



Hálós modell

A 60-as évek közepétől folyamatosan napirenden volt az adatnyilvántartások szabványosítására való törekvés. Mivel ezekben az években a számítógépek alkalmazásának túlnyomó részét az adatbázis-kezelés tette ki, egyre világosabbá vált, hogy az általános célú programnyelvek nem jelentenek megfelelő megoldást, ha nagytömegű adathalmaz feldolgozása a feladat. A hierarchikus és hálós adatmodellek pedig túlságosan nehezékké tették az adatkezelést. Az egységes adatbázis szemlélet kialakulása megkezdődött, de a felmerülő problémák kezeléséhez teljesen új koncepcióra volt szükség.

A relációs adatbázisok elméleti alapelveit elsőként Edward F. Codd, az IBM fejlesztőmérnöke fogalmazta meg egy 1971-ben megjelent tanulmányában. Az új adatmodell matematikai alapokra épült. Codd az azonos tulajdonságokkal jellemezhető egyedek csoportját a halmazelméleti értelemben vett relációként értelmezve, egy táblázatban írja le. Néhány év múlva, 1976-ban P. S. Chen továbbfejlesztette Codd javaslatait, és hamarosan publikálták az új modell alapelveit.

A relációs adatbázis alapfogalmai

A relációs adatbázisban az egy egyedtípushoz tartozó egyedek adatait egy táblázatban tároljuk. Legyen az egyedtípus a könyv:

könyv (ISBN, szerző, cím, kiadó, kiadás éve, példányszám)

a táblázat pedig:

ISBN	Szerző	Cím	Kiadó	Kiadás éve	Példány
963 618 1764	Horváth Tamás	A világháló lehetőségei	ComputerBooks	2016	300
985 561 1234	Varga Tamás	A világ fővárosai	Európa	2017	100
...					
777 231 1231	Kollár István	A filozófia története	Gondolat	2018	234

- **Rekord:** a táblázat egy sora, egy egyed összes tulajdonságát tartalmazza.
- **Mező:** a táblázat egy oszlopa, az összes egyed egy bizonyos tulajdonságát tartalmazza.
- **Mezőérték:** egy cella tartalma, egy egyed egy adott tulajdonsága.
- **Mezőnév:** a táblázat első sora azonosítja az oszlopban megadott tulajdonságokat.
- **Mezőtípus:** egy adott oszlopban csak azonos típusú adatok szerepelhetnek. Például a Szerző oszlop minden cellájába csak szöveget írhatunk.

A fentiek alapján a mező három tulajdonsággal jellemezhető: mezőnév, érték, típus. Az adatok visszakeresése

szempontjából nem szerencsés, ha a táblázat egy cellájában több érték szerepel. Ugyanakkor a könyv egyed típus alapján is kiderül, hogy erre elvileg szükség lehet. Ha például "Az internet világa" c. könyvnek több szerzője van, az nem korrekt megoldás, hogy csak egy szerzőt tárolunk az adott sorban. A korrekt adatokat tartalmazó táblázatnak valahogy így kellene kinéznie:

ISBN	Szerző	Cím	Kiadó	Kiadás éve	Példány
963 618 1764	Kovács Pál Tóth Barna	Az internet világa	ComputerBooks	2016	300
985 561 1234	Varga Tamás	A világ fővárosai	Európa	2017	100
...					
777 231 1231	Kollár István	A filozófia története	Gondolat	2018	234

Ez a megoldás azonban éppen azt rontaná el, ami a relációs adatbázis lényegét jelenti, vagyis azt, hogy egyszerű szerkezetű, áttekinthető táblázatokkal dolgozunk. Az igazi megoldást a későbbiekben tárgyaljuk.

