Homework 1 (by Deepa Parameswaran)
Due: hard copy, in class at 1:30pm, on Tuesday, Feb. 3

VERY IMPORTANT: Deposit hard copy of your answers, in class. For ease of grading, please
1. Separate your answers, on different page(s) for each question (staple additional pages, if needed).
2. Type the full info on each page: your name, Andrew ID, course#, Homework#, Question# on each of the 4 pages.

Reminders:
• Plagiarism: Homework is to be completed individually.
• Typeset all of your answers whenever possible. Illegible handwriting may get zero points, at the discretion of the graders.
• Late homeworks: in that case, please email it
  – to all TAs
  – with the subject line exactly 15-415 Homework Submission (HW 1)
  – and the count of slip-days you are using.

For your information:
• Graded out of 100 points; 4 questions total
• Rough time estimate: ≈6 hours (1-2 hours for each question)

<table>
<thead>
<tr>
<th>Question</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-Relationship Diagram</td>
<td>25</td>
</tr>
<tr>
<td>SQL Tables from the ER Model</td>
<td>25</td>
</tr>
<tr>
<td>Relational Algebra for Job Portal</td>
<td>25</td>
</tr>
<tr>
<td>Relational Calculus</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Question 1: Entity-Relationship Diagram ............. [25 points]

On separate page, with ‘[course-id] [hw#] [question#] [andrew-id] [your-name]’

Consider a database to store information about a Research Organization. The database has the following properties:

- Every department has a title, and a unique department ID (departmentID).
- A department may have zero or more employees.
- Each employee belongs to exactly one department. We store the name of the employee and a unique employee ID for each employee (employeeID).
- Employees can be researchers or managers. For managers we store their annual bonus amount and for researchers their doctorate degree subject.
- People work in projects. Each Project has a unique project ID (projectID).
- Every project has exactly one manager and zero or more researchers.
- A manager can manage one or more projects but a researcher must work on exactly one project.

Figure 1: Almost correct ER diagram

Question 1 continues...
Given this description of the database and its constraints, we have created a mostly correct Entity-Relationship Diagram, shown in Figure [1].

(a) [10 points] Find and correct any mistakes in the given ER diagram. Specifically, number and list them, like, e.g.
   1. delete: arrow, from \( x \to y \)
   2. change to bold line: thin line, from \( z \to w \)
   3. change to bold box: entity \( e \)

(b) [5 points] There may also be some missing element(s). If none, say 'none' - otherwise, add them to the picture, and list them, numbered. E.g.
   1. add: attribute \( a \), to entity \( e \)
   2. add: bold line, arrow, from \( c \to d \).
   3. add: weak entity, \( f \), with attributes . . .

(c) [10 points] List and number all the bold lines and all the arrows that are in the final, corrected version of the diagram. E.g.
   1. bold, line, from Department to Employee
   2. thin, arrow, from \( x \to y \)

Clarifications/Hints:

- List your assumptions, if any. We will accept all reasonable assumptions.
Question 2: SQL Tables from the ER Model ........ [25 points]

On separate page, with ‘[course-id] [hw#] [question#] [andrew-id] [your-name]’

Consider a database for the New York art gallery. It records information about artists, paintings and exhibits. The constraints are exactly as shown in Figure 2. Paintings, Artists and Exhibitions have unique identifiers as shown in the Figure, with binary relationships among them as illustrated. To clarify:

- The line from “Exhibition” to “showcases”, is thick.
- The arrow from “Painting” to “paints”, is also thick.
- No other lines, boxes, or diamonds, are thick.

![ER Diagram for Art Exhibits](image-url)

Figure 2: ER diagram for art exhibits: turn to SQL tables

(a) [20 points] Give the DDL statements, that correspond to the above ER diagram.
- Use proper data types (we’ll accept all reasonable choices).
- Avoid syntax errors (we’ll forgive missing semicolons).
- Specify your decisions with respect to CASCADE deletions. (E.g., ’I decided to reject deletions in Exhibition, when there are still participating artists’)
• Without using CHECK statements, enforce as many as possible of the implied integrity constraints as you can.

