Database Management Systems

Madhavan Mukund

`https://www.cmi.ac.in/~madhavan`

Sai University
Lecture 4, 25 August 2023

- List instructors from Physics department with salary above 90,000

σ dept_name = Physics *Instructor*

$\wedge$ AND $\vee$ OR

salary > 90000

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

- List instructors from Physics department with salary above 90,000

- List names of instructors

$\pi_{\text{name}}(\textit{Instructor})$

1 or more column

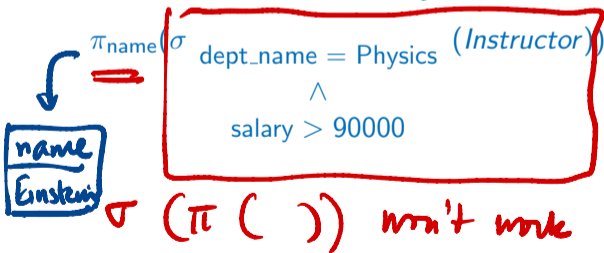$\pi_{\text{ID, name}}(\textit{Instructor})$

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
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

- List instructors from Physics department with salary above 90,000

- List names of instructors

- List names of instructors from Physics department *with salary > 90000*

$$\pi_{\mathsf{name}}\left(\sigma_{\substack{\mathsf{dept\_name} \,=\, \mathsf{Physics} \\ \wedge \\ \mathsf{salary} \,>\, 90000}}(\mathit{Instructor})\right)$$

name
Einstein

$\sigma\ (\pi\ (\ \ \ ))$ won't work

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

- List details of courses offered by 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
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

Teaches

ID	course_id	sec_id	semester	year
10101	CS-101	1	Fall	2017
10101	CS-315	1	Spring	2018
10101	CS-347	1	Fall	2017
12121	FIN-201	1	Spring	2018
15151	MU-199	1	Spring	2018
22222	PHY-101	1	Fall	2017
32343	HIS-351	1	Spring	2018
45565	CS-101	1	Spring	2018
45565	CS-319	1	Spring	2018
76766	BIO-101	1	Summer	2017
76766	BIO-301	1	Summer	2018
83821	CS-190	1	Spring	2017
83821	CS-190	2	Spring	2017
83821	CS-319	2	Spring	2018
98345	EE-181	1	Spring	2017

- Instructor × 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
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
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

- $\sigma_{\text{InstructorID} = \text{course\_id}}(\text{Instructor} \times \text{Teaches})$

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

A query language — join

- Instructor ⋈ $_{InstructorID\ =\ course\_id}$ Teaches — cartesian project with select

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

Select σ_θ

Project $\pi_{c_1, c_2 .. c_k}$

Cartesian Product $\times$

Join (Select + Cartesian Product)

$$T_1 \bowtie_\theta T_2$$

theta
(arbitrary condition)

Revisit join in a bit

$R \subseteq S_1 \times S_2$ R is itself a set (of pairs (s_1, s_2))

$R_1 \subseteq S_1 \times S_2$

$R_2 \subseteq S_1 \times S_2$ $\longrightarrow$ $R_1 \cup R_2$ Set Union

$R_1 \cap R_2$ Set intersection

Instructors in Physics : $\sigma_{dept-name = Physics}$ (Instructor)

Comp. Sci $\sigma_{dept-name = Comp Sci}$ (Instructor)

Either Physics or CS $\sigma \cup \sigma$

$$\sigma_{\text{dept\_name} = \text{Physics} \lor \text{dept\_nam} = \text{comp\_sc}} (\text{Instructor})$$

$\theta_1 \quad \sigma_{\text{deptname} = \text{Physics}} (\text{Instructor})$

$\theta_2 \quad \sigma_{\text{salary} > 90000} (\text{Instructor})$

$\sigma_{\theta_1 \land \theta_2}$

$\sigma \cdot \quad \cap \quad \sigma$

Third operation — Complementation

Squares $\subseteq \mathbb{N} \times \mathbb{N}$ $\qquad \{(i, i^2) \mid i \in \mathbb{N}\}$

$\{(0,0), (1,1), (2,4), (3,9), \dots\}$

NOT (Squares)
$\{(0,1), (0,2) \dots$
$(1,0), (1,2) \dots$

Universe

$\overline{S}$

S

not(s)

not(squares) does not contain $(2, 3.5)$

Instructor

LD	Name	Dept	Sal
101	ABC	X	M
202	DEF	Y	N

not(Instructor) ?

