

Week 10

Relational Algebra, Transaction Processing

[\[Show with no answers\]](#) [\[Show with all answers\]](#)

Relational Algebra

1. Relational algebra operators can be *composed*. What precisely does this mean? And why is it important?

[\[show answer\]](#)

2. The natural join ($R \Join S$) joins two tables on their common attributes. Consider a theta-join ($R \Join[C] S$) where the join condition C is a conjunction of $R.A = S.A$ for each attribute A appearing in the schemas of both R and S (i.e. it joins on the common attributes). What is the difference between the above natural join and theta join?

[\[show answer\]](#)

3. The definition of relational algebra in lectures was based on *sets* of tuples. In practice, commercial relational database management systems deal with *bags* (or *multisets*) of tuples.

Consider the following relation that describes a collection of PCs

PCs				
Model	Speed	RAM	Disk	Price
1001	700	64	10	799
1002	1500	128	60	2499
1003	1000	128	20	1499
1004	1000	256	40	1999
1005	700	64	30	999

Consider a projection of this relation on the processor speed attribute, i.e. $\text{Proj}[\text{speed}](\text{PCs})$.

- What is the value of the projection as a set?
- What is the value of the projection as a bag?
- What is the average speed if the projection is a set?
- What is the average speed if the projection is a bag?
- Is the minimum/maximum speed different between the bag and set representation?

[\[show answer\]](#)

4. Consider the following two relations:

R			S	
A	B	C	B	C
a1	b1	c1	b1	c1
a1	b2	c2	b2	c2
a2	b1	c1		

Give the relation that results from each of the following relational algebra expressions on these relations:

a. $R \text{ Div } S$

[\[show answer\]](#)

b. $R \text{ Div } (\text{Sel}[B \neq b1](S))$

[\[show answer\]](#)

c. $R \text{ Div } (\text{Sel}[B \neq b2](S))$

[\[show answer\]](#)

d. $R \times S - (\text{Sel}[R.C=S.C](R \text{ Join}[B=B] S))$

[\[show answer\]](#)

5. Consider two relations $R1$ and $R2$, where $R1$ contains $N1$ tuples and $R2$ contains $N2$ tuples, and $N1 > N2 > 0$. Give the minimum and maximum possible sizes (in tuples) for the result relation produced by each of the following relational algebra expressions. In each case state any assumptions about the schemas of $R1$ and $R2$ that are needed to make the expression meaningful.

- $R1 \text{ Union } R2$
- $R1 \text{ Intersect } R2$
- $R1 - R2$
- $R1 \times R2$
- $\text{Sel}[a=5](R1)$
- $\text{Proj}[a](R1)$
- $R1 \text{ Div } R2$

[\[show answer\]](#)

6. (Date, exercise 6.11) Let A and B be two arbitrary relations. In terms of the keys of A and B , state the keys for each of the following RA expressions. Assume in each case that A and B meet the requirements of the operation (e.g. they are union-compatible for the Union and Intersect operations).

a. $\text{Sel}[\text{cond}](A)$, where cond is any condition

[\[show answer\]](#)

b. $\text{Proj}[\text{attrs}](A)$, where attrs is any set of attributes

[\[show answer\]](#)

c. $A \times B$

[\[show answer\]](#)

d. $A \text{ Union } B$

[\[show answer\]](#)

e. $A \text{ Intersect } B$

[\[show answer\]](#)

f. $A - B$

[\[show answer\]](#)

g. $A \text{ Div } B$

[\[show answer\]](#)

7. Consider the following relational schema:

```
Suppliers(sid, sname, address)
Parts(pid, pname, colour)
Catalog(supplier, part, cost)
```

Assume that the ids are integers, that `cost` is a real number, that all other attributes are strings, that the **supplier** field is a foreign key containing a supplier id, and that the **part** field is a foreign key containing a part id. Write a relational algebra expression to answer each of the following queries:

a. Find the *names* of suppliers who supply some red part.

[\[show answer\]](#)

b. Find the *sids* of suppliers who supply some red or green part.

[\[show answer\]](#)

c. Find the *sids* of suppliers who supply some red part or whose address is 221 Packer Street.

[\[show answer\]](#)

d. Find the *sids* of suppliers who supply some red part and some green part.

[\[show answer\]](#)

e. Find the *sids* of suppliers who supply every part.

[\[show answer\]](#)

f. Find the *sids* of suppliers who supply every red part.

[\[show answer\]](#)

g. Find the *sids* of suppliers who supply every red or green part.

[\[show answer\]](#)

h. Find the *sids* of suppliers who supply every red part or supply every green part.

[\[show answer\]](#)

i. Find the *pids* of parts that are supplied by at least two different suppliers.

[\[show answer\]](#)

j. Find pairs of *sids* such that the supplier with the first *sid* charges more for some part than the supplier with the second *sid*.

[\[show answer\]](#)

k. Find the *pids* of the most expensive part(s) supplied by suppliers named "Yosemite Sham".

[\[show answer\]](#)

l. Find the *pids* of parts supplied by every supplier at a price less than 200 dollars (if any supplier either does not supply the part or charges more than 200 dollars for it, the part should not be selected).

