Relational Algebra

Content from Database system concepts-Korth and Dr. Bruns

Functions on Relations

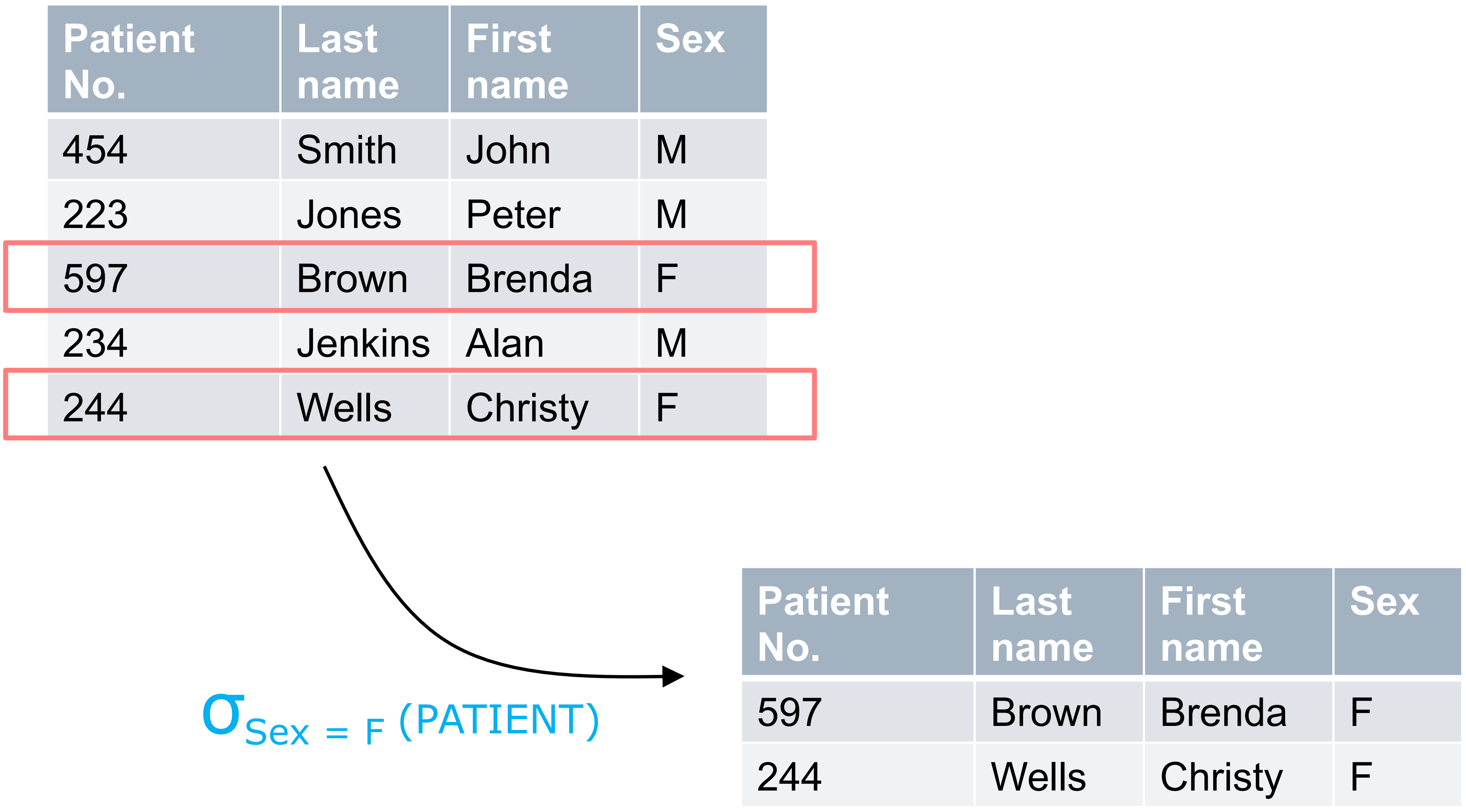
In programming we define objects and methods on those objects.

For example, for strings we have operations like

- substring(string, first, last)
- concat(string1, string2)

In relational databases the objects are relations, and we define functions on relations.

