CS348: INTRODUCTION TO DATABASE MANAGEMENT

(Winter, 2011)

FINAL EXAMINATION

INSTRUCTOR: Grant Weddell TIME: 150 minutes

WRITE YOUR NAME AND ID HERE: _____

NOTE 1: This is a closed book examination. For example, class text, copies of overhead slides and printed notes may not be used. There are 21 pages. Answer all questions in the space provided.

NOTE 2: Some of the questions in this examination are open ended; however, they can be answered to an appropriate level by short organized answers. It is recommended that you spend part of your time organizing your answer, rather than writing down ideas in the order they occur to you. The conciseness and organization of your answers will be taken into consideration in the grading.

NOTE 3: There are 100 marks in total. As a guide to managing your time, the marks awarded for each question are indicated in parenthesis at the start of each question.

NOTE 4: You are also being tested on your ability to understand the questions. In the case of a perceived ambiguity, state a clear assumption and proceed to answer the question.

NOTE 5: Write all your answers on the space provided on the exam.

NOTE 6: Cheating is an academic offense. Your signature below indicates that you understand and agree to the University's policies regarding cheating on exams.

PLEASE SIGN HERE: _____

I. (20 marks; continued on next two pages) Questions on relational database design, on functional dependencies and on normalization theory. Answer each part.

(a) In no more than two sentences, explain each of the following.

1. deletion anomaly

2. completeness of Armstrong's axioms

3. first normal form

(b) Describe an algorithm that efficiently solves each of the following two problems. (Assume that a relation scheme consists of a set of attributes R and a set of functional dependencies F over those attributes.)

1. Given R and F together with a binary decomposition, determine if the latter is a lossless-join decomposition.

2. Given R and F and an alternative set of functional dependencies G, determine if F^+ is the same as G^+.

(c) Indicate whether each of the following statements is true or false. In each case justify your answer with no more than three sentences or examples.

1. Given a set of attributes R and a set of functional dependencies F over those attributes, the number of candidate keys on R is $O(|F|^k)$, where k is some constant and where $|F|$ is the number of occurrences of attributes in F.

2. Primary keys behave differently from candidate keys.

3. Limiting functional dependencies by allowing at most two attributes on their left-hand-side and a single attribute on their right-hand-side does not reduce their expressiveness.

II. (24 marks; continued on next three pages) Consider the following SQL data definition for maintaining information about employees at a hypothetical company.

```
CREATE TABLE emp
   (  num        INTEGER NOT NULL,
      name       VARCHAR(20) NOT NULL,
      dept       VARCHAR(20) NOT NULL,
      salary     INTEGER NOT NULL,
      boss       INTEGER NOT NULL,
   PRIMARY KEY (num),
   FOREIGN KEY (boss) REFERENCES emp (num) );
```

You can assume that there is one president that has herself/himself as the boss, that all other employees have a boss that is someone else and that there are no cycles in the boss hierarchy for anyone other than the president. (A cycle would exist if, for example, Fred was the boss of Mary and Mary was in turn the boss of Fred.) Translate each of the following queries on this schema to SQL. *In each case, also indicate if the query can be expressed in the relational algebra.*

(a) The number and name of each employee, excluding the president, together with the number and name of the employee's boss. The result should be sorted by the name of the boss and then by the name of the employee.

(b) The names of the departments with the highest average salary of their employees.

(c) The numbers and names of employees who have the president as their boss, and that have a salary among the lowest of those employees who do not have the president as their boss.

Now write queries in the relational calculus and relational algebra on the same schema to answer each of the following.

(d) The numbers, names and departments of employees that are not the boss of any other employees.

(e) The numbers, names and departments of employees that have the highest saleries in their respective departments.

III. (14 marks; continued on the next page) Questions on advanced SQL and on programming applications. Answer each of the five parts using no more than three sentences in each case.

(a) Consider developing an application using a SQL DBMS.

1. Consider the assertion: *it is not possible to avoid the overhead of query optimization if one uses the dynamic embedded SQL standard.* Is this true or false? Justify your answer.

