

Student name: _____ Student ID: _____

General Guidelines: Exam duration: 120 min; closed book and no collaboration. This exam has 7 pages, all must be returned and all questions need to be answered. Use the spaces after the questions in this set of pages - write your name and id on all of them (additional pages may be used but should not be necessary). Marked exams will be (tentatively) available on Apr 30 between 1-4pm (at GSB 733). Deadline for appeals is May 02 at 5 pm. Good luck !

Question 1 (15 marks) Suppose that we have a many-to-one binary relationship R between entity sets A and B (i.e., one entity of B can "match" with many entities of A), and also have a many-to-many binary relationship S between entity sets B and C. A has attributes a1 and a2, with a1 being the primary key and a2 being a candidate key; B has attributes b1 and b2, both forming B's primary key; and C has c1 as its single key attribute. R has no descriptive attributes, and S has an attribute s1.

- a) Draw the respective ER-diagram for the scenario above so as to capture as many of the constraints as possible. If you cannot capture some constraint, explain why.
- b) Write SQL statements that create tables corresponding to the ER-model obtained in (a) so as to capture as many of the constraints as possible. If you cannot capture some constraint using only CREATE TABLE statements, explain why.

Student name: _____ Student ID: _____

- d) If a manager manages more than one department, he or she controls the sum of all the budgets for those departments. Find the managerids of managers who control more than \$5,000,000.
- e) For each department, and considering only the employees over 18 years old in those departments, find the average salaries.

Question 3 (6 marks) Consider the following query: Find the names of sailors with a higher rating than all sailors with age < 21. The following two SQL queries attempt to obtain the answer to this question. Does either one compute the correct result? If one does not, explain why.

- a)

```
SELECT S.sname
FROM Sailors S
WHERE NOT EXISTS ( SELECT * Sailors S2
                   WHERE S2.age < 21 And
                   S.rating <= S2.rating )
```
- b)

```
SELECT S.name
FROM Sailors S
WHERE S.rating > ANY ( SELECT S2.rating
                      FROM Sailors S2
                      WHERE S2.age < 21 )
```

Student name: _____ Student ID: _____

Question 4 (15 marks)

Consider the following relation: Emp(*eid*: integer, *ename*: string, *age*: integer) and two B+-trees, B1 and B2, both of order 2 and built on Emp.*eid* and on Emp.*age*, respectively. Assume that the following tuples are inserted, in this order, into Emp:

<i>eid</i>	<i>ename</i>	<i>age</i>
23	I.M.Socool	40
12	Notso Cool	25
67	Donald Mouse	70
69	Mickey Duck	25
68	Dumbo Dog	40
07	Kitefly Andfall	87

- a) Assuming that the leaf entries in both B+-trees contain a key value and a set of pointers to all tuples with that key value, and ignoring the space overhead this imposes (i.e., do assume that a set of pointers require as much space as a single pointer), show both B+-trees after the insertion of these tuples into table Emp. Note that the actual tuples are pointed by the leaf nodes

- b) If the tuples above (and only those) were inserted in a different order would it be possible to have different resulting B+-trees? Justify your answer

Student name: _____ Student ID: _____

- c) Assuming that reading each tuple or a tree node requires one I/O and using B2 (the B+-tree built on attribute Emp.age) obtained in item (a) above, state - justifying your answer: (i) how many I/Os would be needed to find the number of distinct age values in Emp; and (ii) how many I/Os would be needed to find the average age stored in Emp.

Question 5 (12 marks) Consider the following instance of relation R:

<i>A</i>	<i>B</i>	<i>C</i>
1	2	3
4	2	3
5	3	3

and the following functional dependencies:

FD1: $A \rightarrow B$

FD2: $B C \rightarrow A$

FD3: $B \rightarrow C$

Can you say which of the above functional dependencies hold and which do not hold over R? Justify your answer.

Student name: _____ Student ID: _____

Question 6 (10 marks) Assume an extendible hash scheme where the directory is held in main memory at all times, but the data tuples reside in disk. What is the minimum and maximum number of I/Os required to fetch a tuple with a given primary key value? Justify your answer.

Question 7 (12 marks) Assume a hypothetical disk with the following specifications:

Rotational delay: D secs
Seek time: S secs
Transfer time: T cluster/sec
Number of clusters per track: B
There are only two double-sided platters
Each platter face has a R/W head

Further, assume all files have the same size, occupying C clusters each, where $B/C = 2$ (a constant value).

- a) Assuming no parallel I/O is possible, show a formula only for the maximum time required to read all tuples, in no particular order, stored in a single file in this disk.
- b) Assuming no parallel I/O is possible, show a formula only for the minimum time required to read all tuples, in no particular order, stored in three files in this disk.

Student name: _____ Student ID: _____

- c) Redo item (b) but now assuming that parallel I/O is feasible.

Question 8 (10 marks) Chose an option (true or false) for the following statements:

- a) SQL can process queries that cannot be written in Relational Algebra.

TRUE
 FALSE

- b) A view is always materialized as a temporary table.

TRUE
 FALSE

- c) The ISAM indexing structure is initially (before any updates) height-balanced.

TRUE
 FALSE

- d) Only one attribute per relation, the primary key, can have unique values.

TRUE
 FALSE

- e) Enforcing First Normal Form is enough to avoid Insertion/Deletion/Update anomalies in the Relational Model.

TRUE
 FALSE