Relational Algebra

COMP1048: Databases and Interfaces (2024-2025)

Matthew Pike & Yuan Yao

Table of contents

1 Lab Overview

In this lab, you will practice various relational algebra operations. Understanding relational algebra is important as it forms the basis for SQL and the relational data model.

Complete the following exercises and compile your answers into a single PDF document, as detailed in the Submitting your lab work section.

2 Questions

Q1 Unions and Intersections

Consider the relations $R(A, B, C)$ and $S(A, B, C)$ shown in Table 1a and Table 1b respectively.

R		
A	B	C
—	—	—
1	2	3
4	2	3
4	5	6
2	5	3
1	2	6

(a) R Relation

S		
A	B	C
—	—	—
2	5	3
2	5	4
4	5	6
1	2	3
2	7	3

(b) S Relation

Table 1: Relations R and S for Q1.

1. Compute $R \cup S$. Which of the following tuples does **not appear** in the result?

   a. (1, 2, 3)
   b. (4, 5, 3)
   c. (4, 5, 6)
   d. (2, 5, 4)

Solution

b. (4,5,3) - Since this tuple appears in neither of the tables. Duplicates are removed.

$R \cup S$			
In Result	A	B	C
X	1	2	3
	4	2	3
X	4	5	6
	2	5	3
	1	2	6
X	2	5	4
	2	7	3

2. Compute $R \cap S$. Which of the following tuples **appears** in the result?

   a. (2, 5, 3)
   b. (2, 5, 4)
   c. (4, 2, 3)
   d. (1, 2, 6)

Solution

a. (2,5,3) - This tuple appears in both tables.

$R \cap S$			
In Result	A	B	C
	1	2	3
	4	5	6
X	2	5	3

3. Compute $(R - S) \cup (S - R)$. Which of the following tuples **appears** in the result?

   a. (1, 5, 6)
   b. (4, 5, 6)
   c. (2, 5, 4)
   d. (4, 5, 3)

Solution

c. (2,5,4)

$R - S$		
A	B	C
4	2	3
1	2	6

$S - R$		
A	B	C
2	5	4
2	7	3

$(R - S) \cup (S - R)$		
A	B	C
4	2	3
1	2	6
2	5	4
2	7	3

Q2 Theta Join (I)

Consider the relations $R(A, B)$ and $S(B, C, D)$ shown in Table 7a and Table 7b respectively.

R	
A	B
1	2
3	4
5	6

(a) R Relation

S		
B	C	D
2	4	6
4	6	8
4	7	9

(b) S Relation

Table 7: Relations R and S for Q2.

1. Compute $R \bowtie_\theta S$ with the condition θ:

$$R.A < S.C \quad \text{and} \quad R.B < S.D$$

Assume each tuple has the schema $(A, R.B, S.B, C, D)$. Which of the following tuples is in the result?

a. (5, 6, 2, 4, 6)
b. (3, 4, 5, 7, 9)
c. (1, 2, 2, 6, 8)
d. (3, 4, 4, 6, 8)

Solution

d. (3, 4, 4, 6, 8)

$R.A < S.C$ AND $R.B < S.D$				
A	R.B	S.B	C	D
1	2	2	4	6
1	2	4	6	8
1	2	4	7	9
3	4	2	4	6
3	4	4	6	8
3	4	4	7	9
5	6	4	6	8
5	6	4	7	9

Q3 Theta Join (II)

Consider the relations $R(A, B)$ and $S(B, C, D)$ shown in Table 9a and Table 9b respectively.

R	
A	B
—	—
1	a
7	t
2	g
4	c
9	t

(a) R Relation

S		
B	C	D
—	—	—
c	5	6
a	7	8
t	8	9

(b) S Relation

Table 9: Relations R and S for Q3.

1. Compute $R \bowtie_\theta S$ with the condition θ:

$$R.B = S.B \quad \text{and} \quad R.A < S.C$$

Assume each tuple has the schema $(A, R.B, S.B, C, D)$. Which of the following tuples **appears** in the result?

a. (2, g, g, 7, 8)
b. (4, c, c, 7, 8)
c. (2, g, t, 8, 9)
d. (4, c, c, 5, 6)

Solution

d. (4, c, c, 5, 6)

$R.B = S.B$ AND $R.A < S.C$				
A	R.B	S.B	C	D
1	a	a	7	8
7	t	t	8	9
4	c	c	5	6

Q4 Projection

Consider the relation $R(A, B, C)$ shown in Table 11.

R		
A	B	C
—	—	—
1	2	3
4	2	3
4	5	6
2	5	3
1	2	6

Table 11: Relation R for Q4.

1. Compute the projection $\pi_{C,B}(R)$. Which of the following tuples is in the result?

   a. (1, 2, 6)
   b. (6, 5)
   c. (2, 6)
   d. (5, 6)

Solution

b. (6,5)

$\pi C, B(R)$	
C	B
3	2
3	2
6	5
3	5
6	2

Q5 University Challenge

> **i** These are not exhaustive examples
>
> This question is based on the relations discussed in the lecture. The tuples presented in below are provided as examples only. You should not assume that these are the only tuples that exist in these relations.

Consider the relations $Uni(uName, County, Enr)$, $St(SID, sName, GPA, HS)$ and $Ap(SID, uName, Subj, Dec)$ which are shown in Tables 13, 14 & 15 respectively.

Uni

uName	County	Enr
NOTT	Nott	18000
CAM	Cam	22000
UCL	Lon	20000

Table 13: The Uni (University) Relation.

