CS 455 Principles of Database Systems

Homework 2: Relational Algebra

Name:

Instructions: Print this assignment using single-side pages. Fill in your name above, and write in the solutions in the space provided below each question. You are allowed to use the back of each page. If you used any scratch paper to show your work, append those to the end. **Note:** It is important you use this format for gradescope.

Submission: After you've filled in the answers, scan all pages into a PDF, and submit to canvas.

Problems

1. Assume we have two relations $R(A, B)$ and $S(B, C)$. Assume that R = $\{(1, 2), (2, 3), (3, 4)\}$, and $S = \{(2, 2), (2, 3), (5, 1)\}$. How many tuples are in the result of $\pi_A(R \bowtie S)$? (5pts)

2. Assume we have relations $R(A_1, ..., A_n)$ and $S(B_1, ..., B_m)$. Define the right-semi join operation $R \rtimes S$. The right-semi join retains all tuples in S for which there is a tuple in R that is equal on their common attributes. For instance, the right-semi join between R and S as defined in the previous problem produces $\{(2, 2), (2, 3)\}$. (You may use $\bowtie$ in your expression.)

   $R \rtimes S =$? (5pts)

For the remaining problems, consider the relation instances below that model competitions in the world of college sports. Assume that each city may have multiple schools (*e.g.*, both Columbia and NYU are in New York). Schools are split into two conferences: A and B. Within each conference, there are four divisions: North, South, East, and West. Games are played by "home" and "away" (*i.e.*, the traveling) schools, identified by the schoolIDs. For instance, game 7 was between GaTech (24 points) and ASU (6 points) on 9/8/1980.

School

schoolID	city	school	conference	division	spent
0	Seattle	Washington	B	West	500M
1	Cleveland	Case	A	North	100M
2	Pittsburgh	CMU	A	North	-50M
3	San Francisco	Berkeley	B	West	800M
4	Oakland	Mills	A	West	350M
5	Tempe	ASU	A	West	400M
6	Miami	Miami	A	East	200M
7	Houston	Rice	B	South	250M
8	New York	Columbia	A	East	800M
9	New York	NYU	A	East	400M
10	Buffalo	SUNY	B	East	100M
11	Atlanta	GaTech	A	South	200M
12	Baltimore	JHU	A	North	0M
13	Lafayette	Purdue	A	South	90M

Game

gameID	Away	Home	date	year	awayScore	homeScore
0	0	11	9/2	2012	0	3
1	3	13	10/17	2009	23	0
2	5	10	10/10	2012	10	10
3	4	9	11/20	2015	17	7
4	2	7	9/27	2014	7	14
5	9	8	10/30	1990	14	15
6	8	3	8/30	1980	21	9
7	11	5	9/8	1980	24	6
8	4	2	10/28	1981	35	15
9	12	10	11/27	2012	3	40
...	...	...	...	...	...	...

Coach

name	schoolID	title
Carroll	0	Head Coach
Jackson	1	Head Coach
Tomlin	2	Head Coach
Kelly	3	Assistant Coach
Brown	1	Head Coach
Day	0	Assistant Coach
…	...	...

3. Consider the following expression:

$$School - (School \bowtie \pi_{schoolID}(\sigma_{title='HeadCoach'}(Coach)))$$

(a) Draw its *expression tree*. (5pts)

(b) Write in your own words, what is the expression asking for? I am not looking for a step-by-step explanation of each of the operators applied. I just want a succinct high-level comment on what is being requested. (10pts)

4. Write a relational-algebra expression that retrieves the scores between JHU and Purdue from 2000 to 2010. Feel free to use temporary variables for storing intermediate results. (10pts)

5. Write a relational-algebra expression that retrieves all schools (city and name) that do not currently have a coach. Feel free to use temporary variables for storing intermediate results. (10pts)

6. Write a relational-algebra expression to find which year(s) that *'Case'* recorded its fewest wins. Note that there may be multiple years during which *'Case'* recorded their lowest win total. Do not assume you know what Case's schoolID is. Feel free to use temporary variables for storing intermediate results. (10pts)

7. Write a relational-algebra expression to return a list of schools and the year in which that school recorded the same number of wins and losses. (10pts)