

B1 Final Examination

April 18, 2000

It is a close-book examination and the time for the test is 120 minutes. There are eleven (11) questions over two (2) pages. The value of each question is indicated in [] and the total is 100. Good luck to all of you.

1. A well-known program package in C, called the Berkeley DB, provides all the access methods to various hard disk files, including B+trees, hash tables, and sequential record files. The Berkeley DB also contains functions for locking, logging and transaction processing. Is such a system necessarily a database management system? Explain your answer. [5]
2. Give an SQL query to find all drinkers that visit only bars that serve no beer that they like on a database consisting of three relations below:

visits (drinker, bar)
 services (bar, beer)
 likes (drinker, beer)

The first one indicates the bars each drinker visits, the second tells what beers each bar serves, and the last indicates which beers each drinker likes to drink. [9]

3. Consider a database consisting of the following tables.

employee(employee_name, position, salary)
 supervision(employee_name, supervisor_name)

The first table indicates the job title and salary for each employee and the second indicates who is whose boss.

Write a trigger (or two) to enforce the constraint that every employee has a supervisor, except the president (i.e., the position is "president"). [9]

4. Consider $R = ABCDE$. [10]

- (a) Indicate whether the following table (1) violates the FD $AB \rightarrow D$, and (2) violates the MVD $AC \twoheadrightarrow D$:

A	B	C	D	E
1	2	3	4	5
1	3	3	5	4
1	2	2	6	5

- (b) Indicate whether the following table violates the MVD $AC \twoheadrightarrow E$:

A	B	C	D	E
a	a	a	a	f
a	a	b	b	e
a	e	b	b	f
a	e	b	b	e

5. Consider the following set F of functional dependencies on relation schema $R = ABCDE$.

$A \rightarrow BC$
 $CD \rightarrow E$
 $B \rightarrow D$
 $E \rightarrow A$

- (a) List all the candidate keys for R.
- (b) Decompose R into a BCNF database with respect to F.
- (c) Indicate whether your schema obtained in the previous step is dependency preserving with respect to F. [10]

6. Given a relation r on the schema $R = ABC$. Write an SQL query to test whether the functional dependency $B \rightarrow C$ holds on relation r. You need to give an SQL query and then specify under what condition the answer to the query indicate the FD holds on r. [9]



04825
 CMPUT 391 (B1)
 YUAN, L.-Y.
 APR 00 FINAL
 PAGES: 2

7. The SQL3 specifies four (4) different levels of isolation, i.e., READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ, and (ANOMALY) SERIALIZABLE. Present two sample schedules S_1 and S_2 such that S_1 observes READ UNCOMMITTED but not READ COMMITTED, and S_2 observes REPEATABLE READ but not SERIALIZABLE. [9]
8. The following is a schedule with one action missing:

$T_1,$	T_2
R(A)	
	R(B)
???	
W(C)	
	W(A)

You are asked to figure out what actions could replace the ??? and make the schedule not serializable. List all possible non-serializable replacements.

Note that a possible replacement is an action of the form $R(T_i, Q)$ (or $W(T_i, Q)$), where $i = 1, 2$, and Q is one of A, B, C, indicating that transaction T_i reads from (or, write onto) data item Q . A non-serializable replacement is a replacement $R(T_i, Q)$ (or $W(T_i, Q)$) such that the given schedule is not serializable when ??? is replaced by $R(T_i, Q)$ (or $W(T_i, Q)$). [10]

9. Consider the relations $r_1(A, B)$, $r_2(B, C)$, and $r_3(C, D)$. Let $V(r_1, A)$ be 20, $V(r_1, B)$ be 80, $V(r_2, B)$ be 50, $V(r_2, C)$ be 60, $V(r_3, C)$ be 120, $V(r_3, D)$ be 60, Further assume that r_1 has 100 tuples, r_2 has 200 tuples, and r_3 has 500 tuples. Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$, and give an efficient strategy (meaning, in which order) for computing the join.

Note that $V(r, A)$ denotes the selection cardinality of r on A , that is, the number of distinct values for attribute A in r , and that the uniform distribution is assumed for all relations. [9]

10. Consider the following database consisting of four transactions of items.

transaction	items
t_1	pen, ink, diary, soap
t_2	pen, ink, diary
t_3	pen, diary
t_4	pen, ink, soap

- (a) Simulate the iterative algorithm (the Apriori algorithm discussed in the class) for finding the large item-sets on the given database with support $\geq 40\%$. (You may show step by step how the algorithm works for this example).
- (b) Find all the association rules with confidence $\geq 70\%$ and support $> 70\%$. [10]
11. Consider the following training set. Note that each object has three attributes, i.e., A, B, and C, and is classified as either P or N. Try to construct two different decision trees such that one is much simpler than the other. [10]

No	A	B	C	Class
1	a1	b2	c1	P
2	a2	b1	c1	N
3	a2	b2	c3	P
4	a1	b3	c3	P
5	a3	b2	c6	P
6	a2	b1	c1	N
7	a1	b1	c3	N
8	a3	b1	c3	N
9	a1	b2	c6	P
10	a2	b1	c6	N
11	a3	b2	c1	P
11	a2	b2	c1	P