Databases – Introduction to Relational Model (Chapter 2)

Jaeyong Choi
Dept. of AI-Software, Gachon University

Relational Algebra

- A **procedural language** consisting of **a set of operations** that take **one or two relations as input** and **produce a new relation as their result**.

- **Six basic operators**
  - select: σ
  - project: $\prod$
  - union: $\cup$
  - set difference: –
  - Cartesian product: x
  - rename: ρ

- **Additional operators**
  - intersection: $\cap$
  - join: $\bowtie$
  - Division : /

Selection Operation

- ***Select*** – selection of rows (tuples)
  - Syntax: $\sigma_{\theta}(r)$ (θ: condition)

A	B	C	D
a	a	1	7
a	b	5	7
b	b	12	3
b	b	23	10

Relation *r*

A	B	C	D
a	a	1	7
b	b	23	10

$\sigma_{A=B \wedge D>5}(r)$

Conjunction (and): ∧

Disjunction (or): ∨

Negation (not): ¬

Implication (if..then): →

Equivalence (if and only if): ↔

Selection Operation

- We allow comparisons using
  - $=, \neq, >, \geq. <. \leq$ in the selection predicate.

- We can combine several predicates into a larger predicate by using the connectives:
  - $\wedge$ (and), $\vee$ (or), $\neg$ (not)

Selection Operation

- select those tuples of the instructor relation where the instructor is in the "Physics" department.
- Query

$$\sigma_{\text{dept\_name} = \text{"Physics"}}(\text{instructor})$$

$\sigma_\theta(r)$

- Result

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Instructor

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

Selection Operation

- Find the instructors in Physics with a salary greater $90,000 σ [ ]
- Find all departments whose name is the same as their building name: σ [ ]

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
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Instructor

dept_name	*building*	*budget*
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

Department

Projection Operation

- ***Project*** – selection of columns (attributes)
  - Syntax: $\Pi_A(r)$ (*A*: attributes)
  - Deletes attributes that are not in *projection* list
    - Eliminate *duplicates*

A	B	C
a	10	1
a	20	1
b	30	1
b	40	2

Relation *r*

A	C
a	1
a	1
b	1
b	2

=

A	C
a	1
b	1
b	2

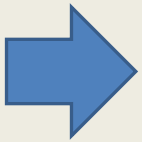
$\Pi_{A,C}(r)$

Projection Operation

- Eliminate the *dept_name* attribute of instructor
- Query: Π
- Result:

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
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

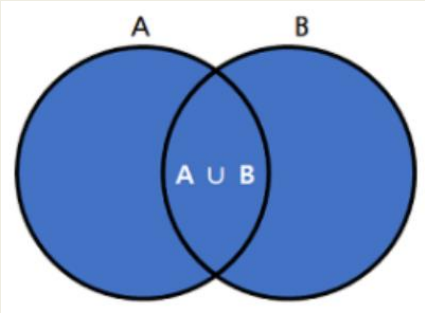
Instructor



ID	*name*	*salary*
10101	Srinivasan	65000
12121	Wu	90000
15151	Mozart	40000
22222	Einstein	95000
32343	El Said	60000
33456	Gold	87000
45565	Katz	75000
58583	Califieri	62000
76543	Singh	80000
76766	Crick	72000
83821	Brandt	92000
98345	Kim	80000

Union operation

- ***Union*** of two relations



A	B
a	1
a	2
b	1

Relation *r*

A	B
a	2
b	3

Relation *s*

A	B
a	1
a	2
b	1
b	3

$r \cup s$

- *Union compatibility*
  - *r* and *s* are *union-compatible*, if they have the same # of attributes and each attribute is from the same domain

Union Operation

- Find all courses taught in the Fall 2017 semester, or in the Spring 2018 semester, or in both

$\sigma_\theta(r)$ $\Pi_A(r)$

$$\Pi_{\text{course\_id}}(\sigma_{\text{semester= ``Fall'' } \wedge \text{ year=2017}}(\text{section})) \cup \Pi_{\text{course\_id}}(\sigma_{\text{semester= ``Spring'' } \wedge \text{year=2018}}(\text{section}))$$

