Assignments 4

Due Friday, October 6, 2017

In this assignment, you will be required to use PostgreSQL. Your solutions should include the PostgreSQL statements for solving the problems. Submit a file Assignment4.sql with your solutions and a file OutputAssignment4.txt with your outputs to the queries.

Remark: Consider a relation $R(A, B)$ and a relation $S(C)$ and consider the following RA (Relational Algebra) expression F:

$$\pi_A(R) - \pi_A(\sigma_{B=1}(R \bowtie_{B=C} S))$$

Then we can write this query in SQL in a way that closely mimics its RA formulation. This can be done using the WITH statement of SQL as follows.[1]

First, we break the RA expression F up into sub-expressions as follows. (Notice that each sub-expression corresponds to the application of a single RA operation.)

Expression Name	RA expression
E_1	$\pi_A(R)$
E_2	$R \bowtie_{B=C} S$
E_3	$\sigma_{B=1}(E_2)$
E_4	$\pi_A(E_3)$
F	$E_1 - E_4$

Then we write the following SQL query. Notice how the expressions $E1$, $E2$, $E3$, and $E4$ occur as separate queries in the WITH statements and that the final query gives the result for the expression F.[2]

If you use overloading of the relational name, the query becomes

```
WITH
E1 AS (SELECT DISTINCT A FROM R),
E2 AS (SELECT A, B, C FROM R INNER JOIN S ON (B = C)),
E3 AS (SELECT A, B, C FROM E2 WHERE B = 1),
E4 AS (SELECT DISTINCT A FROM E3)
(SELECT A FROM E1) EXCEPT (SELECT A FROM E4);
```

[1] For more information about the WITH statement see `https://www.postgresql.org/docs/9.1/static/queries-with.html`.

[2] For better readability, I have used relational-name overloading. Sometimes, you may need to introduce new attribute names in SELECT clauses using the AS clause. Also, use DISTINCT where needed.

In your answer to a problem, you should write the resulting RA expression as such an SQL query. (Your SQL query should of course execute correctly in PostgreSQL). In a separate file you should also submit the text for the RA expressions in their standard notation, just as show for the expression F above.

In the following questions, use the data provided for the student, majors, book, cites, buys relations.

Write the following queries as RA expressions. For each such RA expression, write a SQL query (using the WITH statement) that mimics this expression.

Each problem is worth 10 points.

We will create some expressions that will be useful throughout.

$$\begin{aligned} \mathcal{S} &= \pi_{Sid}(Student) \\ \mathcal{B} &= \pi_{Bookno}(Book) \end{aligned}$$

So $\mathcal{S}$ is the set of all Sids of students and $\mathcal{B}$ is the set of all Booknos of Books.

1. Find the bookno of each book that is cited by at least one book that cost less than \$50.

   $$\pi_{Citedbookno}(Cites \bowtie \sigma_{Price<50}(Book))$$

   Or, equivalently, with semi-join $\ltimes$

   $$\pi_{Citedbookno}(Cites \ltimes \sigma_{Price<50}(Book))$$

2. Find the bookno and title of each book that was bought by a student who majors in CS and in Math.

   $$M = \pi_{Sid}(\sigma_{Major=\text{`CS'}}(Major)) \cap \pi_{Sid}(\sigma_{Major=\text{`Math'}}(Major))$$

   $$\pi_{bookno,title}(book \bowtie (buys \bowtie M))$$

   Or, alternatively, with semi-join $\ltimes$

   $$\pi_{bookno,title}(book \ltimes (buys \ltimes M))$$

3. Find the sid-bookno pairs (s, b) pairs such student s bought book b and such that book b is cited by at least two books that cost less than \$50.

   $$\begin{aligned} E_1 &= Cites \ltimes \pi_{Bookno}(\sigma_{Price<50}(Book)) \\ E_2 &= \pi_{E_{1_1}.citedbookno}(\sigma_{E_{1_1}.bookno \neq E_{1_2}.bookno \wedge E_{1_1}.citedbookno = E_{1_2}.citedbookno}(E_{1_1} \times E_{1_2})) \end{aligned}$$

   $$\pi_{sid,bookno}(buys \bowtie_{buys.bookno=E_2.citedbookno} E_2)$$

4. Find the bookno of each book with the next to highest price.

$$\begin{aligned} E_1 &= \pi_{Book_1.bookno, Book_1.price}(\sigma_{Book_1.price < Book_1.price}(Book_1 \times Book_2)) \\ E_2 &= \sigma_{E_{1_1}.price < E_{1_2}.Price}(E_{1_1} \times E_{1_2}) \end{aligned}$$

$$\pi_{E_2.E_{1_1}.bookno}(E_2) - \pi_{Book_1.bookno}(E_1)$$

5. Find the sid of each student who bought all books that cost more than $50.

$$booksover50 = \pi_{bookno}(\sigma_{price>50}(book))$$

$$\mathcal{S} - \pi_{sid}(\mathcal{S} \times booksover50 - buys)$$

6. Find the Bookno of each book that was not bought by any student who majors in CS.

$$\mathcal{B} - \pi_{bookno}(buys \ltimes \pi_{sid}(\sigma_{major=\text{`CS'}}(Major)))$$

7. Find the Bookno of each book that was not bought by all students who majors in CS.

$$\pi_{bookno}(\pi_{sid}(\sigma_{major=\text{`CS'}}(Major)) \times \mathcal{B} - buys)$$

8. Find sid-bookno pairs (s, b) such that not all books bought by student s are books that cite book b.

$$\begin{aligned} buyspadded &= buys \times \mathcal{B} \\ citespadded &= \mathcal{S} \times cites \end{aligned}$$

$$\pi_{sid, citedbookno}(buyspadded - citespadded)$$

9. Find sid-bookno pairs (s, b) such student s only bought books that cite book b.

$$\begin{aligned} buyspadded &= buys \times \mathcal{B} \\ citespadded &= \mathcal{S} \times cites \end{aligned}$$

$$\mathcal{S} \times \mathcal{B} - \pi_{sid, citedbookno}(buyspadded - citespadded)$$

10. Find the bookno of each book that cites all but two books. (In other words, for such a book, there exists only two books that it does not cite.)

$$
\begin{aligned}
E &= \mathcal{B} \times \mathcal{B} - cites \\
F &= \sigma_{E_1.bookno_1 = E_2.bookno_1 \wedge E_1.bookno_2 \neq E_2.bookno_2}(E_1 \times E_2) \\
H &= \pi_{F.E_1.bookno_1, F.E_1.bookno_2, F.E_2.bookno_1}(F) \\
I &= \sigma_{F.E_1.bookno_1 = E_3.bookno_1 \wedge F.E_1.bookno_2 \neq E_3.bookno_2 \wedge F.E_2.bookno_1 \neq E_3.bookno_2}(H \times E_3)
\end{aligned}
$$

$$\pi_{F.E_1.bookno_1}(H) - \pi_{F.E_1.bookno_1}(I)$$