Database Management Systems

(COP 5725)

Fall 2019

Instructor: Dr. Markus Schneider

TA: Kyuseo Park

Homework 2 Solutions

Name:	
UFID:	
Email Address:	

Pledge (Must be signed according to UF Honor Code)

On my honor, I have neither given nor received unauthorized aid in doing this assignment.

____

Signature

For scoring use only:

	Maximum	Received
Exercise 1	25	
Exercise 2	25	
Exercise 3	22	
Exercise 4	28	
Total	100	

Exercise 1 (Relational Algebra) [25 points]

Consider the following relations. The primary keys are underlined. All attributes are of type string if not indicated otherwise.

- Student (<u>s_ID</u>, s_name, s_degree: integer, advisorID, d_ID)
- Lecture (<u>l_ID</u>, l_name, l_degree: integer, p_ID, d_ID)
- Register (<u>s_ID</u>, <u>l_ID</u>, score: integer, Semester)
- Professor (<u>p_ID</u>, p_name, d_ID)
- Department (<u>d_ID</u>, d_name, address)

1. [5 points] Find the names of professors who have taught in every semester.
   - $\rho_{LectureProfessorSemester}(\pi_{l\_ID,\ semester}(Register) \bowtie \pi_{l\_ID,\ p\_ID}(Lecture))$
   - $\rho_{ProfessorTaughtInSemester}(\pi_{p\_ID,\ semester}(LectureProfessorSemester)))$
   - $\rho_{ProfessorTaughtInEverySemester}(ProfessorTaughtInSemester \div \pi_{semester}(Register))$
   - $\pi_{p\_name}(ProfessorTaughtInEverySemester) \bowtie Professor$

2. [5 points] List the names of lectures that the CISE department offers but that are taught by a professor whose department is not CISE.
   - $\rho_{CISE}\ (\pi_{d\_ID}(\sigma_{d\_name='CISE'}(Department)$
   - $\rho_{CISETeaching}\ ((\pi_{l\_name,p\_ID}(Lecture \bowtie CISE)) \bowtie Professor)$
   - $\pi_{l\_name}(\sigma_{CISETeaching.d\_ID<>CISE.d\_ID}\ (CISETeaching \times CISE)$

3. [5 points] Find the names of students who got the highest score in the lecture ‘Databases’.
   - $\rho_{DB}\ (\pi_{s\_ID,score,s\_name}\ (\sigma_{l\_name='Databases'}\ (Student \bowtie Register \bowtie Lecture)))$
   - $\rho_A(DB), \rho_B(DB)$
   - $\pi_{s\_name}\ (DB - (\pi_{B.s\_ID,B.score,B.s\_name}\ (A \bowtie_{A.score>B.score} B)))$

4. [5 points] Find the names of students who have registered every lecture of the CISE department.
   - $\pi_{s\_name}\ ((Register \div \pi_{l\_ID}\ (\sigma_{d\_name='CISE'}(Lecture \bowtie Department))) \bowtie Student)$

5. [5 points] Find the names of students who got more than 90 in the ‘DB’ lecture and less than 70 in the ‘Algorithm’ lecture.

- $\pi_{s\_name}\left(\sigma_{l\_name='DB' \wedge score>90}\ (Student \bowtie Register \bowtie Lecture)\right)$
  $\cap$
  $\pi_{s\_name}\ (\sigma_{l\_name='Algorithm' \wedge score<70}(Student \bowtie Register \bowtie Lecture))$

Exercise 2 (Relational Algebra) [25 points]

Consider the following relations for an online bookstore. The primary keys are underlined. All attributes are of type string if not indicated otherwise.

- Book(<u>bID</u>, title, author, publisher, year, price)
- Customer(<u>cID</u>, name, ssn, gender, age: integer, email, street, city, state, zipcode)
- Credit_Card(<u>ccID</u>, cID, cardNumber, exprDate)
    - cID is a foreign key to Customer relation.
- Book_Order(<u>bID, ccID</u>, quantity: integer, time)
    - bID is a foreign key to Book relation.
    - ccID is a foreign key to Credit_Card relation.

