Question	Mark
1	
2	
3	
4	
5	
6	
Total	

Surname: .....

Given name: .....

Id number: .....

CMPUT 291 (B2) - Final Examination (36%)

Time: 120 minutes
Questions: 6
Pages: 8

April 17, 2001
Closed Book
Calculators Allowed

- ***No questions during exam time.***
- ***If you are unsure, write down your assumptions.***
- ***Answer each question in the space given on this form.***
- ***No additional sheets are allowed.***

Question 1 `[18 marks]` Consider the following relational schema. Attribute `w_e_id` is a foreign key into `employee`.

employee (`e_id`, `e_name`, `e_loc`)
works_on (`w_project_id`, `w_e_id`, `w_percent_time`)

1. `[4 marks]` Give an SQL `create` statement for relation `works_on`.

2. [7 marks] Consider the dedicated employees who work on all projects, serving at least 10% of their time on each project. Create a suitable view listing the names of these dedicated employees, and the percentage of time they spend on each project.

3. [7 marks] Write an SQL query that is equivalent to the query below but does not use the `minus` operation.

```
select  e_loc
from    employee, works_on
where   e_id = w_e_id   and
        w_project_id = 'P1'
minus
select  e_loc
from    employee
group by e_loc
having  count(*) < 100;
```

Question 2 [12 marks] Consider the following relational schema that models students and student marks for a particular course:

student (sid, name, major)
marks (type, number, sid, mark)

The **student** relation includes one tuple for each student in the course, with the `sid` number as the key. The **marks** relation includes one tuple for each exam or assignment mark received by each student. The `type` attribute is either `'exam'` or `'assignment'`, and the `number` field holds a code that identifies which exam or assignment the mark is for. For example, the tuple `('exam', 'midterm', 0991234, 80)` indicates that the student whose `sid` number is `0991234` received a mark of 80 on the midterm exam, while `('assignment', '2', 0991234, 90)` says that the same student received a mark of 90 on assignment 2. Attributes `type`, `number`, and `sid` are the key for **marks**, and `sid` is a foreign key referencing the **student** relation.

Write an SQL query that returns the `sid` numbers and names of `'CS'` majors whose average assignment mark is at least 10 points higher than their average exam mark. `CS` majors who do not have any assignment marks, or who do not have any exam marks, should not be included in the result. The query should not report a student more than once.

Question 3 [12 marks]

1. Translate the following relational calculus query into relational algebra, assuming the existence of relations $R(A, B, C)$, and $S(D, E)$:

$$\{t \mid t \in R \ \wedge \ t.A = 5 \ \wedge \exists x \ (x \in S \ \wedge (t.B = x.D \ \vee \ t.C = x.D))\}$$

2. Consider the following two relations:

$$R(A_1, A_2, \cdots, A_m)$$
$$S(A_m, A_{m+1}, \cdots, A_n)$$

where $1 < m < n$. Prove or disprove the following algebraic equivalence defined over these two relations:

$$\sigma_{(A_1=k) \ OR \ (A_n=k)}(R \bowtie S) \quad = \quad ((\sigma_{A_1=k}(R)) \bowtie (\sigma_{A_n=k}(S))$$

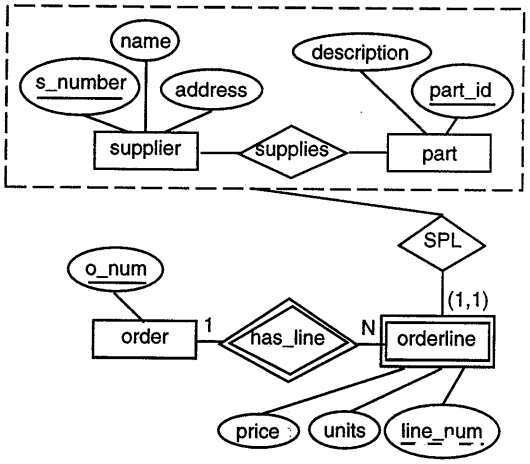
Notes: The $\bowtie$ denotes the *natural-join* operator (where equalities are specified on all fields having the same name). If you prove the equivalence, your proof should be clear and brief. To disprove the equivalence, provide a specific counter example.

Question 4 [8 marks] The following describes information to be maintained in a parts/suppliers/orders database:

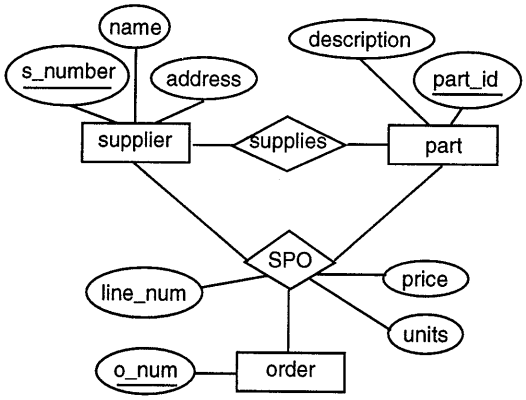
- Each part has a unique part identifier and a description. A part may have many possible suppliers.
- Suppliers have unique supplier numbers. They also have names and addresses.
- Orders have unique order numbers. An order includes one or more order lines, numbered consecutively starting with 1. Each order line identifies a single part, the unit price for that part, and the number of units of the part that have been ordered. Each line also indicates which supplier (from among the possible suppliers for the line's part) will be supplying the parts described on that line.

Consider the following two entity-relationship models for the above requirements:

Model A



Model B



Which of these two models best describes the parts/suppliers/orders? For the model which is *not* best, identify (in point form) the shortcomings. Be brief, and be specific.

Question 5 [12 marks]

Translate the E-R diagram of Model A in the previous question into a relational schema. Identify all primary, and foreign key constraints. Use as few schema as possible.

Question 6 [12 marks]

1. True or false: for equality queries, B^+–trees are faster than hash-based indexes on the average? Explain your answer.

2. Consider a B^+–tree in which an internal node can hold up to 3 keys, and a leaf can hold up to 2 records. Starting from an empty tree, perform the following operations in the given order (show the final tree in each part).

   (a) Insert 33*, 44*

   (b) Insert 65*, 25*

   (c) Insert 100*, 50*

   (d) Insert 60*

   Use the following notation in your tree diagrams: denote an entry that contains both a search key k, and a data record (or a pointer to a data record) by k^*. An entry that contains just a search key k (with no data record) has no $'*'$.