

Introduction to Operating Systems and SQL for Data Science

Practice 8 - Relational Algebra

RA - Marks

Unary operators

- Selection σ
- Projection π
- Aggregate function \mathfrak{A}

Binary operators

- Union \cup
- Intersection \cap
- Difference $-$
- Cartesian product \times
- Division \div

Logical symbols

- Logical AND \wedge
- Logical OR \vee
- Logical NOT \neg

- Left Outer Join \bowtie
- Right Outer Join \bowtie
- Full Outer Join \bowtie

RA – tips for questions

1. Read the question carefully and notice which tables we should use
2. Understand common space of these tables (for \bowtie operations)
3. Find the conditions in the questions (for σ operations)
4. Find which fields we need (for π operations)
5. Notice (from start) which field should be in the final result

Pay attention:

1. Field declared in project operations exists in the table
2. Tables in \bowtie operations have at least one common field

DB Schema

Teachers

<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
100000000	Peretz Shoval	1945	Ph.D.	Information...	372
200000000	Yael Edan	1900	Ph.D.	Computer...	364
300000000	Helman Stern	1900	Ph.D.	Transportation...	364
400000000	Michael Codish	2000	Ph.D.	Semantic-...	202
500000000	Klara Kedem	2000	Ph.D.	Computational...	202
600000000	Shlomi Dolev	2000	Ph.D.	Distributed...	202

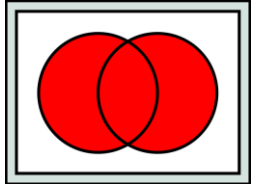
Departments

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

Heads_Of_Departments

<u>Teacher_ID</u>	<u>Department_ID</u>	<u>Start_Date</u>	<u>Stop_Date</u>
500000000	202	01-Jan-2001	01-Feb-2001
600000000	202	02-Feb-2001	02-Mar-2001

Union (\cup)



- Operation is possible if both tables are compatible.

Departments1

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600

Departments2

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

Result \leftarrow Departments1 \cup Department2

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

No duplications!

Outer Union (\cup^*)

Departments1

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

Departments2

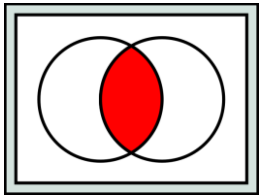
<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students	Budget
999	X-Science	900000000	900	₪9,000,000

An Extra Field

Departments1 \cup^ Departments2*

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students	Budget
372	Information Systems Engineering	100000000	500	
364	Industrial Engineering and Management	200000000	600	
202	Computer Science	400000000	700	
999	X-Science	900000000	900	₪9,000,000

Intersection (\cap)



- Operation is possible if both tables are compatible.

Departments1

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600

Departments2

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

Result \leftarrow *Departments1* \cap *Department2*

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
364	Industrial Engineering and Management	200000000	600

Difference (−)

- Operation is possible if both tables are compatible.

Departments

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600
202	Computer Science	400000000	700

Departments1

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
372	Information Systems Engineering	100000000	500
364	Industrial Engineering and Management	200000000	600

Result ← Departments − Department1

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students
202	Computer Science	400000000	700

Selection (σ)

Query: all the data of teachers born in the 20th century:

Selection (σ)

Query: all the data of teachers born in the 20th century:

$\sigma_{(Year_Of_Birth \geq 1900 \wedge Year_Of_Birth \leq 1999)} (Teachers)$

<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
100000000	Peretz Shoval	1945	Ph.D.	Information...	372
200000000	Yael Edan	1900	Ph.D.	Computer...	364
300000000	Helman Stern	1900	Ph.D.	Transportation...	364

Projection (π)

Query: IDs and names of all teachers with Ph.D degree and belongs to CS department (ID 202)

Projection (π)

Query: IDs and names of all teachers with Ph.D degree and belongs to CS department

1. **Result1** $\leftarrow \sigma_{(Last_Degree = 'Ph.D.' \wedge Department_ID = 202)} (Teachers)$

2. **Result** $\leftarrow \pi_{ID, Name} (Result1)$

\Leftrightarrow

1. **Result** $\leftarrow \pi_{ID, Name} (\sigma_{(Last_Degree = 'Ph.D.' \wedge Department_ID = 202)} (Teachers))$

Result

<u>ID</u>	Name
4000000000	Michael Codish
5000000000	Klara Kedem
6000000000	Shlomi Dolev

Cartesian product (\times)

R		
A	C	D
a1	c1	d1
a2	c2	d2
a3	c3	d3

S	
A	B
a1	b1
a2	b2



RxS				
R.A	C	D	S.A	B
a1	c1	d1	a1	b1
a1	c1	d1	a2	b2
a2	c2	d2	a1	b1
a2	c2	d2	a2	b2
a3	c3	d3	a1	b1
a3	c3	d3	a2	b2

Cartesian product (×)

