

Bánné Varga Gabriella

FOGALMI ADATMODELLEZÉS  
ÉS RELÁCIÓS ADATBÁZIS-TERVEZÉS



Bánné Varga Gabriella

FOGALMI  
**ADAT**MODELLEZÉS  
ÉS RELÁCIÓS  
**ADAT**BÁZIS-TERVEZÉS

Elmélet és gyakorlat

© Bánné Varga Gabriella, Typotex, Budapest, 2022

Engedély nélkül semmilyen formában nem másolható!

A könyv megjelenését a Case Partners Kft. és a WEBváltó Kft. támogatta.

Lektorálta: Gollnhofer Gábor

ISBN 978 963 493 189 8

Kedves Olvasó!

Köszönjük, hogy kínálatunkból választott olvasnivalót!  
Újabb kiadványainkról és akcióinkról a [www.typotex.hu](http://www.typotex.hu)  
és a [facebook.com/typotexkiado](https://facebook.com/typotexkiado) oldalakon értesülhet.

Typotex Kiadó

Alapította Votisky Zsuzsa, 1989

A kiadó az 1795-ben alapított Magyar Könyvkiadók  
és Könyvterjesztők Egyesülésének tagja.

Felelős kiadó: Németh Kinga

Felelős szerkesztő: Kótai Kata

Borítóterv és tördelés: Szalay Éva

Készült a Multiszolg Bt. nyomdájában

Felelős vezető: Kajtor Bálint

# Tartalom

Köszönetnyilvánítás .....	11
Előszó.....	13

## I. RÉSZ

<b>FOGALMI ADATMODELLEZÉS .....</b>	<b>17</b>
<b>1. fejezet Bevezetés .....</b>	<b>19</b>
Egy kis történelem.....	19
Az adatmodellezés fogalma, értelmezése .....	21
A fejlesztés 3 szintje .....	23
Barker-féle egyed-kapcsolat modellezés .....	26
<b>2. fejezet Egyedek, tulajdonságok, kapcsolatok.....</b>	<b>27</b>
Egyedek .....	27
Tulajdonságok.....	28
Attribútum vagy egyed?.....	29
Származtatott adatok.....	29
Kapcsolatok.....	30
Kapcsolatok megnevezése .....	30
1:M kapcsolat .....	33
M:N kapcsolat .....	34
Feloldó kapcsolatok.....	34
1:1 kapcsolat .....	35
„Kapcsolatot megvalósító attribútumok” .....	35
Esettanulmány: Nobel-díjasok kezdő modell.....	36
<b>3. fejezet Modell haladó fogalmai .....</b>	<b>43</b>
Egyedi azonosító .....	43
Nem áthelyezhető és korlátozott kapcsolat.....	48
Rekurzív kapcsolatok.....	49
Szupertípusegyedek .....	50
Íves kapcsolatok.....	53
Ortogonalis szubtípusok .....	54
Esettanulmány: Nobel-díjasok továbbfejlesztett modell.....	59

<b>4. fejezet</b>	<b>Egyéb üzleti szabályok és megszorítások .....</b>	<b>65</b>
	Struktúra és megkötések .....	65
	Procedurális vs. definitív ellenőrzés .....	66
	Domain .....	66
	Egyedek definiálása domain érvényes értékekre .....	67
	Domain definiálása szövegesen .....	69
	„Üres” értékek.....	70
	Állapotátmenetek .....	70
	Mátrix.....	70
	Állapotátmeneti diagram.....	71
	Egyedéletrajz .....	72
	Grafikusan nem ábrázolható egyéb megszorítások .....	74
	Peremfeltételek .....	76
	Esettanulmány: Nobel-díjasok modell szöveges megszorítások.....	77
<b>5. fejezet</b>	<b>Idő kezelése és változások modellezése.....</b>	<b>79</b>
	Dátum attribútum és NAP egyed.....	79
	Attribútumok változása, történet egyedek.....	80
	Kapcsolatok változása .....	82
	Szabvány kódok változásának követése.....	83
	Bitemporális modellek.....	86
	Esettanulmány: Nobel-díjasok modell, idő és változások figyelembevételével .....	88
<b>6. fejezet</b>	<b>Fázis vége.....</b>	<b>91</b>
	META modell .....	91
	CASE eszközök.....	93
	Egyed-kapcsolat diagramok elrendezése .....	94
	Csapdák elkerülése.....	98
	Legyezőcsapda .....	98
	Szakadékcsapda .....	100
	Módosítás helye.....	101

