# CS 303 - Databases & Information Systems
## End Sem Exam

Weightage: 25%

*Instructions:*

1. *There are 11 questions in this question paper. All questions are compulsory.*
2. *For each question, please justify how you arrived at your answer. Incomplete answers will lead to loss of marks.*
3. *Assumptions made in each question should be clearly stated.* ***The assumption should be rational and reasonably accurate. Wild assumptions should not be considered for arriving at an answer.***
4. *DO NOT COPY.*

1. Suppose there is a relation r (A, B, C), with a B + -tree index with search key (A, B). [3 Marks] 9min
   a. What is the worst-case cost of finding records satisfying $10 < A < 50$ using this index, in terms of the number of records retrieved **n1** and the height **h** of the tree?
   b. What is the worst-case cost of finding records satisfying $10 < A < 50 \wedge 5 < B < 10$ using this index, in terms of the number of records **n2** that satisfy this selection, as well as **n1** and **h** defined above?
   c. Under what conditions on n1 and n2 would the index be an efficient way of finding records satisfying $10 < A < 50 \wedge 5 < B < 10$?

2. Assume that you have just built a dense B+ tree index on a heap file containing 20,000 records. The key field for this B+ tree index is a 40-byte string, and it is a candidate key. Pointers (i.e., record ids and page ids) are (at most) 10-byte values. The size of one disk page is 1,000 bytes. The index was built in a bottom-up fashion using the bulk-loading algorithm, and the nodes at each level were filled up as much as possible. Answer the following with proper explanations:
   [5 Marks] 15
   a. How many levels does the resulting tree have?
   b. For each level of the tree, how many nodes are at that level?
   c. How many levels would the resulting tree have if key compression is used and it reduces the average size of each key in an entry to 10 bytes?
   d. How many levels would the resulting tree have without key compression, but with all pages 70 percent full?

3. Let relations r1 (A, B, C) and r2 (C, D, E) have the following properties: r1 has 20,000 tuples, r2 has 45,000 tuples, 25 tuples of r1 fit on one block, and 30 tuples of r2 fit on one block. Estimate the number of block transfers and seeks required, using each of the following join strategies for $r1 \bowtie r2$:
   [4 Marks] 12
   a. Nested-loop join.
   b. Block nested-loop join.

4. Consider the relations r1 (A, B, C), r2 (C, D, E), and r3 (E, F ), with primary keys A, C, and E, respectively. Assume that r1 has 1000 tuples, r2 has 1500 tuples, and r3 has 750 tuples. Estimate the size of $r1 \bowtie r2 \bowtie r3$ , and give an efficient strategy for computing the join. [2 Marks] 6

45min

5. Show that the following equivalences hold. Explain how you can apply them to improve the efficiency of certain queries: [4 Marks]

a. $E1 \bowtie_\Theta (E2 - E3) = (E1 \bowtie_\Theta E2) - (E1 \bowtie_\Theta E3)$.

b. $\sigma_\Theta (E1 \bowtie E2) = \sigma_\Theta (E1) \bowtie E2$, where $\theta$ uses only attributes from E1.

6. Consider the scenario given below and answer the questions that follow with respect to ER Data Model: [10 Marks]

Scenario:

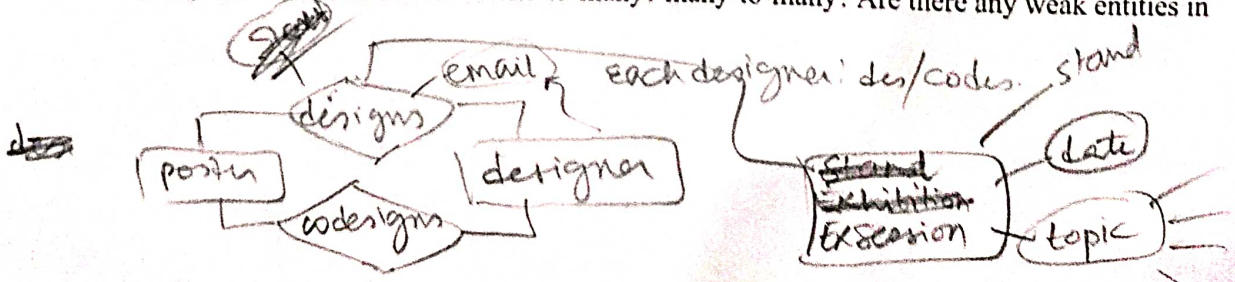
You are one organizer of a poster exhibition, and you must design a database to keep track of the administration of the exhibition. You are required to design an Entity Relationship Model for a database to be used by the organizers of a poster exhibition in order to keep track of three phases in the exhibition: submission, selection and presentation.

*Submission Phase:* Graphic designers create posters for the exhibition to illustrate one of the chosen global problems. Relevant information on designers includes their name and their affiliation, i.e. the organization they work for. A poster has a title and is assigned an identification number, and it may be created by several graphic designers; although each individual designer may only be involved with one poster. Where a group of graphic designers create a poster, we distinguish between the main designer and the co-designers. In the case of a single graphic designer, that person is considered to be the main designer of the poster. The main designer is always the point of contact, so should provide an email address.

*Selection Phase:* All posters created for this exhibition are judged by members of a jury. A judge is a graphic design expert with experience in communication for raising public awareness and for public benefit. Judge information that is of relevance to the organizing committee includes the judge's name, their affiliation and email. Each poster is judged by three different judges. When judging a poster, a judge gives a decision: accept or reject. A poster is selected for the exhibition only if all three judges give an "accept" decision. Note that judges are not allowed to compete in the exhibition themselves.