Result ← *Departments* × *Heads_Of_Departments*

<u>ID</u>	<u>Name</u>	<u>Head_Of_Department_ID</u>	<u>Number_Of_Students</u>	<u>Teacher_ID</u>	<u>Department_ID</u>	<u>Start_Date</u>	<u>Stop_Date</u>
372	Information Systems Engineering	100000000	500	500000000	202	01-Jan-2001	01-Feb-2001
364	Industrial Engineering and Management	200000000	600	500000000	202	01-Jan-2001	01-Feb-2001
202	Computer Science	400000000	700	500000000	202	01-Jan-2001	01-Feb-2001
372	Information Systems Engineering	100000000	500	600000000	202	02-Feb-2001	02-Mar-2001
364	Industrial Engineering and Management	200000000	600	600000000	202	02-Feb-2001	02-Mar-2001
202	Computer Science	400000000	700	600000000	202	02-Feb-2001	02-Mar-2001

Join(\bowtie)

R		
A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

S	
A	D
a1	d1
a2	d2
a2	d3
a3	d4
a3	d5
a3	d6
a6	d7
a7	d8



$R \bowtie_{R.A=S.A} S$				
R.A	B	C	S.A	D
a1	b1	c1	a1	d1
a2	b2	c2	a2	d2
a2	b2	c2	a2	d3
a3	b3	c3	a3	d4
a3	b3	c3	a3	d5
a3	b3	c3	a3	d6

$$R \bowtie S \Leftrightarrow S \bowtie R$$

Join(⋈)

- *Theta (Inner) join (⋈)*
- *Query: details of the departments and their head of department.*

Join(⋈)

- Theta (Inner) join (\bowtie)
- Query: details of the departments and their head of department.

$Result \leftarrow \sigma_{Head_Of_Department_ID = Teachers.ID} (Departments \times Teachers)$

\Leftrightarrow

$Result \leftarrow Departments \bowtie_{Head_Of_Department_ID = Teachers.ID} Teachers$

Result

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students	<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
372	Information ...	100000000	500	100000000	Peretz Shoval	1945	Ph.D.	Information...	372
364	Industrial...	200000000	600	200000000	Yael Edan	1900	Ph.D.	Computer...	364
202	Computer ...	400000000	700	400000000	Michael Codish	2000	Ph.D.	Semantic-...	202

Left Outer Join(\bowtie)

R		
A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

S	
A	D
a1	d1
a2	d2
a2	d3
a3	d4
a3	d5
a3	d6
a6	d7
a7	d8



$R \bowtie_{R.A=S.A} S$				
R.A	B	C	S.A	D
a1	b1	c1	a1	d1
a2	b2	c2	a2	d2
a2	b2	c2	a2	d3
a3	b3	c3	a3	d4
a3	b3	c3	a3	d5
a3	b3	c3	a3	d6
a4	b4	c4	null	null
a5	b5	c5	null	null

Left outer Join(\bowtie)

- *Query: codes and names of all departments with no head of department.*

Left outer Join(\bowtie)

- Query: codes and names of all departments with no head of department.

Result $\leftarrow \pi_{\text{Departments.ID, Departments.Name}} (\sigma_{\text{Teachers.ID} = \text{NULL}} (\text{Departments} \bowtie$

$\text{Head_Of_Department_ID} = \text{Teachers.ID} \text{ Teachers}))$

Outer_Left_Join_Temporal

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students	<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
372	Information ...	100000000	500	100000000	Peretz Shoval	1945	Ph.D.	Information...	372
364	Industrial...	200000000	600	200000000	Yael Edan	1900	Ph.D.	Computer...	364
202	Computer ...	400000000	700	400000000	Michael Codish	2000	Ph.D.	Semantic-...	202
999	X-Science	900000000	900						

Result

<u>ID</u>	Name
999	X-Science

תוצאה סופית:

Right Outer Join(\bowtie)

R		
A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

S	
A	D
a1	d1
a2	d2
a2	d3
a3	d4
a3	d5
a3	d6
a6	d7
a7	d8



$R \bowtie_{R.A=S.A} S$				
R.A	B	C	S.A	D
a1	b1	c1	a1	d1
a2	b2	c2	a2	d2
a2	b2	c2	a2	d3
a3	b3	c3	a3	d4
a3	b3	c3	a3	d5
a3	b3	c3	a3	d6
null	null	null	a6	d7
null	null	null	a7	d8

Right outer Join(\bowtie)

Query: ids and names of all teachers that are not head of departments.

Right outer Join(\bowtie)

Query: ids and names of all teachers that are not head of departments (and never were).

