DEGREE EXAMINATIONS — 1999

End of Trimester 1

COMP 302

Database Systems

Time Allowed: 3 Hours

Instructions: Answer all questions.
Make sure your answers are clear and to the point.
Calculators and foreign language dictionaries are allowed.
No reference material is allowed.
There are 180 possible marks on the exam.

Question 1. Short answer [12 marks]

a. Briefly describe the structure of a database management system (DBMS). [3 marks]

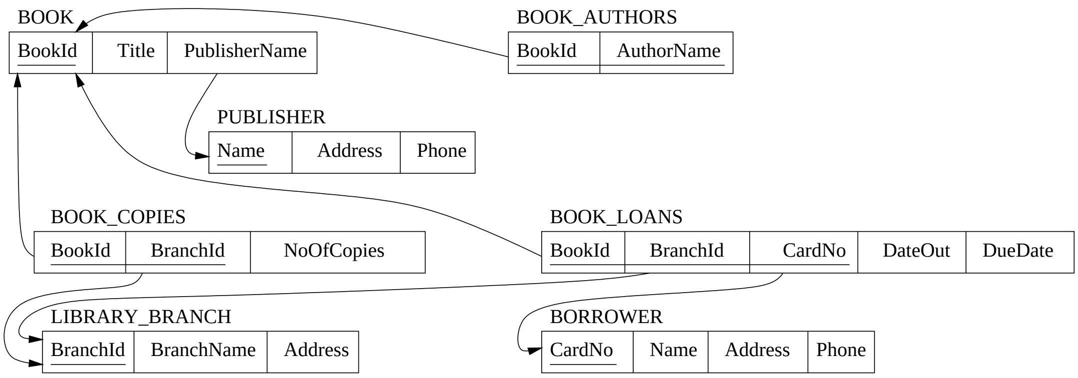
b. If your operating system is upgraded to support some new functions on the operating system files (e.g., the ability to force some sequence of bytes to a disk), which layer(s) of the DBMS would you have to rewrite in order to take advantage of these new functions? [3 marks]

c. What is a *transaction*? In what ways is it different from an ordinary program (in a language such as C)? [3 marks]

d. In two sentences describe what the *phantom problem* is. Can it occur in a database where the set of database objects is fixed and only the values of objects can be changed? [3 marks]

Question 2. Relational databases [45 marks]

a. Part of the reverse engineering process involves creating a conceptual schema from an existing implemented database. Consider the relational schema below for a **Library** database which is used to keep track of books, borrowers and book loans. Referential integrity constraints are shown as directed arcs. Map the relational schema to an ER schema. State any assumptions you make. [15 marks]



b. Write SQL queries for the following queries on the **Library** database.

(i) For each book that is loaned out from the **Victoria** branch and whose due date is today, retrieve the *book title*, the *borrower's name*, and the *borrower's address*. [7 marks]

(ii) For each library branch retrieve the *branch name* and the *total number of books* loaned out from that branch. [7 marks]

(iii) Retrieve the *names*, *addresses*, and *number of books* checked out for all borrowers who have more than **5** books checked out. [8 marks]

(iv) Retrieve the *name* of all borrowers who do not have a book checked out. [8 marks]

Question 3. Normalisation [40 marks]

a. Suppose you have the following three tuples in a legal instance of a relation schema S with three attributes ABC (listed in order): (1,2,3), (4,2,3), and (5,3,3).

(i) Which of the following dependencies can you infer does **not** hold over schema S?

1. $A \rightarrow B$
2. $\{B, C\} \rightarrow A$
3. $B \rightarrow C$ [4 marks]

(ii) Can you identify any dependencies that hold over S? Explain your answer. [3 marks]

b. Give a set of functional dependencies for the relation schema $R(A, B, C, D)$ with primary key AB under which R is in 1NF but not in 2NF. [6 marks]

c. Give a set of functional dependencies for the relation schema $R(A, B, C, D)$ with primary key AB under which R is in 2NF but not in 3NF. Is this relation with the functional dependencies that you specified in BCNF? Explain your answer. [6 marks]

d. Suppose you are given a relation R with four attributes, $ABCD$. For each of the following sets of functional dependencies, assuming those are the only dependencies that hold for R, do the following:

1. Identify the candidate key(s) for R.
2. Identify the highest normal form that R satisfies (1NF, 2NF, 3NF or BCNF).
3. If R is not in BCNF, decompose it into a set of BCNF relations that preserve the dependencies. If a decomposition isn't possible, state why it isn't possible.

(i) $C \rightarrow D, C \rightarrow A, B \rightarrow C$ [7 marks]

(ii) $B \rightarrow C, D \rightarrow A$ [7 marks]

(iii) $A \rightarrow B$, $\{B, C\} \rightarrow D, A \rightarrow C$ [7 marks]

Question 4. Object oriented databases [38 marks]

a. Give 2 example applications where it would be more appropriate to use an object oriented database than a relational database. Describe why it is more appropriate. [5 marks]

