# Modelul Relațional

#### Modele de date

- Modelul ierarhic (1965)
- Modelul rețea (1965)
- Modelul relațional (1NF) (1970s)
- Model relaţional imbricat (1970s)
- Obiecte complexe (1980s)
- Model obiectual (1980)
- Model relaţional-obiectual (1990s)
- XML (DTD), XML Schema (1990s)

## Model relațional - idei

- Utilizează o structură de date simplă: Tabela
  - simplu de înțeles
  - utilă în modelarea multor situații/entități din lumea reală
  - conduc la interogări de o complexitate redusă
- Utilizeză matematica în descrierea/reprezentarea
   înregistrărilor și a colecțiilor de înregistrări: Relația
  - pot fi modelate formal
  - permit utilizarea de limbaje de interogare formale
  - au proprietăți ce pot fi modelate și demostrate matematic

# Relația - definiție formală

- O relație sau structura unei relații R este o listă de nume de atribute  $[A_1, A_2, ..., A_n]$ .
- **Domeniu** = mulțime de valori scalare (tipuri atomice intreg, text, dată, etc)
- $D_i = Dom(A_i)$  domeniul lui  $A_i$ , i=1..n
- Instanța unei relații ([R]) e o submulțime a

$$D_1 \times D_2 \times ... \times D_n$$

# Relația - definiție formală

- **Grad** (aritate) = numărul tuturor atributelor din structura unei relații
- **Tuplu** = un element al instanței unei relații, o înregistrare. Toate tuplurile unei relații sunt distincte!
- Cardinalitate = numărul tupluri unei relații

## Exemplu de relație

Students(sid:integer; name:string;email:string; age:integer; gr:integer)

field name field type (domain)

sid	name	email	age	gr
2833	Jones	jones@scs.ubbcluj.ro	19	231
2877	Smith	smith@scs.ubbcluj.ro	20	232
2976	Jones	jones@math.ubbcluj.ro	21	233
2765	Mary	mary@math.ubbcluj.ro	22	233

relation
schema
relation
instance

relation tuple

cardinalitate = 4, grad = 5, toate tuplurile distincte!

## Baze de date relaționale

O bază de date este o mulțime de relații

 Structura unei baze de date este mulțimea structurilor relațiilor acesteia

■ Instanța (starea) unei baze de date este mulțimea instanțelor relațiilor acesteia

## Repezentarea grafică a relațiilor

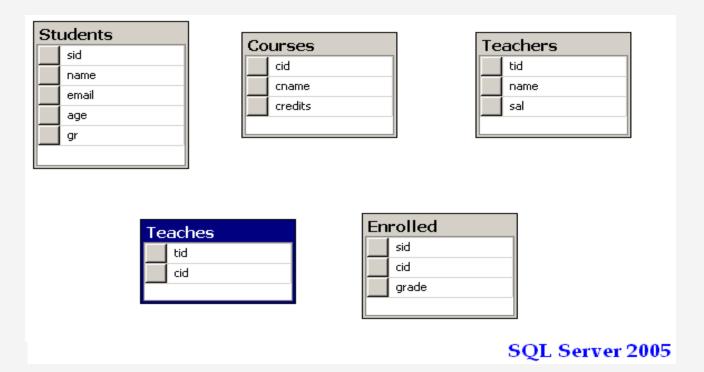
Students(sid:string, name:string, email:string, age:integer, gr:integer)

Courses(cid: string, cname: string, credits:integer)

Enrolled(sid:string, cid:string, grade:double)

Teachers(tid:integer; name: string; sal: integer)

Teaches(tid:integer; cid:string)



## Constrângeri de integritate (CI)

- **CI**: sunt condiții ce trebuie să fie îndeplinite de către *orice* instanță a unei baze de date
  - specificate la momentul definirii structurii relației
  - verificate la modificarea conținutului relației
- O instanță a unei relații că este legală dacă satisface toate CI specificate
  - SGBD nu va permite instanțe ilegale

## Constrângeri de integritate - exemple

- Students(sid:string, name:string, email:string, age:integer, gr:integer)
  - Constrângere de domeniu: *gr:integer*
  - Constrângere de interval: 18 ≤ age ≤ 70
- TestResults(sid:string, TotalQuestions:integer, NotAnswered:integer, CorrectAnswers:integer, WrongAnswers:integer)
  - TotalQuestions = NotAnswered + CorrectAnswers + WrongAnswers nu e o CI!