## II. RÉSZ

	<b>RELÁCIÓS ADATBÁZIS-TERVEZÉS.....</b>	<b>103</b>
<b>7. fejezet</b>	<b>Adatbázisterv készítése .....</b>	<b>105</b>
	Relációs adatbázis alapfogalmak.....	106
	Táblák.....	106
	Deklaratív megszorítások (Constraintek) .....	107
	Adatbázis-objektumok grafikus ábrázolása.....	109
	A transzformálás alapelve, filozófiája.....	112
	Szabvány névkonvenciók alkalmazása.....	112
	Alapvető transzformációk 1:N kapcsolattal .....	112
	M:N kapcsolat .....	116

	1:1 kapcsolat .....	116
	Ívek .....	117
	Általános ív megoldás .....	118
	Szuper- és szubtípusok transzformálása .....	119
	Szupertípustábla .....	120
	Szubtípustáblák .....	121
	Szuper- és szubtípustáblák együtt (ívmegoldás) .....	121
	Domain kódtáblák megvalósításai .....	124
	Mesterséges egyedi azonosítók .....	126
	Üzleti megszorítások egyéb megoldási lehetőségei .....	127
	Nem áthelyezhető kapcsolat .....	127
	Korlátozott kapcsolat .....	127
	1:M kapcsolatok kötelezősége az 1 oldalon .....	128
	Állapotátmenetek .....	129
	Audit adatok a tervben .....	130
	És mi marad a fizikai adatbázis tervre? .....	130
	Módosításokról még egyszer .....	131
	Esettanulmány: Nobel-díjasok adatbázisterv .....	132
<b>8. fejezet</b>	<b>Normalizálás .....</b>	<b>137</b>
	A normalizálás és célja .....	137
	1. normálforma értelmezési problémák .....	139
	Normálformák .....	139
	Normalizálás helye a modellalkotásban .....	145
	Esettanulmány: Nobel-díjasok listájának normalizálása .....	148
<b>9. fejezet</b>	<b>Denormalizálás .....</b>	<b>153</b>
	Mit miért és mikor denormalizáljunk? .....	153
	Leggyakoribb denormalizálási technikák .....	154
	Szármasztott adatok tárolása .....	154
	Táblák teljes vagy részleges „összekapcsolása” .....	154
	Detail adatok tárolása Master adatokkal .....	155
	„Rövidre záró” kapcsolatok .....	156
	Történeti táblákkal kapcsolatos technikák .....	156
	Aktuális detail adat tárolása Master adatokkal .....	156
	Időtartam vége oszlop bevezetése .....	157
	Aktuális jelző .....	157
	Domain kódtáblák kiiktatása .....	158
	Táblák „összekapcsolása” .....	158
	Érvényes értékek „bedrótozása” .....	158
	Egyéb megoldások .....	159
	Hierarchiaszint-mutató .....	159
	Táblák „kapcsolása” kiegészítő FK oszloppal .....	159
	Esettanulmány: Nobel-díjasok adatbázisterv denormalizálás .....	161

### III. RÉSZ

<b>MODELLEK AZ ADATTÁRHÁZBAN .....</b>	<b>163</b>
<b>10. fejezet Adattárház és dimenzionális modell .....</b>	<b>165</b>
Adattárházak kialakulása .....	165
Az adattárház első szintje .....	167
Dimenzionális modellek az adatpiacokban .....	169
Csillag modell, dimenziók, tényadatok .....	170
Néhány bevett denormalizálás csillag sémában .....	171
<b>11. fejezet Data Vault modell .....</b>	<b>175</b>
Alapvető komponensek .....	175
3NF vs. Csillag modell vs. Data Vault modell .....	178