101 ABC Y M
101 ABC X N

Unmanageable list of
entries

Therefore   No "not" in relational algebra

Courses Offered in Fall 2017 that are <u>not</u>
Offered in Spring 2018

Fall 2017

Spring 2018

So far
$\cup$, $\cap$

$\sigma_{semester = Fall \wedge year = 2017}$ (Teaches)

Set Difference   X − Y   X \ Y

X \ Y   All elements in X that are not in Y

$Y \setminus X$

$X - Y, X \setminus Y$
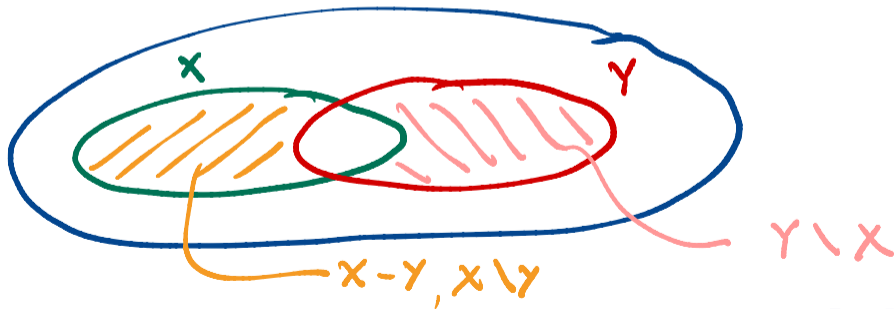
Set Difference

Fall 2017 \ Spring 2018

Select
Project
Cartesian Product

Join

Union
Intersection
Set Difference

(No set complement)

Intermediate Names — Assignment

$$\text{Fall 2017} \leftarrow \sigma_{Semester = Fall \land Year = 2017} (Teaches)$$

$$\text{Spring 2018} \leftarrow \ - \ - \ .$$

$$\text{Fall 2017} \ \searrow \ \text{Spring 2018}$$

Renaming

Rename a relation

Instructor $\leadsto$ I

$$\rho_{I}(\text{Instructor})$$

rho

Identify all teachers teaching 2 courses

for each row r1 in Teaches
    for each row r2 in Teaches
        r1 teachs.id = r2.teacher.id
        r1 course.id ≠ r2 course.id
           If so, record r1.teaches.id

Same column name in T1 & T2

$\left. \begin{array}{l} \text{T2. column name} \\ \text{T1. column name} \end{array} \right\}$ to distinguish

For my nested loop

Teaches ⋈ Teaches

R1.inst =R2.inst

How to distinguish ?

$$\rho_{T1} (Teaches) \bowtie_{\substack{T1.teachid \\ =T2.teachid}} \rho_{T2} (Teaches)$$

$$\land$$

$$T1.courseid \neq T2.courseid$$

Without renaming ||

$$Teaches.teachid = Teaches.teachid$$

More generally, can also rename columns

$$\rho_{T_1} \ (Teaches)$$

$$\rho_{T_1}(Id, C, S, Y) \ (Teaches)$$

Teal Id	Conac Id	Sec	Ya

$\downarrow$

Id	C	S	Y

Select
Project
Cartesian Product

Join

Union
Intersect

Set Diff

Assignment
Renaming

Very common version of Join

— Same Column Name matches across tables

Personal

ID	Name	DoB

Accounts

ID	Salary

Personal ⋈ Accounts
$$\text{Personal} \bowtie_{\text{Personal.ID} = \text{Accounts.ID}} \text{Accounts}$$

"NATURAL JOIN"

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

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 $\{A, B, C\}$ $\{C, B, A\}$ Instructor, unsorted

$\{A, A, B, C\}$ ✗

Π_{dept_name} (Instructor)

Comp Sci   ↓ Skeleton
Finance
Music
Physic
History
Biology

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, unsorted

Tax query   People from Chennai with income > 5cr

$$\sigma_{\substack{Dist=Chennai \ \wedge \\ Income > 5cr \ \wedge \\ TN.name = IT.name}} (\text{TN Data} \times \text{IT Data})$$

Unnecessarily large

Instead

Much smaller table

$$\sigma_{TN.name=IT.Name}\left(\sigma_{\substack{Dist= \\ Chennai}}(\text{TN Data}) \times \sigma_{\substack{Income \\ >5cr}}(\text{IT Data})\right)$$

Second form is more efficient

Declarative programming — Say what you want to compute, not how to compute it

```
fact(n) {
    ans = 1
    for i = 1 to n
        ans = ans * i
    return (ans)
}
```

$fact(0) = 1$

$fact(n) = n * fact(n-1)$
$\quad n > 0$

- "Programming language" for relational algebra

Structured Query Language

IBM's Project R

SQL

- "Programming language" for relational algebra

- Querying tables — select, project, join, ...

SQL

- "Programming language" for relational algebra

- Querying tables — select, project, join, . . .

- Data definition — describing relational schema

SQL

- "Programming language" for relational algebra

- Querying tables — select, project, join, . . .

- Data definition — describing relational schema

- Data manipulation — populating and modifying rows in tables

SQL