(b) [5 points] Which of the implied IC (integrity constraints) of Figure 2 need CHECK statements to be enforced? List them all, or say none. For example, a (possibly, correct) answer could be:
• thin line, from “Painting” to “showcases”
Question 3: Relational Algebra for Job Portal ....... [25 points]

On separate page, with ‘[course-id] [hw#] [question#] [andrew-id] [your-name]’

Consider the relations of a Job Portal database as shown in Table 1. They describe a
Job Portal, recording people, their skills and their endorsements (as in LinkedIn(TM)).

We have the following tables:

- **Member**: For each member we record the userID and name.
- **Skill**: For each Skill we record the skillID and the skillName.
- **Endorsement**: Each row shows which person has what skill, at what proficiency
  level prof (eg., months of experience), and the count n of endorsements from other
  members.

For example, the first row of Table 1(a) means that user 'M103' (='John') masters skill
'S1' (='JAVA') with proficiency level '18', and he has received n=5 endorsements from
other users.

<table>
<thead>
<tr>
<th>userID</th>
<th>name</th>
<th>user ID</th>
<th>skillID</th>
<th>skillName</th>
<th>prof</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>M101</td>
<td>Jack</td>
<td>M103</td>
<td>S1</td>
<td>JAVA</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>M102</td>
<td>Jill</td>
<td>M103</td>
<td>S2</td>
<td>.Net</td>
<td>46</td>
<td>5</td>
</tr>
<tr>
<td>M103</td>
<td>John</td>
<td>M103</td>
<td>S3</td>
<td>Oracle</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>M104</td>
<td>Jane</td>
<td>M105</td>
<td>S4</td>
<td>SAP</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>M105</td>
<td>Job</td>
<td>M106</td>
<td>S5</td>
<td>R</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>M106</td>
<td>Jay</td>
<td>M104</td>
<td>S4</td>
<td>R</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M105</td>
<td>S5</td>
<td>R</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M106</td>
<td>S4</td>
<td>R</td>
<td>45</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 1: Relations of Job Portal database.

Given this database instance, answer the following questions:

(a) [2 points] Which of the following is the meaning of the expression
\(\sigma_{n<5}(Endorsement)\)

1. It lists all the n values, that are less than 5, eliminating duplicates (i.e., \(\{1,3,4\}\)
in our case).
2. It lists all Endorsement tuples ((userID, skillID, prof, and n) with less than
5 endorsements.
3. It lists the n value for each Endorsement tuple, and it rounds it down to 5, if
higher than 5.
4. None of the above. The real answer is ..............

(b) [2 points] We want to list the mature skills, that is, the skillNames, for which
there is at least one veteran (defined as prof > 36 months of experience). Which,
if any, of the following expressions achieves that?
1. $\sigma_{\text{skillName}}(\pi_{\text{prof}>36}(\text{Skill} \bowtie \text{Endorsement}))$
2. $\pi_{\text{skillName}}(\sigma_{\text{prof}>36}(\text{Skill} \bowtie \text{Endorsement}))$
3. $\pi_{\text{skillName}}((\sigma_{\text{prof}>36}(\text{Skill})) \bowtie \text{Endorsement})$
4. $\sigma_{\text{skillName}>36}(\pi_{\text{prof}}(\text{Skill} \bowtie \text{Endorsement}))$
5. None of the above. The real answer is .............

(c) For the following expression:

$$\sigma_{n<4}(\text{Member} \bowtie \text{Endorsement})$$

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [3 points] How many tuples are in the answer?
iv. [3 points] List all the tuples in the answer, as a table.

(d) For the following expression:

$$\pi_{\text{userID}, \text{skillID}}(\text{Endorsement}) \div \pi_{\text{skillID}}(\sigma_{\text{userID}=\'M105\'}(\text{Endorsement}))$$

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [3 points] How many tuples are in the answer?
iv. [3 points] List all the tuples in the answer, as a table.