2. When writing application programs using C and the embedded SQL standard, explain how code written in C communicates with code written in SQL.

(b) Explain the purpose of each of the following.

1. an embedded SQL preprocessor

2. a cursor

3. the ROLLBACK WORK command

(c) Explain why it might be worthwhile to code SQL stored procedures when developing an application. For example, is there any incentive relating to performance?

IV. (12 marks; includes next page) Your company is starting the development of a patient billing system to be marketed to private medical practices in Ontario. The system is to be called PATMAN (short for PATient billing MANager), and is to run as a remote client-server system using several commodity grade personal computers and a local area network. An initial analysis phase of the project has resulted in the following description of the relevant data for PATMAN.

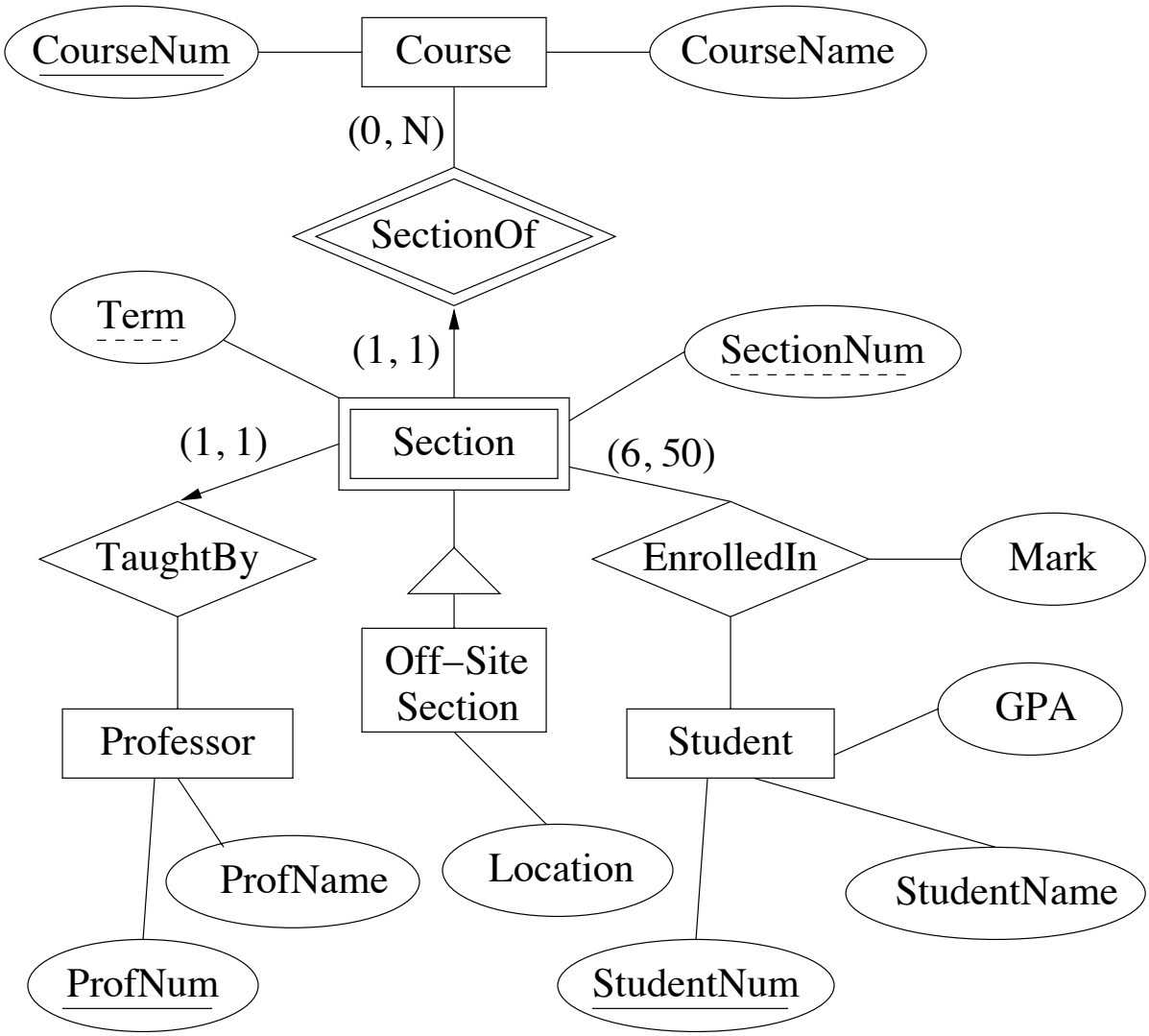
- A practice has a number of patients and doctors.
- Doctors are identified by name.
- Each patient has a number used to identify the patient called the OHIP number, a name and an age.
- Each patient is either a male or female and has a next of kin identified by name.
- Each medical procedure paid by the government of Ontario is identified by a procedure code and has a description and a charging category.
- Each charging category has a dollar value.
- Each patient has a number of billing records, with each billing record recording the medical procedure, the date on which the procedure was performed, the examining doctor and some additional comments on the part of the examining doctor.
- Billing records are either outstanding or paid in full.