#### Chei Primare

- O mulțime de atribute reprezintă o cheie a unei relații dacă:
  - 1. Nu există două tuple care au aceleași valori pentru toate atributele **ŞI**
  - 2. Aceste lucru nu este adevărat pentru nici o submuțime a cheii
- Dacă a 2-a afirmație este falsă → **super cheie**
- Daca există >1 cheie pentru o relație → chei candidat
- Una dintre cheile candidat este selectată ca cheie primară

### Chei străine (externe)

- O cheie străină (externă) este o mulțime de câmpuri a unei relații utilizate pentru a `referi' un tuplu al unei alte relații (un fel de `pointer logic').
  - Aceasta trebuie să corespundă cheii primare din a doua relație.

#### De exemplu pentru

Enrolled (sid: string, cid: string, grade: double)

sid este cheie externă referind Students

## Integritate referențială

■ Integritate referențială = nu sunt permise valori pentru cheia străină care nu se regăsesc în tabela referită.

Exemplu de model de date fără integritate referențială:

Link-uri HTML



## Integritate referențială

- Fie *Students* și *Enrolled*; *sid* in *Enrolled* este o cheie străină ce referă o înregistrări din *Students*.
- Adaugarea in *Enrolled* a unui tuplu cu un id de student inexistent, acesta va fi respins de SGBD.

sid	cid	grade	Studer	its			
1234	Alg1	9 •	sid	name	email	age	gr
1235	Alg1	10	1234	John	j@cs.ro	21	331
1234	DB1	10	1235	Smith	s@cs.ro	22	331
1237	DB2	9	1236	Anne	a@cs.ro	21	332

## Integritate referențială

- Dacă o înregistrare din *Students* este ştearsă dar ea este referită din *Enrolled*:
  - se şterg toate înregistrările ce o refera din *Enrolled*.
  - nu se permite ştergerea înregistrării din *Students*
  - sid din *Enrolled* va avea asignată o valoare implicită.
  - sid din *Enrolled* va avea asignată valoarea *null*.

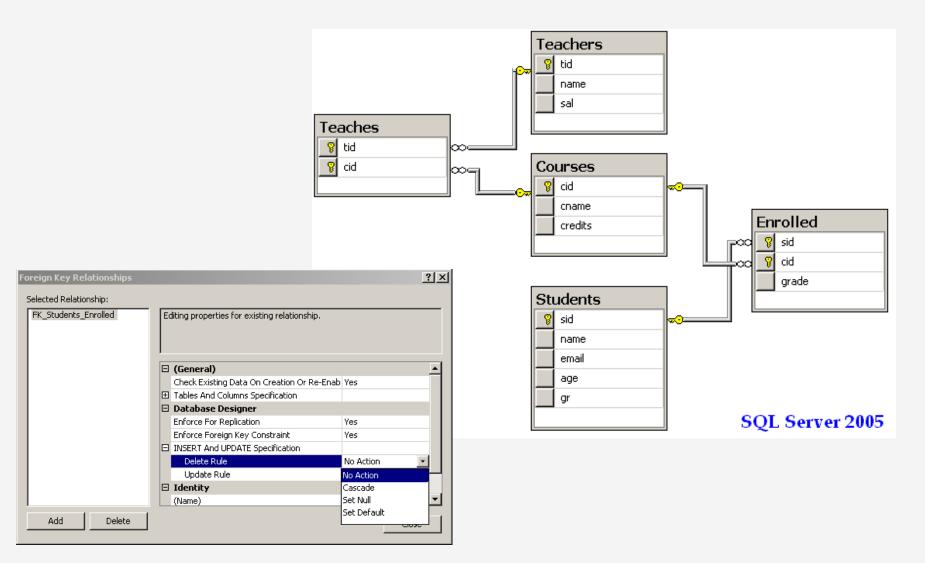
#### **Enrolled**

	sid	cid	grade
$\Rightarrow$	1234	Alg1	9
	1235	Alg1	10
$\Rightarrow$	1234	DB1	10

#### Students

Storter				
sid	name	email	age	gr
1234	John	j@cs.ro	21	331
1235	Smith	s@cs.ro	22	331
1236	Anne	a@cs.ro	21	332