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2017	Painter	514	B
BIO-301	1	Summer	2018	Painter	514	A
CS-101	1	Fall	2017	Packard	101	H
CS-101	1	Spring	2018	Packard	101	F
CS-190	1	Spring	2017	Taylor	3128	E
CS-190	2	Spring	2017	Taylor	3128	A
CS-315	1	Spring	2018	Watson	120	D
CS-319	1	Spring	2018	Watson	100	B
CS-319	2	Spring	2018	Taylor	3128	C
CS-347	1	Fall	2017	Taylor	3128	A
EE-181	1	Spring	2017	Taylor	3128	C
FIN-201	1	Spring	2018	Packard	101	B
HIS-351	1	Spring	2018	Painter	514	C
MU-199	1	Spring	2018	Packard	101	D
PHY-101	1	Fall	2017	Watson	100	A

Figure 2.6 The *section* relation.

course_id
CS-101
CS-315
CS-319
CS-347
FIN-201
HIS-351
MU-199
PHY-101

Difference operation

- ***Difference*** of two relations

A	*B*
a	1
a	2
b	1

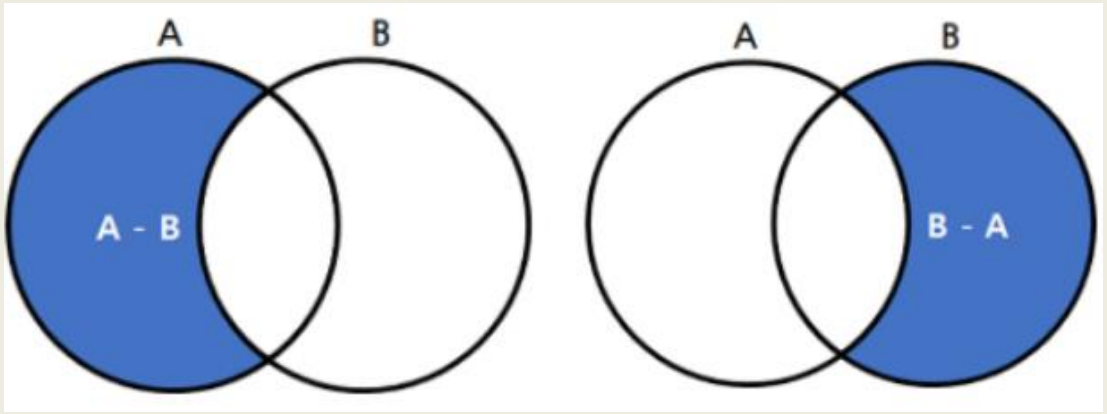
Relation *r*

A	*B*
a	2
b	3

Relation *s*

A	*B*
a	1
b	1

$r - s$



Intersection operation

- ***Intersection*** of two relations
  - Note: $r \cap s = r - (r - s)$

A	B
a	1
a	2
b	1

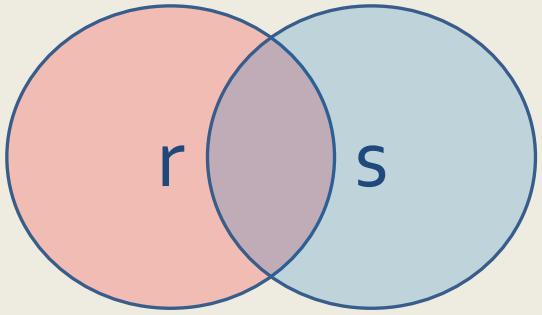
Relation *r*

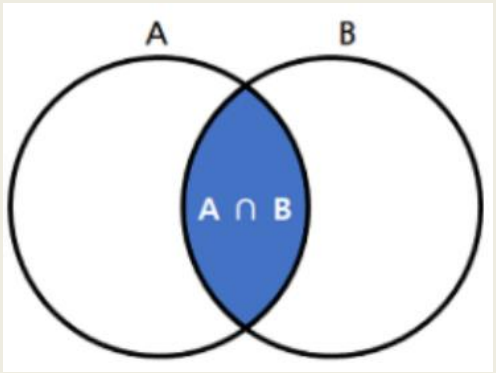
A	B
a	2
b	3

Relation *s*

A	B
a	2

$r \cap s$





Set-Intersection Operation

- Find the set of all courses taught in both the Fall 2017 and the Spring 2018 semesters.

$$\Pi_{\text{course\_id}} (\sigma_{\text{semester= ``Fall'' } \wedge \text{year=2017}} (\text{section})) \cap \Pi_{\text{course\_id}} (\sigma_{\text{semester= ``Spring'' } \wedge \text{year=2018}} (\text{section}))$$

