

Adatbázis-kezelő rendszerek I.



A RELÁCIÓS ADATMODELL

Adatmodellek- Bevezetés



Adatmodellezés



- Az **adatmodellezés** a szoftvermérnöki munkában maga az adatmodell létrehozásának *folyamata* egy információs rendszer számára, amely *formális adatmodellező technikákon* alapul.
- Az adatok tárolásának módja meghatározó jelentőségű az adatok hozzáférése és manipulációja szempontjából.
- Egy jó adatmodell olyan, hogy:
 - Könnyen írhatóak hozzá jó, közérthető lekérdezések.
 - A modellezett világ kis mértékű változása nem igényli a modell változását.
 - Ha a modellezett terület nagy mértékben változik, akkor az adatmodellt és a kapcsolódó programokat könnyű módosítani.

Adatbázis – milyen célból?



Adatmodell –milyen célból?

- **OLTP- Online Transaction Processing**
 - Rövid tranzakciók
 - Egyszerű lekérdezések
 - Adatok kis részét érinti
 - Gyakori módosítások
- **OLAP- Online Analytical Processing**
 - Hosszú tranzakciók
 - Összetett lekérdezések
 - Nagytömegű adatot érint
 - Módosítások csak ritkán
- **BigData**

Adatmodellek



- Az **adatmodell** egy eszkörendszer, amely leírja:
 - az adatokat,
 - az adatok kapcsolatait,
 - az adatok szemantikáját,
 - az adatokra vonatkozó korlátozásokat.
- Az adatmodellek fő kategóriái:
 - Koncepcionális (magas szintű, vagy objektum alapú) adatmodellek
 - Logikai (implementációs vagy rekord alapú) adatmodellek
 - Fizikai (alacsony szintű) adatmodellek
- Félig-strukturált adatmodell (XML, JSON)

Koncepcionális adatmodellek

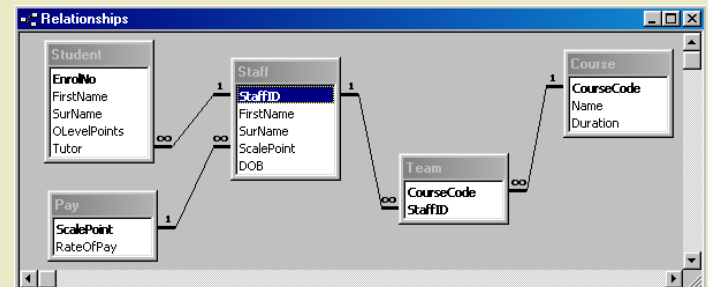


- Olyan fogalmakon alapul, amely közérthetőek.
- Az adatokat fogalmi szinten írja le.
- Meglehetősen rugalmas struktúra.
- Több mint 30 modell, mint pl:
 - Egyed –Kapcsolat modell (Entity-Relationship – ER)
 - Kiterjesztett Egyed-Kapcsolat modell (EER)
 - Adatstruktúra Diagram (Data Structure Diagram - DSD)

Logikai adatmodellek



- Fő alkalmazási területük:
 - Az adatbázis teljes logikai leírása, és
 - az implementáció magas szintű leírása.
- Olyan fogalmak használata, melyek még viszonylag érthetőek, de nem állnak messze a fizikai megvalósítástól sem.
- Legelterjedtebb modellek:
 - Relációs adatmodell
 - Hópehely és csillag séma
 - Objektum-orientált adatmodell
 - Objektum-relációs adatmodell
 - Hálós modell
 - Hierarchikus modell



Fizikai adatmodellek



- Az adattárolás részleteit írják le (pl. indexek, adattípusok, fájlok, korlátozások, ...).
- Nem egy átlagos felhasználónak szól, hanem inkább informatikusoknak.

The screenshot shows the 'Staff : Table' design view in Microsoft Access. The top section is a table with three columns: Field Name, Data Type, and Description. The fields listed are StaffID (Number), FirstName (Text), SurName (Text), ScalePoint (Number), and DOB (Date/Time). Below this table is the 'Field Properties' section, which is divided into 'General' and 'Lookup' tabs. The 'General' tab is active, showing various properties for the selected field (StaffID). The properties include Field Size (Integer), Format, Decimal Places (Auto), Input Mask, Caption, Default Value, Validation Rule (>=1001 And <=1199), Validation Text (Must be >=1001 And <=1199), Required (No), and Indexed (Yes (No Duplicates)). A blue text box on the right side of the Field Properties section provides a note: 'A field name can be up to 64 characters long, including spaces. Press F1 for help on field names.'