## Repezentarea grafică a CI

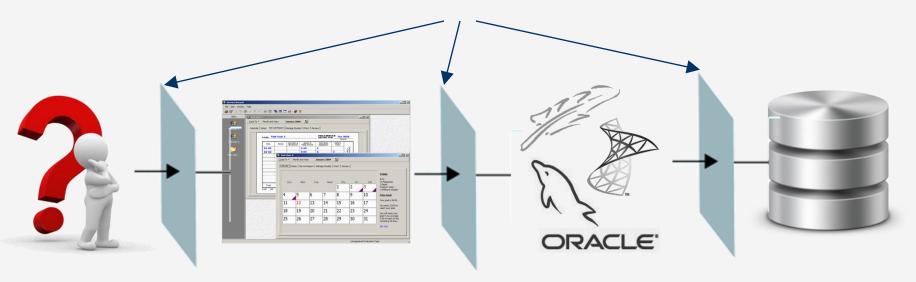


## Cum apar CI?

- CI se bazează pe semantica entităților din lumea reală / conceptuală modelate.
- Putem verifica dacă o CI este incălcată de instanța unei tabele, însă NU vom putea deduce dacă o CI este adevărată doar consultând o singură instanță.
  - OCI se referă la *toate instanțele* posibile ale unei tabele
- Cheile primare şi externe sunt cele mai comune CI;

#### Nivelele de abstractizare

Nivele diferite de abstractizare



User

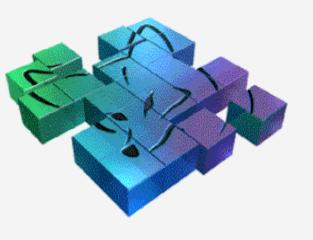
Application software

Database management system

Actual database

## **STUDENT**

Name
Date of birth
Sex
CNP
Group



## Structura fizică

#### Faculty.dbc

		_					
42	53	54	42	20	ze	30	37
20							
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
03	50	81	01	f0	06	4 f	52
4c	45	2d	49	44	01	14	3c
54	2d	4e	41	4d	45	01	13
45	54	24	4e	55	41	42	45

