Carnegie Mellon University

15-415 Database Applications

Spring 2012, Faloutsos

Assignment 1 : ER + Formal Q.L.

Due: 2/07, 1:30 pm, in class – hard copy

Reminders

- Weight: 5% of the homework grade.

- The points of this homework add up to 100.

- Like all homeworks, this has to be done individually.

- Lead TA: U Kang.

- Rough time-estimates: 2~4 hours (1 to 2 hours for questions 1~3, and 1 to 2 hours for questions 4~6).

- Please type your answers. Illegible handwriting may get no points, at the discretion of the graders. Only drawings may be hand-drawn, as long as they are neat and legible.

Remember that:

- Whenever you are making an assumption, please state it clearly.

Question 1 : Olympic Database [10 points]

We want to design a database for storing the information of U.S. athletes and their participations in Olympic games.

- An athlete has his/her SSN, the name, and the birth date.
- Each Olympiad has the year, the country, and the city. For example, the 2008 Olympic game was held in Beijing, China.
- An athlete can participate in several sports in an Olympiad. For example, Michael Armstrong participated in the swimming, diving, and water polo events.
- Every athlete in our database participated in at least one Olympiad. Every Olympiad has at least one athlete (and actually, several thousands).

Q1.1 Draw an ER diagram for this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. [5 points]

Q1.2 Turn the ER diagram into tables: give SQL statements to create the tables. Make sure to indicate primary keys, foreign keys, if any. Specify the appropriate integrity constraints to enforce the total/partial coverage, and overlap constraints, if any. [5 points]

Question 2 : Cuisine Database [10 points]

The student administrator of the Mini University wants to design a database for profiling students' preferences for the nearby restaurants and the dishes. For example, the student Jack Smith likes the "rib eye steak" at the "TGI Fridays", not the "rib eye steak" at the "Olive Garden", while the student Nancy Graham likes the Italian spaghetti at the "Union Grill".

- Each student has his/her SSN, the name, and the department.
- Each restaurant has its name, and the mailing address.
- The name of a restaurant may not be unique, but the mailing address is unique.
- Each dish has the name and the price.
- The price of a dish might differ on different restaurants. For example, "rib eye steak" is $15 at "TGI Fridays", but $20 at the "Olive Garden".
- The name of a dish is unique within a restaurant. Two restaurants may have the same dish name.
- A dish is offered by at least one restaurant, and a restaurant offers at least a dish.

Q2.1 Draw an ER diagram for this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. [10 points]

Question 3 : Student Database [20 points]

We want to design a student database for the Mini University.

- A student has his/her SSN, the name, and the age.
- Some students belong to the School of Engineering, and we want to store their E-SAT (Engineering SAT) scores.
- Some students belong to the School of Fine Art, and we want to store their FA-SAT (Fine Art SAT) scores.
- There are other schools (e.g. School of Humanities and Social Sciences) at the Mini University, but we don't explicitly record the school, for anybody.
- The Mini University allows students to have double majors. Thus, a student could be in both the School of Engineering (with an E-SAT) as well as in the School of Fine Arts (with an FA-SAT).

Q3.1 Draw an ER diagram showing the class hierarchies. Make sure to indicate primary keys. [5 points]

Q3.2 Are the constraints in this hierarchy overlapping constraints? Give your answers and the reasons. [5 points]

Q3.3 Are the constraints in this hierarchy covering constraints? Give your answers and the reasons. [5 points]

Q3.4 Turn the ER diagram into tables: give SQL statements to create the tables. Make sure to indicate primary keys and foreign keys, if any. Specify the appropriate integrity constraints to enforce the total/partial coverage, and overlap constraints, if any. [5 points]

Question 4 : Relational Algebra [20 points]

A network intrusion refers to an activity where an attacker sends malicious packets to a victim to do malicious things (e.g. crash the victim machine, or gain root access of the victim machine.) A network security company monitors the network and records the network intrusion events. For the purpose, it has the following two tables in its database.

The table MACHINE contains the attacker or victim information and it has the following three fields:

- mname (the machine name of the attacker or the victim; e.g. jupiter)
- ipaddr (the ip address of the attacker or the victim; e.g. 9.9.9.9)
- os (the operating system of the attacker or the victim; e.g. Linux, Windows).

The primary key of the MACHINE table is ipaddr.

The table EVENT contains the attack information and it has the following three fields:

- a-ipaddr (the ip address of the attacker; e.g. 9.9.9.9)
- v-ipaddr (the ip address of the victim; e.g. 10.10.10.2)
- date (the date the attack was performed; e.g. 2/7/2012).

The primary key of the EVENT TABLE includes all three fields.