A reláció fogalma

A reláció fogalmát egy példán keresztül mutatjuk be. Jelöljük a véges számú elemeket tartalmazó halmazokat az abc nagybetűivel: A, B, C,... A halmaz elemeit adjuk meg felsorolással, vagy meghatározással. Például:

$$A = \{a, b, c\}$$

$$B = \{1, 2\}$$

Vegyük az összes lehetséges olyan rendezett elempárt, amelynek első eleme az A halmazból, második eleme a B halmazból való. Például: (a,2), (c,1), (c,2),... Az összes lehetséges elempár az A és B halmazok Descartes-féle szorzata (vagy direkt szorzata). Jele: $A \times B$.

A reláció a halmazok Descartes-féle szorzatának valamely részhalmaza. A Descartes-féle szorzatot és a relációt táblázatban könnyebb áttekinteni:

$A \times B$

A	B
a	1

a	2
b	1
b	2
c	1
c	2

Az $A \times B$ halmaz egy részhalmaza: egy reláció.

A	B
a	1
b	1
c	1
c	2

Az egyedtípus megadható a tulajdonsághalmazokkal. Legyen az egyedtípus egy boltban kapható termék:

Termék (kód, megnevezés, egységár, mennyiségi egység, mennyiség)

A tulajdonsághalmazok:

kód: $K = \{1,2,3,\dots\}$

megnevezés: $A = \{\text{Legfeljebb 30 karakter hosszú szöveg}\}$

egységár: $B = \{100000\text{-nél kisebb, pozitív egész szám}\}$

mennyiségi egység: $C = \{\text{db, ív, csomag}\}$

mennyiség: $D = \{1000\text{-nél kisebb pozitív egész szám}\}$

A Descartes-féle halmaz:

$H = A \times B \times C \times D$

A H halmaz egy lehetséges eleme: (1, radír, 120, db, 1500).

A boltban kapható termékek táblázata:

Kód	Megnevezés	Egységár	Mennyiségi egység	Mennyiség
1	radír	120	db	1500
2	toll	520	db	100
3	füzet	55	db	1000
4	csomagoló	35	ív	10000
5	füzet	60	db	1000
6	füzet	60	db	430

A táblázat minden sora eleme a H halmaznak, de ez fordítva nem igaz. A Descartes-féle halmaznak van olyan eleme, amely nem lehet leírása egyetlen terméknek sem. Például: (1856, agdagffgd, csomag, 2, 1). A táblázat sorainak halmaza tehát a H halmaz részhalmaza, vagyis matematikai értelemben reláció. A relációt jellemezhetjük attribútumai felsorolásával:

Termék(Kód, Megnevezés, Egységár, Mennyiségi egység, Mennyiség)

Ezt a jellemzést relációsémának nevezzük. Az adatbázisséma az adatbázist alkotó relációk sémáinak összessége.

Az egyedeket az különbözteti meg egymástól, hogy legalább egy tulajdonságban különböznek egymástól. (Az egyed, az "egyedi"!)

Azt, vagy azokat a tulajdonságokat, amelyek az egyed (rekord) egyértelműen meghatározzák, egyedi azonosítónak nevezzük. Ilyen például az **autó** egyednél a rendszám, a **termék** egyednél a kód. A meghatározásból következik, hogy két különböző egyednek nem lehet azonos az egyedi azonosítója, az azonosító tulajdonság tehát minden rekordban más értéket vesz fel. Az azonosító alapján bármely egyed megkereshetjük a relációban. Az azonosító lehet természetes (rendszám) vagy mesterséges (kód), lehet egyszerű (rendszám), összetett (a személy neve, anyja neve, születési idő, születési hely). A relációkra vonatkozó első kikötés alapján kimondhatjuk, hogy minden relációban van egyedi azonosító (legrosszabb esetben az azonosító a teljes rekord). A reláció sémában az azonosító tulajdonságot aláhúzással jelöljük.