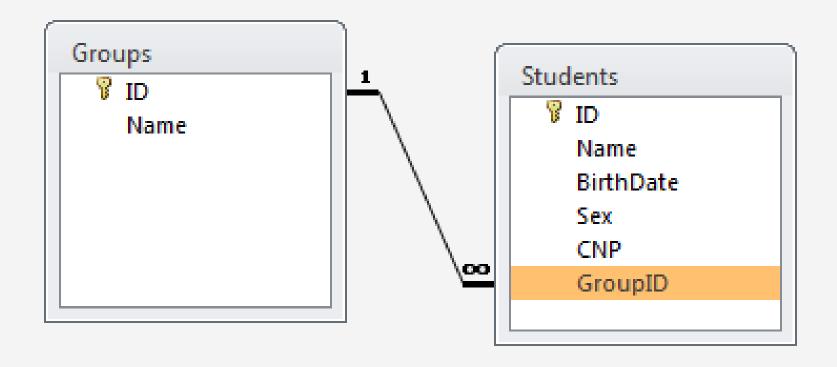
#### Students.dbf

			56	31	2e	30
31	39	32	44	65	66	61
67	65	20	53	65	74	20
20	20	20	20	20	20	20
20	20		20			20
						00
						53
						34
						20
53	54	42	20	2e	30	37
20	20	20	20	20	20	20
20	20	20	20	20	20	20
	20	20	20	20	20	20
50	81	01	f0	06	4 f	52
		49	44	01		3с
		41	4d	45		13
54	24	4e	55	41	42	45
	20 20 41 55 21 53 20 20 45 20 20 45 2d	31 39 67 65 20 20 20 20 00 ff 41 44 55 4d 21 0a 53 54 20 20 20 20 50 81 45 2d 2d 4e	31 39 32 67 65 20 20 20 20 00 ff 01 41 44 45 55 4d 42 21 0a 20 53 54 42 20 20 20 20 20 20 20 20 20 50 81 01 45 2d 49 2d 4e 41	31 39 32 44 67 65 20 53 20 20 20 20 20 20 20 20 00 ff 01 00 41 44 45 52 55 4d 42 45 21 0a 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 50 81 01 f0 45 2d 49 44 2d 4e 41 4d	31 39 32 44 65 67 65 20 53 65 20 20 20 20 20 20 20 20 00 ff 01 00 7c 41 44 45 52 34 55 4d 42 45 52 21 0a 20 20 20 20 20 20 20 20 20 20 20 20 20	31 39 32 44 65 66 67 65 20 53 65 74 20 20 20 20 20 20 20 20 20 20 00 ff 01 00 7c 80 41 44 45 52 34 0f 55 4d 42 45 52 14 21 0a 20 20 20 20 20 20 20 20 20 20 20 20 20

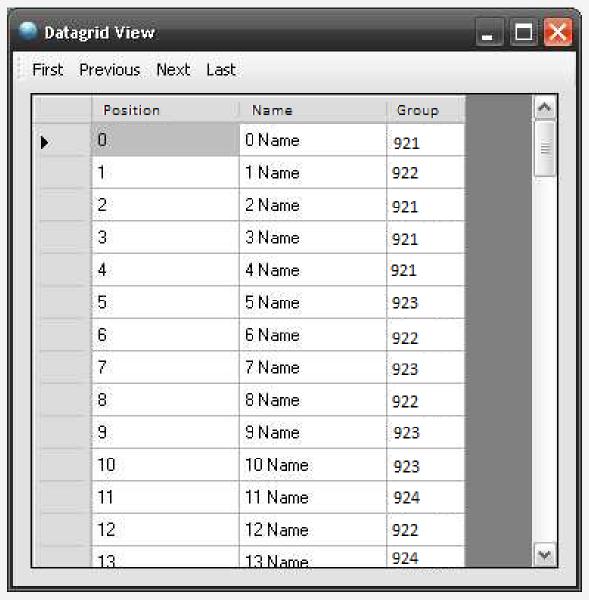
#### Groups.dbf

20	20	20	31	56	31	2e	30
38	31	39	32	44	65	66	61
61	67	65	20	53	65	74	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	00	ff	01	00	7с	80	00
45	41	44	45	52	34	0f	53
4e	55	4d	42	45	52	14	34
34	21	0a	20	20	20	20	20
42	53	54	42	20	2e	30	37
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
03	50	81	01	f0	06	4 f	52
4c	45	2d	49	44	01	14	3c
54	2d	4e	41	4d	45	01	13
45	54	24	4e	55	41	42	45

