egy borszaküzlet nyilvántartó rendszerének adatbázisát kell megterveznünk. Tárolni kell a bor azonosítóját, nevét, színét, árát, aktuális raktárkészletét, a termelő adószámát, nevét, székhelyét, és telefonszámát. Egy termelő többféle bort is termel, viszont egy konkrét azonosítóval ellátott bort csak egy termelő termel.
2. Országos autós túraverseny adatbázisát kell megterveznünk. Tárolni kell a versenyzők nevét, lakcímét, rajtszámát, autók rendszámát, típusát, színét, hengerűrtartalmát, gyártási évét. Egy versenyző egy autóval indul, minden versenyző, és minden autó részt vesz a versenyen.
3. Egy vállalat dolgozóit különféle tanfolyamokra küldi. Egy nyilvántartó rendszer segítségével követhető nyomon, hogy ki milyen tanfolyamot, milyen eredménnyel végzett el. A tárolási igények: az alkalmazottak igazolványszáma, neve, születési dátuma, lakcíme, a tanfolyam kódja, neve, összóraszám, a vizsga típusa (szóbeli, írásbeli) és a vizsga eredménye. Egy ember több tanfolyamon is részt vehet, egy tanfolyamra pedig több ember is beiratkozhat.
4. Egy vállalat szolgálati autóit szeretné nyilvántartani. Tárolni kell az autók rendszámát, motor- és alvázszámát, típusát, gyártási évét, fajtakódját, fajtáját (teher, személy, pótkocsi, ...), a fajta maximális össztömegét, a járművet használó dolgozó igazolványszámát, nevét, címét és telefonszámát. Egy ember több autót is használhat, de egy autónak csak egy gazdája van. Az adatbázisban minden dolgozót tárolunk függetlenül attól, hogy van e szolgálati autója vagy sem.