Select σ: filter tuples of a relation



Patient No.	Last name	First name	Sex
454	Smith	John	M
223	Jones	Peter	M
597	Brown	Brenda	F
234	Jenkins	Alan	M
244	Wells	Christy	F

$\sigma_{Sex\ =\ F}$ (PATIENT)

Patient No.	Last name	First name	Sex
597	Brown	Brenda	F
244	Wells	Christy	F

Project Π : slice attributes of relation

Patient No.	Last name	First name	Ward No.
454	Smith	John	6
223	Jones	Peter	8
597	Brown	Brenda	3
234	Jenkins	Alan	7
244	Wells	Chris	6

The operation takes a table as input and produces a table as output

$$\Pi_{\text{“patient no.”,”first name”}}\ (\text{PATIENT})$$

Patient No.	First name
454	John
223	Peter
597	Brenda
234	Alan
244	Chris

Union: add the tuples of two relations

Class_101	Student_1
Class_205	Student_4
Class_205	Student_5

Class_101	Student_2
Class_205	Student_3

Relation1 U Relation2
Expression U Expression

Class_101	Student_1
Class_205	Student_4
Class_205	Student_5
Class_101	Student_2
Class_205	Student_3

1. Duplicate tuples eliminated
2. Schema of relations must match exacly

Question about project

Patient No.	Last name	First name	Ward No.
454	Smith	John	6
223	Jones	Peter	8
597	Brown	Brenda	3
234	Jenkins	Alan	7
244	Brown	Chris	6

What happens if we perform a project on this relation using attribute 'Last name'?

$\pi_{\text{last name}}$ (PATIENT)

Last name
Smith
Jones
Brown
Jenkins
Brown

Problem?

Removing duplicate rows

As part of applying an operation of relation algebra, duplicate rows should be removed.

Duplicate rows aren’t technically allowed in a valid table.

In practice a DB system may allow duplicate rows.

$\Pi_{lastname}$ (PATIENT) will remove duplicate last names from the resultant relation.

Relation Algebra: Sets (duplicates eliminated)

SQL: Multisets, Bags (duplicates allowed)

Answering questions with Project/Select

What are the names of students with more than 100 credits?

To get the answer:

1. select the rows of the ‘student’ table with tot_cred > 100
2. get the ‘name’ attribute

STUDENT

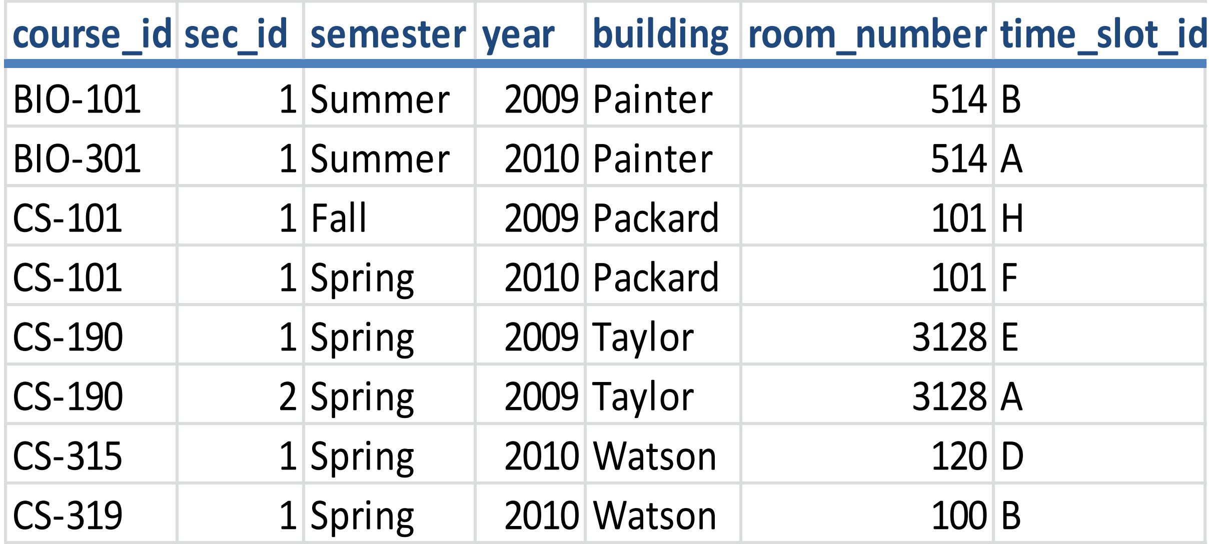
ID	name	dept_name	tot_cred
128	Zhang	Comp. Sci.	102
12345	Shankar	Comp. Sci.	32
19991	Brandt	History	80
23121	Chavez	Finance	110
44553	Peltier	Physics	56
45678	Levy	Physics	46

$\Pi_{\text{Name}} (\sigma_{\text{tot\_cred}>100} (\text{STUDENT}))$

Another example

What courses are offered in Spring '09, and what buildings are they in?