# Structura conceptuală

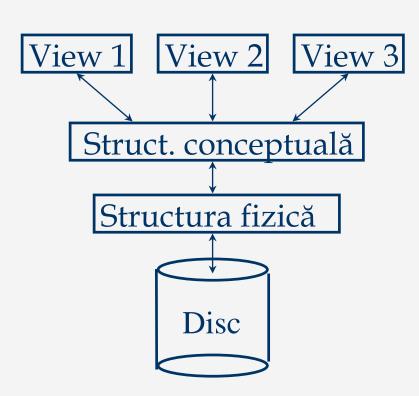


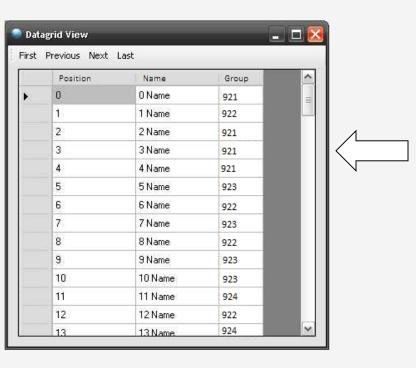
# Vizualizare pentru utilizator

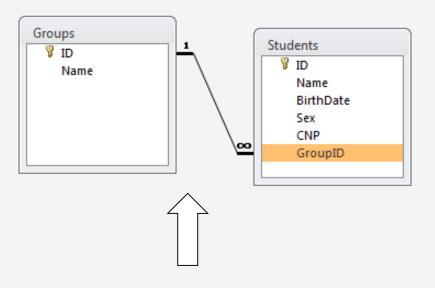


#### Nivelele de abstractizare

- Mai multe <u>structuri externe (views)</u>, câte o singură <u>structură conceptuală (logică)</u> și o <u>structură fizică (internă)</u>.
  - *Views* cum văd utilizatorii datele.
  - *Conceptual -* modelul logic compus din relații, atribute, etc
  - *Fizic* fișierele de date și indecși







#### Faculty.dbc

					_		
42	53	54	42	20	ze	30	37
20	20	20	20	20	2e 20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
03	50	81	01	f0	20 06	4 f	52
4c	45	2d	49	44	01	14	3c
54	2d	4e	41	4d	45 4d	01	13
45	5.4	24	4e	55	4.1	42	45

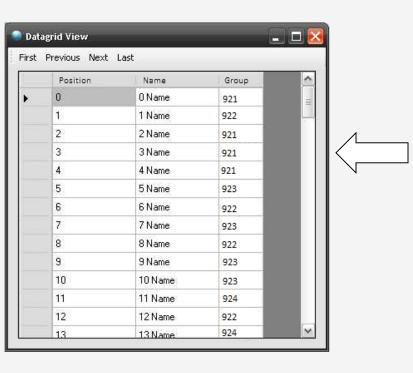
#### Students.dbf

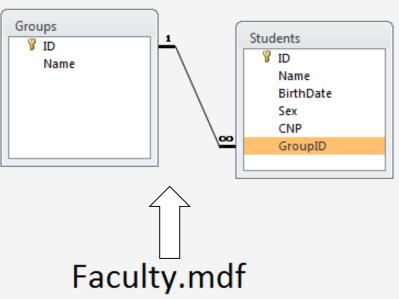
20	20	20	31	56	31	2e	30
38	31	39	32	44	65	66	61
61	67	65	20	53	65	74	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	00	ff	01	00	7с	80	00
45	41	44	45	52	34	0f	53
4e	55	4d	42	45	52	14	34
34	21	0a	20	20	20	20	20
42	53	54	42	20	2e	30	37
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
03	50	81	01	f0	06	4 f	52
4c	45	2d	49	44	01	14	3с
54	2d	4e	41	4d	45	01	13
45	54	24	4e	55	4.1	42	45

#### Groups.dbf

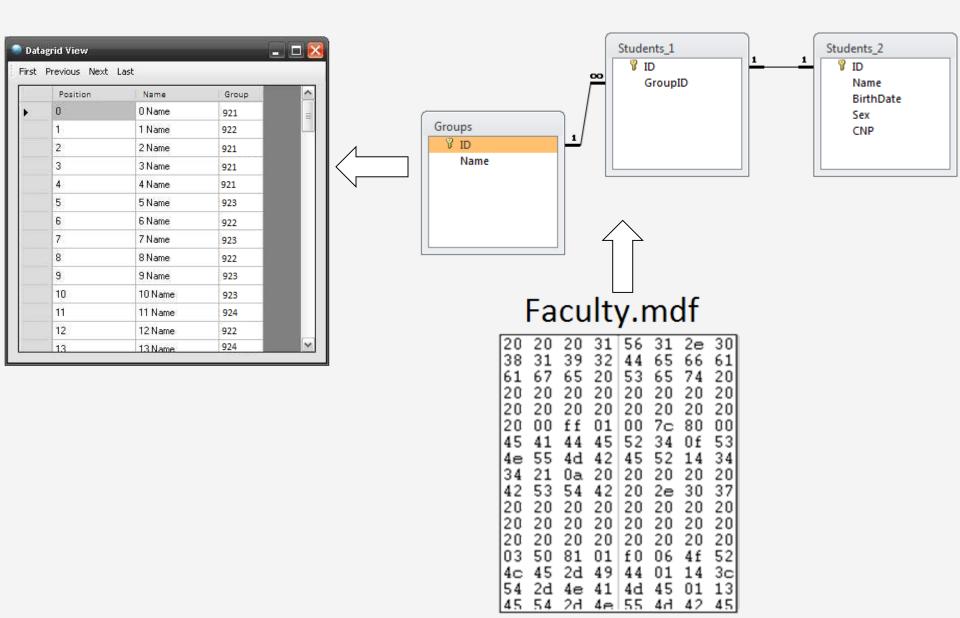
	_		٦٢	٠.,	~~	•	
20	20	20	31	56	31	2e	30
8	31	39	32	44	65	66	61
1	67	65	20	53	65	74	20
20	20	20	20	20	20	20	20
0.9	20	20	20	20	20	20	20
20	00	ff	01	00	7с	80	00
15	41	44	45	52	34	0f	53
lе	55	4d	42	45	52	14	34
34	21	0a	20	20	20	20	20
12	53	54	42	20	2e	30	37
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
3	50	81	01	f 0	06	4 f	52
lc.	45	2d	49	44	01	14	3с
4	2d	4e	41	4d	45	01	13
15	54	24	4e	55	41	42	45

