

Adatbázisok 10. – Normalizálás példa

Horváth Árpád <horvath.arpad@amk.uni-obuda.hu>

Óbudai Egyetem
Alba Regia Műszaki Kar (AMK)
Székesfehérvár

2015. október 15.

Ebben a segédletben a német Normalisierung Wikipédia-cikkben szereplő példán mutatjuk be a normalizálás lépéseit. Minden normálforma esetén szerepel egy példa, amelyik nincs az adott formában, majd a leírás, hogy miért, legvégül a normalizálás után kapott táblák.

Az elsődleges kulcs mezői a vetített változatban **vastagítva szerepelnek**.

Az adatbázisunkban tárolni szeretnénk az albumok egyes számait sorrendben, az album címét, előadóját, az előadó együttes alapításának az évét (egyéni énekes esetén, a zenei működésének kezdetét).

Ez még nincs 1NF-ben

CD_id	album	alapítás dátuma	címlista
4711	Anastacia – Not That Kind	1999	{1. Not That Kind, 2. I'm Outta Love, 3. Cowboys & Kisses}
4712	Pink Floyd – Wish You Were Here	1964	{1. Shine On You Crazy Diamond}
4713	Anastacia – Freak of Nature	1999	{1. Paid my Dues}

1. táblázat: CD_dal tábla

Miért nincs?

Az előző tábla nincs első normálformában, mert

- az egyes mezőknek tovább nem oszthatóaknak kell lenniük, más néven atomiaknak. Ha az album mező tartalmazza az előadót és az album nevét is, akkor nem tudunk csak az egyik szerint keresni. Persze, hogy mi számít eleminek részben a feladattól is függ. A munkáltató szempontjából a teljes név lehet elemi, az adóhivataléból az elő- és utónév számít csak eleminek.
- Az egyes mezők nem tartalmazhatnak felsorolást.

Ez már 1NF

CD_id	albumcím	előadó	alapítás éve	sáv	cím
4711	Not That Kind	Anastacia	1999	1	Not That Kind
4711	Not That Kind	Anastacia	1999	2	I'm Outta Love
4711	Not That Kind	Anastacia	1999	3	Cowboys & Kisses
4712	Wish You Were Here	Pink Floyd	1964	1	Shine On You Crazy Diamond
4713	Freak of Nature	Anastacia	1999	1	Paid my Dues

2. táblázat: A CD_dal tábla új változata

- Az album címét és az előadót külön szedtük, hogy atomi legyen az adatunk.
- Az utolsó sorban levő lista egyes elemeit külön-külön sorba kell szerepeltetni úgy, hogy a többi adat azonos maradjon.
- Ha az egyes dalok sorrendjét is rögzíteni szeretnénk, akkor egy új mezőt kell hozzáadni a táblázathoz, amely a sorszámot (a sáv sorszámát) tartalmazza. Ebben az esetben a CD azonosítószáma (CD_id) önmagában már nem lesz kulcs, hiszen több sorban azonos az értéke. A (CD_id, sáv) mezőpárt célszerű elsődleges kulcsnak választani.

Ez még nincs 2NF-ben

CD_id	albumcím	előadó	alapítás_éve	sáv	cím
4711	Not That Kind	Anastacia	1999	1	Not That Kind
4711	Not That Kind	Anastacia	1999	2	I'm Outta Love
4711	Not That Kind	Anastacia	1999	3	Cowboys & Kisses
4712	Wish You Were Here	Pink Floyd	1964	1	Shine On You Crazy Diamond
4713	Freak of Nature	Anastacia	1999	1	Paid my Dues

3. táblázat: A CD_dal tábla

Miért nincs?

Az előző tábla nincs második normálformában, mert

- az album címe csak az egyik kulcsmezőtől függ, a másiktól nem.

Ez már 2NF-ben van

Egyik tábla

CD_id	album	előadó	alapítás_éve
4711	Not That Kind	Anastacia	1999
4712	Wish You Were Here	Pink Floyd	1964
4713	Freak of Nature	Anastacia	1999

4. táblázat: CD tábla

Ez már 2NF-ben van

Másik tábla

CD_id	sáv	cím
4711	1	Not That Kind
4711	2	I'm Outta Love
4711	3	Cowboys & Kisses
4712	1	Shine On You Crazy Diamond
4713	1	Paid my Dues

5. táblázat: Dal tábla

Ez még nincs 3NF-ben

CD_id	album	előadó	alapítás_éve
4711	Not That Kind	Anastacia	1999
4712	Wish You Were Here	Pink Floyd	1964
4713	Freak of Nature	Anastacia	1999

6. táblázat: CD tábla

Miért nincs?

Az elsődleges kulcstól függ az album mező, attól pedig az előadó mező, tehát itt tranzitív függőség van.

Ezek már 3NF-ben vannak

CD_id	album	előadó_id
4711	Not That Kind	311
4712	Wish You Were Here	312
4713	Freak of Nature	311

7. táblázat: CD tábla

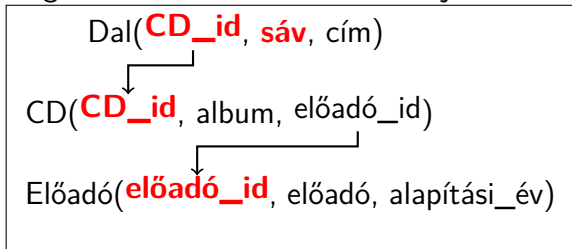
előadó_id	előadó	alapítás_éve
311	Anastacia	1999
312	Pink Floyd	1964

8. táblázat: Előadó tábla

- Az előadó_id bevezetése helyett az előadó nevét is használhatnánk idegen kulcsként, azonban általában célszerűbb egy azonosító számot alkalmazni, mivel kevesebb helyet foglal, mint egy szöveg típusú mező. (Esetleg még kétbetűs rövidítések praktikusak lehetnek idegen kulcsként, mint a mondial adattáblában.)
- Természetesen a kiinduló adatainkat csak a Dal tábla ismeretében tudjuk előállítani. Meggyőződhetünk róla, hogy abban nincsen tranzitív függőség, tehát az is 3NF-ben van.

A következő tábláink vannak

A zárójelen belül a mezők nevei, az egyes táblák elsődleges kulcsai **vastagítva vannak** (írásban aláhúzással célszerű jelölni). A zárójel előtt a tábla neve áll. A nyilak az idegenkulcs-hivatkozásokat mutatják.



Valójában a normalizálást nem szoktuk a konkrét adatokon követni, azt a megértés kedvéért tettük most. A normalizásához a táblák fenti ábrázolása jól használható.

A normalizásán kívül az adatbázis tervezésének másik (később kifejlesztett) módszere az egyed-kapcsolat modellen alapuló.