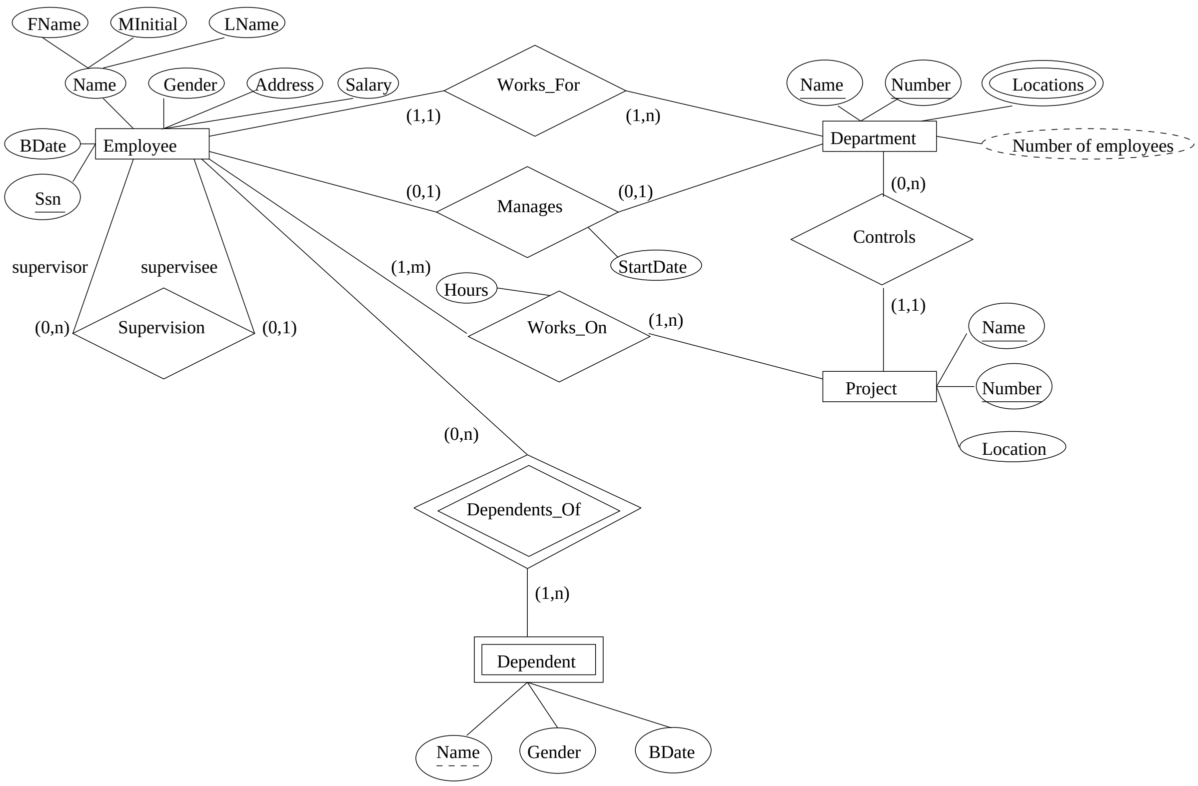
b. A restriction of relational datbases is that all attribute values must be atomic. This is not the case for object oriented database systems. What are two kinds of data types supported in object oriented database systems that aren't supported in relational systems? Give an example of each. Describe how each example would be represented if only a relational database management system was available. [7 marks]

c. Object oriented database management systems (OODBMSs) differ in how an object becomes persistent. In fact, a single OODBMS may offer more than one way to make objects persistent:

1. by type: an object may be made persistent when it is created based on its type,
2. by explicit call: the user may specify explicitly at any time that an object is persistent,
3. by reference: some systems automatically determine persistence of objects by reachability from certain globally known persistent root objects.

Which of the above does **Objectivity** use? Give an example which demonstrates how objects are made persistent in **Objectivity**. [6 marks]

d. Consider the **company** schema represented in the following ER diagram.



Map the **company** schema into **Objectivity** classes. [15 marks]

e. Give pseudocode which outlines how you would find *all employees whose salary is greater than $100,000* in **Objectivity**. [5 marks]

Question 5. Query processing and query optimisation [45 marks]

a. List the 5 steps that are followed when a query is processed by a relational database management system. [3 marks]

b. Consider a relation schema **flight{flight_number, start, destination, departure_time, arrival_time}** where **start** is the town where the flight starts, **destination** is the town where the flight finishes, **departure_time** is the time that the flight leaves start and **arrival_time** is the time that the flight arrives at destination. The primary key of the relation is **flight_number**. Consider the query:

select f.start, e.destination, f.departure_time, e.arrival_time
from flight f, flight e
where f.destination = e.start

(i) Describe in English what this query is asking. [4 marks]

(ii) One query optimisation technique reorganises a query tree using heuristic rules. Draw the canonical query tree for this query then show how the query tree is optimised. [10 marks]

c. Another query optimisation technique compares the costs of executing a query using different execution strategies. Why is this approach more suitable for compiled queries than interpreted queries? [4 marks]

d. Consider the schema described above and the query:

select f.start, e.destination, f.departure_time, e.arrival_time
from flight f, flight e
where f.start = 'Wellington' and f.destination = e.start

where there are 1000 **flight** records. The blocking factor of the **flight** relation is 10 records/block. There are 2 memory blocks available as buffers. There are 10 distinct values for **start** and the same 10 distinct values for **destination**. The blocking factor of the result is 10 records/block. State any assumptions you make and describe your working as you answer the following questions:

(i) What is the cost of executing this query if there are no indexes on the file? [12 marks]

(ii) What is the cost of executing this query if there is a clustering index on **start**? Assume there are 2 levels in the index. [12 marks]

****