## Independența fizică a datelor





## Independența logică a datelor



### Interogări

- Posibile informații pe care dorim sa le obținem din baza de date anterioară (*Faculty Database*) :
  - Care este numele studentului cu *sid* = 2833?
  - Care este salarul profesorilor care predau cursul *Alg*100?
  - Câți studenți sunt înscriși la cursul *Alg*100?
- Astfel de întrebări referitoare la datele stocate într-un SGBD se numesc *interogări*.
- → limbaj de interogare

## Limbaje SGBD

- Data Definition Language (DDL)
  - Definesc structura **conceptuală**
  - Descriu constrângerile de integritate
  - Influențează **structura fizică** (în anumite SGBD-uri)
- Data Manipulation Language (DML)
  - Operații aplicate instanțelor unei baze de date
  - DML procedural (cum?) vs. DML declarative (ce?)
- Limbaj gazdă
  - Limbaj de programare obișnuit ce permite utilizatorilor să includă comenzi DML în propriul cod

# Limbaje de interogare pentru BD relaționale

SQL (Structured Query Language)
SELECT name FROM Students WHERE age > 20

Algebra

 $\pi_{name}(\sigma_{age > 20} (Students))$ 

Domain Calculus

 $\{\langle X \rangle \mid \exists V \exists Y \exists Z \exists T : Students(V, X, Y, Z, T) \land Z \geq 20\}$ 

T-uple Calculus

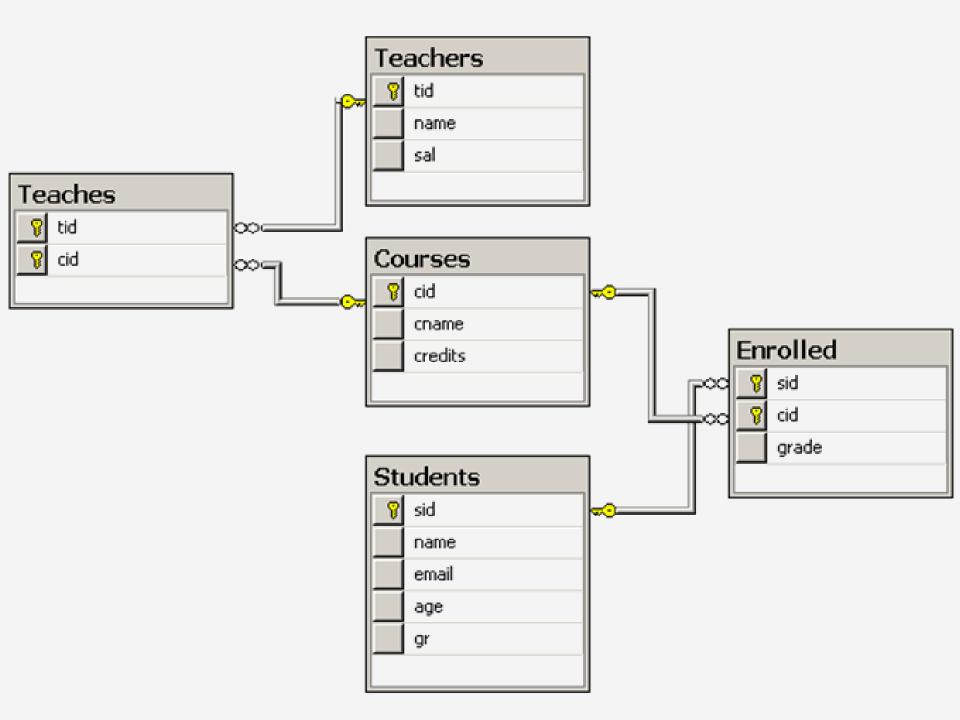
 $\{X \mid \exists Y : Y \in Students \land Y.age > 20 \land X.name = Y.name\}$ 