1. $Past \leftarrow \pi_{ID, Name} (Heads_Of_Departments \bowtie_{Teacher_ID = ID} Teachers)$
2. $Not_Present(ID, Name) \leftarrow \pi_{Teachers.ID, Teachers.Name} (\sigma_{Departments.ID = NULL} (Departments \bowtie_{Head_Of_Department_ID = Teachers.ID} Teachers)))$
1. $Result \leftarrow Not_Present - Past$

Outer_Right_Join_Temporal

<u>ID</u>	Name	Head_Of_Department_ID	Number_Of_Students	<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
372	Information ...	100000000	500	100000000	Peretz Shoval	1945	Ph.D.	Information...	372
364	Industrial...	200000000	600	200000000	Yael Edan	1900	Ph.D.	Computer...	364
				300000000	Helman Stern	1900	Ph.D.	Transportation...	364
202	Computer ...	400000000	700	400000000	Michael Codish	2000	Ph.D.	Semantic-...	202
				500000000	Klara Kedem	2000	Ph.D.	Computational...	202
				600000000	Shlomi Dolev	2000	Ph.D.	Distributed...	202

Not_Present

<u>ID</u>	Name
300000000	Helman Stern
500000000	Klara Kedem
600000000	Shlomi Dolev

Past

<u>ID</u>	Name
500000000	Klara Kedem
600000000	Shlomi Dolev

Result

<u>ID</u>	Name
300000000	Helman Stern

Full Outer Join(\bowtie)

R		
A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

S	
A	D
a1	d1
a2	d2
a2	d3
a3	d4
a3	d5
a3	d6
a6	d7
a7	d8



$R \bowtie_{R.A=S.A} S$				
R.A	B	C	S.A	D
a1	b1	c1	a1	d1
a2	b2	c2	a2	d2
a2	b2	c2	a2	d3
a3	b3	c3	a3	d4
a3	b3	c3	a3	d5
a3	b3	c3	a3	d6
a4	b4	c4	null	null
a5	b5	c5	null	null
null	null	null	a6	d7
null	null	null	a7	d8

Full outer Join(\bowtie)

Departments \bowtie *Head_Of_Department_ID = Teachers.ID* **Teachers**

<u>ID</u>	Name	Head_Of_De partment_ID	Number_Of _Students	<u>ID</u>	Name	Year_Of Birth	Last_Deg ree	Specialization	Departme nt_ID
372	Information ...	100000000	500	100000000	Peretz Shoval	1945	Ph.D.	Information...	372
364	Industrial...	200000000	600	200000000	Yael Edan	1900	Ph.D.	Computer...	364
				300000000	Helman Stern	1900	Ph.D.	Transportation...	364
202	Computer ...	400000000	700	400000000	Michael Codish	2000	Ph.D.	Semantic-...	202
				500000000	Klara Kedem	2000	Ph.D.	Computational...	202
				600000000	Shlomi Dolev	2000	Ph.D.	Distributed...	202
999	X-Science	900000000	900						

Natural join (*)

R		
A	B	C
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4
a5	b5	c5

S	
A	D
a1	d1
a2	d2
a2	d3
a3	d4
a3	d5
a3	d6
a6	d7
a7	d8




R*S				
R.A	B	C	S.A	D
a1	b1	c1	a1	d1
a2	b2	c2	a2	d2
a2	b2	c2	a2	d3
a3	b3	c3	a3	d4
a3	b3	c3	a3	d5
a3	b3	c3	a3	d6

Natural join (*)

- *Syntactic sugar for apply join by a field with same name in both tables, no need to write the fields we join by them.*
- *$Departments1(\underline{Dep_ID}, Dep_Name, ID, Number_Of_Students) \leftarrow \pi_{ID, Name, Head_Of_Department, Number_Of_Students}(Departments)$*
- *$Result \leftarrow Departments1 * Teachers$*

Result



<u>Dep_ID</u>	Dep_Name	Number_Of_Students	<u>ID</u>	Name	Year_Of_Birth	Last_Degree	Specialization	Department_ID
372	Information ...	500	100000000	Peretz Shoval	1945	Ph.D.	Information...	372
364	Industrial...	600	200000000	Yael Edan	1900	Ph.D.	Computer...	364
202	Computer ...	700	400000000	Michael Codish	2000	Ph.D.	Semantic-...	202

Division (\div)

- A will appear in the result only if it showed on R with every b value of S

R			S		R \div S
A	B				
a1	b1	→	b1		a2
a1	b3		b2		a3
a2	b1				
a2	b2				
a2	b3				
a3	b1				
a3	b2				

חילוק מחזיר את השורות שבמונה, רק אם בחלק מהשורות שלו הוא כולל את כל הערכים של המכנה. למשל ל a3, a2, b2 יש גם b1, b2. ולכן הם מופיעים.

Division (÷)

Query:

IDs of all teachers who have been head of departments for both
CS and Industrial Engineering and Management

Numerator=

Result ← Numerator ÷ Denominator

חילוק מחזיר את השורות שבמונה, רק אם בחלק מהשורות שלו הוא כולל את כל הערכים של המכנה. למשל ל5מיליון פה, יש גם 202 וגם 364. ולכן הוא מופיע.

Numerator

<u>Teacher_ID</u>	<u>Department_ID</u>
500000000	202
600000000	202
500000000	364

÷

Denominator

<u>Department_ID</u>
202
364

→

Result

<u>Teacher_ID</u>
500000000

Aggregation (Σ)

- Query: For each department calc average age of all its head department

Result(*Department_ID*, *Age*) \leftarrow *Department_ID* (Σ Average
(*CURRENT_YEAR* - *Year_Of_Birth*) (*Teachers* \bowtie *ID = Teacher_ID*
Heads_Of_Departments))

<u>ID</u>	Age
202	22