Database Management Systems (COP 5725)

Spring 2017

Instructor: Dr. Markus Schneider

TAs: Keke Zhai, Man Mohan Devineni

Homework 2 Solutions

<table>
<thead>
<tr>
<th>Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UFID</td>
<td></td>
</tr>
<tr>
<td>Email Address</td>
<td></td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Maximum</th>
<th>Received</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise 1</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Exercise 2</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Exercise 3</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Exercise 1 (Knowledge Questions) [25 points]

Answer the following questions:

1. Briefly describe the terms relation, key, and primary key.[5 points]

   Relation – Given n domains D1, D2..., Dn(which may contain only atomic values),
   A relation (instance) \( r_R \) is defined as a subset of the Cartesian product of n domains:

   \[
   r_R \subseteq D_1 \times D_2 \times \ldots \times D_n \quad (r_R \text{ finite})
   \]

   Key – Given \( R(A_1, A_2, \ldots, A_n) \) and let \( X \subseteq \{A_1, A_2, \ldots, A_n\} \), X is called key, if the following conditions are fulfilled:
   - uniqueness: for all relation instances \( r_R \) of \( R \) holds:
     \[
     \forall t_1, t_2 \in r_R : t_1[X] = t_2[X] \Rightarrow t_1 = t_2 \cdots \cdots
     \]
   - minimality: there is no \( Y \subsetneq X \), so that the uniqueness is fulfilled.

   Primary Key – Selected candidate key that identifies tuples uniquely within a relation. No two tuples within a relation can have the same primary key.

2. What is a 1: m relationship set \( R \) between two entity sets E1 and E2? (5 points)

   Each member of E1 can be associated with any number of members from E2. However, each member of E2 can be associated with at most one member of E1.

3. Describe the definition of relational algebra and the basic operations in relational algebra with their name and their correct symbolic notation. [5 points].

   A relational algebra is a procedural language which allows specifying how a query has to be evaluated. It consists of a set of operations that take one or two relations as input and produce a new relation as their output.

   The basic operations in relational algebra are union \( \cup \), difference \( - \), Cartesian product \( \times \), projection \( \pi \), select \( \sigma \), and rename \( \rho \).
4. What is the difference between a set and a list? Use relation and tuple as simple examples. [5 points]

List is an ordered sequence of elements whereas Set is a distinct list of elements which is unordered. A relation is defined as a set of tuples in which no tuples are identical within one relation. A tuple is defined as a list of attribute values. These values are ordered according to the relation schema, but there is no restriction of the qualities between these values.

5. Express the Quotient (Division) operator $\div$ in terms of other basic operations. [5 points]

$$R \div S = \pi_{R \setminus S}(R) - \pi_{R \setminus S}(\pi_{R \setminus S}(R \times S) - R)$$
Exercise 2 (Relational Algebra) [25 Points]

Assume the following database schema for this homework:

Suppliers(sid: integer, sname: string, address: string)

Parts(pid: integer, pname: string, color: string)

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

Give an expression in the relational algebra to express each of the following queries.

(a) Find the names of suppliers who supply some blue part. [5 points]

\[ \pi_{\text{sname}}(\sigma_{\text{color}='\text{red}'}(\text{Parts} \bowtie \text{Catalog} \bowtie \text{Suppliers})) \]

(b) Find the sids of suppliers who supply some red part or are at 102 Pine Street. [5 points]

\[ \rho_{\text{R1}}(\pi_{\text{sid}}((\sigma_{\text{color}='\text{red}'}(\text{Parts} \bowtie \text{Catalog}))) \]
\[ \rho_{\text{R2}}(\pi_{\text{sid}}(\sigma_{\text{address}='102 Pine Street'}(\text{Suppliers}))) \]
\[ \text{R1} \cup \text{R2} \]

(c) Find the sids of suppliers who supply every blue part. [5 points]

\[ \pi_{\text{sid}, \text{pid}}(\text{Catalog}) \sqcup \pi_{\text{pid}}(\sigma_{\text{color}='\text{blue'}(\text{Parts}))} \]

(d) Find pairs of sids such that the supplier with the first sid charges more for some part than the supplier with the second sid. [5 points]

\[ \rho_{\text{R1}}(\text{Catalog}) \]
\[ \rho_{\text{R2}}(\text{Catalog}) \]
\[ \pi_{\text{R1}.\text{sid}, \text{R2}.\text{sid}}(\sigma_{\text{R1}.\text{pid}=\text{R2}.\text{pid}} \land \text{R1}.\text{sid}=\text{R2}.\text{sid} \land \text{R1}.\text{cost}>\text{R2}.\text{cost}(\text{R1} \times \text{R2})) \]