## Structured Query Language (SQL)

- Dezvoltat de IBM (*system R*) în anii 1970
- Ulterior a apărut nevoia de standardizare
- Standarde (ANSI):
  - SQL-86
  - SQL-89 (minor revision)
  - SQL-92 (major revision) 1,120 pagini
  - SQL-99 (major extensions) 2,084 pagini
  - SQL-2003 (sectiuni SQL/XML) 3,606 pagini
  - SQL-2006
  - SQL-2008
  - SQL-2011

### Nivele SQL

- *Data-definition language (DDL):* 
  - Creare / stergere / modificare *tabele* și *views*.
  - Definire constrangeri de integritate (CI's).
- Data-manipulation language (DML)
  - Permit formularea de interogari
  - Inserare /ştergere / modificare înregistrări.
- Controlul accesului:
  - Asignează sau elimină drepturi de acces si modificare a *tabelelor* și a *view*-urilor.



#### Students

sid	name	email	age	gr
1234	John	j@cs.ro	21	331
1235	Smith	s@cs.ro	22	331
1236	Anne	a@cs.ro	21	332

#### Enrolled

sid	cid	grade
1234	Alg1	9
1235	Alg1	10
1237	DB2	9

#### Courses

cid	cname	credits
Alg1	Algorithms1	7
DB1	Databases1	6
DB2	Databases2	6

## SELECT

#### Studenții cu vârsta de 21 de ani:

```
SELECT *
FROM Students S
WHERE S.age = 21
```

sid	name	email	age	gr
1234	John	j@cs.ro	21	331
1236	Anne	a@cs.ro	21	332

#### Returnează doar numele și adresele de e-mail:

```
SELECT S.name, S.email
FROM Students S
WHERE S.age = 21
```

name	email
John	j@cs.ro
Anne	a@cs.ro

# Interogare SQL simplă

SELECT [DISTINCT] target-list FROM relation-list WHERE qualification

- *relation-list* lista de nume de relații/tabele.
- <u>target-list</u> listă de attribute ale relațiilor din relation-list
- *qualification* comparații logice (Attr *op* const sau Attr1 *op* Attr2, unde *op* is one of <, >, =,  $\le$ ,  $\ge$ ,  $\ne$ ) combinate cu AND, OR sau NOT.
- <u>DISTINCT</u> (optional) indică faptul că rezultatul final nu conține duplicate.

## Evaluare conceptuală

```
SELECT [DISTINCT] target-list
FROM relation-list
WHERE qualification
```

- Calcul produs cartezian al tabelelor din relation-list.
- Filtrare înregistrări ce nu verifică qualifications
- Ştergere atribute ce nu aparțin target-list.
- Dacă **DISTINCT** e prezent, se elimină înregistrările duplicate.

#### 1. PRODUS CARTEZIAN

2. ELIMINA LINII

3. ELIMINA COLOANE

4. ELIMINA DUPLICATE

# Această strategie e doar la nivel conceptual!

Modul actual de evaluare a unei interogări e **mult** optimizat

## Range Variables

SELECT S.name, E.cid FROM Students S, Enrolled E

WHERE S.sid=E.sid AND E.grade=10



SELECT name, cid

FROM Students, Enrolled

WHERE Students.sid=Enrolled.sid

AND grade=10

## Interogare: Studenții care au cel puțin o notă

```
SELECT S.sid
FROM Students S, Enrolled E
WHERE S.sid=E.sid
```

Rezultatul e diferit cu DISTINCT?