St

SID	sName	GPA	HS
0135	John	18.5	100
0025	Mary	19.3	1000
0423	Mary	17.5	300

Table 14: The St (Student) Relation.

Ap

SID	uName	Subj	Dec
0135	CAM	CS	A
0135	NOTT	CS	A
0423	NOTT	ENG	R

Table 15: The Ap (Apply) Relation.

> **i** Relation Descriptions
>
> We use shorthands for the attributes and values in the relations. Here are the descriptions:
>
> - Uni: University relation
> - `uName`: University Name
> - `County`: County where the University is located (e.g., Lon = London)
> - `Enr`: Enrollment
> - St: Student relation
> - `SID`: Student ID
> - `sName`: Student Name
> - `GPA`: Grade Point Average
> - `HS`: High School size
> - Ap: Apply relation
> - `SID`: Student ID
> - `uName`: University Name
> - `Subj`: Subject
> - `Dec`: Decision

Q5.1

1. Which of the following describes the result of this expression?

$$\pi_{uName(Uni)} - \pi_{uName}(Ap \bowtie (\pi_{SID}(\sigma_{GPA>19}(St)) \cap \pi_{SID}(_{\sigma_{Subj}='CS'}(Ap))))$$

   a. All Universities with no $GPA > 19$ Applicants who applied for CS at that University.
   b. All Universities with no $GPA > 19$ Applicants who applied for CS at any University.
   c. All Universities where all Applicants either have $GPA > 19$ or applied for CS at that University.
   d. All Universities where no Applicants have $GPA > 19$ or no Applicants applied for CS at that University

> Solution
>
> **b.**
> The intersection finds the `sID` of students with a $GPA > 19$ and applied for CS at any university. The Natural join finds all universities that these students applied for. The negation leaves all universities that students did not apply to.

Q5.2

1. Which of the following describes the result of this expression?

$$\pi_{sName,uName}(\sigma_{HS>Enr}(\sigma_{County='Lon'}(Uni \bowtie St \bowtie (\sigma_{subj='CS'}(Ap)))))$$

   a. All Student-University name pairs, where the student is applying to CS at the University, the University is in London, and the University is smaller than some High School.
   b. Students paired with all London Universities to which the Student applied to CS, where at least one of those Universities is smaller than the Student's High School.
   c. Students paired with all Universities smaller than the Student's high school to which the Student applied to CS, where at least one of those Universities is in London.
   d. Students paired with all London Universities smaller than the Student's High School to which the Student applied to CS.

> Solution
>
> **d.**
> The inner natural join combines `Students` with `Universities` they have applied for (where the subject is 'CS'). From this, only universities in the county of london are selected, before again being filtered by $HS > Enr$. Finally, the student's and university names are projected.

Q5.3

> **Note**
>
> For this question you will need to ignore the tuples presented in Table 13, Table 14 and Table 15.

1. Suppose that the Student relation has 20 tuples. ρ is the Rename operator. What is the minimum and maximum number of tuples in the result of the following expression?

$$\rho_{s1(i1,n1,g,h)}St \bowtie \rho_{s2(i2,n2,g,h)}St$$

   a. minimum = 0, maximum = 400
   b. minimum = 20, maximum = 20
   c. minimum = 20, maximum = 400
   d. minimum = 40, maximum = 40

> **Solution**
>
> **c.** minimum = 20, maximum = 400
> If every student has a unique $G - H$ combination, then they combine only with themselves (Minimum of 20 - number of students). If every student has the same $G - H$ combination, then all pairs join - $20 * 20 = 400$.

Q5.4

> **Note**
>
> For this question you will need to ignore the tuples presented in Table 13, Table 14 and Table 15.

1. Assume that relations `Uni`, `St`, and `Ap` have 5, 20, and 50 tuples respectively. Assume that `uName` is a key for `Uni`. Do not assume `sName` is a key for `St`. Assume that university names in `Ap` also appear in `Uni`. What is the minimum and maximum number of tuples in the result of this expression:

$$\pi_{uName}(Uni) \cup \rho_{(uName)}(\pi_{sName}(St)) \cup \pi_{uName}(Ap)$$

   a. minimum = 5, maximum = 25
   b. minimum = 5, maximum = 75
   c. minimum = 25, maximum = 45
   d. minimum = 75, maximum = 75

> **Solution**
>
> **a.** minimum = 5, maximum = 25
> Remember - duplicates are eliminated.
> If all student have names that are also university names, then there are 5 names - this is the minimum, since there are 5 universities and duplicates are eliminated. If every student has a unique name, then there are 5 universities and 20 students - 25 maximum.

3 Submitting your lab work

Compile your answers into a single PDF document. Your submission should:

- Be neat and easy to read
- Must include a cover page with:
  - Your name
  - Student ID
  - University email address
  - Module code (COMP1048) and title (Databases and Interfaces)
  - Lab number (002) and title (Relational Algebra)
  - Date of submission

Name your PDF file using the following format - `DBI_lab002-<student_id>.pdf`, where, `<student_id>` is your student ID. For example, if your student ID is `z123456`, you should name your PDF file `DBI_lab002-z123456.pdf`.

Please ensure you submit your work by the deadline - 21 October 2024 at 15:00. Late submissions will not be accepted, as stated on the coursework issue sheet.

Submissions that are unreadable, corrupted, missing the required cover page, or do not demonstrate a reasonable attempt to answer all questions will receive a mark of zero for the lab. Your submission must be less than 10MB in size.

This lab contributes 1% of your overall module grade.