1. [4 points] Find the titles of books that were purchased by the customers Ahn or Berr.
    - $\pi_{\text{title}}\left(\sigma_{name='Ahn'}(Book \bowtie Customer \bowtie CreditCard \bowtie Book\ Order)\right)$
    $\cup$
    $\pi_{\text{title}}\left(\sigma_{name='Berr'}(Book \bowtie Customer \bowtie CreditCard \bowtie Book\ Order)\right)$

2. [6 points] Find the names of customers who ordered the book "DB" the most.
    - $\rho_{DB\ BUYER}\left(\pi_{\text{name,quantity}}\left(\sigma_{title='DB'}(Book \bowtie Customer \bowtie CreditCard \bowtie Book\ Order)\right)\right)$
    - $DB\ BUYER - \left(\pi_{\text{B.name,B.quantity}}\left(\rho_A(DB\ BUYER) \bowtie_{A.quantity>B.quantity} \rho_B(DB\ BUYER)\right)\right)$

3. [4 points] Find the names of publishers whose books are not sold yet.
    - $\rho_{Publishers}\left(\pi_{publisher}(Book \bowtie Book\ Order)\right)$
    - $\pi_{publisher}(Book) - Publishers$

4. [3 points] Find the genders of customers who ordered the "Machine Learning" book authored by Chris Bishop.
    - $\pi_{gender}\left(Customer \bowtie \left(CreditCard \bowtie \left(BookOrder \bowtie \sigma_{title='Machine\ Learning' \wedge author='Chris\ Bishop'}(Book)\right)\right)\right)$

5. [4 points] List the titles of books that were purchased by customers whose credit cards are expiring in 2018.
    - $\pi_{title}\left(Book \bowtie \left(BookOrder \bowtie \left(CreditCard \bowtie \left(\pi_{cId}(CreditCard) - \pi_{cId}\left(\sigma_{exprDate<2018 \vee exprDate>2018}(CreditCard)\right)\right)\right)\right)\right)$

6. [4 points] List the emails of customers who have not purchased any book.

- $\pi_{email}((\pi_{cId}(Customer) - \pi_{cId}(Customer \bowtie CreditCard \bowtie BookOrder)) \bowtie \pi_{cId,email}(Customer))$

Exercise 3 (Relational Algebra) [22 points]

Consider the following relations. The primary key is underlined. All attributes are of type string if not indicated otherwise.

- branch (<u>branch_name</u>, branch_city, assets: integer)
- customer (<u>customer_id</u>, customer_name, customer_street, customer_city)
- account (<u>account_number,</u> branch_name, balance)
- depositor (<u>customer_id, account_number</u>)

1. [6 points] Find the names of customers who have an account in the branch with highest assets.
   - $\rho_{HighestBranch}\ (\pi_{branch\_name,assets}\ (branch) - \pi_{B.branch\_name,B.assets}\ (\rho_A(branch) \bowtie_{A.assets > B.assets}\ \rho_B(branch)))$
   - $\pi_{customer\_name}(customer\ \bowtie\ depositor\ \bowtie\ account\ \bowtie\ HighestBranch)$

2. [4 points] Find the names of customers who live in the same city where the branches are located that the customers use.
   - $\pi_{customer\_name}(\sigma_{branch\_city=\ customer\_city}\ (customer\ \bowtie\ account\ \bowtie\ depositor\ \bowtie\ branch))$

3. [4 points] Find the names of customers whose account balance is less than \$1,000 and the branch that has the account is located in Gainesville.
   - $\pi_{customer\_name}(\ \sigma_{branch\_city\ =\ 'Gainesville'\ \wedge\ balance\ <\ 1000}\ (customer\ \bowtie\ depositor\ \bowtie\ account\ \bowtie\ branch)\ )$

4. [4 points] Find the names of customers that have accounts in every branch in Gainesville.
   - $\pi_{customer\_name}\ (\ (customer\ \bowtie\ account\ \bowtie\ depositor)\ \div\ \pi_{branch\_name}(\sigma_{branch\_city\ =\ 'Gainesville'}(branch)\ ))$

5. [4 points] Find the names of customers who do not have an account. (assume the bank keeps the customer information even if the account is closed.).
   - $\pi_{customer\_name}((\pi_{customer\_id}\ (customer)\ - \pi_{customer\_id}\ (depositor)\ )\ \bowtie\ customer)$

Exercise 4 (Relational Algebra) [28 points]

The following questions let you think deeper about the concepts of the Relational Algebra.

1. [6 points] Let $\mathbb{R}$ be the schema of a relation R, and let $A_1, A_2, \ldots, A_n \subseteq \mathbb{R}$. Is the term $\pi_{A_1}(\pi_{A_2}(\ldots(\pi_{A_n}(R))\ldots)) = \pi_{A_1}(R)$ correct, in general? If yes, argue why. If not, argue why not. If your answer is no, are there any restrictions that could make the statement true? If so, what are these restrictions in mathematical notation?