### IV. RÉSZ

<b>FELADATOK.....</b>	<b>183</b>
<b>12. fejezet Feladatok és megoldások .....</b>	<b>185</b>
Feladat: Heti és napi időjárás-jelentés .....	187
Feladat: Tanfolyami katalógus .....	189
Feladat: Nyugta.....	191
Feladat: Házi könyvtár.....	194
Feladat: Szervezeti egységek .....	197
Feladat: Postai címek .....	198
Feladat: Recepttár .....	205
Továbbfejlesztés: Mértékegységek átváltása .....	210
Feladat: Étlap.....	212
Továbbfejlesztés/1: Menü bevezetése .....	214
Továbbfejlesztés/2: Helyettesítő tétel bevezetése .....	216
Továbbfejlesztés/3: Termékfajta menüételként.....	216
Adatbázis feltöltése .....	219
További megszorítások.....	222
Megszorítások a felhasználói felületen.....	223
Feladat: Pizza étlap normalizálása.....	225
Feladat: Munkatárs tábla normalizálása.....	227
Feladat: Termék tábla denormalizálása .....	228
Feladat: Nobel-díjasok összefüggő specifikáció .....	229
<b>1. MELLÉKLET: Mi a címem? .....</b>	<b>233</b>
A probléma.....	233
Honnan származnak a megváltoztatott adatok? .....	233
Címképzési és adatátvételi folyamat .....	235
Konklúzió .....	236



---

<b>Címregiszter tervezése .....</b>	<b>237</b>
Címregiszter tartalmi specifikáció .....	237
Értelmező rendelkezések.....	238
Címképzés.....	239
Lakóépületek típusai.....	239
Elméleti modell .....	242
Elméleti adatmodell attribútumokkal .....	244
Adatbázisterv.....	246
Címelemek domain tábláinak bevezetése .....	247
Épület típus és Állapot bevezetése .....	249
És a hiányzó üzleti logika.....	249
Címek szabványos megjelenítése .....	250
Tovább is van... .....	250
<b>2. MELLÉKLET: Egyéb modellezési konvenciók.....</b>	<b>251</b>
Chen-féle modell.....	252
Bachman-modell.....	253
Information Engineering/Martin.....	253
<b>Név- és tárgymutató .....</b>	<b>255</b>
<b>Irodalomjegyzék.....</b>	<b>260</b>



# Köszönetnyilvánítás

Köszönetet szeretnék mondani elsősorban a könyv lektorának, Gollnhofer Gábornak, aki rengeteg munkája mellett vállalta a lektorálás időigényes feladatát. Nagy elméleti tudás és komoly gyakorlati tapasztalat birtokában tett javaslatai, megjegyzései sokat segítettek a könyv végleges verziójának kialakításában.

Köszönet illeti továbbá azokat a partnereket, akik megbíztak ilyen jellegű projektjeik megvalósításával; valamint az Oracle-nek és számos oktatási cégnek, hogy alkalmat adott a témában tanfolyamok megtartására Magyarországon és külföldön, és a hallgatóimnak, akik kérdéseikkel, felvetett problémáikkal szintén inspiráltak a könyv megírásában. Úgy gondolom, hogy az elmúlt 25 év, amikor párhuzamosan oktattam a legújabb tananyagokat, alkalmaztam az ott javasoltakat projektjeimben, majd a valós munka tapasztalatait visszaforgattam az oktatásba, szakmai életem leggyümölcsözőbb időszaka volt.



# Előszó

A 2010-es évek elején fogalmazódott meg bennem, hogy sok-sok éves oktatói és fejlesztői munkám összefoglalásaként két könyvet megírok. Kiadómnak is vázoltam a koncepciót: jó lenne egy részletes adatmodellezést tárgyaló kötet, és egy másik az adattárházak tervezéséről.

Direkt fogalmaztam most kicsit lazán, szeretném megmagyarázni az okát. Amikor 30 évvel ezelőtt elkezdtem foglalkozni adatmodellezéssel, adatbázis-tervezéssel, még nem léteztek adattárházak, és nem volt külön modellezési technikájuk. Ezért aztán lazán adatmodellezésről beszéltünk. Később, hogy megkülönböztessük a kétféle megközelítést, beszéltünk már operatív rendszerek modelljeiről, szakmai zsargonban 3NF (3. normálformájú) modellekről, illetve az adattárházi dimenzionális modellekről.

Mivel az említett időben adattárházakat terveztem, sok tapasztalat, gyakorlati eredmény állt rendelkezésemre, ezért úgy döntöttem, hogy ezzel a témával indítsunk. A könyv megjelenése óta eltelt időszakban ismét sok új tapasztalatot gyűjtöttem adattárház témában. Kollégáimnak szoktam mondani: akármennyi tapasztalatom van egy területen, egy új projekt mindig hoz új kihívásokat is a már ismerős rutinszerű megoldások mellett. Első gondolatom az volt, hogy egészítsük ki a 2012-ben kiadott, már régen elfogyott verzióját a könyvnek, és csináljunk egy új kiadást.

