TDBC86 Database Concepts Fall 2001
Obligatory Exercise 5
Due date: December 14, 2001 at 1700

The following table applies to Problems 1 and 2.

Number of tuples in the relation	5×10^9
Size of one tuple	160 bytes
Size of pointer to B-tree node	4 bytes
Page size	2048 bytes

1. A B-tree is used as the data structure for a relation. Assume that each B-tree node is stored in one page, and that each tuple is stored as one record. Answer the following.

   (a) Compute the maximum number of records (tuples) which may be stored in a single page.

   (b) Compute the maximum height that such a B-tree may have. (The height is defined as the length of a path from the root to a leaf.)

   (c) Compute the minimum height that such a B-tree may have.

2. This question investigates record distributions in B-trees, using the values computed in Question 1.

   (a) For a tree of the maximum height computed in 1(b), compute the average number of records in non-root nodes, assuming that the root contains exactly one record. State whether or not this value is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. If it is consistent, give also the total number of nodes in the tree.

   (b) For a tree of the maximum height computed in 1(b), compute the average number of records in a node, assuming a uniform distribution of records in all nodes, including the root. State whether or not this value is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. If it is consistent, give also the total number of nodes in the tree.

   (c) For a tree of the maximum height computed in 1(b), assume a uniform distribution of records in all nodes but the root. Compute the minimum number of records which is possible for the root node, under the condition that the average number of records in the other nodes is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. Give also the total number of nodes in the tree.

(d) For a tree of the maximum height computed in 1(b), assume a uniform distribution of records in all nodes but the root. Compute the maximum number of records which is possible for the root node, under the condition that the average number of records in the other nodes is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. Give also the total number of nodes in the tree.

(e) For a tree of the minimum height computed in 1(c), compute the average number of records in non-root nodes, assuming that the root contains exactly one record. State whether or not this value is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. If it is consistent, give also the total number of nodes in the tree.

(f) For a tree of the minimum height computed in 1(c), compute the average number of records in a node, assuming a uniform distribution of records in all nodes, including the root. State whether or not this value is consistent with the constraints on the minimum and maximum number of records in a node of a B-tree. If it is consistent, give also the total number of nodes in the tree.

The following table applies to Problems 3 and 4.

Number of tuples in the relation	5×10^9
Size of one tuple	160 bytes
Size of pointer to a B^+-tree node	4 bytes
Size of internal B^+-tree key	16 bytes
Page size	2048 bytes
Time to access a single page	6 ms.
Sequential-access pointers in the leaf nodes	8 bytes total

3. Suppose now that the relation will be stored in a B^+-tree. Assume that each node, interior or leaf, is stored in one page, and that each tuple is stored as one record. Answer the following.

(a) Compute the maximum number of indices per non-leaf node.

(b) Compute the maximum number of records per leaf node.

(c) Compute the maximum height that such a B^+-tree may have. (The height is defined as the length of a path from the root to a leaf.)

(d) Compute the minimum height that such a B^+-tree may have.

TDBC86, Obligatory Exercise 5, page 3

4. This question investigates record distributions in B^+-trees, using the values computed in Question 3.

   (a) For a B^+-tree of the minimum height computed in 3(d), compute the average number of indices in the interior nodes other than the root, assuming that the root contains exactly one index. State whether or not this value is consistent with the constraints on the minimum and maximum number of indices in an interior node of a B^+-tree. In addition, if this value is consistent, compute the number of interior nodes and the number of leaf nodes, as well as the average number of records in each leaf node.

   (b) For a B^+-tree of the minimum height computed in 3(d), compute the average number of indices in the interior nodes, assuming a uniform distribution of indices throughout all interior nodes, and the maximum number of records per leaf node. State whether or not this value is consistent with the constraints on the minimum and maximum number of indices in an interior node of a B^+-tree. In addition, if this value is consistent, compute the number of interior nodes and the number of leaf nodes.

   (c) For a B^+-tree of the minimum height computed in 3(d), assume a uniform distribution of indices in all interior nodes but the root. Compute the minimum number of indices which is possible for the root node, under the constraint that the average number of indices in the other interior nodes is consistent with the constraints on the minimum and maximum number of entries in the nodes of a B^+-tree. In addition, compute the number of interior nodes and the number of leaf nodes. For this computation, set the number of records in each leaf node to be the minimum possible.

   (d) Repeat the computation of (c), this time setting the number of records in each leaf node to be the maximum possible.

   (e) For a B^+-tree of the minimum height computed in 3(d), assume a uniform distribution of indices in all interior nodes but the root. Compute the maximum number of indices which is possible for the root node, under the constraint that the average number of indices in the other interior nodes is consistent with the constraints on the minimum and maximum number of entries in the nodes of a B^+-tree. In addition, compute the number of interior nodes and the number of leaf nodes. For this computation, set the number of records in each leaf node to be the maximum possible.

Notes:

- As stipulated in the course syllabus, this exercise may be done either individually, in a group of two, or in a group of three.
- Remember that there are point penalties for late submission. See the course syllabus.

- It is strongly recommended that you use a text-processing tool to display your results. If you write them by hand, they must be very neat.
- It is not allowed to copy the work of others. The submission must be the original work of the individual(s) in the working group.
- The grader reserves the right to interview members of the working group about the solution.
- So that solutions may be discussed in a class meeting, students and/or groups that are very late in preparing a solution may be required to solve an alternate problem to receive credit for this exercise.
- If you have solved this problem for a previous offering of the course, you may re-use your old solution, subject to the following conditions: (a) You may not work with any partners, except possibly those with whom you worked to prepare the solution in the previous course. (b) You must explicitly note any partners from the previous course with whom you submitted a joint solution for that course. Note that grading criteria may not be identical between years, so that a solution which was found to be satisfactory last year may not be evaluated similarly this year. Note also that the problems themselves may be different from those of previous years. You must in any case solve the problems for this year!