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
BIO-101	1	Summer	2017	Painter	514	B
BIO-301	1	Summer	2018	Painter	514	A
CS-101	1	Fall	2017	Packard	101	H
CS-101	1	Spring	2018	Packard	101	F
CS-190	1	Spring	2017	Taylor	3128	E
CS-190	2	Spring	2017	Taylor	3128	A
CS-315	1	Spring	2018	Watson	120	D
CS-319	1	Spring	2018	Watson	100	B
CS-319	2	Spring	2018	Taylor	3128	C
CS-347	1	Fall	2017	Taylor	3128	A
EE-181	1	Spring	2017	Taylor	3128	C
FIN-201	1	Spring	2018	Packard	101	B
HIS-351	1	Spring	2018	Painter	514	C
MU-199	1	Spring	2018	Packard	101	D
PHY-101	1	Fall	2017	Watson	100	A

Figure 2.6 The *section* relation.

course_id
CS-101

Cartesian product

- Joining two relations – ***Cartesian product***

A	B
a	1
b	2

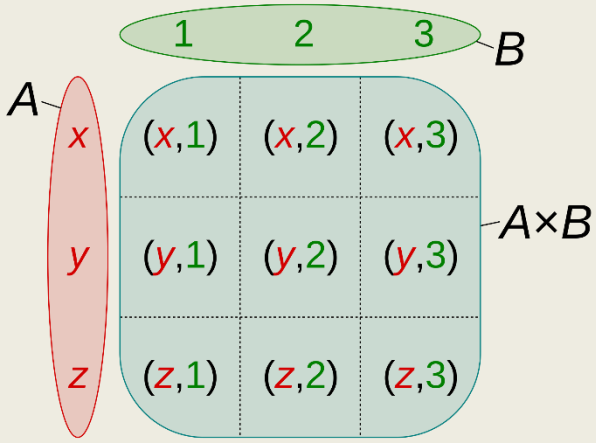
Relation *r*

C	D	E
a	10	x
b	10	x
b	20	y
c	10	y

Relation *s*

A	B	C	D	E
a	1	a	10	x
a	1	b	10	x
a	1	b	20	y
a	1	c	10	y
b	2	a	10	x
b	2	b	10	x
b	2	b	20	y
b	2	c	10	y

$r \times s$



Cartesian product

- Cartesian product – naming issue

A	B
a	1
b	2

Relation *r*

A	D	E
a	10	x
b	10	x
b	20	y
c	10	y

Relation *s*

rAA	B	sAA	D	E
a	1	a	10	x
a	1	b	10	x
a	1	b	20	y
a	1	c	10	y
b	2	a	10	x
b	2	b	10	x
b	2	b	20	y
b	2	c	10	y

$r \times s$

Renaming a table

- ***Renaming*** a table
  - Allows us to refer to a relation by more than one name
  - Syntax: $\rho_x(E)$ – returns the expression E under the name X

A	B
a	1
b	2

Relation r

r.A	r.B	s.A	s.B
a	1	a	1
a	1	b	2
b	2	a	1
b	2	b	2

$r \times \rho_s(r)$

$$\rho_{(r.A,r.B,s.A,s.B)}(r \times s)$$

Composition Operation

- ***Composition*** of operations
  - Can build expressions using multiple operations
  - Note: the result of an operation is a table

r.A	B	s.A	D	E
a	1	a	10	x
a	1	b	10	x
a	1	b	20	y
a	1	c	10	y
b	2	a	10	x
b	2	b	10	x
b	2	b	20	y
b	2	c	10	y

$r \times s$

r.A	B	s.A	D	E
a	1	a	10	x
b	2	b	10	x
b	2	b	20	y

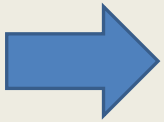
$\sigma_{r.A=s.A}(r \times s)$

Composition Operation

- Find the names of all instructors in the Physics department.