Ahogy ezt fontolgattam, eszembe jutott, hogy az adattárházaknál nemcsak az új dimenzionális modelleket tervezzük az elemzésekhez (ez a munka szebbik része), hanem a különböző forrásrendszerekből előbb létre kell hozni egy integrált céges/vállalati adatbázist, amelynek szerkezete az operatív rendszerekéhez hasonló, és finoman szólva is vegyes minőségű adatokat kapunk általában.

Az adattárházaknál így – és nem vagyok egyedül a tapasztalatommal – az adatokat illetően kétféle problémával szoktunk szembesülni:

- az adatszerkezetek zavarosak, redundánsak, rosszul strukturáltak, nem dokumentáltak és
- az adatok minősége sem megfelelő.

Most, hogy erről írok, szívesen nekiállnék furcsa adatanomáliákat bemutatni, de hát ez nem illik egy bevezetőbe, és különben is túl hosszú lenne (3. kötet?).

Az adattárházaknál cél minden adatot betölteni, és megvizsgálni, hogy az eleget tesz-e az üzleti elvárásoknak, s ha nem, regisztrálni a hiba tényét. Így történhetett, hogy az adattárházak mellékleteként előállt általában egy Hiba-adatpiac, amiből részletes elemzések készíthetők a forrásrendszerek minőségéről, az adatok hibáiról. Ezeket megpróbáltuk visszacsatolni a forrásrendszerek gazdáinak, általában

vegyes sikerrel. Voltak olyan hibák, melyek javításához ragaszkodtunk. Közben pedig mondogattuk: ezt is a forrásoldalon kellett volna megoldani...

Emellett az elmúlt években sokszor tapasztaltam új informatikai rendszerek használata során a felhasználói felület logikátlan voltát, valamint hogy megenged hibás adatfelvitelt. Amikor felhasználói felületről írok, akkor csak feltételezem, hogy azok valamilyen szinten tükrözik az adatbázis szerkezetét, mert azokba persze nincs beelátásom.

Így jutottam arra a felismerésre, hogy inkább az elsőnek tervezett kötettel kellene folytatni. Mire a tartalma összeállt, eljött a koronavírus-karantén, amikor a kényszert kihasználtam, és nekifogtam az írásnak. Az elhatározásomban megerősített egy valós életbeli saját problémám, amikor lakásunk címének furcsa, ráadásul nem egyértelmű változásáról értesültem. Próbáltam kielemezni az anomáliákat, nevezetesen, hogy ugyanannak az ingatlannak időben miért volt, illetve van négy különböző címe. Ebből született az 1. melléklet, amely körüljárja a problémát, és próbál megoldást bemutatni egy új cím modellel, ahol az üzleti logikán nagy hangsúly van.

Egy bekezdés erejéig vissza kell térnem az adatmodell kifejezésre, és már az Előszóba egy kis terminológiát becsempészni. Elsőre az *Adatmodellezés és relációs adatbázis-tervezés* címet akartam adni a könyvnek. Aztán rájöttem, hogy adatmodellezés alatt az irodalomban sokan mást és mást írnak le, értenek. Írnak logikai, fizikai, fogalmi modellezésről, elemzésről, relációs adatmodellről, és sok gyakorlati szakember fejében is maximum az adatbázis strukturális tervezése villan be a kifejezés hallatán.

A „legelején” szeretném kezdeni a könyvben, vagyis a fogalmi modellezéssel, amely során a szervezet vagy a vizsgált terület információigényét fogalmazzuk meg **függetlenül** az adatok tárolásától, hozzáférési mechanizmusoktól, a valóságnak megfelelő természetes fogalmakat használva. Egy ilyen modellt „vihetünk tovább”, és átültethetjük az adatbázis logikai szerkezetébe, ahol már figyelembe vesszünk technikai, hatékonysági, hozzáférési követelményeket. Bizonyos esetekben nincs semmi akadálya annak, hogy ez a két szerkezet (szinte) azonos legyen. Esetünkben ez a második szint a relációs adatbázis terve. A harmadik szint a tényleges fizikai adatbázis leírása. Ezzel nem foglalkozunk, itt már a konkrét adatbázis-kezelő tulajdonságai is szerepet játszanak. A modellek megfogalmazására többféle konvenció alakult ki az idők folyamán, és azok részletezettsége sem azonos szintű. Könyvünkben a Richard Barker által kidolgozott jelölésrendszert használjuk, melyet az *Oracle CASE\*Method Entity Relationship Modelling* könyvében publikált először.

A könyv a fentiek alapján több részre tagolódik.