Draw an ER diagram on the *next page* that represents the PATMAN data. In addition, include comments below to explain any unusual representation decisions.

V.(15 marks) Further questions on ER diagrams. Do each part.

(a) Draw an ER diagram with an entity set E that is over constrained, that is, for which it is not possible for E to have any entities.

(b) Your company has been awarded a contract to implement an information system for a foreign company. The contract includes a specification of the relevant information in the form of the following ER diagram.



Convert the ER diagram to a relational schema on the *next page*. Your schema can be described by a set of "CREATE TABLE" commands or by using the UML-like box notation used in class. Using a single sentence for each case, explain any constraints reflected in the ER diagram that are *not* captured by your relational schema.

VI. (15 marks; continue on next three pages) Questions on transactions, concurrency control and recovery. Answer each part. Assume there is a `part` table given by the following when relevant.

pnum	pname	color	price	dept
1	Disk	Red	200.00	Machining
2	Axle	Blue	120.00	Machining
4	Cog	Blue	30.00	Milling
5	Block	Green	500.00	Milling

Consider the following sequence of operations on the database, assumming all four clients are operating at isolation level 0, that access and update tuples in the `part` table.

T_1: begin
T_1: read the price of part 4
T_2: begin
T_2: change the color of part 4 to 'yellow'
T_3: begin
T_3: read the price of part 4
T_1: read the color of part 1
T_4: begin
T_4: change the name of part 1 to 'Block'
T_2: change the price of part 5 to 150.00
T_3: read the price of part 2
T_2: commit
T_4: change the department of part 5 to 'Machining'
T_1: commit
T_3: commit
T_4: commit

(a) Assume a database engine accomplishes concurrency control by a strict two-phase locking protocol, with locks being granted on individual tuples. For example, the fourth request requires that transaction T_2 successfully acquires an exclusive lock on the first tuple before executing. List the requests in a possible order of actual execution by the engine when all clients are operating at isolation level 3.

(b) Suppose that a system failure occurs after the requests shown above have been processed (according to the strict 2PL protocol). Indicate the fate (commit or abort) of each of the transactions whose requests appear in the sequence above. Also, show the contents of the `part` table after the system has recovered from the failure, and before any new transactions are allowed to run.

(c) Briefly explain why the SQL standard enables applications to set isolation levels. Also, outline an example of a circumstance in which it makes sense for a client on behalf of some application to set the isolation level to 0.

(d) Indicate whether each of the following statements is true or false. In each case, justify your answer with no more than three sentences.

1. Given an arbitrary schedule for interleaving the execution of a given set of transactions, there is an efficient algorithm to determine if the schedule is serializable.

2. Given that a system failure had occurred, there is an efficient algorithm to recover all transactions that were aborted by using a log.