DB Qual 2005; Part C

Problem 1 (8 points)

You want to build a hash-table index for the following scenario:

- The file contains 1024 fixed-size records;
- Each disk block can hold up to 8 records;
- The search keys are hashed into a 13-bit string, and no two records yield the same 13-bit string.

(a) If you build an extensible hash table, in the *best* case what will be the (in-memory) directory size?

(b) If you build an extensible hash table, in the *worst* case what will be the (in-memory) directory size?

(c) If you build a linear hash table, in the *best* case how many blocks will be used?

Problem 2 (8 points)

A given transaction T updates database objects A and B, in this order. At some point during or after the execution of T, the system crashes. For each of the two objects, there are four possible cases:

1. T's update was not written to the log, and it was not propagated to the database. (The update could have been reflected in the in-memory log-tail, or could have been made to a copy of the object in the buffer pool, but these changes did not make it to disk.)
2. T's update was written to the log (on disk), but it was not propagated to the database.
3. T's update was not written to the log, but it was propagated to the database (on disk).
4. T's update was written to the log (on disk), and it was propagated to the database (on disk).

Thus, we can represent the state of affairs (after the crash) by two numbers (one for each object) referring to the cases above. For example, $A:1, B:2$ means that A's update was neither logged nor propagated to the database, while B's update was logged but not propagated to the database.

(a) Assume that pure *undo* logging is used and that T did *not* commit before the crash. List all possible states using the $A:n, B:m$ notation.

(b) Assume that pure *undo* logging is used and that T *did commit* before the crash. List all possible states.

(c) Assume that *undo-redo* logging is used and that T did *not* commit before the crash. List all possible states using the $A : n, B : m$ notation.

(d) Assume that pure *undo-redo* logging is used and that T *did commit* before the crash. List all possible states.

Problem 3 (4 points)

Consider a database with two objects, A and B. Give a schedule that can be produced by a two-phase locking protocol but cannot be produced by a validation protocol. Briefly explain your answer.