- ◆ Az I. rész foglalkozik az adatmodellezés kialakulásával, a fogalmi adatmodellezés céljaival, feladataival, a különböző szintű modellek értelmezésével, és mint a cím is mondja, részletesen magával a fogalmi adatmodellezéssel.
- ◆ A II. rész mutatja be a relációs adatbázis-tervezést, kezdve azon, hogy hogyan lehet átültetni a fogalmi modellt relációs adatbázistervbe. Megvizsgáljuk, hogyan lehet ellenőrizni a modell minőségét normalizálással, milyen lehetőségek állnak rendelkezésre, ha hatékonysági szempontok miatt el kell térnünk a „tisztá” 3. normálformájú szerkezetektől (denormalizálás). Az intuitív egyed-kapcsolat modellezéssel szembeállítjuk a relációs adatelemzés módszerét és a kétféle megközelítés eredményének összevetését.

- ◆ Az első két részben tárgyalt lépéseket, technikákat egy esettanulmányon vesszük végig. A feladat a Nobel-díjasok nyilvántartásával foglalkozik. Minden fejezet után a feladathoz újabb részleteket hozunk nyilvánosságra (mint egy igazi projekt során), és módosítjuk a modellt az új ismereteknek megfelelően.
- ◆ A III. részben kitekintünk az adattárház-tervezési alapelvekre, röviden bemutatjuk a dimenzionális modellezés lényegét, majd áttekintjük a Data Vault technikát, amelyet az adattárházak építése során használhatunk a dimenzionális adatpiacok kialakítását megelőző adatbázisrétegben.
- ◆ A IV. rész feladatokat tartalmaz. A feladatok különböző fázisokhoz (szintekhez) tartozhatnak, és különböző technikák gyakorlására, bemutatására szolgálnak. Vannak közöttük olyanok is, amelyek a fogalmi modellt és az adatbázis tervét is tartalmazzák. Néhány feladatnál megmutatjuk az adatbázis feltöltését, illetve a deklaratív ellenőrzések érvényesítését adatbázisszinten és felhasználói felület esetén is.
- ◆ A Nobel-díjasok nyilvántartásának specifikációját itt egyben is közöljük, azzal a céllal, ha valaki a nulláról akarja a feladatot megoldani.
- ◆ Az 1. melléklet foglalkozik egy valós életbeli probléma bemutatásával és a lehetséges megoldással.
- ◆ A 2. mellékletben néhány elterjedt modellezési jelölésrendszert mutatunk be.

A könyvben végig hangsúlyosan kezeljük a fogalmi és adatbázis-struktúrák tervezése mellett az adatokra vonatkozó minden ismeret, megkötés megadását, majd megvalósítását. Ezzel szeretnénk az előzőekben felvetett adatminőség-kérdésekre reflektálni. Így számos olyan megoldást, technikát is bemutatunk, amelyek nem részei közvetlenül a modellezés bemutatott módszerének, de az elemzést, tervezést segítik, kiegészítik, főleg az adatminőség kérdésében. Ilyenek például az egyed-életrajzok, az állapotátmeneti diagramok, bitemporális modellek.

Szeretném hangsúlyozni, hogy mind az elméletet, mind a gyakorlatot, a feladatok megoldását nagyon fontosnak tartottuk. Remélem, a két aspektus kellően támogatja egymást végig a könyvben.

Végül el kell még mondanom, hogy első „rendszeremet” én is az adatszerkezetek kialakításával kezdtem, és körülbelül a harmadik adatbázis tervezése során alkalmaztam magától értetődően, módszeresen és könnyedén a könyvben ismerttetett technikákat. Minden egyes új projekt ma is nagy kihívás, amely sok izgalmat, kalandot és komoly felfedezési folyamatot jelent.

Nagyon izgalmas látni, hogyan bomlanak ki a részletek, és kerülnek egymással összhangba. Ebben a grafikus ábrázolási technika rengeteget segít. Szoktam hangsúlyozni, hogy ne mi döntsünk a felhasználó helyett, hanem „ássunk minél mélyebbre”, nyomozzunk kitartóan a tények megismerésében.

Kívánom, hogy olvasóim karanténtól és vírustól szabad körülmények között olvassák a könyvet. Szeretném, ha minél több modellezési, tervezési tanácsot alkalmaznának projektjeikben, és ne adják fel még akkor sem, ha ez esetleg más módszer, mint ahogyan eddig dolgoztak.