Field Name	Data Type	Description
StaffID	Number	StaffIdentifiers (1001 to 1199)
FirstName	Text	PersonNames (String <=12 Characters)
SurName	Text	PersonNames (String <=12 Characters)
ScalePoint	Number	ScalePoints (Number, from 1 to 8)
DOB	Date/Time	StaffBirthDates (Date, >= 21 years before today)

Field Properties

General | Lookup

Field Size: Integer

Format:

Decimal Places: Auto

Input Mask:

Caption:

Default Value:

Validation Rule: >=1001 And <=1199

Validation Text: Must be >=1001 And <=1199

Required: No

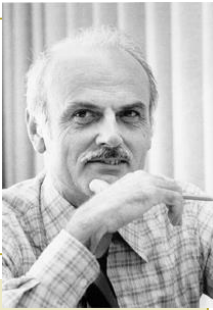
Indexed: Yes (No Duplicates)

A field name can be up to 64 characters long, including spaces. Press F1 for help on field names.

A relációs adatmodell



A relációs adatmodell



- 1970-ben Dr. E.F. Codd vetette fel a relációs adatmodell gondolatát:
 - "A Relational Model of Data for Large Shared Data Banks," Communications of the ACM, June 1970.
- A relációs modell a matematikai *reláció* fogalmán nyugszik.
 - Egy reláció megjelenési formája: *táblázat*.
- *Relációs adatbázis-kezelő rendszer (RDBMS - Relational Database Management System)*: relációs modellen alapuló DBMS
 - Számos kereskedelmi és nyílt forráskódú termék alapul a relációs modell elvén, köztük: IBM's DB2, Oracle Database, Microsoft SQL Server, PostgreSQL, MySQL
 - Ezek közül legtöbb az SQL-t használja adatdefiníciós és adatmanipulációs nyelvként.

Reláció



- Kétdimenziós táblázat, amely **oszlopokból** és **sorokból** áll.
- Az adatbázisban minden **táblázat** egyedi névvel rendelkezik.
- **Rekord = Sor**: egymással kapcsolatban lévő adatok együttese.
 - Minden táblázatban megtalálható egy *elem (oszlop)*, vagy *elemhalmaz (oszlopok)*, melynek értékei egyértelműen azonosítják a sorokat.
 - Esetenként mesterséges *sor azonosítók* vagy *szekvenciális számozás* tölti be az azonosító szerepét.
- **Attribútum = Oszlop**: hivatkozhatunk rá az *attribútum nevével* (oszlop fejléc).

Példa



Táblázat neve

ÜZLET

Attribútumok (vagy oszlopok)

<u>ÜzletID</u>	Város	Utca_hsz
Boo1	Veszprém	Tó u. 11.
Boo2	Balatonfüred	Parti u. 12.
Boo3	Veszprém	Fő tér 3.

*Rekordok
(vagy sorok)*

Domain (értelmezési tartomány)



- Minden attribútum, meghatározott értékkészletből vehet fel értékeket.
- **Domain (értelmezési tartomány)**, amely megadja az **értékkészletet**: az attribútum által felvehető értékek halmaza
 - Atomi értékek
- Például domain lehet:
 - magyar irányítószámok halmaza
 - Magyarországi felsőoktatási intézmények halmaza
 - Fizetés: 67000 és 2000000 közti értékek

Fokszám és kardinalitás



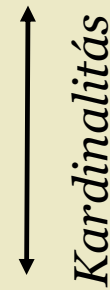
- **Fokszám:** az attribútumok száma
- **Kardinalitás:** a rekordok száma
 - Új rekordok beszúrásával, vagy törlésével változik

ÜZLET

<u>ÜzletID</u>	Város	Utca_hsz
B001	Veszprém	Tó u. 11.
B002	Balatonfüred	Parti u. 12.
B003	Veszprém	Fő tér 3.