A relációra vonatkozó szabályok

1. A reláció nem tartalmazhat két azonos sort, a reláció tartalmaz azonosítót, amely a sorokat egyértelműen meghatározza
2. A reláció sorainak sorrendje lényegtelen.
3. A reláció oszlopainak sorrendje lényegtelen.
4. A reláció egy adott oszlopába csak az oszlopban tárolt attributum értéktartományába eső adat kerülhet. Az oszlop csak azonos típusú adatokat tartalmaz.

5. A reláció oszlopait az oszlopok neve azonosítja.

Relációs algebra

Mivel a reláció egy halmaz (a tulajdonsághalmazok direkt szorzatának részhalmaza) a relációkkal végezhető műveletek halmazműveletek. A relációkkal végezhető műveletek eredménye is reláció.

Válogatás (szelekció)

A szelekció a rekordok bizonyos szempont szerinti kiválogatását jelenti. A szempont valamely mezőre, vagy mezők kombinációjára vonatkozó feltétel, egy logikai kifejezés.

Az **Egységár < 100** feltétel szerinti szelekció eredménye:

Kód	Megnevezés	Egységár	Mennyiségi egység	Mennyiség
4	füzet	55	db	1000
5	csomagoló	35	ív	10000
6	füzet	60	db	430

A (Megnevezés="füzet") AND (Mennyiség < 500) feltétel szerinti szelekció eredménye:

Megnevezés	Egységár	Mennyiségi egység	Mennyiség
füzet	60	db	430

Vetítés (projekció)

A vetítés a reláció oszlopainak egy részhalmazát adja. Például:

Megnevezés	Mennyiségi egység	Mennyiség
radír	db	1500
toll	db	100
füzet	db	1000
csomagoló	ív	10000
füzet	db	1000
füzet	db	430

A projekció eredménye lehet olyan reláció, amely azonos sorokat tartalmaz. A legtöbb adatbázis-kezelőben paraméterrel megadhatjuk, hogy az azonos sorokat törölje, vagy benne hagyja az eredmény táblázatban.

Egyesítés és metszet

Két azonos sémájú reláció egyesítése az a reláció, amely mindkét reláció sorait tartalmazza. Az eredmény reláció szerkezete értelemszerűen azonos az operandusok szerkezetével. Jelölése: $R1 \cup R2$.

Két azonos sémájú reláció metszete az a reláció, amely csak azokat a sorokat tartalmazza, amely mindkét relációban benne van. Az eredmény reláció szerkezete értelemszerűen azonos az operandusok szerkezetével. Jelölése: $R1 \cap R2$.

Két könyvesboltból érkeznek a megrendelések a központi raktárba. A megrendelések adatait a következő relációséma írja le:

Megrendelés(ISBN, Megrendelő, Mennyiség, Dátum).

R1 és R2 a két bolt megrendeléseit tartalmazó reláció.

R1:

ISBN	Megrendelő	Mennyiség	Dátum
112 242 3334	Barna Mária	3	2018.01.02
811 232 3433	Prompt	4	2018.01.04
111 222 3333	Wagner Bea	1	2018.01.11

R2:

ISBN	Megrendelő	Mennyiség	Dátum
114 142 3334	Farkas Mária	1	2018.01.01
811 232 3433	Prompt	4	2018.01.03
111 222 3333	Wagner Bea	1	2018.01.11

$R1 \cup R2$:

ISBN	Megrendelő	Mennyiség	Dátum
112 242 3334	Barna Mária	3	2018.01.02

811 232 3433	Prompt	4	2018.01.04
111 222 3333	Wagner Bea	1	2018.01.11
114 142 3334	Farkas Mária	1	2018.01.01

R1 \cap R2:

ISBN	Megrendelő	Mennyiség	Dátum
811 232 3433	Prompt	4	2018.01.03
111 222 3333	Wagner Bea	1	2018.01.11

Különbség

Az azonos szerkezetű, R1 és R2 relációk különbsége az a reláció, amelyben az R1 mindazon sorai szerepelnek, amik R2-ben nincsenek benne.

