CSCC43 Introduction to Databases
(Summer 2009)

Assignment 1 - Relational Algebra

Electronic copy is due at: Sunday, May 31, at 11:59pm

Hard Copy is due in tutorial at: Monday, June 1, at 1:20pm

Total Marks: 100

Weight: 10% of your final grade

The last 48 hours before the deadline are considered a silent period.

Student 1

NAME (LAST, FIRST)   : _____

STUDENT NUMBER   : _____

Student 2

NAME (LAST, FIRST)   : _____

STUDENT NUMBER   : _____

UTSC login name: _____

Schema

Consider the following (simplified) relational schema describing online product orders of a company.

- *Customer(cid,clname,cfname,phone)*, which records information about the customers of the company, including a customer's id (*cid*), last name (*clname*), first name (*cfname*) and contact phone number (*phone*). The primary key of this relation schema is (*cid*).

- *Address(aid,street,postal,city)*, which records information about customers' addresses, including an address's id (*aid*), street number and name (*street*), postal code (*postal*) and city name (*city*). The primary key of this relation schema is (*aid*).

- *CustomerAddress(cid,aid)*, which records the current address of customers, including the id of a customer (*cid*), the id of the current address (*aid*) the customer has/had. The primary key of this relation schema is (*cid*).

- *Employee(eid,elname,efname)*, which records information about the employees of the company, including an employee's id (*eid*), last name (*elname*) and first name (*efname*). The primary key of this relation schema is (*eid*).

- *Product(pid,desc,price)*, which records information of products of the company, including the id of a product (*pid*), the description of the product (*desc*) and the unit price of the product (*price*).

- *Order(num,cid,eid,date)*, which records information about the orders placed by the customers, including the order number (*num*), the id of the customer who placed the order (*cid*), the id of the employee who processed the order (*eid*), and the date when the order was placed (*date*). The primary key of this relation schema is (*num*).

- *OrderDetail(num,pid,quantity)*, which records information about order details, including the order number (*num*), the id of a product in the order (*pid*), and quantity of the product in the order. An order may contain more than one product. The primary key of this relation schema is (*num,pid*).

Queries

Write the following queries in Relational Algebra. Extended Relational Algebra operators, such us max and count, are not permited.

Q1 (5 points): Find the first and last names of customers who placed an order after April 1, 2009 (using '04/01/2009' to represent this date in your query).

Sample Solution

$\pi_{cfname,clname}\sigma_{date>'04/01/2009'}(\text{Customer} \bowtie \text{Order})$.

Q2 (10 points): Find the ids, first and last names of employees who processed orders placed by different customers who have the same current address.

Sample Solution

$\pi_{eid,fname,lname}(\text{Employee} \bowtie \text{B})$, where:

- $\text{B} = \rho_{A_1}(A) \bowtie_{A_1.num \neq A_2.num \wedge A_1.aid=A_2.aid \wedge A_1.eid=A_2.eid} \rho_{A_2}(A)$
- $\text{A} = (\text{Order} \bowtie \text{Customer} \bowtie \text{CustomerAddress})$

Q3 (5 points) Find the orders in which the total amount for some product exceeds \$1,000; Show the order numers only.

Sample Solution

$\pi_{num}\sigma_{price \times quantity>1000}(\text{Order} \bowtie \text{OrderDetail} \bowtie \text{Product})$.

Q4 (10 points) Find the ids of employees who processed **all** the orders in which the total amount for some product exceeds \$1,000 (hint: you may use your result for Q3).

Sample Solution

B / A, where:

- $\text{B} = \pi_{eid,num}(\text{Employee} \bowtie \text{Order})$
- $\text{A} = \pi_{num}\sigma_{price \times quantity>1000}(\text{Order} \bowtie \text{OrderDetail} \bowtie \text{Product})$

Q5 (15 points) Find the first and last names of customers who placed an order with at least two products or who placed an order containing a product with unit price more than \$100.

Sample Solution

D $\cup$ B, where:

- $\text{D} = \pi_{cfname,clname}(\text{Customer} \bowtie \text{Order} \bowtie \text{C})$
- $\text{C} = \pi_{O_1.num}(\rho_{O_1}(OrderDetail) \bowtie_{O_1.num=O_2.num \wedge O_1.pid \neq O_2.pid} \rho_{O_2}(OrderDetail))$

- B $=\pi_{cfname,clname}$(Customer $\bowtie$ Order $\bowtie$ A)
- A $= \sigma_{price>'100'}$(OrderDetail $\bowtie$ Product)

Q6 (15 points) Find the id and description of the product with the highest unit price in order '100001' (note: there may be more than one product with the same highest unit price).

