Carnegie Mellon University
Department of Computer Science
15-415/615 - Database Applications
C. Faloutsos & A. Pavlo, Spring 2015

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)

Revision : 2015/02/01 14:04

Question	Points	Score
Entity-Relationship Diagram	25	
SQL Tables from the ER Model	25	
Relational Algebra for Job Portal	25	
Relational Calculus	25	
Total:	100	

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.

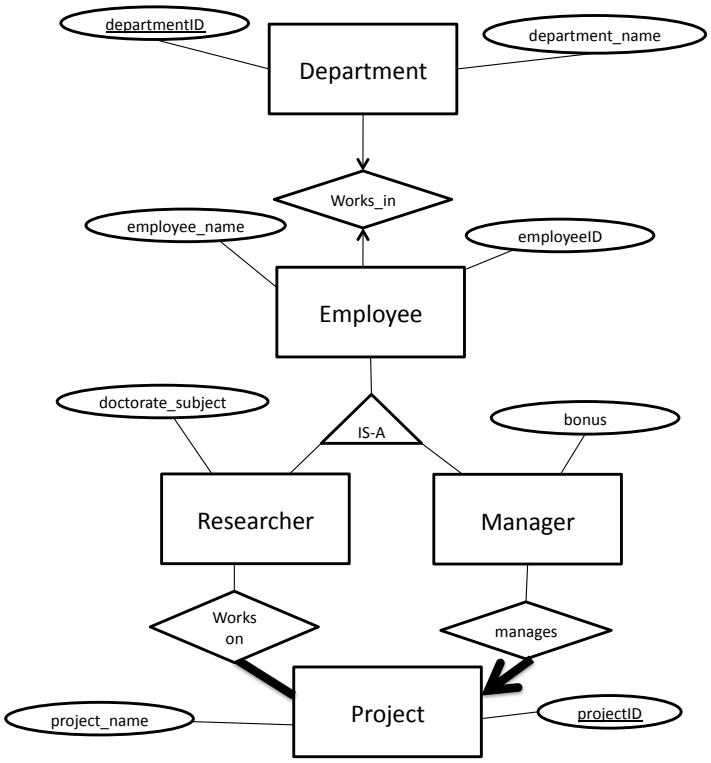

Figure 1: Almost correct ER diagram

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.

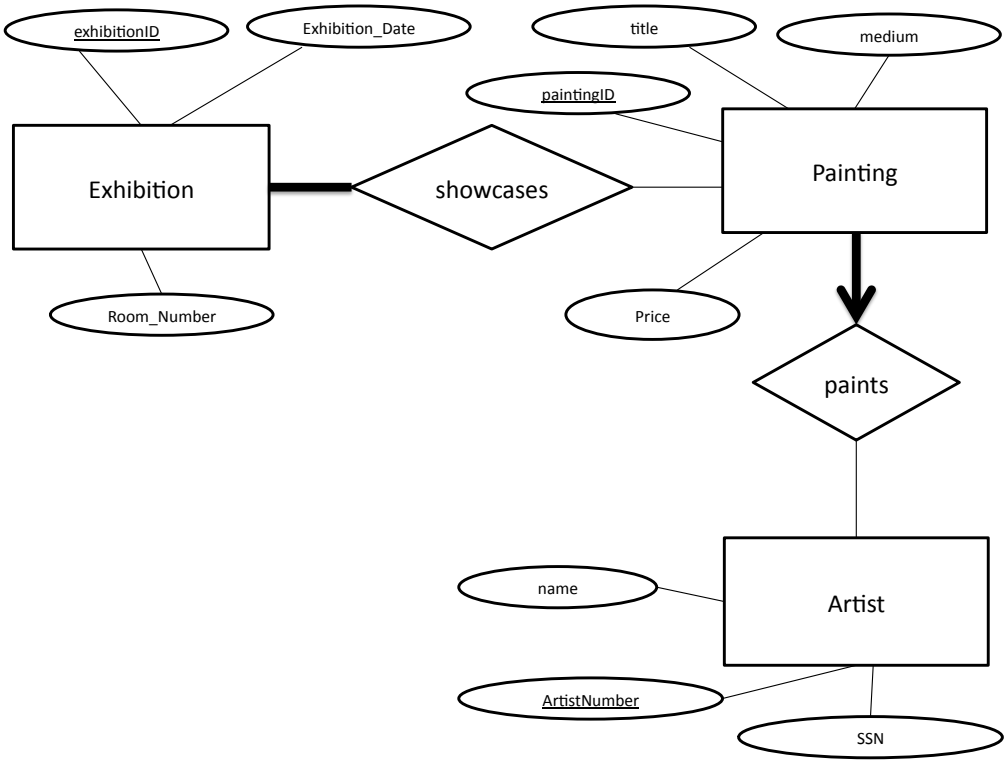

Figure 2: ER diagram for art exhibits: turn to SQL tables

(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.