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Instructor



Name
Einstein
Gold

Natural join

- Joining two relations – *Natural join*
  - Let r and s be relations on schemas R and S respectively
  - The "natural join" of relations r and s is a relation on schema $R \cup S$ obtained as follows:
    - Consider each pair of tuples t_r from r and t_s from s
    - If t_r and t_s have the same value on each of the attributes in $R \cap S$, add a tuple t to the result
    - t has the same value as t_r on R; t has the same value as t_s on S
  - *c.f.* theta join

Natural join

- Natural join example 1

$$r \bowtie s = \Pi_{A,B,D,E}\left(\sigma_{r.A=s.A}(r \times s)\right)$$

A	B
a	1
b	2

Relation *r*

A	D	E
a	10	x
b	10	x
b	20	y
c	10	y

Relation *s*

r.A	B	s.A	D	E
a	1	a	10	x
a	1	b	10	x
a	1	b	20	y
a	1	c	10	y
b	2	a	10	x
b	2	b	10	x
b	2	b	20	y
b	2	c	10	y

$r \times s$

r.A	B	s.A	D	E
a	1	a	10	x
b	2	b	10	x
b	2	b	20	y

$\sigma_{r.A=s.A}(r \times s)$

A	B	D	E
a	1	10	x
b	2	10	x
b	2	20	y

$r \bowtie s$

Natural join

- Natural join example 2

A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

r

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ϵ

s

$r \bowtie s$

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	γ
δ	2	β	b	δ

Natural join

- To get only those tuples of *"instructor X teaches"* that pertain to instructors and the courses that they taught

  $\sigma_{\text{instructor.id} = \text{teaches.id}}$ (instructor x teaches)

- Can equivalently be written as with natural join

  $instructor \bowtie_{Instructor.id = teaches.id} teaches$.

- The result of this expression, shown in the next slide

instructor

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

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-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

The *instructor* x *teaches* table

instructor.ID	*name*	*dept_name*	*salary*	*teaches.ID*	*course_id*	*sec_id*	*semester*	*year*
10101	Srinivasan	Comp. Sci.	65000	10101	CS-101	1	Fall	2017
10101	Srinivasan	Comp. Sci.	65000	10101	CS-315	1	Spring	2018
10101	Srinivasan	Comp. Sci.	65000	10101	CS-347	1	Fall	2017
10101	Srinivasan	Comp. Sci.	65000	12121	FIN-201	1	Spring	2018
10101	Srinivasan	Comp. Sci.	65000	15151	MU-199	1	Spring	2018
10101	Srinivasan	Comp. Sci.	65000	22222	PHY-101	1	Fall	2017
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
12121	Wu	Finance	90000	10101	CS-101	1	Fall	2017
12121	Wu	Finance	90000	10101	CS-315	1	Spring	2018
12121	Wu	Finance	90000	10101	CS-347	1	Fall	2017
12121	Wu	Finance	90000	12121	FIN-201	1	Spring	2018
12121	Wu	Finance	90000	15151	MU-199	1	Spring	2018
12121	Wu	Finance	90000	22222	PHY-101	1	Fall	2017
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
15151	Mozart	Music	40000	10101	CS-101	1	Fall	2017
15151	Mozart	Music	40000	10101	CS-315	1	Spring	2018
15151	Mozart	Music	40000	10101	CS-347	1	Fall	2017
15151	Mozart	Music	40000	12121	FIN-201	1	Spring	2018
15151	Mozart	Music	40000	15151	MU-199	1	Spring	2018
15151	Mozart	Music	40000	22222	PHY-101	1	Fall	2017
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
22222	Einstein	Physics	95000	10101	CS-101	1	Fall	2017
22222	Einstein	Physics	95000	10101	CS-315	1	Spring	2018
22222	Einstein	Physics	95000	10101	CS-347	1	Fall	2017
22222	Einstein	Physics	95000	12121	FIN-201	1	Spring	2018
22222	Einstein	Physics	95000	15151	MU-199	1	Spring	2018
22222	Einstein	Physics	95000	22222	PHY-101	1	Fall	2017
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