(e) For the following expression:

$$\pi_{E.\text{userID}, E1.\text{userID}}(\rho_{E}(\text{Endorsement}) \bowtie E.\text{skillID}=E1.\text{skillID} \land E.\text{userID}>E1.\text{userID} \rho_{E1}(\text{Endorsements}))$$

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [3 points] How many tuples are in the answer?
iv. [3 points] List all the tuples in the answer, as a table.

Homework 1 continues...
Question 4: Relational Calculus .......................... [25 points]

On separate page, with ‘[course-id] [hw#] [question#] [andrew-id] [your-name]’

We will again use the Job portal database from the last question (Table 1). We start with questions on relational tuple calculus (RTC).

(a) For the following RTC expression
   \{t | \exists e \in \text{Endorsement} ( e.\text{skillID} = "S1" \land e.\text{userID} = t.\text{userID}) \}
   i. [0 points] Optional: describe in English what the expression does
   ii. [1 point] How many, and which are the columns (= attributes) in the answer?
   iii. [1 point] How many tuples are in the answer?
   iv. [2 points] List all the tuples in the answer, as a table.

(b) For the following RTC expression
   \{t | \exists e_1 \in \text{Endorsement}, \exists e_2 \in \text{Endorsement} \}
   \{e_1.\text{skillID} = e_2.\text{skillID}
   \land e_1.\text{userID} > e_2.\text{userID}
   \land t.\text{user1} = e_1.\text{userID}
   \land t.\text{user2} = e_2.\text{userID}) \}
   i. [0 points] Optional: describe in English what the expression does
   ii. [1 point] How many, and which are the columns (= attributes) in the answer?
   iii. [1 point] How many tuples are in the answer?
   iv. [2 points] List all the tuples in the answer, as a table.

(c) For the following RTC expression
   \{t | \exists e \in \text{Endorsement}, \exists m \in \text{Member} \}
   \{e.\text{userID} = m.\text{userID}
   \land t.\text{name} = m.\text{name}
   \land e.\text{skillID} = "S1"
   \land e.\text{prof} > 20) \}
   i. [0 points] Optional: describe in English what the expression does
   ii. [1 point] How many, and which are the columns (= attributes) in the answer?
   iii. [1 point] How many tuples are in the answer?
   iv. [2 points] Give, as a table, all of the tuples returned by the query.

Question 4 continues...
The next questions are on relational domain calculus (RDC).

(d) For the following RDC expression
\[
\{ \langle u \rangle | \exists s, \exists p, \exists n (\langle u, s, p, n \rangle \in \text{Endorsement} \land n > 4) \}
\]

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [1 point] How many tuples are in the answer?
iv. [2 points] List all the tuples in the answer, as a table.

(e) For the following RDC expression:
\[
\{ \langle u_1, u_2 \rangle : | \exists s, \exists p_1, \exists n_1, \exists p_2, \exists n_2 (\langle u_1, s, p_1, n_1 \rangle \in \text{Endorsement} \\
\quad \land \langle u_2, s, p_2, n_2 \rangle \in \text{Endorsement} \\
\quad \land u_1 > u_2) \}
\]

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [1 point] How many tuples are in the answer?
iv. [2 points] List all the tuples in the answer, as a table.

(f) For the following RDC expression:
\[
\{ \langle u_1 \rangle | \exists p_1, \exists n_1 (\langle u_1, "S4", p_1, n_1 \rangle \in \text{Endorsement} \\
\quad \land \forall u_2 (\exists p_2, \exists n_2 ( (\langle u_2, "S4", p_2, n_2 \rangle \in \text{Endorsement}) \Rightarrow (n_1 \geq n_2))) \}
\]

i. [0 points] Optional: describe in English what the expression does
ii. [1 point] How many, and which are the columns (= attributes) in the answer?
iii. [2 points] How many tuples are in the answer?
iv. [2 points] List all the tuples in the answer, as a table.

End of Homework 1