B $/\ \pi_{pid}$(A)

B $= \pi_{A_1.pid1,A_1.desc1,A_2.pid}(\rho_{A_1(pid1,desc1,price1)}(A) \bowtie_{A_1.price1 \geq A_2.price} \rho_{A_2}(A))$

A $= \pi_{pid,desc,price}\sigma_{num='100001'}$(OrderDetail $\bowtie$ Product)

Questions

Answer following questions.

Q7 (10 points) Consider Q1 again. Explain in detail how your query for Q1 will produce the desired results. Will your query give the same result when some customer (i) did not place any order at all, (ii) placed one order, or (iii) placed more than one orders?

This query first find, for each customer, all the orders the customer placed (through the natural join); let's call this intermediate result A. If some customer did not place any order, he/she will not appear in A. Then it selects from A all the customers who has placed at least one order after April 1, 2009; let's call this intermediate result B. Finally, it projects B onto *cfname* and *clname*; let's call the final result C.

So if a customer did not place any order, he/she will not appear in A (and hence in C). If a customer placed more than one order, he/she will appear multiple times in A (possibly in B); however, since in Relational Algebra we are dealing with sets, any duplication will be eliminated in C. In summary, (ii) and (iii) will always give the same result, while (i) will produce different result.

Q8 (10 points) Having a separate *CustomerAddress(cid,aid)* is not necessary. Explain why? If we want to store for a customer both current and past addresses, is the schema suffient? If yes, explain why; if not, modify the schema to allow this.

Since *Customer(cid,clname,cfname,phone)* and *CustomerAddress(cid,aid)* share the same primary key (*cid*); they can be merged into a single relation schema *Customer(cid,clname,cfname,phone,aid)*.

The given schema cannot store multiple addresses for a customer. To allow this, the minimal change would be introducing a new attribute into *AddressHistory* as in *AddressHistory(cid,aid,start)*, where the new attribute *start* represents the effective date of an address. A customer may be associated with more than one address with different effective date; the address of the most recent effective date is the active addresss of the customer. The primary key of this relation schema is (*cid,aid,start*).

Q9 (10 points) Explain what the following query will return?

$$\pi_{num}(\pi_{cid,eid}(A) \bowtie \text{Order})$$

where $A = \rho_{C(cid,ln,fn,phone)}(\text{Customer}) \bowtie \rho_{E(eid,ln,fn)}(\text{Employee})$.

Sample Solution

This query finds the order number of all the orders which has been ordered by a customer who has exact the same name as the employee who processed the order.

Q10 (10 points) Explain what the following query will return?

$$A - B$$

where

- $A = \pi_{eid}(\text{Employee} \bowtie \text{Order})$
- $B = \pi_{eid}(C - \pi_{eid,cid}(\text{Employee} \bowtie \text{Order}))$
- $C = \pi_{eid}(\text{Employee} \bowtie \text{Order}) \times \pi_{cid}(\text{Customer})$

Sample Solution

This query finds the ids of employees who has processed at least one order for **every** customer. If you cannot see why, here is more detailed explanation:

- "$A = \pi_{eid}(\text{Employee} \bowtie \text{Order})$" finds the ids of employees who have processed some order.
- "$C = \pi_{eid}(\text{Employee} \bowtie \text{Order}) \times \pi_{cid}(\text{Customer})$" finds all the possible pairs of ids of employees (who has processed some order) and ids of customers.
- "$B = \pi_{eid}(C - \pi_{eid,cid}(\text{Employee} \bowtie \text{Order}))$" find the ids of employees who has not processed any order for some customer.
- "$A - B$" find the ids of employees who has processed at least one order for every customer.

4

Submission instructions

You can work in groups of 2 people, with only one submission per group. Your submission must be typed. Handwritten assignments will not be accepted. The electronic submission must contain the following files (with these exact names):

1. **group.txt** - text file, with information (first and last names, student numbers) about all members of the group, and your UTSC login account; you can download a template of this file on the assignement page on the course website at here
2. **a4.<extension>** - your solution to this assignment
   - you can use one of the acceptable formats (**.ps, .pdf or plain ASCII file**); we prefer Postscript or PDF submissions (you may lost points for clarity and readability reasons). Submissions in other format will not be marked.
   - It is very important that you include in this file the information from group.txt as well (suggestion: include it in the first page of your file).
3. Check the assignments section on the course website for using *submit* command.

Suggestions

You can use any editor you like, or you are familiar with. If you want to edit your assignment in LaTeX, you can find some useful information on the course website. You can use the example file provided there as a reference. If you choose to copy some examples from there, please acknowledge this in your assignment.