(e) Find the pids of the least expensive parts supplied by suppliers named Ramebit Asrim. [5 points]

\[ \rho_{\text{R1}}(\pi_{\text{sid}}(\sigma_{\text{sname}='\text{Ramebit Asrim'}(\text{Suppliers}))) \]
\[ \rho_{\text{R2}}(\text{R1} \bowtie \text{Catalog} \]
\[ \rho_{\text{R3}}(\text{R2}) \]
\[ \rho_{\text{R4}}(\pi_{\text{R3}.\text{pid}}(\sigma_{\text{R3}.\text{cost}>\text{R2}.\text{cost}} \land \text{R3}.\text{sid}=\text{R2}.\text{sid})(\text{R3} \times \text{R2})) \]
\[ \pi_{\text{pid}}(\text{R2}) \end{array} \]
\[ \rightarrow \text{R4} \]
Exercise 3 (Relational Algebra) [50 points]

Consider the following database schema:

**EMPLOYEE** (emp_name: string, address_street: string, address_city: string, phone_no: integer)

**DEPARTMENT** (dept_name: string, mgr_name: string, mgr_start_date: date)

**WORKS_ON** (emp_name: string, industry_name: string, salary: integer)

**INDUSTRY** (industry_name: string, location_city: string)

**MANAGES** (emp_name: string, mgr_name: string)

Express the following colloquial queries in Relational Algebra:

1. Find the names of all employees who work for ‘Verizon Wireless’. [7 points]
   \[ \pi_{\text{emp} \_\text{name}} (\sigma_{\text{industry} \_\text{name} = \text{Verizon Wireless}} (\text{Employee} \bowtie \text{Works} \_\text{on})) \]

2. Find the phone number of employee with the highest salary. [7 points]
   \[ \pi_{\text{phone} \_\text{no}} ((\text{Employee}) \bowtie (\pi_{\text{emp} \_\text{name}} (\text{Works} \_\text{on}) - \pi_{\text{emp} \_\text{name}} (\sigma_{\text{w}1.\text{salary} > \text{w}2.\text{salary}} (\rho_{\text{w}1} (\text{Works} \_\text{on}) \times \rho_{\text{w}2} (\text{Works} \_\text{on})))))) \]

3. Find the names and cities of residence of all employees who work for ‘IBM’ and earn more than $80,000 per year. [7 points]
   \[ \pi_{\text{emp} \_\text{name}, \text{address} \_\text{city}} (\sigma_{(\text{industry} \_\text{name} = \text{IBM}) \land (\text{salary} > 80000)} (\text{Works} \_\text{on}) \bowtie \text{Employee}) \]

4. Find the names of all employees who live in the same city and on the same street as their managers. [9 points]
   \[ \pi_{\text{e1.emp} \_\text{name}} (\sigma_{\text{e1.address_stress} = \text{e2.address_stress} \land \text{e1.address_city} = \text{e2.address_city}} (\rho_{\text{e1}} (\text{Employee}) \bowtie \text{e1.emp} \_\text{name} = \text{m.emp} \_\text{name} \rho_{\text{m}} (\text{Manages}) \bowtie \text{e2.emp} \_\text{name} = \text{m.mgr} \_\text{name} \rho_{\text{e2}} (\text{Employee}))) \]

5. Find the department name that every employee working in it lives in the ‘Genius’ street. [7 points]
   \[ \pi_{\text{dept} \_\text{name}} ((\text{Department} \bowtie \text{Manages}) \div (\pi_{\text{emp} \_\text{name}} (\sigma_{(\text{address_stress} = \text{Genius})} (\text{Employee})))) \]

6. Find the manager name with the earliest start time at department CISE. [7 points]
   \[ \rho_{\text{d1}} (\sigma_{\text{depart} \_\text{name} = \text{CISE}} (\text{Department})) \]

   \[ \rho_{\text{d2}} (\text{d1}) \]
7. Find the phone number of all employees who are not managers. [6 points]

\[ \pi_{\text{phone_no}}(\text{Employee} \bowtie (\pi_{\text{emp_name}}(\text{Employee}) - \pi_{\text{emp_name}}(\text{Manages}))) \]