SECTION



course_id	sec_id	semester	year	building	room_number	time_slot_id
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E
CS-190	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B

$$R = \sigma_{semester=Spring\ \wedge\ year=2009}(SECTION)$$

$$\Pi_{course\_id,\ building}(R)$$

Exercise

Write the operations to get the IDs of students who took course CS-101 in Fall 2009.

Remember the structure of the ‘takes’ table:

TAKES

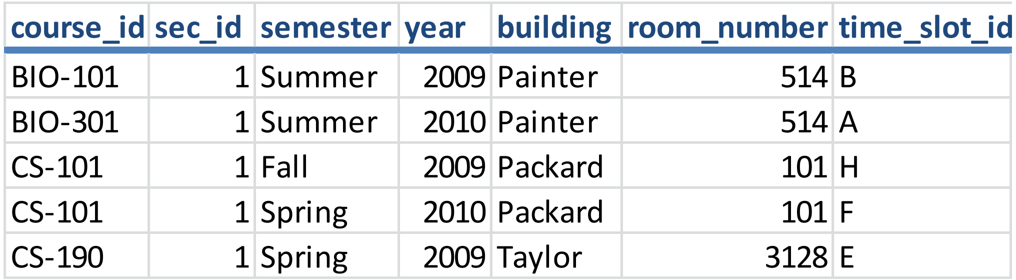
ID	course_id	sec_id	semester	year	grade
128	CS-101	1	Fall	2009	A
128	CS-347	1	Fall	2009	A-
12345	CS-101	1	Fall	2009	C
12345	CS-190	2	Spring	2009	A
12345	CS-315	1	Spring	2010	A
12345	CS-347	1	Fall	2009	A

$$\Pi_{\text{ID}} \; (\sigma_{\text{course\_id=“CS-101” ^ semester=“Fall” ^ year=2009}} \; (\text{TAKES}))$$

A ‘union’ example

Get the ID of courses taught in Fall 2015 or Fall 2016

SECTION



course_id	sec_id	semester	year	building	room_number	time_slot_id
BIO-101	1	Summer	2009	Painter	514	B
BIO-301	1	Summer	2010	Painter	514	A
CS-101	1	Fall	2009	Packard	101	H
CS-101	1	Spring	2010	Packard	101	F
CS-190	1	Spring	2009	Taylor	3128	E

$$R1 = \sigma_{\text{semester=“Fall”} \wedge \text{year=2015}}(\text{SECTION})$$

$$R2 = \sigma_{\text{semester=“Fall”} \wedge \text{year=2016}}(\text{SECTION})$$

$$\Pi_{\text{“course\_ID”}}\ (R1 \cup R2)$$

Last line could alternatively be written:

$$\Pi_{\text{“course\_ID”}}\ (R1) \cup \Pi_{\text{“course\_ID”}}\ (R2)$$

Cartesian Product – combine 2 relations

R1

id	name
1	Jack
2	Sally
3	Trudy

R2

course	id	time
DB	1	10
OS	2	4

Schema of resultant relation is schema(R1) UNION schema(R2)

R1.id	name	course	R2.id	time
1	Jack	DB	1	10
1	Jack	OS	2	4
2	Sally	DB	1	10
2	Sally	OS	2	4
3	Trudy	DB	1	10
3	Trudy	OS	2	4

Question: if you take the product of a table with 5 rows and a table with 8 rows, how many rows do you get?

Using cross product

What are the names of CS instructors and what classes do they teach?

INSTRUCTOR

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000

TEACHES

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009

$$R1 = \text{INSTRUCTOR} \times \text{TEACHES}$$

$$R2 = \sigma_{\text{instructor.ID = teaches.ID} \wedge \text{dept\_name = “Comp. Sci.”}} (R1)$$

$$\Pi_{\text{“name”, “course\_id”}} (R2)$$

Natural Join ⋈

- Enforce equality on all attributes with same name
  - Performs a cross-product
  - Enforces equality on all attributes with the same name
- Eliminates one copy of duplicate attributes
  - Don’t need to keep two copies of the duplicate column because the values are always going to be equal

Using Natural Join ⋈

What are the names of CS instructors and what classes do they teach?