   The answer is no, in general. Since the attribute sets A_i are arbitrary subsets of R, it could be that attributes do not exist any more for a projection due to their elimination in an earlier projection. In other words, the statement is true if $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \mathbb{R}$ holds. This ensures that each subsequent projection has access to the attributes it needs for performing this operation. Since A_1 is the smallest of all attribute sets, we can directly take it for projection. In other words, the right side of the expression above is an optimization of the left side.

2. [6 points] Let $\mathbb{R}$ be the schema of a relation R, and let $A \subseteq \mathbb{R}$. What is the condition in mathematical notation such that $\pi_A(\sigma_F(R)) = \sigma_F(\pi_A(R))$ holds where F is assumed to be a correct predicate on R?

   We must ensure that the attributes in A are the only ones used in F. Otherwise, the right side could access attributes in F that have been removed by the preceding projection. Let *attr* be a function that extracts all attributes used in F into a set. Then the condition for the fulfilment of the statement is: $attr(F) \subseteq A$.

3. [16 points] Let $R(A, B)$ be a relation with $r > 0$ tuples, and let $S(B, C)$ be a relation with $s > 0$ tuples. We assume that A, B, and C have the *same data type*. For each of the following Relational Algebra expressions, in terms of r and s, determine the *minimum* and *maximum number of tuples* that the result relation can have. In other words, we are interested in the number of tuples the following Relational Algebra expressions can have *at least* and *at most*. The numbers have to be given by using the two variables r and s. Please note that you have to give precise explanations for your answers.

   a. [4 points] $R \cup \rho_{T(A,B)}(S)$

      - $\rho_{T(A,B)}(S)$ renames S into T and the attributes B and C of S into A and B to make both relation schemas compatible.
      - Minimum number of possible tuples in the result
        - Case 1: $R = T \rightarrow R \cup T = R \rightarrow$ Result: r $[r = s]$
        - Case 2: $R \subset T \rightarrow R \cup T = T \rightarrow$ Result: s $[r < s]$
        - Case 3: $T \subset R \rightarrow R \cup T = R \rightarrow$ Result: r $[r > s]$
        - Overall result: $\max(r, s)$
      - Maximum number of possible tuples in the result
        - $R \cap T = \varnothing \rightarrow$ Result: $r + s$

b. [3 points] $(R \bowtie R) \bowtie R$

- $R \bowtie R = R \rightarrow (R \bowtie R) \bowtie R = R$
- Minimum number of possible tuples in the result: r
- Maximum number of possible tuples in the result: r

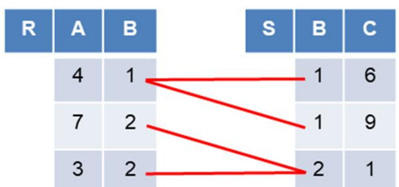
c. [4 points] $\pi_{A,\,C}(R \bowtie S)$

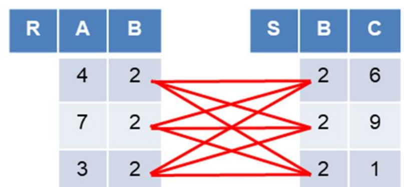
- Minimum number of possible tuples in the result

  $\pi_B(R) \cap \pi_B(S) = \varnothing \rightarrow R \bowtie S = \varnothing \rightarrow \pi_{A,\,C}(R \bowtie S) = \varnothing \rightarrow$ Result: 0

- Maximum number of possible tuples in the result

  $\pi_B(R) = \pi_B(S) \wedge |\pi_B(R)| = |\pi_B(S)| = 1 \rightarrow$ Result: $r \cdot s$





d. [5 points] $\sigma_{A=C}(R \bowtie S)$

- Minimum number of possible tuples in the result
  - Case 1: $\pi_B(R) \cap \pi_B(S) = \varnothing \rightarrow R \bowtie S = \varnothing$
    $\rightarrow \sigma_{A=C}(R \bowtie S) = \varnothing \rightarrow$ Result: 0
  - Case 2: $\pi_B(R) \cap \pi_B(S) \neq \varnothing \wedge \forall\, t \in R \bowtie S : t.A \neq t.C$
    $\rightarrow \sigma_{A=C}(R \bowtie S) = \varnothing \rightarrow$ Result: 0
- Maximum number of possible tuples in the result
  - $\pi_B(R) = \pi_B(S) \wedge |\pi_B(R)| = |\pi_B(S)| = 1 \wedge (R \subseteq \pi_{C,B}(S) \vee R \supset \pi_{C,B}(S))$
    $\rightarrow$ Result: $\min(r, s)$