Homework 5 | Query Execution & Society

Updates made to the assignment spec after release are highlighted in red.

Objectives: To translate fluidly from SQL to RA. To understand the statistics used by SQL engines and how they inform modifications to RA trees. To identify the likelihood of harming society using the impact/scope/opacity rubric.

Due date: Tue, October 31st @ 10:00pm 🎃🦇🧛

***Median completion time (23sp):* 5 hours**

Resources

For this assignment, you will need:

- Pen and paper, or any drawing tools you prefer (e.g., PowerPoint, draw.io).

Problem Set

The technical questions (i.e., all questions except the last two) refer to the following tables:

```
CREATE TABLE TrickOrTreaters (
    id          INTEGER PRIMARY KEY,
    costume     VARCHAR(64) NOT NULL   -- eg, Patricia Selinger, TI-84, etc
);
CREATE TABLE CandyBags (
    id          INTEGER PRIMARY KEY,
    tid         INTEGER REFERENCES TrickOrTreaters,
    decoration  VARCHAR(64)            -- eg, pumpkin, skeleton, etc
);
CREATE TABLE Candies (
    bid         INTEGER REFERENCES CandyBag,
    name        VARCHAR(64) NOT NULL,  -- eg, Snickers, candy corn, etc
    quantity    INTEGER NOT NULL,
    isFullSize  INTEGER NOT NULL       -- everybody prefers full-sized candy!
);
```

1. Convert the following SQL query to a relational algebra tree. If relevant, please:

- Draw the tables from left to right in the same order in which they appear in the FROM clause
- Order any joins in the same order in which they appear in the query

```
   SELECT b.tid
     FROM CandyBags b, Candies c
    WHERE b.id = c.bid
      AND c.isFullSize = 1
 GROUP BY b.tid
   HAVING COUNT(DISTINCT c.name) > 10;
```

2. Consider the following *correlated* query:

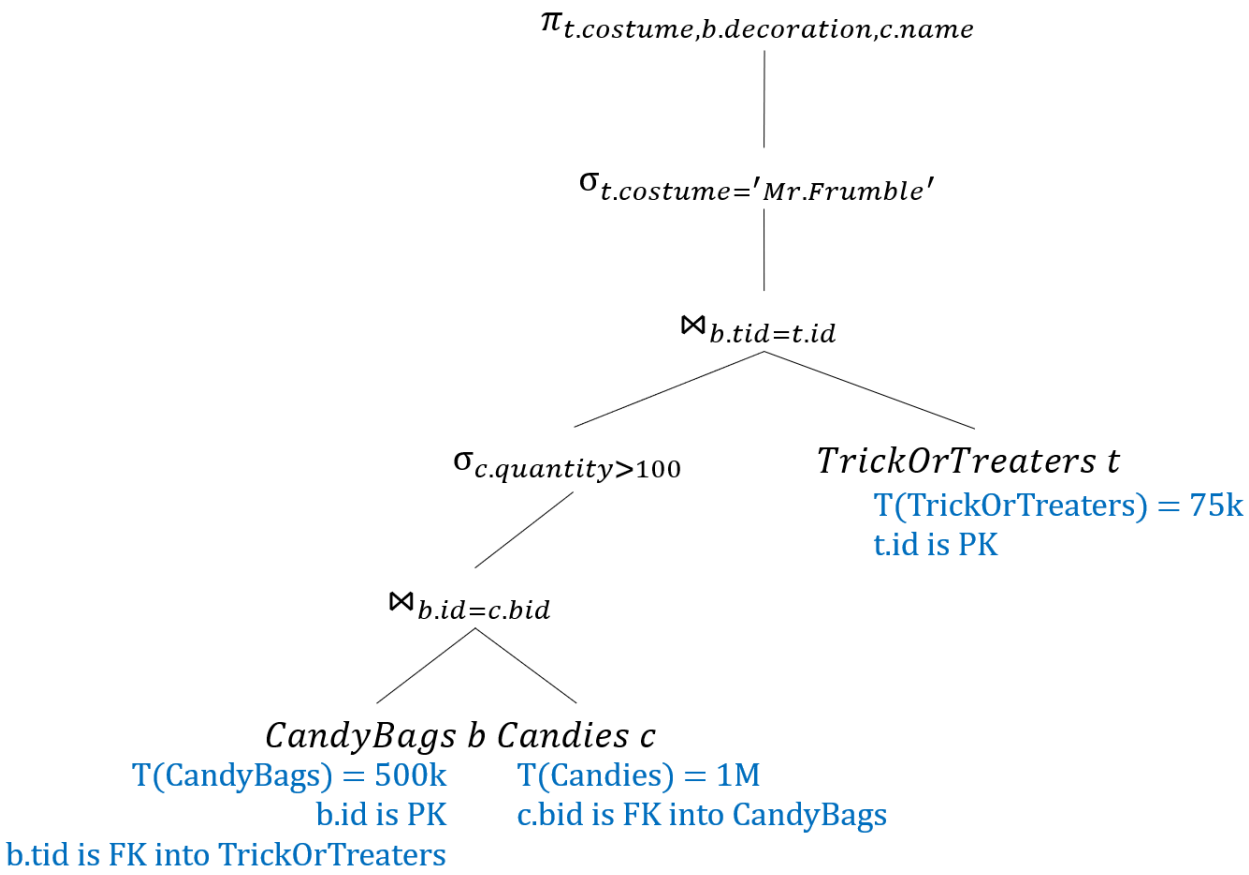
```
SELECT DISTINCT c.name
  FROM CandyBags b, Candies c, TrickOrTreaters t
 WHERE t.id = b.tid AND b.id = c.bid
   AND t.costume = 'Bojack Horseman'
   AND NOT EXISTS (
        SELECT *
          FROM CandyBags b2, Candies c2, TrickOrTreaters t2
         WHERE t2.id = b2.tid AND b2.id = c2.bid
           AND t2.costume = 'Princess Caroline'
           AND c2.isFullSize = c.isFullSize
   );
```