userID	name
M101	Jack
M102	Jill
M103	John
M104	Jane
M105	Job
M106	Jay

(a) `Member`

skillID	skillName
S1	JAVA
S2	.Net
S3	Oracle
S4	SAP
S5	R

(b) `Skill`

userID	skillID	prof	n
M103	S1	18	5
M103	S2	46	5
M103	S4	17	5
M102	S4	10	4
M101	S4	15	1
M106	S4	21	1
M105	S4	31	3
M104	S5	45	4

(c) `Endorsement`

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 $\sigma_{\texttt{n}<5}(\texttt{Endorsement})$

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 `skillName`s, 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. $\sigma_{\texttt{skillName}}(\pi_{\texttt{prof}>36}(\texttt{Skill} \bowtie \texttt{Endorsement}))$
2. $\pi_{\texttt{skillName}}(\sigma_{\texttt{prof}>36}(\texttt{Skill} \bowtie \texttt{Endorsement}))$
3. $\pi_{\texttt{skillName}}(\quad(\sigma_{\texttt{prof}>36}(\texttt{Skill}))\quad \bowtie \texttt{Endorsement})$
4. $\sigma_{\texttt{skillName}>36}(\pi_{\texttt{prof}}(\texttt{Skill} \bowtie \texttt{Endorsement}))$
5. None of the above. The real answer is .....

(c) For the following expression:

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

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:

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

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:

$$\pi_{\texttt{E.userID},\ \texttt{E1.userID}}\,(\rho_{\texttt{E}}(\texttt{Endorsement}) \bowtie_{\texttt{E.skillID}=\texttt{E1.skillID} \wedge \texttt{E.userID}>\texttt{E1.userID}} \rho_{\texttt{E1}}(\texttt{Endorsements}))$$

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.

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

$\{t \mid \exists e \in \texttt{Endorsement}\ (\ e.\texttt{skillID} = "S1" \quad \wedge \quad e.\texttt{userID} = t.\texttt{userID})\}$

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

$$\begin{aligned}\{t \mid \exists e1 \in \texttt{Endorsement}, \exists e2 \in \texttt{Endorsement}\\ (\ e1.\texttt{skillID} = e2.\texttt{skillID}\\ \wedge\ e1.\texttt{userID} > e2.\texttt{userID}\\ \wedge\ t.\texttt{user1} = e1.\texttt{userID}\\ \wedge\ t.\texttt{user2} = e2.\texttt{userID})\}\end{aligned}$$

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

$$\begin{aligned}\{t \mid \exists e \in\texttt{Endorsement}, \exists m \in \texttt{Member}\\ (e.\texttt{userID} = m.\texttt{userID}\\ \wedge\ t.\texttt{name} = m.\texttt{name}\\ \wedge\ e.\texttt{skillID} = "S1"\\ \wedge\ e.\texttt{prof}{>}20)\}\end{aligned}$$

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.

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

(d) For the following RDC expression
$\{\langle u\rangle \mid \exists s, \exists p, \exists n(\langle u, s, p, n\rangle \in \texttt{Endorsement} \wedge\ n{>}4)\}$

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:

$$\begin{aligned}\{\langle u1, u2\rangle :\ \mid \exists s, \exists p1, \exists n1, \exists p2, \exists n2(\\ &\langle u1, s, p1, n1\rangle \in \texttt{Endorsement}\\ &\wedge \langle u2, s, p2, n2\rangle \in \texttt{Endorsement}\\ &\wedge u1{>}u2)\}\end{aligned}$$

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:

$$\begin{aligned}\{\langle u1\rangle \mid \exists p1, &\exists n1(\\ &\langle u1, \text{"}S4\text{"}, p1, n1\rangle \in \texttt{Endorsement}\\ &\wedge \forall u2(\exists p2, \exists n2\\ &\quad (\ (\langle u2, \text{"}S4\text{"}, p2, n2\rangle \in \texttt{Endorsement})\ \Rightarrow\ (n1 \geqslant n2)))\}\end{aligned}$$

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.