Fokszám



Kardinalitás

Descartes-szorzat



- Adott két halmaz : D_1, D_2 .
- **Descartes-szorzat:** D_1 és D_2 halmaz Descartes-szorzatán (vagy direkt szorzatán) azt a halmazt értjük, melynek azon rendezett párok az elemei, amiknek első eleme D_1 -beli, második eleme pedig D_2 -beli és a szorzat minden lehetséges párt tartalmaz. Jelölése: $D_1 \times D_2$
- Példa: $D_1 = \{2, 4\}$, $D_2 = \{1, 3, 5\}$

$$D_1 \times D_2 = \{(2,1), (2,3), (2,5), (4,1), (4,3), (4,5)\}$$

A reláció formális definíciója



Adottak a következő domain-ek: D_1, D_2, \dots, D_n

- Az r **reláció** a D_1, D_2, \dots, D_n domain-eken értelmezett n -ed fokú matematikai reláció, amely az r -et definiáló domain-ek Descartes-szorzatának részhalmaza:

$$r \subseteq D_1 \times D_2 \times \dots \times D_n$$

Így a reláció n elemű sorok halmaza (a_1, a_2, \dots, a_n) , ahol minden $a_i \in D_i$

Formális def. (folyt.)



- Egy R **relációs sémát** a reláció neve (R) és az attribútumok listája határozza meg. Jelölése: $R(A_1, A_2, \dots, A_n)$
 - A relációs séma a relációs struktúráját határozza meg.
 - Példa:

HALLGATÓ(NeptunID, Név, Szüldat, Lakcím)

- **Relációs adatbázis séma:** Relációs sémák olyan halmaza, ahol minden egyes relációs séma egyedi névvel rendelkezik.

A relációk tulajdonságai



- Egy reláció a következő tulajdonságokkal rendelkezik:
 - A reláció neve egy relációs adatbázis sémán belül mindig különbözik a többitől.
 - A relációs egy-egy mezője pontosan egy atomi értéket tárolhat (vagy üres).
 - A relációban minden egyes attribútum egyedi névvel rendelkezik.
 - Egyazon attribútum értékei egyazon domain-ből származnak.
 - A reláció nem tartalmazhat 2 azonos sort.
 - Az attribútumok sorrendje tetszőleges.
 - A rekordok sorrendje tetszőleges.
 - NULL érték kiemelt szerepe (nem adtunk meg semmilyen értéket).

A relációs modell



KULCSOK ÉS KORLÁTOZÁSOK

Kulcsok



- Az R reláció **szuperkulcsa**:
 - Az R reláció SK szuperkulcsa az R reláció olyan attribútumhalmaza, amelyre igaz, hogy bármely érvényes $r(R)$ reláció előforduláson nem létezik két olyan rekord, amely azonos értékeket venne fel az SK -n. Vagyis, bármely egymástól különböző $r(R)$ -beli t_1 és t_2 rekordok esetén $t_1[SK] \neq t_2[SK]$.
 - A szuperkulcs magában foglalhat extra (nem feltétlen szükséges) attribútumokat is.
- **Kulcs** vagy **kulcsjelölt**:
 - Egy minimális szuperkulcs; vagyis olyan K szuperkulcs amelyből tetszőleges attribútumot elvéve a fennmaradó attribútumhalmaz már nem alkot szuperkulcsot.
- Ha egy relációban számos kulcsjelölt van, akkor közülük tetszőlegesen kiválasztva **elsődleges kulcshoz** jutunk. A továbbiakban ezt fogjuk használni a rekordok egyértelmű azonosítására.
 - Jelölése: aláhúzás.

Kulcsok (folyt.)



- **Alternáló kulcs:** Az a kulcsjelölt, amely jó lenne elsődleges kulcsnak, de nem őt választottuk ki.

Bármelyik kulcs lehet egyszerű vagy összetett:

- **Összetett kulcs:** Ha egy kulcs több mint egy attribútumból áll, akkor összetett kulcsnak nevezzük.
- **Egyszerű kulcs:** A kulcsot egyetlen attribútum alkotja

Példa



- Példa: A HALLGATÓ relációs séma:

HALLGATÓ(NeptunID, Szígszám, Név, Telefon, Szüldat, Lakcím)

SK1 = {NeptunID},

SK2 = {Szígszám, Név}

SK3 = {Név, Lakcím, Szüldat},

szuperkulcsok.