*Presentation Phase:* All selected posters are then presented in the exhibition by their main graphic designers. The poster presentation is allocated a stand and an exhibition session. Each exhibition session takes place at a specific date, and 4 session topics have been announced: human rights, environmental pollution, poverty, and war.

(a). What are the candidate entity sets you would consider in order to build your ER model?
(b). What are the attributes you can assign to each of these given the available information? What are the attribute domains?
(c). Define relationships between the entities you defined in the first question. What kinds of relationships are these? one-to-one? one-to-many? many-to-many? Are there any weak entities in



your model? Are there any ISA hierarchies? Are there any key constraints? Do any of these relationships involve total participation?

(d). Choose primary keys for each entity in your model. Which ones did you choose, and why?

(e). Draw the diagram representing your ER model.

7. Suppose that we have a ternary relationship R between entity sets A, B, and C such that A has a key constraint and total participation and B has a key constraint; these are the only constraints. A has attributes a1 and a2, with a1 being the key; B and C are similar. R has no descriptive attributes. Write SQL statements that create tables corresponding to this information so as to capture as many of the constraints as possible. If you cannot capture some constraint, explain why. [4 Marks]

8. Given two relations R1 and R2, where R1 contains N1 tuples, R2 contains N2 tuples, and N2 > N1 > 0, give the minimum and maximum possible sizes (in tuples) for the result relation produced by each of the following relational algebra expressions. In each case, state any assumptions about the schemas for R1 and R2 that are needed to make the expression meaningful: [6 Marks]

(1) R1 ∪ R2, (2) R1 ∩ R2, (3) R1 − R2, (4) R1 × R2, (5) $\sigma_{a=5}(R1)$, (6) $\pi_a(R1)$

9. Consider the following relational schema and briefly answer the questions that follow: [4 Marks]

```
Emp(eid: integer, ename: string, age: integer, salary: real)
Works(eid: integer, did: integer, pct time: integer)
Dept(did: integer, budget: real, managerid: integer)
```

a. Define a table constraint on Emp that will ensure that every employee makes at least $10,000.

b. Define a table constraint on Dept that will ensure that all managers have age > 30.

c. Write SQL statements to delete all information about employees whose salaries exceed that of the manager of one or more departments that they work in. Be sure to ensure that all the relevant integrity constraints are satisfied after your updates.

10. a. Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of functional dependencies F = { {A, B}→{C}, {A}→{D, E}, {B}→{F}, {F}→{G, H}, {D}→{I, J} }. What is the key for R? Decompose R into 2NF and then 3NF relations. [3 Marks]

b. Repeat the above activities for the following different set of functional dependencies G = {{A, B}→{C}, {B, D}→{E, F}, {A, D}→{G, H}, {A}→{I}, {H}→{J} }. [3 Marks]

11. Does the relational model, as seen by an SQL query writer, provide physical and logical data independence? Explain. [2 Marks]

your model? Are there any ISA hierarchies? Are there any key constraints? Do any of these relationships involve total participation?

(d). Choose primary keys for each entity in your model. Which ones did you choose, and why?

(e). Draw the diagram representing your ER model.

7. Suppose that we have a ternary relationship R between entity sets A, B, and C such that A has a key constraint and total participation and B has a key constraint; these are the only constraints. A has attributes a1 and a2, with a1 being the key; B and C are similar. R has no descriptive attributes. Write SQL statements that create tables corresponding to this information so as to capture as many of the constraints as possible. If you cannot capture some constraint, explain why.

[4 Marks]

8. Given two relations R1 and R2, where R1 contains N1 tuples, R2 contains N2 tuples, and N2 > N1 > 0, give the minimum and maximum possible sizes (in tuples) for the result relation produced by each of the following relational algebra expressions. In each case, state any assumptions about the schemas for R1 and R2 that are needed to make the expression meaningful:

[6 Marks]

(1) R1 ∪ R2, (2) R1 ∩ R2, (3) R1 − R2, (4) R1 × R2, (5) $\sigma_{a=5}(R1)$, (6) $\pi_a(R1)$

9. Consider the following relational schema and briefly answer the questions that follow: [4 Marks]

```
Emp(eid: integer, ename: string, age: integer, salary: real)
Works(eid: integer, did: integer, pct time: integer)
Dept(did: integer, budget: real, managerid: integer)
```

a. Define a table constraint on Emp that will ensure that every employee makes at least $10,000.

b. Define a table constraint on Dept that will ensure that all managers have age > 30.

c. Write SQL statements to delete all information about employees whose salaries exceed that of the manager of one or more departments that they work in. Be sure to ensure that all the relevant integrity constraints are satisfied after your updates.

10. a. Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of functional dependencies F = { {A, B}→{C}, {A}→{D, E}, {B}→{F}, {F}→{G, H}, {D}→{I, J} }. What is the key for R? Decompose R into 2NF and then 3NF relations. [3 Marks]

b. Repeat the above activities for the following different set of functional dependencies G = {{A, B}→{C}, {B, D}→{E, F}, {A, D}→{G, H}, {A}→{I}, {H}→{J} }. [3 Marks]

11. Does the relational model, as seen by an SQL query writer, provide physical and logical data independence? Explain. [2 Marks]