

CSE 444 Midterm Test

Autum 2008

Name: _____

Total time: 50'

Question 1	/40
Question 2	/30
Question 3	/30
Total	/100

1. **SQL [40 points]**

We have a database of documents. Each document consists of several sections, each section contains several words:

<code>Doc(<u>docID</u>, docTitle)</code>	– documents
<code>Section(<u>docID</u>, <u>secNumber</u>, secTitle)</code>	– sections
<code>WordOcc(<u>docID</u>, <u>secNumber</u>, word)</code>	– word occurrences

`Section(docID)` is a foreign key to `Doc(docID)`; `WordOcc(docID, secNumber)` is a foreign key to `Section(docID, secNumber)`. Each document has at least one section; each section has at least one word.

- (a) **[10 points]** Write a SQL query that computes for each document the total number of distinct words used in that document. The answer should consist of triples: document id, document title, word-count.

- (b) **[5 points]** Write a SQL query that finds all documents containing both keywords 'midterm' and 'solution' in the same section. For each such document you should return its document id and title.

- (c) **[20 points]** Next, you are given a table $QW(\text{word})$ of keywords. Write a SQL query that finds all documents that contain all the keywords listed in QW . For each such document you should return its document id and title.

- (d) **[5 points]** Let Answers_{1b} be the set of answers returned by the query you wrote for question (1b). Assume QW contains exactly two keywords, 'midterm' and 'solution', and let Answers_{1c} be the set of answers returned by the query you wrote for question (1c). Circle all the statements below that are guaranteed to be true:

$\text{Answers}_{1b} \subseteq \text{Answers}_{1c}$ $\text{Answers}_{1b} = \text{Answers}_{1c}$ $\text{Answers}_{1b} \supseteq \text{Answers}_{1c}$

2. [30 points] **Conceptual Design**

(a) [10 points] Design an E/R diagram for a database of documents, authors, and readers. The database has the following entities and relationships:

- **Document**: has a **docID** (the key), **title**.
- **Section**: each document consists of several sections; each a section has a **title**, and a **number** (1, 2, 3, ...). The section number is unique within each document.
- **Author**: each author is the creator of several documents. Each document was created by exactly one author.
- **Reader**: each reader reads several documents, and each document is read by several readers.
- **Person**: has **name**, **address**. Authors and readers are persons.

You have to turn in an E/R diagram.

- (b) [10 points] Database Normalization. Consider a relation $R(A, B, C, D, E, F, G)$ with the following functional dependencies:

$$\begin{aligned}A &\rightarrow B \\C &\rightarrow AD \\CE &\rightarrow B \\EF &\rightarrow C\end{aligned}$$

Compute the Boyce-Codd Normal Form (BCNF) decomposition of R . Indicate each step you make in your computation, by showing the relation to which you apply the step and the violation of BCNF that you use during that decomposition step. Indicate clearly your end result: the relations, their attributes, and their keys.

- (c) [10 points] Consider a relation $R(A, B, C, D, E)$. A set of attributes X is called "closed" if $X^+ = X$. Given an example of functional dependencies on R such that the only closed sets are A, B, AC, BD, ABE .

3. [30 points] Transactions

- (a) [20 points] After a systems failure, the undo-redo recovery log has the following entries:

```
<START T1>
<T1 A 1 2>
<START T2>
<COMMIT T1>
<START T3>
<T3 A 2 3>
<START T4>
<CKPT(T2,T3,T4)>
<T2 B 10 20>
<COMMIT T2>
<START T5>
<T5 D 1000 2000>
<T4 C 100 200>
<COMMIT T5>
<START T6>
<END CKPT>
<T6 D 2000 3000>
```

An entry $\langle T, X, u, v \rangle$ means that transaction T has updated the value of X from u (the old value) to v (the new value). $\langle \text{CKPT}(\dots) \rangle$ denotes the beginning of a checkpoint and lists the currently active transactions. $\langle \text{END CKPT} \rangle$ is written to disk once all dirty pages of the active transactions have been flushed to disk. The redo phase precedes the undo phase during the recovery.

- i. [5 points] Which are the transactions whose actions the recovery manager needs to redo ?

- ii. [5 points] Which are the transactions whose actions the recovery manager needs to undo ?

- iii. [10 points] Indicate the actions of the recovery manager on all the elements, separately during the Redo and the Undo phase.

	Redo	Undo
A=		
B=		
C=		
D=		
E=		

- (b) [10 points] Indicate for each statement below if it is true or false. You do not have to justify your answer:

- i. Every schedule that is possible under a timestamp based concurrency control scheduler is also possible under a multiversion concurrency control scheduler.
Yes or No ?
- ii. Every schedule that is possible under a multiversion concurrency control scheduler is also possible under a timestamp based concurrency control scheduler.
Yes or No ?
- iii. Suppose that every transaction T_i has the form $ST_i, RD_i(X), WT_i(Y), CO_i$, for some elements X and Y. Then a 2-phase locking scheduler that uses shared locks for read operations and exclusive locks for write operations will never result in a deadlock.
Yes or No ?
- iv. Suppose that all transactions are either read-only (i.e. they have the form $ST_i, RD_i(X), RD_i(Y), \dots, CO_i$), or they consists of a single write (i.e. they have the form $ST_i, WT_i(X), CO_i$). Then a 2-phase locking scheduler that uses shared locks for read operations and exclusive locks for write operations will never result in a deadlock.
Yes or No ?