$$\sigma_{instructor.id = teaches.id}(instructor \times teaches)$$

$$instructor \bowtie_{Instructor.id = teaches.id} teaches.$$

instructor.ID	*name*	*dept_name*	*salary*	*teaches.ID*	*course_id*	*sec_id*	*semester*	*year*
10101	Srinivasan	Comp. Sci.	65000	10101	CS-101	1	Fall	2017
10101	Srinivasan	Comp. Sci.	65000	10101	CS-315	1	Spring	2018
10101	Srinivasan	Comp. Sci.	65000	10101	CS-347	1	Fall	2017
12121	Wu	Finance	90000	12121	FIN-201	1	Spring	2018
15151	Mozart	Music	40000	15151	MU-199	1	Spring	2018
22222	Einstein	Physics	95000	22222	PHY-101	1	Fall	2017
32343	El Said	History	60000	32343	HIS-351	1	Spring	2018
45565	Katz	Comp. Sci.	75000	45565	CS-101	1	Spring	2018
45565	Katz	Comp. Sci.	75000	45565	CS-319	1	Spring	2018
76766	Crick	Biology	72000	76766	BIO-101	1	Summer	2017
76766	Crick	Biology	72000	76766	BIO-301	1	Summer	2018
83821	Brandt	Comp. Sci.	92000	83821	CS-190	1	Spring	2017
83821	Brandt	Comp. Sci.	92000	83821	CS-190	2	Spring	2017
83821	Brandt	Comp. Sci.	92000	83821	CS-319	2	Spring	2018
98345	Kim	Elec. Eng.	80000	98345	EE-181	1	Spring	2017

÷ Division Operation Example

- Retrieve the studnos of students who are enrolled on all the courses that Capon lectures on
- Small_ENROL ÷ Capon_TEACH

Small_ENROL

studno	courseno
s1	cs250
s1	cs260
s1	cs280
s2	cs250
s2	cs270
s3	cs270
s4	cs280
s4	cs250
s6	cs250

÷

Capon_TEACH

courseno
cs250
cs280

result
s1
s4

- Notes on relational languages
  - Each query input is a table (or set of tables)
  - Each query output is a table
  - All data in the output table appears in one of the input tables

Equivalent Queries

- There is more than one way to write a query in relational algebra.
- Example: Find instructors in the Physics department with salary greater than 90,000

ID	*name*	*dept_name*	*salary*
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

Equivalent Queries

- There is more than one way to write a query in relational algebra.
- Example: Find information about courses taught by instructors in the Physics department
- Query 1

$$\sigma_{\text{dept\_name= ``Physics''}} (\text{instructor}) \bowtie_{\text{instructor.ID = teaches.ID}} \text{teaches}$$

- Query 2

$$\sigma_{\text{dept\_name= ``Physics''}} (\text{instructor} \bowtie_{\text{instructor.ID = teaches.ID}} \text{teaches})$$

1 $\sigma_{\text{dept\_name= ``Physics''}}$ (instructor) $\bowtie_{\text{instructor.ID = teaches.ID}}$ teaches

2 $\sigma_{\text{dept\_name= ``Physics''}}$ (instructor $\bowtie_{\text{instructor.ID = teaches.ID}}$ teaches)

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
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Figure 2.1 The *instructor* relation.

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-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

Figure 2.7 The *teaches* relation.

ID	Name	Dept_name	Salary
22222	Einstein	Physics	95000
33456	Gold	Physics	87000

ID	Name	Dept_name	Salary	Course_id	Sec_id	semester	Year
22222	Einstein	Physics	95000	PHY-101	1	Fall	2009

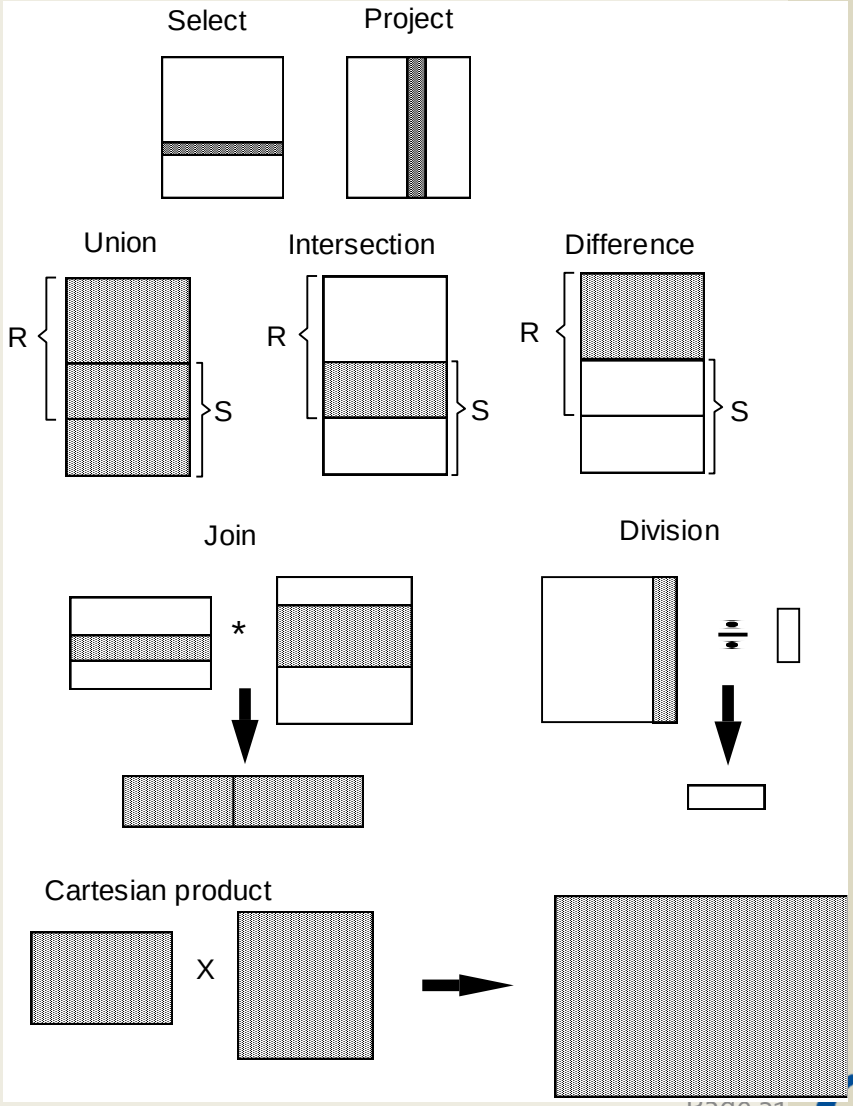
1 $\sigma_{\text{dept\_name= ``Physics''}}$ (instructor) $\bowtie_{\text{instructor.ID = teaches.ID}}$ teaches

