Victoria University of Wellington

DEGREE EXAMINATIONS — 1998

MID-YEAR

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 [35 marks]

a. What is data abstraction? What is the relationship between data abstraction and conceptual data models? [2 marks]

b. What is the difference between the conceptual level and the view level? [2 marks]

c. What is the difference between a query language and a data manipulation language? [2 marks]

d. What is the difference between a data definition language (DDL) and a data manipulation language (DML)? Is the `create table` command part of SQL's DDL, DML, both, or neither? [2 marks]

e. Give an example of a semantic integrity constraint on a database (in English – you don't need to invent any notation) and a reasonable action that might be taken when that constraint is violated. [3 marks]

f. If relation r_1 contains 10 tuples and relation r_2 contains 5 tuples, how many tuples does $r_1 \times r_2$ contain? [2 marks]

g. Why isn't set intersection a fundamental operation in relational algebra? (You don't need to prove that it isn't fundamental – just explain why it isn't.) Give a definition of set intersection in terms of set difference and set union. [3 marks]

h. A company that manufactures lava lamps uses a relational database to keep track of what parts it has in stock, suppliers of parts, and which suppliers supply what parts. The schemas of these relations are:

```
part = (part-num, description, groovyness, quantity)
supplier = (supplier-name, address, phone-number)
supplies = (part-num, supplier-name, address, cost)
```

The primary key for relations on part is part-num, and the primary key for relations on supplier is {supplier-name, address}.

(i) What referential integrity constraint should hold between the supplier and supplies relations? [3 marks]

(ii) Is there any referential integrity constraint between the part and supplies relations? Why or why not? [3 marks]

(iii) Give a functional dependency expressing the fact that two different addresses do not share the same phone-number. [3 marks]

i. What is the DBMS system log used for? What are the typical kinds of entries in a system log? [4 marks]

j. What are checkpoints and why are they important? [3 marks]

k. What are transaction commit points and why are they important? [3 marks]

Question 2. Object oriented databases [65 marks]

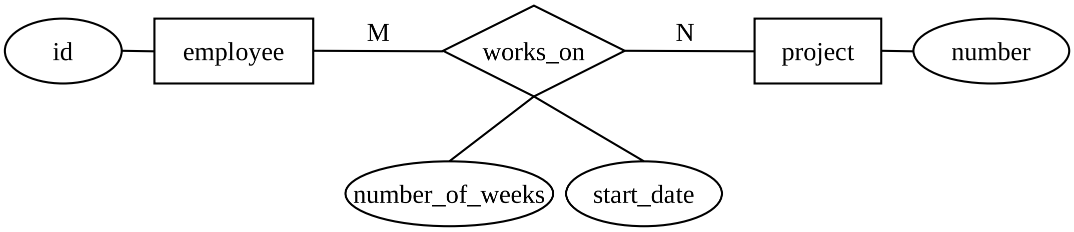
a. Describe the mechanism of persistence used in Objectivity. [3 marks]

b. Give one advantage of a declarative language over a navigational language. [2 marks]

(i) Write and describe a segment of **declarative** code which retrieves the videos a customer is currently hiring. You will also have to describe the schema you are working with. [14 marks]

(ii) Write and describe a segment of **navigational** code which retrieves the videos a customer is currently hiring. Again describe the schema you are working with. [14 marks]

c. Consider the following entity relationship diagram.



Write the code to model this information in Objectivity's DDL. Marks will not be deducted for minor errors in syntax. [12 marks]

d. Describe, in English, how you would add an index to the `employee` class in Objectivity. Give one advantage of an index in this situation. [10 marks]

e. Give a description and example of class (or schema) versioning and object versioning. [10 marks]

Question 3. Relational databases [45 marks]

You are a technical expert employed by the police department. The department has just found the secret hide-out of a gang of bank robbers. The leader of the gang has been using a relational database to store information about banks, robberies, and the members of the gang. You are called in to extract information from the database that will help in catching the robbers.

The database has relations bank-robbed (which stores information about banks that the gang has already robbed), bank-to-rob (which stores information about banks that the gang plans to rob), member (which stores information about gang members), and has-robbed (which stores information about which gang members have participated in robberies of what banks). Note that the gang may be planning to rob a bank that they have already robbed one or more times in the past. The schemas of these relations are:

```
bank-robbed=(name:string, address:string, number-of-depositors:integer)
bank-to-rob=(name:string, address:string, number-of-depositors:integer)
member = (name:string, address:string, years-in-prison:integer,
          banks-at:string, skill1:skilltype, skill2:skilltype)
has-robbed = (member-name:string, bank-name:string, date:string,
              amount:real)
```

where attribute banks-at of member is the name of the bank where the gang member does his or her banking, and skilltype is an enumerated type consisting of {lock-picking, driving, counting-money, public-relations}.

a. Give relational algebra, tuple relational calculus, domain relational calculus and SQL expressions for the following queries:
(i) Find the names and addresses of all gang members who have spent more than 10 years in prison. [7 marks]

(ii) Find the names and addresses of all banks that the gang has never robbed, but plans to rob someday. [8 marks]

(iii) Find all skills of all members of the gang. The resulting relation should have only one attribute (column). [8 marks]

b. If {name} is a candidate key for member and we have tuple variables t and s such that t∈member and s∈member and t[name] = s[name], what must the relationship between tuple t and tuple s be? [2 marks]

c. Comment on the design of the relational database described above. In your answer, consider design issues such as the potential for duplication of information, how well the design permits storage of information that appears to be of interest, if null values would be required very often, and any other relevant design issues. [8 marks]

d. Redesign the relational database described above so that the resulting relations are in 3NF. Show the keys and state any assumptions. [12 marks]

Question 4. Query processing and query optimization [35 marks]

a. Describe the 5 steps required to process queries on a relational database system. [5 marks]

b. Describe two query optimization techniques. [5 marks]

c. Consider the following relation schemas:

```
student = (id:integer, name:string, address:string)
course = (id:integer, name:string, dept:string)
student_course = (st_id:integer, course_id:integer)
```

There is a primary index on id in course. Consider the following query:

SELECT	s.name, s.address, c.name, c.dept
FROM	student s, course c, student_course
WHERE	s.name = 'Smith' **AND** s.id = st_id **AND** course_id = c.id

Choose one of the query optimization techniques you described in part (b) and give a detailed description of how the above query would be optimized using that technique. [25 marks]

****