■ Ce efect are înlocurea *S.sid* cu *S.sname* în clauza SELECT?

Rezultatul e diferit cu DISTINCT în acest caz?

## Expresii și string-uri

■ Obține triplete (cu vârsta studenților + alte două expresii) pentru studenții al căror nume începe și se termină cu B și conțone cel puțin trei caractere.

```
SELECT S.age, age1=S.age-5, 2*S.age AS age2
FROM Students S
WHERE S.name LIKE 'B_%B'
```

- AS şi = sunt două moduri de redenumire a câmpurilor în rezultat.
- LIKE e folosit pentru comparatii pe siruri de caractere. `\_' reprezinta orice caracter si `%' stands reprezinta 0 sau mai multe caractere arbitrare.

#### INNER JOIN

SELECT S.name, C.cname

FROM Students S,

Enrolled E, Courses C

WHERE S.sid = E.sid

AND E.cid = C.cid

SELECT S.name, C.cname
FROM Students S
INNER JOIN Enrolled E ON
S.sid = E.sid,

INNER JOIN Courses C ON

E.cid = C.cid

#### sid email name age gr 1234 John 21 331 j@cs.ro Students 1235 Smith 22 331 s@cs.ro 1236 a@cs.ro 21 332 Anne

Courses

cid	cname	credits
Alg1	Algorithms1	7
DB1	Databases1	6
DB2	Databases2	6

sid	cid	grade
1234	Alg1	9
1235	Alg1	10
1237	DB2	9



name	cname
John	Algorithms1
Smith	Algorithms1

#### LEFT OUTER JOIN

 Daca dorim sa regasim şi studentii fără nici o notă la vreun curs: SELECT S.name, C.cname
FROM Students S
LEFT OUTER JOIN Enrolled E
ON S.sid = E.sid,
LEFT OUTER JOIN Courses C
ON E.cid = C.cid

#### Students

	sid	name	email	age	gr
	1234	John	j@cs.ro	21	331
S	1235	Smith	s@cs.ro	22	331
	1236	Anne	a@cs.ro	21	332

Courses

	cid	cname	credits
,	Alg1	Algorithms1	7
,	DB1	Databases1	6
	DB2	Databases2	6

sid	cid	grade
1234	Alg1	9
1235	Alg1	10
1237	DB2	9



name	cname
John	Algorithms1
Smith	Algorithms1
Anne	NULL

#### RIGHT OUTER JOIN

Pentru a gasi notele asignate unor studenti inexistenti: SELECT S.name, C.cname

FROM Students S

RIGHT OUTER JOIN Enrolled E

ON S.sid = E.sid,

INNER JOIN Courses C ON

E.cid = C.cid

#### Students

	sid	name	email	age	gr
•	1234	John	j@cs.ro	21	331
5	1235	Smith	s@cs.ro	22	331
	1236	Anne	a@cs.ro	21	332

Courses

cid	cname	credits
Alg1	Algorithms1	7
DB1	Databases1	6
DB2	Databases2	6

sid	cid	grade
1234	Alg1	9
1235	Alg1	10
1237	DB2	9



name	cname
John	Algorithms1
Smith	Algorithms1
NULL	Databases2

#### FULL OUTER JOIN

- LEFT+RIGHT OUTER JOIN
- In majoritatea SGBD OUTERe optional

SELECT S.name, C.cname
FROM Students S
FULL OUTER JOIN Enrolled E
ON S.sid = E.sid,
FULL OUTER JOIN Courses C
ON E.cid = C.cid

Students

	sid	name	email	age	gr
	1234	John	j@cs.ro	21	331
7	1235	Smith	s@cs.ro	22	331
	1236	Anne	a@cs.ro	21	332

Courses

cid	cname	credits
Alg1	Algorithms1	7
DB1	Databases1	6
DB2	Databases2	6

sid	cid	grade
1234	Alg1	9
1235	Alg1	10
1237	DB2	9



name	cname
John	Algorithms1
Smith	Algorithms1
NULL	Databases2
NULL	Databases1
Anne	NULL



A SQL query walks into a bar.

He approaches two tables and says:

"Mind if I join you?"