R1-R2:

ISBN	Megrendelő	Mennyiség	Dátum
112 242 3334	Barna Mária	3	2018.01.02

Szorzat (join)

Az R1 x R2 az R1 és R2 relációk sorainak direkt szorzata. A szorzat reláció oszlopainak száma az R1 és R2 relációk oszlopszámának összege, sorainak száma az R1 és R2 reláció sorai számának szorzata. A relációk szorzatát relációk összekapcsolásának is nevezik.

Legyen R1 a könyvek, R2 a megrendelések adatait tartalmazó reláció.

R1:

ISBN	Cím	Kiadó	Kiadás éve
112 242 3334	Napkelte	Európa	2018
811 232 3433	Tenger	Kossuth	2016

111 222 3333	Filozófia	Gondolat	2000
114 142 3334	Matematika	Műszaki	2018

R2:

ISBN	Megrendelő	Mennyiség	Dátum
112 242 3334	Tóth Pál	1	2018.01.01
811 232 3433	Kovács Béla	2	2018.01.04

R1 x R2

R1.ISBN	Cím	Kiadó	Kiadás éve	R2.ISBN	Megrendelő	Mennyiség	Dátum
112 242 3334	Napkelte	Európa	2018	112 242 3334	Tóth Pál	1	2018.01.01
112 242 3334	Napkelte	Európa	2018	811 232 3433	Kovács Béla	2	2018.01.04
811 232 3433	Tenger	Kossuth	2016	112 242 3334	Tóth Pál	1	2018.01.01
811 232 3433	Tenger	Kossuth	2016	811 232 3433	Kovács Béla	2	2018.01.04
111 222 3333	Filozófia	Gondolat	2000	112 242 3334	Tóth Pál	1	2018.01.01
111 222 3333	Filozófia	Gondolat	2000	811 232 3433	Kovács Béla	2	2018.01.04
114 142 3334	Matematika	Műszaki	2018	112 242 3334	Tóth Pál	1	2018.01.01
114 142 3334	Matematika	Műszaki	2018	811 232 3433	Kovács Béla	2	2018.01.04

Ha a szorzat operandusai olyan relációk, amelyek tartalmazznak azonos oszlopokat, akkor a szorzat az oszlopainak nevében jelezni kell, hogy az adott oszlop melyik relációból származik. Megszokott jelölés az, hogy ilyenkor a mezőnevek elé, attól ponttal elválasztva beírjuk a reláció nevét. A fenti példában: R1.ISBN, R2.ISBN. Az utóbbiak úgynevezett minősített nevek.

A szorzat relációnak a fenti példában látszólag nem sok értelme van. Mégis fontos, mert az adatbázis-kezelés egyik leggyakrabban alkalmazott műveletének alapja. A szorzat tartalmaz értelmes adatokat, hiszen azokban a sorokban, ahol az R1 és R2 táblázatban azonosak az ISBN számok, látjuk, hogy mi a címe adott megrendelő által rendelt könyvnek. A relációs adatbázisok egyik legfontosabb művelete, a természetes összekapcsolás, ami a szorzatból vezethető le.

Természetes összekapcsolás (natural join)

Ha az R1 és R2 reláció tartalmaz azonos típusú oszlopot (egy mező mindkét reláció sémájában szerepel), akkor értelmezhetjük a két reláció természetes összekapcsolását. A természetes összekapcsolás a két reláció szorzatának azon sorait tartalmazza, amelyekben a megegyező mező értékei azonosak. Az eredmény reláció a közös mezőt csak egyszer tartalmazza.

Legyen R a könyvek R1 és megrendelők R2 relációjában az ISBN mező közös. Az R1 és R2 relációk természetes összekapcsolása.

R:

ISBN	Cím	Kiadó	Kiadás éve	Megrendelő	Mennyiség	Dátum
112 242 3334	Napkelt e	Európa	1999	Tóth Pál	1	2018.01.01
811 232 3433	Tenger	Kossuth	1998	Kovács Béla	2	2018.01.04