UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

COMPUTER SCIENCE THREE

DATABASE SYSTEMS

Wednesday 24 May 2000

09:30 to 11:00

Examiners:

Prof. K. Turner (External)

Prof. G. Brebner (Chairman)

INSTRUCTIONS TO CANDIDATES

Answer any TWO questions.

All questions carry equal weight.

1. (a) In a multidimensional database, the quantity of summary information can vastly exceed the quantity of raw data. Explain how this can happen. [*3 marks*]

   (b) Suppose that an OLAP system is to be implemented using a relational database.

      i. What is meant by the term 'consolidated dimensional hierarchy'? [*3 marks*]

      ii. Why might there be advantages to using consolidated dimensional hierarchies, rather than snowflake schemata, in the design of the underlying relational database? [*6 marks*]

   (c) The MDX language is used to query multidimension databases.

      i. The `CrossJoin()` function takes two sets of members as arguments. Explain the role of this function in an axis specification in an MDX query. [*3 marks*]

      ii. Explain the role of the slicer specification in an MDX query with respect to the measures dimension and the other dimensions in the cube. [*3 marks*]

   (d) In OLTP systems, 'modification' anomalies can occur when a table is updated.

      i. With the aid of small examples, explain what is meant by the term 'modification anomaly'. [*3 marks*]

      ii. How does Third Normal Form normalisation produce a database design that can avoid undesirable features such as modification anomalies? [*4 marks*]

2. (a) Under what circumstances would the creation of an index be disadvantageous for overall system performance? Your answer should include a characterisation of columns on which it is inadvisable to create indexes. *[3 marks]*

(b) In a B^+-tree, an internal node has the following layout:

$$< P_1,\ K_1,\ P_2,\ K_2,\ \ldots\ ,\ P_{q-1},\ K_{q-1},\ P_q >$$

where p is the order of the B^+-tree, $q \leq p$, each K_i is a search key field and each P_i is a tree pointer.

i. With the aid of a diagram, give the layout of a leaf node in a B^+-tree. *[2 marks]*

ii. Calculate the maximum order of an internal node and of a leaf node in a B^+-tree, given the following data:

Block size, B = 256 bytes.

Search key field size, V = 4 bytes.

Record pointer, R = 8 bytes.

Tree node pointer, P = 6 bytes. *[6 marks]*

iii. Suppose that an internal tree node can contain 2 search key fields and a leaf node contains two search key fields. Starting from an empty B^+-tree, the following sequence of values is to be inserted in the index:

$$9, 4, 2, 7, 3$$

Use diagrams to depict the state of the B^+-tree after each of these inserts, labelling the diagrams with the tree pointers and search key field values. Give brief explanations where nodes are split and where the tree is re-structured. *[8 marks]*

(c) Consider the three transactions, T_1, T_2 and T_3 given below:

$$T_1:\ r_1(A); r_1(B); w_1(A);$$

$$T_2:\ r_2(A); r_2(C); w_2(B);$$

$$T_3:\ r_3(C); r_3(D); w_3(C); w_3(B);$$

Coonsider the following schedule for T_1, T_2 and T_3.

$$r_1(A); r_3(C); r_2(A); r_1(B); r_3(D); r_2(C); w_1(A); c1; w_3(C); w_2(B); w_3(B); c2; c3;$$

Determine the strictest recoverability condition that this schedule satisfies out of strict, cascadeless, recoverable and non-recoverable. *[6 marks]*

3. (a) The following three inference rules constitute a sound and complete set of rules for inferring functional dependencies.

IR1 If $Y \subseteq X$ then $X \rightarrow Y$

IR2 $\{X \rightarrow Y\} \models XZ \rightarrow YZ$

IR3 $\{X \rightarrow Y \, , \, Y \rightarrow Z\} \models X \rightarrow Z$

Now consider the rule

IR4 $\{X \rightarrow Y \, , \, Y \rightarrow Z\} \models X \rightarrow YZ$

Using rules IR1, IR2 and IR3, prove that rule IR4 is also a sound inference rule for functional dependencies. [*6 marks*]

(b) Let $R = \{A, B, C\}$, $R_1 = \{A, B\}$ and $R_2 = \{B, C\}$ be sets of attributes.

i. Give a small example of a relation over R with the property that when the relation is projected onto R_1 and R_2 and the natural join of the projected relations is taken, the resulting relation over R is different from the original relation over R, i.e. the join is lossy.
Your answer should include the original relation, the projected relations and the joined relation. [*3 marks*]

ii. Prove that if the original relation, say r, over R obeys the functional dependency $B \rightarrow A$ then $\{R_1, R_2\}$ is a lossless decomposition of R. [*8 marks*]

(c) Let $R = \{A, B, C, D, E, F, G, H, I\}$ be a set of attributes and let $\Phi = \{ABC \rightarrow D, ABI \rightarrow D, A \rightarrow EH, B \rightarrow F, EF \rightarrow G, G \rightarrow C\}$ be a set of functional dependencies.

Find a minimal cover for Φ. [*8 marks*]