- {Szígszám, Név} szuperkulcs de *nem* kulcsjelölt.
- SK3 = {Név, Lakcím, Szüldat} egy összetett kulcs.
- Ha a SK1 = {NeptunID} választjuk az elsődleges kulcsnak, akkor a SK3 = {Név, Lakcím, Szüldat} egy összetett alternáló kulcs.

Példa



CAR	<u>LicenseNumber</u>	EngineSerialNumber	Make	Model	Year
	Texas ABC-739	A69352	Ford	Mustang	96
	Florida TVP-347	B43696	Oldsmobile	Cutlass	99
	New York MPO-22	X83554	Oldsmobile	Delta	95
	California 432-TFY	C43742	Mercedes	190-D	93
	California RSK-629	Y82935	Toyota	Camry	98
	Texas RSK-629	U028365	Jaguar	XJS	98

A CAR reláció 2 kulcsjelölttel :

LicenseNumber és EngineSerialNumber

Az elsődleges kulcs jellemzői



- Egy relációnak mindig van elsődleges kulcsa.
 - Egy reláció nem tartalmazhat két azonos sort, így mindig lehetséges egyértelműen azonosítani az egyes sorokat.
- *Az elsődleges kulcsattribútumot mindig kötelező kitölteni, nem vehet fel NULL értéket.*
 - Ez a sorok egyértelmű azonosításához szükséges.
 - Megj: Egy relációs séma tetszőleges attribútumára meghatározhatjuk a kötelező kitöltést, még akkor is, ha az nem része az elsődleges kulcsnak.
- **Összetett elsődleges kulcs:** az elsődleges kulcsot több attribútum alkotja együttesen

Idegen kulcs



- **Idegen kulcs:** Előfordulhat, hogy egy reláció (R_1) attribútumai között megjelenik egy másik reláció (R_2) elsődleges kulcsa. Az R_1 reláció ezen attribútuma (vagy attribútumai) **idegen kulcs**, amely hivatkozik az R_2 relációra
 - R_1 : **hivatkozó reláció**
 - R_2 : **hivatkozott reláció**
- Az R_1 reláció t_1 rekordja **hivatkozik** az R_2 reláció t_2 rekordjára, ha $t_1[\text{FK}] = t_2[\text{PK}]$.
- A hivatkozási korlátozás (idegen kulcs) jelölése $R_1.\text{FK}$ -tól kiinduló *nyíl* az R_2 reláció felé.
- Az idegen kulcs magába foglalhat több mint egy attribútumot is: ekkor összetett **idegen kulcsnak** nevezzük.

Idegen kulcs – Példa



ÜZLET



<u>ÜzletID</u>	Város	Utca_hsz
Boo1	Veszprém	Tó u. 11.
Boo2	Balatonfüred	Parti u. 12.
Boo3	Veszprém	Fő tér 3.

ALKALMAZOTT

<u>AlkID</u>	Név	Pozíció	Neme	Szüldat	Fizetés	ÜzletID
1	Tóth József	Üzletvezető	F	1951-01-02	200000	Boo1
2	Nagy Piroska	Eladó	N	1970-01-20	96000	Boo1
3	Varja Jenőné	Pénztáros	N	1968-11-23	110000	Boo1
4	Nagy János	Eladó	F	1961-10-10	99500	Boo2
5	Lakat Aranka	Eladó	N	1955-06-12	100200	Boo3

Az ÜZLET relációban az *ÜzletID* **elsődleges kulcs**,
az ALKALMAZOTT relációban a *ÜzletID* attribútum **idegen kulcs**.

Hivatkozó reláció:
ALKALMAZOTT

Hivatkozott reláció:
ÜZLET

Az idegen kulcs értékei (hivatkozási integritási korlátozás)



- Az R_1 idegen kulcsa a következő értékeket veheti fel:
 - (1) az R_2 relációban szereplő egy már meglévő elsődleges kulcs értéke, vagy
 - (2) NULL.
- A (2)-es esetben a FK nem lehet része az R_1 reláció elsődleges kulcsának.
- Ha a hivatkozott elsődleges kulcs összetett, akkor a rá hivatkozó idegen kulcsnak is összetettnek kell lennie!