Q4.1 (**Victim list**) The company wants to find the victim ip addresses that were attacked on 2/7/2012. Write the relational algebra expression. [5 points]

Q4.2 (**Attacker profile**) The company wants to find the operating systems used for attacking the victim '10.10.10.2'. Write the relational algebra expression. [5 points]

Q4.3 Consider the following EVENT table:

EVENT	a-ipaddr	v-ipaddr	date
	9.9.9.1	10.10.10.2	2/1/2012
	9.9.9.2	10.10.10.1	2/2/2012
	9.9.9.2	10.10.10.3	2/2/2012
	9.9.9.9	10.10.10.4	2/3/2012
	9.9.9.9	10.10.10.3	2/4/2012
	9.9.9.10	10.10.10.3	2/5/2012
	9.9.9.10	10.10.10.4	2/6/2012

Consider the following expression:

$$\pi_{\text{a-ipaddr,v-ipaddr}}(\text{EVENT}) \div \pi_{\text{v-ipaddr}}[\sigma_{\text{a-ipaddr}='9.9.9.9'}(\text{EVENT})]$$

Q4.3.1 Describe in English what the query tries to do. A (possibly incorrect) answer is 'It gives all the attackers on Jan 2/1/2012'. [2 points]

Q4.3.2 How many columns are returned in the output? [2 points]

Q4.3.3 What are the column headers of the output? [2 points]

Q4.3.4 How many rows are returned in the output? [2 points]

Q4.3.5 Give the contents of the returned rows. [2 points]

Question 5: Relational Tuple Calculus [20 points]

We use the same schemas MACHINE and EVENT as in the Question 4.

Q5.1 For the task of Q4.1, write the relational tuple calculus expression. [5 points]

Q5.2 For the task of Q4.2, write the relational tuple calculus expression. [5 points]

Q5.3 Consider the following EVENT table instance (this is different from the EVENT table instance in Q 4.3):

EVENT	a-ipaddr	v-ipaddr	Date
	9.9.9.1	10.10.10.2	2/1/2012
	9.9.9.2	10.10.10.1	2/2/2012
	10.10.10.1	10.10.10.3	2/2/2012
	9.9.9.9	9.9.9.1	2/3/2012
	9.9.9.9	9.9.9.10	2/4/2012

Consider the following relational tuple calculus expression:

$\{t | \exists x \in EVENT \ \wedge \ \exists y \in EVENT$

$(\ t.\text{ipaddr} = x.v - \text{ipaddr}, t.\text{ipaddr} = y.a - \text{ipaddr})\}$

Q5.3.1 Describe in English what the query tries to do. A (possibly incorrect) answer is 'It gives all the attackers who attacked themselves'. [2 points]

Q5.3.2 How many columns are returned in the output? [2 points]

Q5.3.3 What are the column headers of the output? [2 points]

Q5.3.4 How many rows are returned in the output? [2 points]

Q5.3.5 Give the contents of the returned rows. [2 points]

Question 6: Relational Domain Calculus [20 points]

We use the same schemas MACHINE and EVENT as in the Question 4.

Q6.1 For the task of Q4.1, write the relational domain calculus expression. [5 points]

Q6.2 For the task of Q4.2, write the relational domain calculus expression. [5 points]

Q6.3 Consider the following EVENT table instance,

EVENT	a-ipaddr	v-ipaddr	date
	9.9.9.1	10.10.10.2	2/1/2012
	9.9.9.1	9.9.9.1	2/2/2012
	9.9.9.2	10.10.10.3	2/2/2012
	9.9.9.2	10.10.10.4	2/2/2012
	9.9.9.2	9.9.9.10	2/3/2012
	9.9.9.9	9.9.9.1	2/3/2012
	9.9.9.9	9.9.9.10	2/4/2012

and the following MACHINE table instance.

MACHINE	mname	ipaddr	os
	Jupiter	9.9.9.1	Windows
	Sun	9.9.9.2	Linux
	Mars	9.9.9.9	Linux
	Earth	9.9.9.10	Windows
	Moon	10.10.10.2	Linux
	Pluto	10.10.10.3	Linux
	Neptune	10.10.10.4	Linux

Consider the following relational domain calculus expression:

$\{< am > | \exists aip, \exists aos, \exists vip, \exists d, \exists vm(< am, aip, aos >\in MACHINE \wedge$

$< aip, vip, d >\in EVENT \wedge < vm, vip,' Windows' >\in MACHINE\}$

Q6.3.1 Describe in English what the query tries to do. A (possibly incorrect) answer is 'It gives all the ip addresses of the victims'. [2 points]

Q6.3.2 How many columns are returned in the output? [2 points]

Q6.3.3 What are the column headers of the output? [2 points]

Q6.3.4 How many rows are returned in the output? [2 points]

Q6.3.5 Give the contents of the returned rows. [2 points]