CSCB20 A1

Winter 2022

Points: 59

Release Date: February 1st

Due Date: February 17th, 11:59pm(with 3 day extension)

Submission Platform: MarkUs

Note: if a piece of text has been highlighted, that means it has been modified compared to the original
Last Updated: February 13th 7:40pm

Context
In this assignment, you are a SQL Developer Intern on the University of Toruneto Health and Safety Team. Your team is working on projects related to Uni 2.0, a hybrid online and in person method of education during the pandemic. The senior database developer has written out some relationship schemas and would like you to write out RA and/or SQL queries for the following.

Instructions & Important Notes
In General,

- It is possible that not all subparts of a question will be marked. However, attempt all the questions.

For RA,

- Type (**Do not handwrite**) the queries below in relational algebra.
- You must use the same relational algebra notation as we have used in class and on the slides.
- It is possible that the relational algebra query cannot be expressed, in that case, write 'cannot be done in RA' and explain why not.
- Since there are many ways to write RA, explanations on the more complicated RA queries may help the TA understand what is going on.

For SQL,

- You are required to create the defined schema as a table in your database in order to test your queries only(i.e. Don't submit your DDL)
- It is essential your database tables do not deviate from the defined schema above, not doing so will result in some, or all of your SQL queries not being able to run on our end.
- Enter your completed queries in the A1.sql starter code, your submitted file should be able to compile from top to bottom without compile errors
- Any SQL query that does not compile will have a mark of 0.
- Any SQL query that runs but does not match the result of our test data will receive at most 70%.

Q1. Relational Algebra

The following is a safety supplies schema used to order various types of supplies.

Subsuppliers (sid: integer, subid: integer, sname: text, saddress: text, scountry: text, subname: text, subaddress: text, subcountry: text)

- A tuple in this relation represents two suppliers, of which one is a sub-supplier of the other. Any column related to the sub-supplier starts with 'sub'.

Suppliers (sid: integer, sname: text, saddress: text, scountry: text)

- A tuple in this relation represents a supplier, and the supplier's information.

Product (pid: integer, pname: text)

- A tuple in this relation represents the products available to the University.

ProductTag (tid: integer, pid: integer, tagname: text)

- A tuple in this relation represents what category each product is tagged with. It is possible for a product to have multiple tags associated with it.

Catalog (sid: integer, pid: integer, cost: real)

- A tuple in this relation represents the products that each supplier offers, at their respective costs.

Inventory (pid: integer, quantity: integer)

- A tuple in this relation represents how much of a product the University has in stock.

a) (5 points) Convert these relation algebra queries into simple English sentences

i)

$$\pi_{sname}(\pi_{sid}((\sigma_{tagname='PPE'}ProductTag) \bowtie (\sigma_{cost<6}Catalog)) \bowtie Suppliers)$$

ii)

$$\pi_{sname}(\pi_{sid}((\sigma_{tagname='PPE'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers))$$

iii)

$$\pi_{sname}((\sigma_{tagname='PPE'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers) \cap$$
$$\pi_{sname}((\sigma_{tagname='SuperTech'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers)$$

iv)

$$\pi_{sid}((\sigma_{tagname='PPE'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers) \cup$$
$$\pi_{sid}((\sigma_{tagname='SuperTech'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers)$$

v)

$$\pi_{sname}((\pi_{sid,sname}((\sigma_{tagname='PPE'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers)) \cap (\pi_{sid,sname}((\sigma_{tagname='SuperTech'}ProductTag) \bowtie (\sigma_{cost<6}Catalog) \bowtie Suppliers)))$$

*(an extra bracket was added at the end to correct the question)

b) (10 points) Write the relational algebra for these following questions

i) Find the names of the suppliers who supply some 'PPE' or 'Testing' product.

ii) Find the sids of the suppliers who supply 'PPE' lower than $10 and greater than $420.

iii) Find the sids of the suppliers who do not supply 'PPE' lower than $10 or greater than $1337.

iv) Find the sids of the suppliers who supply every type of 'Cleaning' products in their catalog.

v) Find pairs of sids such that the supplier with the first sid charges 20% or more for some product than the supplier with the second sid.

vi) Find the pids of products supplied by at least two different suppliers.

vii) Find the sid selling the most expensive 'Super Tech' located in 'USA'.

viii) Find the sid selling the second most expensive 'Super Tech' located in 'USA'.

ix) Find the pids of products supplied by every supplier at less than $69.

x) Find the pids of products of which we do not have any in our inventory

c) (8 points) Write the relational algebra for these harder questions with comments

i) For each pair of suppliers that have a "business relationship" (*[1]) with each other, find pids they both offer in their catalog, but which we do not have inventory of. Return the columns as pid, sid1, sid2, cost1, cost2.

ii) For each pid, find the suppliers that have products listed in their catalog at the exact same price. Return the columns as pid, sid, cost.

iii) Find the pids that have been listed as at least 3 different tags. However, one of the tags must be 'PPE', and one of them must not be 'Super Tech'. Return columns containing the pid, pname, cost.

iv) For each pair of "reciprocal subsuppliers"(*[2]), find all of their "uncommon subsuppliers"(*[3]). Every uncommon subsupplier of the pair should have only one row. Return the sid of the

reciprocal subsuppliers, along with the sid, name and address of the uncommon subsupplier.