[\[show answer\]](#)

8. Using the Suppliers-Parts-Catalog schema from the previous question, state what the following relational algebra expressions compute:

a.

```
Proj[sname] (
    Proj[pid] (Sel[colour='red'] (Parts))
    Join
    Sel[cost<100] (Catalog)
```

```
        Join  
        Suppliers  
    )
```

[\[show answer\]](#)

b.

```
    Proj[sname] (  
        Proj[sid] (  
            Sel[colour='red'] (Parts)  
            Join  
            Sel[cost<100] (Catalog)  
            Join  
            Suppliers  
        )  
    )
```

[\[show answer\]](#)

c.

```
    Proj[sname] (  
        Sel[colour='red'] (Parts)  
        Join  
        Sel[cost<100] (Catalog)  
        Join  
        Suppliers  
    )  
    Intersect  
    Proj[sname] (  
        Sel[colour='green'] (Parts)  
        Join  
        Sel[cost<100] (Catalog)  
        Join  
        Suppliers  
    )
```

[\[show answer\]](#)

d.

```
    Proj[sid] (  
        Sel[colour='red'] (Parts)  
        Join  
        Sel[cost<100] (Catalog)  
        Join
```

```

        Suppliers
    )
Intersect
Proj[sid](
    Sel[colour='green'](Parts)
    Join
    Sel[cost<100](Catalog)
    Join
    Suppliers
)

```

[\[show answer\]](#)

e.

```

Proj[sid,sname](
    Proj[sid,sname](
        Sel[colour='red'](Parts)
        Join
        Sel[cost<100](Catalog)
        Join
        Suppliers
    )
    Intersect
    Proj[sid,sname](
        Sel[colour='green'](Parts)
        Join
        Sel[cost<100](Catalog)
        Join
        Suppliers
    )
)

```

[\[show answer\]](#)

9. Consider the following relational schema containing flight, aircraft and pilot information for an airline:

```

Flights(flno, from, to, distance, departs, arrives)
Aircraft(aid, aname, cruisingRange)
Certified(employee, aircraft)
Employees(eid, ename, salary)

```

The `Flights` relation describes a particular timetabled flight from a source (city) to a destination (city) at a particular time each week. Note that this schema doesn't model which specific aircraft makes the flight. The `Aircraft` relation describes individual planes, with the `aname` field containing

values like 'Boeing 747', 'Airbus A300', etc. The `Certified` relation indicates which pilots (who are employees) are qualified to fly which aircraft. The **`employee`** field contains the *eid* of the pilot, while the **`aircraft`** field contains the *aid* of the aircraft. Finally, the `Employees` relation contains information about all of the employees of the airline, including the pilots.

Write, where possible, a relational algebra expression to answer each of the following queries. If it is not possible to express any query in relational algebra, suggest what extra mechanisms might make it possible.

- a. Find the *ids* of pilots certified for 'Boeing 747' aircraft.

[\[show answer\]](#)

- b. Find the *names* of pilots certified for 'Boeing 747' aircraft.

[\[show answer\]](#)

- c. Find the *ids* of all aircraft that can be used on non-stop flights from New York to Los Angeles.

[\[show answer\]](#)

- d. Identify the flights that can be piloted by a pilot whose salary is more than \$100,000.

[\[show answer\]](#)

- e. Find the names of pilots who can operate planes with a range greater than 3000 miles, but are not certified on 'Boeing 747' aircraft.

[\[show answer\]](#)

- f. Find the total amount paid to employees as salaries.

[\[show answer\]](#)

- g. Find the *ids* of employees who make the highest salary.

[\[show answer\]](#)

- h. Find the *ids* of employees who make the second highest salary.

[\[show answer\]](#)

- i. Find the *ids* of employees who are certified for the largest number of aircraft.

[\[show answer\]](#)

- j. Find the *ids* of employees who are certified for exactly three aircraft.

[\[show answer\]](#)

k. Find if there is any non-stop or single-stop way to travel by air from Sydney to New York.

[\[show answer\]](#)

l. Is there a sequence of flights from Sydney to Timbuktu (somewhere in the middle of Africa)? Each flight in the sequence is required to depart from the destination of the previous flight; the first flight must leave Sydney, and the last flight must arrive in Timbuktu, but there is no restriction on the number of intermediate flights.

[\[show answer\]](#)

Transaction Processing

10. [based on Ramakrishnan, ex.17.1]

Give a brief definition for each of the following terms:

- a. transaction
- b. serializable schedule
- c. conflict-serializable schedule
- d. view-serializable schedule

[\[show answer\]](#)

11. Draw the precedence graph for the following schedule (where C means "commit"):

T1:	R(A)	W(Z)		C
T2:		R(B)	W(Y)	C
T3:	W(A)		W(B)	C

[\[show answer\]](#)

12. [based on Ramakrishnan, ex.17.2]

Consider the following incomplete schedule S:

T1:	R(X)	R(Y)	W(X)		W(X)
T2:			R(Y)		R(Y)
T3:			W(Y)		

- a. Determine (by using a precedence graph) whether the schedule is conflict-serializable
- b. Modify S to create a complete schedule that is conflict-serializable

[\[show answer\]](#)

13. [based on Ramakrishnan, ex.17.3]

For each of the following schedules, state whether it is conflict-serializable and/or view-serializable. If you cannot decide whether a schedule belongs to either class, explain briefly. The actions are listed in the order they are scheduled, and prefixed with the transaction name.

a. T1:R(X) T2:R(X) T1:W(X) T2:W(X)

b. T1:W(X) T2:R(Y) T1:R(Y) T2:R(X)

c. T1:R(X) T2:R(Y) T3:W(X) T2:R(X) T1:R(Y)

d. T1:R(X) T1:R(Y) T1:W(X) T2:R(Y) T3:W(Y) T1:W(X) T2:R(Y)

e. T1:R(X) T2:W(X) T1:W(X) T3:W(X)

[\[show answer\]](#)

14. Is the following schedule serializable? Show your working.

T1:	R(X)W(X)W(Z)	R(Y)W(Y)
T2:	R(Y)W(Y)R(Y)	W(Y)R(X)W(X)R(V)W(V)

[\[show answer\]](#)