Comprehensive Exam
Databases
Fall 2009

1) A crucial step during the database design process is to *decompose* a table into a collection of tables that in some sense are "better" than the original table. For example, the following table:

loan-info = (loan#, borrower-name, branch-name, amount)

Could be decomposed into the following tables:

borrower = (borrower-name, loan#)

loan = (loan#, branch-name, amount)

(a) (9 points) List three properties that such a decomposition, and the resulting tables, are required to satisfy.

(b) (8 points) Give an example of an SQL query where such a decomposition would improve query performance. Be sure to explain how the decomposition improves the performance of the query.

(c) (8 points) Give an example of an SQL query where such a decomposition would degrade (or hurt) query performance. Be sure to explain how the decomposition would degrade the performance of the query.

2) (30 points) Consider the following relational schemes. Note that attributes forming the primary key for each relation have been underlined.

student(<u>student-name</u>, street, city)	-- Basic student information.
offering(<u>department</u>, <u>number</u>, <u>section</u>, time, population)	-- Courses currently offered; for CSE5260 department is “CSE” and number is 5260. Section would be D-1, D-2, E-1, E-2, etc.
titles(<u>department</u>, <u>number</u>, title)	-- Course titles; “CSE5260” is “Database Systems”
enrollment(<u>student-name</u>, <u>department</u>, <u>number</u>, <u>section</u>)	-- Indicates which students are enrolled in which courses and sections.

Give an <u>SQL query</u> for part (a).

(a) The department, number and title for every course that has at least 5 sections. Also include the total number of students in all sections of the course combined.

department	number	title	students
CSE	5260	Database Systems	76
MTH	5100	Discrete Math	87
PSY	4260	Abnormal Psychology	65
CHM	2035	Introduction to Chemistry	98
:			

Give a <u>relational algebraic expression</u> for part (b).

(b) For each psychology (PSY) course, list the section that has the largest population. Include the course number, title, section, and population in the result.

number	title	section	population
1001	Introduction to Psychology	D-5	57
3100	Abnormal Psychology	D-3	25
4260	Child Development	E-2	62
5264	Statistics in Psychology	E-1	43
:			

Give a <u>tuple calculus expression</u> for part (c).

(c) A list of the names of students who are currently enrolled in every class that "Jones" is currently enrolled in.

3) (10 points) Recall that when an ER diagram is converted to tables, there are frequently several different ways to do the conversion.

(a) List and describe two different ways to convert a many-to-one relationship to tables.

____

____

(b) For each option in part (a), explain the advantages and disadvantages of the option. Also describe the circumstances under which the option is preferable. Specifically, what types of queries or data would make the option desirable?

4) (20 points)

(a) Define 1NF.

(b) Define 2NF.

(c) Define 3NF.

(d) Define BCNF.

5) Consider the following set F of functional dependencies for the relational scheme R=(A,B,C,D,E,F,G).

A=>B
A=>D
C=>EA
F=>G

(a) (3 points) Give the attribute closure of C.

(b) (3 points) Give a candidate key for R. Prove that your answer is correct, i.e., that your candidate key is a candidate key.

(c) (3 points) Give one non-trivial functional dependency that is in F^+, but not in F. Use Armstrong's axioms to prove that the functional dependency is in F^+.

(d) (3 points) Now consider the decomposition of R into R1=(A,B,F,G) and R2=(A,C,D,E). Does this decomposition have a lossless join (yes or no)? If so, then explain why, and if not, then explain why not.

(e) (3 points) Now suppose that the set of functional dependencies, for the same relational scheme, is:

A=>B
A=>D
C=>EA
F=>G
C=>B

Does the decomposition of R into R1=(A,B,C, F,G) and R2=(A,C,D,E) have a lossless join (yes or no)? If so, then explain why, and if not, then explain why not.