Can this query be rewritten as an *uncorrelated* query? If so, please decorrelate; if not, explain why in no more than 5 sentences.

3. This and the next two problems require the following statistics:

T(TrickOrTreaters) = 75,000 T(CandyBags) = 500,000 T(Candies) = 1,000,000 V(TrickOrTreaters, costume) = 25,000 min(Candies, quantity) = 1 max(Candies, quantity) = 201

and refer to this RA tree:



For each of the 5 stages, determine its selectivity factor. Then, compute the output cardinality of those stages; as a hint, the final projection produces 20 tuples.

4. ~~*Using*~~ *Starting from* the same tree as ~~*before*~~ *question 3*, push the selection operators down as low as possible. Submit your modified tree to Gradescope.

5. For each of the 5 stages *in ~~your~~ the tree you modified ~~tree~~ in question 4*, determine its selectivity factor and cardinality. Then, compute the cardinality of those stages. Since

these trees are semantically identical, your modified tree's final projection still produces 20 tuples.

6. Please read the following two sections:
    - How we rank Feed
    - How you can influence what you see

    of this article describing Facebook's algorithm for ranking Instagram posts. Then, describe the scale, impact, and opacity of this algorithm. You should be able to answer each section in a maximum of 3-5 sentences, and you may find this example helpful.

7. Next, please reflect on your own personal experience. Specifically:
    - What is one thing that you *learned* while doing this assignment?
    - What is one thing that *surprised you* while doing this assignment?
    - What is one *question that you still have* after doing this assignment?

8. Unlike usual, the following questions are *required*. This homework is newish to CSE 344, and we need your feedback to help us improve it for future students:
    - How many hours did it take you to finish this assignment, including time to set up your computer (if necessary)?
    - How many of those hours did you feel were valuable and/or contributed to your learning?
    - Did you collaborate with other students on this assignment? If so, approximately how many people did you collaborate with? Do not include yourself or course staff in the count.

Submission Instructions

You should submit the questions on Gradescope under HW5. You will need to submit two different RA trees, which we recommend uploading in PDF format.