Korlátozások



- A **korlátozások** olyan *feltételek* amelyeknek *minden érvényes* reláció-előforduláson teljesülniük kell.
 - **Invalid (érvénytelen) státusz:** adatbázis állapot, amelyen nem érvényesül minden korlátozás
- Korlátozások 3 típusa:
 - **Domain** korlátozás
 - **Egyed szintű** korlátozás
 - **Hivatkozási integritás** korlátozás
- Felhasználó által definiált korlátozások
 - ✦ Pl. egy irodában legfeljebb 20 alkalmazott dolgozhat

Korlátozások (folyt.)



- **Domain** korlátozások
 - a reláció minden elemének meg kell felelnie, a saját attribútum korlátozásainak és típusának
 - Pl. Név NOT NULL, Kor ≥ 18 (CHECK)
- **Egyed szintű** korlátozások
 - PRIMARY KEY: Minden táblázat kell hogy rendelkezzen elsődleges kulccsal, ezen elsődleges kulcsnak egyedinek kell lennie és nem vehet fel NULL értéket.
 - UNIQUE korlátozások

Korlátozások (folyt.)



- **Hivatkozási integritás** korlátozás
 - Két reláció között jön létre
 - Csak meglévő, tárolt rekordra lehet hivatkozni (nem hivatkozhatunk olyan rekordra, ami nincs (nem tároljuk))
 - Kapcsolódó technikák: **kaszkádolt módosítás** és **kaszkádolt törlés** (lásd később)

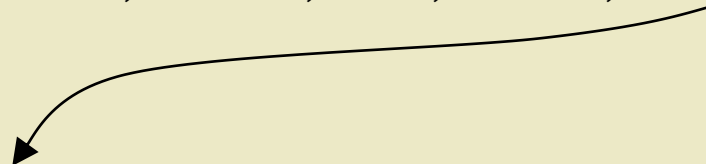
Példa – Hivatkozási integritás



- Tekintsük a következő 2 relációt:
ALKALMAZOTT és OSZTÁLY.
 - Az ALKALMAZOTT táblában az *OsztKód* mező idegen kulcs, amely azon OSZTÁLYRA hivatkozik, ahol az alkalmazott dolgozik.

ALKALMAZOTT (Szicsz, VezNév, KerNév, Szüldat, Lakcím, Nem, Fizetés, OsztKód)

OSZTÁLY (OsztKód, Osztálynév)



Adatmanipulációs műveletek a reláción



- Rekord **BESZÚRÁSA**
- Rekord **TÖRLÉSE**
- Rekord **MÓDOSÍTÁSA**

A korlátozásokat az adatmanipulációs műveletek nem sérthetik meg!

Korlátozások megőrzése



- Ha sérül a korlátozás számos lehetőség van:
 - A művelet visszavonása (REJECT).
 - Járulékos műveletek végrehajtása a korlátozás érvényre juttatásához:
 - ✦ CASCADE lehetőségek,
 - ✦ SET NULL,
 - ✦ SET DEFAULT,
 - ✦ NO ACTION (halasztó hatályú ellenőrzés),
 - ✦ RESTRICT (nem halasztó hatályú ellenőrzés)
 - Felhasználó által definiált korlátozás érvényesítését biztosító rutin futtatása.

Kaskádolási technikák



- ***Kaskádolt törlés:*** Amennyiben a hivatkozott táblában törlődik egy sor, akkor a hivatkozó tábla azon rekordjai is automatikusan törlődnek, amelyek a hivatkozott tábla törölt rekordjára hivatkoztak.
- ***Kaskádolt módosítás:*** Ha a hivatkozott tábla valamely rekordjának elsődleges kulcsa módosul, akkor az erre a rekordra hivatkozó rekord(ok) idegen kulcsának értéke is automatikusan frissül.

Példa



- **Kaskádolt módosítás:** Ha az OSZTÁLY táblában valamely rekord *OsztKód* mezője változik, akkor az azon az osztályon dolgozó összes dolgozó esetében ennek megfelelően frissül az *OsztKód* mező értéke az ALKALMAZOTT táblában.
- **Kaskádolt törlés** Ha az OSZTÁLY táblából törlődik egy rekord (osztály), akkor minden olyan dolgozó adatai is törlődnek (teljes rekordok), akik az az adott osztályon dolgoznak.

ALKALMAZOTT (Szigszám, VezNév, KerNév, Szüldat, Lakcím, Nem, Fizetés, OsztKód)

OSZTÁLY (OsztKód, Osztálynév)