INSTRUCTOR

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000

TEACHES

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009

$$\Pi_{\text{“name”, “course\_id”}} (\sigma_{\text{dept\_name=“Comp. Sci.”}} (\text{INSTRUCTOR} \bowtie \text{TEACHES}))$$

Theta Join $\bowtie_\theta$

- $\text{Exp1} \bowtie_\theta \text{Exp2} = \sigma_\theta(\text{Exp1} \times \text{Exp2})$
  - Performs a cross-product (combine all tuples)
  - Keep tuples that pass θ the condition

Lab – Use Case 1

What are the names of students and instructors?

INSTRUCTOR

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000

STUDENT

ID	name	dept_name	tot_cred
128	Zhang	Comp. Sci.	102
12345	Shankar	Comp. Sci.	32
19991	Brandt	History	80
23121	Chavez	Finance	110
44553	Peltier	Physics	56
45678	Levy	Physics	46

$$\Pi_{\text{“name”}} (\text{INSTRUCTOR}) \cup \Pi_{\text{“name”}} (\text{STUDENT})$$

Lab – Use Case 2

What are the names of students and instructors?

INSTRUCTOR

ID	iName	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000

STUDENT

ID	sName	dept_name	tot_cred
128	Zhang	Comp. Sci.	102
12345	Shankar	Comp. Sci.	32
19991	Brandt	History	80
23121	Chavez	Finance	110
44553	Peltier	Physics	56
45678	Levy	Physics	46

$\Pi_{\text{“iName”}}$ (INSTRUCTOR) $\cup$ $\Pi_{\text{“sName”}}$ (STUDENT) ?

Rename ρ

- Lets say relation R1(C1,…,Cn)
- $\rho_{R2(C1,\ldots,Cn)}$ (R1)
- $\rho_{R2(A1,\ldots,An)}$ (R1)
- $\rho_{R3(A1,\ldots,An)}$ (Exp)
  - Exp is relation algebra expression
  - Remember: result of expression is relation
- Rename operator reassigns the schema in the result of Exp

Lab – Use Case 3

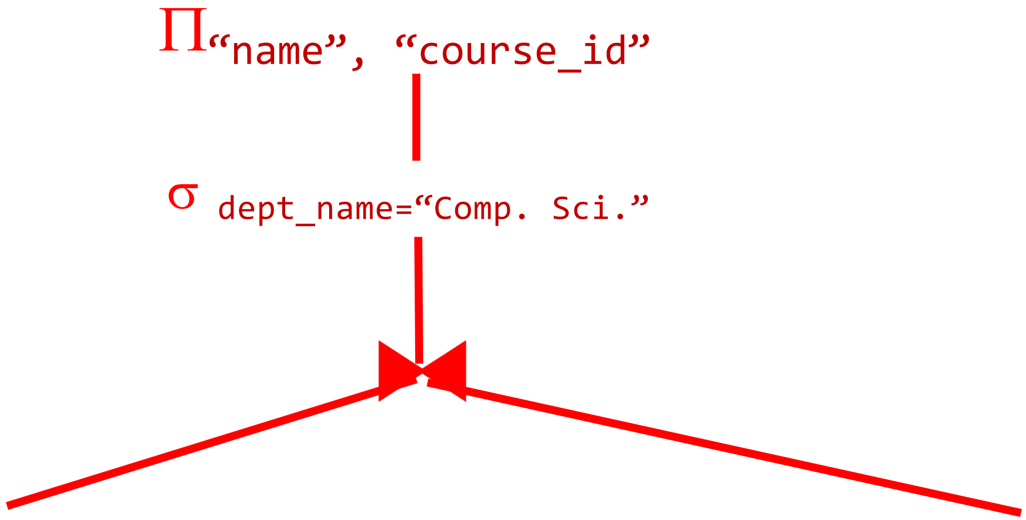
Employees (Employee_ID, Name, Salary, Manager_ID)

Find name of each employee’s manager?

$$\Pi_{\text{“ename”,”mname”}} (\sigma_{\text{eMID = mID}} (\rho_{\text{emp(eID,eName,eMID)}}(\text{Employees}) \times \rho_{\text{mgr(mID,mName,mMID)}}(\text{Employees})))$$

Expression Tree

SQL is compiled into an expression tree


$\Pi_{\text{“name”, “course\_id”}}$
σ dept_name=“Comp. Sci.”
$\bowtie$

INSTRUCTOR

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000

TEACHES

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-101	1	Spring	2010
15151	MUS-199	1	Spring	2010
22222	PHY-101	1	Fall	2009