2 $\sigma_{\text{dept\_name= ``Physics''}}$ (instructor $\bowtie_{\text{instructor.ID = teaches.ID}}$ teaches)

instructor.ID	*name*	*dept_name*	*salary*	*teaches.ID*	*course_id*	*sec_id*	*semester*	*year*
10101	Srinivasan	Comp. Sci.	65000	10101	CS-101	1	Fall	2017
10101	Srinivasan	Comp. Sci.	65000	10101	CS-315	1	Spring	2018
10101	Srinivasan	Comp. Sci.	65000	10101	CS-347	1	Fall	2017
12121	Wu	Finance	90000	12121	FIN-201	1	Spring	2018
15151	Mozart	Music	40000	15151	MU-199	1	Spring	2018
22222	Einstein	Physics	95000	22222	PHY-101	1	Fall	2017
32343	El Said	History	60000	32343	HIS-351	1	Spring	2018
45565	Katz	Comp. Sci.	75000	45565	CS-101	1	Spring	2018
45565	Katz	Comp. Sci.	75000	45565	CS-319	1	Spring	2018
76766	Crick	Biology	72000	76766	BIO-101	1	Summer	2017
76766	Crick	Biology	72000	76766	BIO-301	1	Summer	2018
83821	Brandt	Comp. Sci.	92000	83821	CS-190	1	Spring	2017
83821	Brandt	Comp. Sci.	92000	83821	CS-190	2	Spring	2017
83821	Brandt	Comp. Sci.	92000	83821	CS-319	2	Spring	2018
98345	Kim	Elec. Eng.	80000	98345	EE-181	1	Spring	2017

Summary

Symbol (Name)	Example of Use
σ (Selection)	$\sigma_{\text{salary}>=85000}(instructor)$
	Return rows of the input relation that satisfy the predicate.
Π (Projection)	$\Pi_{ID,salary}(instructor)$
	Output specified attributes from all rows of the input relation. Remove duplicate tuples from the output.
$\bowtie$ (Natural join)	$instructor \bowtie department$
	Output pairs of rows from the two input relations that have the same value on all attributes that have the same name.
$\times$ (Cartesian product)	$instructor \times department$
	Output all pairs of rows from the two input relations (regardless of whether or not they have the same values on common attributes)
$\cup$ (Union)	$\Pi_{name}(instructor) \cup \Pi_{name}(student)$
	Output the union of tuples from the two input relations.



Assignment #2 (150pt)

- **Do Exercises (p. 62):**
  - 2.10, 2.11, 2.12, 2.13, 2.15, 2.18
- **Due: One Week Later**
  - Before the lecture – 9/21 (Wed)
- **Method: upload your report in Cyber Campus**
  - Questions are uploaded in Assignment 2 folder
  - Answers must be written in English !