(*1)business relationship: supplier A is a subsupplier of supplier B. Here, suppliers A and B have a business relationship with each other.
(*2)reciprocal sub suppliers: supplier A is a subsupplier for supplier B, and supplier B is also a subsupplier for supplier A. Here, Supplier A and B are reciprocal subsuppliers with each other. It does not matter for which products.
(*3)uncommon subsuppliers: supplier C is a subsupplier for supplier A, but supplier C is *not* a subsupplier for supplier B. Here, supplier C is the uncommon subsupplier of supplier A,B.

d) (2 marks) Revise the Schema

In your opinion, this database schema is pretty bad. Show off your skills to the manager and rewrite the schema. There are no right or wrong answers. Marks are based on the quality of answers, and any theories, guidelines or practices you've incorporated. Feel free to take some time to research the topic, but remember to cite your sources.

e) (1 bonus mark) Meme

After showing your schema in part d to the manager, he/she decides to fire either you, or the senior database developer. Who gets fired? Make up a funny meme or short story. If you can make the TA grin, you get the point.

(more on next page)

Q2. Relational Algebra Continued

This is an enhanced contact tracing schema used to log community members' travel within campus. Every time someone enters a room, they are required to scan their contact tracing QR code. We assume that everyone follows the rules.

Occupancy (occupancyid: integer, utorid: integer, roomid: integer, alertlevel: integer, date: text)

- A tuple in this relation represents each time a member enters a room and scans their QR code. The alertlevel represents the danger level of the member, where 0 is the lowest with no exposure, and 5 is highest meaning they are an active spreader. The alertlevel of an individual can be updated at any time by the database analysts to see which room they've exposed. The date is in text format, represented as 'YYYY-MM-DD'.

Room (roomid: integer, roomname: text, alertthreshold: integer)

- A tuple in this relation represents a room in the University. The alertthreshold is the maximum exposure allowed from levels 0 to 5, similar to alertlevel. For example, if the room is a rooftop terrace, the alertthreshold would be 5. Meaning that even if someone has covid entered this room in the past, we would not need to notify others.

Approved (utorid: integer, roomid: integer)

- A tuple in this relationship represents the members approved to be in a certain room. For example, a student would not be able to enter the janitor's office.

Member (utorid: integer, name: text, email: text, vaxstatus: integer)

- A tuple in this relationship represents the types of roles each member associates as at the University. We assume a member cannot be associated with more than one role.

Student (utorid: integer, cgpa: integer)

- A tuple in this relationship represents a student.

Employee (utorid: integer, position: text, salary: integer)

- A tuple in this relationship represents employees and their annual salary.

a) (8 points) Write the relational algebra for these following questions

i) Find the utorids of students not approved to be in room 'IC404'.

ii) Find the utorids of employees that are approved for at least 3 rooms.

iii) Find the utorids of employees that are approved for exactly 3 rooms.

iv) Find the utorids of employees that are approved for at most 3 rooms

v) Find the roomids where the student 'Oscar Lin' was approved for, and where the alert level has exceeded the alert threshold between 2021-09-01 and 2021-12-31.

vi) Find utorids of members that traveled between 2020-03-17 and 2021-12-31 but not approved to be in at least one room.

vii) Find the total amount paid to employees as salaries.

viii) Find the utorids and emails of students with vaxstatus ‘0’ that exceeded any room’s alert threshold.

Q3: SQL Queries

a) (10 marks) For each subpart in question 1B above, write down the SQL query

b) (8 marks) For each subpart in question 1C above, write down the SQL query

c) (8 marks) For each subpart in question 2 above, write down the SQL query

(more on next page)

Style and formatting requirements

To make your algebra more readable, and to minimize errors, we are including these style and formatting requirements:

- In your assignment statements in RA, you must include names for all attributes in the intermediate relation you are defining.
- Use meaningful names for intermediate relations and attributes, just as you would in a program.
- If you want to include comments, put them before the algebra that they pertain to, not after. Make them stand out from the algebra, for example by using a different font.

  For example, this looks reasonable:

  /* Students who had very high grades in any offering of a csc */

  $$course.High(sID) := \pi_{sID}\sigma_{dept='csc' \wedge grade>95}(Department \bowtie Student)$$

  A modest portion of your mark will be for good style and formatting.

Submission Instructions:

- Your RA questions **must be typed; handwritten assignments will not be marked**. You may use any word-processing software you like. Many academics use LaTeX. It produces beautifully typeset text and handles mathematical notation well. If you would like to learn LaTeX, there are helpful resources online.
- Whatever you choose to use, you need to produce a final document in pdf format. You must declare your team (whether it is a team of two or three students) and hand in your work electronically using the MarkUs online system.
- Well before the due date, you should declare your team and try submitting with MarkUs.
- For this assignment, hand in two files: **A1.pdf**, and **A1.sql** (for the sql file, please use the format provided in the starter A1.sql file)
- If you are working in a group, only one of you should hand it in. Check that you have submitted the correct version of your file by downloading it from MarkUs.
- Late submissions with 10% penalty will be accepted until day 3 (72 hours) past the original submission deadline. Later submissions will not be graded unless valid reason for the late submission is provided.