Given name:_____ **Family name**:_____

Student number:_____ **Signature**:_____

UNIVERSITY OF VICTORIA
Faculty of Engineering

Department of Computer Science

Examinations December 2002

CSC 370 F02 (Database Systems)
Instructor: Daniel M. German

Duration: 3 hrs

This is a closed-book exam. You are only allowed two letter-size sheets of paper.

This examination paper consists of **16** pages and **6** questions. Please bring any discrepancy to the attention of an invigilator. The number in brackets at the start of each question is the number of points the question is worth.

Answer all questions.

Please write your answers clearly.

For instructor's use:

	Score
1 (10)	
2 (16)	
3 (15)	
4 (21)	
5 (10)	
6 (28)	
Total (100)	

1. General questions

(a) [2] Briefly explain what is the *durability* property of a transaction.

(b) [2] Briefly describe *pipelining* (when applied to query evaluation).

(c) [2] Briefly explain why is I/O cost important in a DBMS.

(d) [2] Give 2 examples of database operations that require the use of *sorting*.

(e) [2] Define the term *functional dependency*.

2. SQL

For this question, consider the following relation and its instance. Attributes with an empty value should be assumed to be NULL.

```
Students(sid:integer, name: string, login: string,
         age: integer, gpa: real)
```

sid	*name*	*login*	*age*	*gpa*
53666	Jones	Jones@cs	18	7.4
53668	Smith	smith@ee	18	7.8
53650	Smith	smith@math	19	7.4
53831	Madayan	madayan@music	11	8
53832	Guldy	guldu@music	12	

(a) [3] Write a query that computes the average of the attribute `gpa` using `AVG`. Compute the result of the query.

(b) [3] Write a query that computes the average value of the attribute `gpa` using `COUNT(gpa)`, and `SUM(gpa)`. In other words, you are not allowed to use `AVG(gpa)`. Compute the result of the query.

(c) [2] In your answer (b), replace `COUNT(age)` by `COUNT(*)`. Compute the result of the new query.

(d) [2] Write a query that computes the name of those students who do not have a gpa yet.

(e) [3] Write a query that groups students who are at least 18 years old by their age. For each of these groups compute their average gpa and order the result by age (in ascending order).

(f) [3] Write a query that computes the name of those students who are at least as old as the average age.

3. Transaction Management

Consider a database with objects X and Y and assume that there are two transactions T1 and T2. T1 reads objects X and Y and then writes X. T2 reads objects X and Y and then writes X and Y.

(a) [3] Give an example of a schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-read conflict.

(b) [3] Give an example of a schedule with actions of transactions T1 and T2 on objects X and Y that results in a read-write conflict.

(c) [3] Give an example of a schedule with actions of transactions T1 and T2 on objects X and Y that results in a write-write conflict.

(d) [6] For each schedule, show that Strict 2PL disallows the schedule.

4. Query Evaluation

For this section, assume the following schema:

```
Reserves(Sid : integer,
         Bid : integer,
         Day : date,
         CreditCard: char(16))
```

where the key is $(Sid, Bid, Date)$. The relation is kept in an unsorted heap file.

We also have indexes in the following attributes: a hash index on $\langle bid \rangle$, and an unclustered B+Tree index on $\langle Sid, Bid, Date \rangle$.

Finally, you are given the following query:

```
SELECT * FROM Reserves
WHERE Sid = 10 AND Bid = 10
```

And you are given the following statistics about the relation: each page can hold 100 Reserves tuples. For each attribute in the relation, you can assume that its values are uniformly distributed.

Finally the fan out of the B+ Tree is 100.

(a) [6] List all the possible access paths available and, if necessary, their corresponding primary conjunct(s).

(b) [6] You are also given the following information: the relation contains 1 million rows, 1,000 different values for sid and 10 different values for bid. Which one of the access paths you listed is likely to be the least expensive? Show all your work. If necessary, clearly state any assumptions you make.

(c) [3] Assume that the number of different values for the sid and bid attributes remains the same, but the number of rows in the relation is now 1,000. Which one of the access paths you listed is likely to be the least expensive? Show all your work. If necessary, clearly state any assumptions you make.

(d) [3] Replace the original query with the following one. Would your answer for (b) change?

```
SELECT  Bid FROM Reserves
WHERE Sid = 10 AND Bid = 10
```

Explain your answer and show all your work. If necessary, clearly state any assumptions you make.

(e) [3] Using the relation $Reserves$, give an example of a query that is likely to be implemented using the iteration technique. Explain your answer. If necessary, clearly state any assumptions you make.

5. Crash Recovery

(a) [6] List the 3 phases of the ARIES recovery algorithm. Briefly explain each one of them.

(b) [4] How does the recovery manager ensure atomicity and durability of transactions?

6. Normalization

(a) [10] Prove the following statement: Let R be a relation and F be a set of FDs that hold over R. **If** the decomposition of R into relations R_1, R_2 is lossless-join **then** F^+ contains either the FD $R_1 \cap R_2 \rightarrow R_1$ or the FD $R_1 \cap R_2 \rightarrow R_2$. Show all your work.

(b) [8] Consider the attribute set R= ABCDEFGH and the FD set

$$F = \{AB \rightarrow C, AD \rightarrow E, B \rightarrow D, BC \rightarrow A, E \rightarrow G\}$$

Is this decomposition dependency preserving? Is it lossless-join? Show all your work.

$$\{AB, BC, ABDE, EG\}$$

Is this decomposition dependency preserving? Is it lossless-join? Show all your work.

$$\{ABC, ACDE, ADG\}$$

(c) [5] Let relation R = ABCE, and assume the following set of FDs:

$$F = \{A \rightarrow BC, BC \rightarrow A, B \rightarrow E, CE \rightarrow B\}$$

Find all the keys for this relation. Show all your work.

(d) [5] You are given a relation $R = ABCD.$ with set of FDs:

$$F = \{AB \rightarrow C, AB \rightarrow D, C \rightarrow A, D \rightarrow B\}$$

You are also given the keys for this relation: AB, AD, BC, CD
Is R in BCNF, in 3NF, in both, or in neither? Show all your work.

If R is not in BCNF, then provide a lossless-join, dependency preserving decomposition for it. Show all your work.

End of examination
Total pages: 16
Total marks: 100