INFS2200/7903 – Relational Database Systems

School of Information Technology and Electrical Engineering (ITEE), UQ

Tutorial 5

Question 1 Consider the following configuration:

- A disk with block size (B) = 512 bytes
- A block pointer (P) = 6 bytes

An EMPLOYEE file has the following fields:

NAME (30 bytes), SSN (9 bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes), BIRTHDATE (8 bytes), SEX (1 byte), JOBCODE (4 bytes), SALARY (4 bytes)

The EMPLOYEE file has N = 30,000 EMPLOYEE records of fixed-length format and there is an additional byte used as a deletion marker in each record.

A. Calculate the record size (R) in bytes.

B. Calculate the blocking factor (bfr) and the number of file blocks (b) assuming an unspanned organization.

C. Suppose the file is ordered by the key field SSN and we want to construct a primary index on SSN.
   a. Calculate the index blocking factor (bfri).
   b. Calculate the number of index entries and the number of index blocks.

D. Suppose the file is not ordered by the key field SSN and we want to construct a secondary index on SSN.
   a. Calculate the index blocking factor (bfri).
   b. Calculate the number of index entries and the number of index blocks.

E. Suppose the file is ordered by the non-key field DEPARTMENTCODE and we want to construct a clustering index on DEPARTMENTCODE that uses block anchors (every new value of DEPARTMENTCODE starts at the beginning of a new block). Assume there are 1000 distinct values of DEPARTMENTCODE, and the EMPLOYEE records are evenly distributed among these values.
   a. Calculate the index blocking factor (bfri).
   b. Calculate the number of index entries and the number of index blocks.

F. Now we want to extend this single-level clustering index to a multi-level index.
   a. Calculate the number of levels needed if we make it a multi-level index.
   b. Calculate the total number of blocks required by the multi-level index.
   c. Calculate the number of block accesses needed to search for and retrieve all records in the file having a specific DEPARTMENTCODE value using the multi-level index. Assume that multiple blocks in a cluster are contiguous. That is, data records with the same DEPARTMENTCODE are stored in adjacent blocks.

Answers for Question 1 are given below:

A. R = (30 + 9 + 9 + 40 + 9 + 8 + 1 + 4 + 4) + 1 = 115 bytes.

B. Blocking factor (bfr) = floor (B/R) = floor (512/115) = 4 records per block.

Number of file blocks (b) = ceiling (N/bfr) = ceiling (30000/4) = 7500 blocks.

C. For a primary index on SSN:

a. Index entry size (R_i) = (SSN+P) = (9+6) = 15 bytes.

Index blocking factor (bfr_i) = floor (B/R_i) = floor (512/15) = 34 index entries per block.

b. Number of index entries (N_i) = number of file blocks (b) = 7500 entries.

Number of index blocks (b_i) = ceiling (N_i/bfr_i) = ceiling (7500/34) = 221 blocks.

D. For a secondary index on SSN:

a. Index entry size (R_i) = (SSN+P) = (9+6) = 15 bytes.

Index blocking factor (bfr_i) = floor (B/R_i) = floor (512/15) = 34 index entries per block.

b. Number of index entries (N_i) = number of file records (N) = 30000 entries.

Number of index blocks (b_i) = ceiling (N_i/bfr_i) = ceiling(30000/34) = 883 blocks.

E. For a clustering index on DEPARTMENTCODE:

a. Index entry size (R_i) = (DEPARTMENTCODE+P) = (9+6) = 15 bytes.

Index blocking factor (bfr_i) = floor (B/R_i) = floor (512/15) = 34 index entries per block.

b. Number of index entries (N_i) = Number of distinct DEPARTMENTCODE values = 1000 entries.

Number of index blocks (b_i) = ceiling (N_i/bfr_i) = ceiling (1000/34) = 30 blocks.

F. For a multi-level index on DEPARTMENTCODE:

a. We can calculate the number of levels step by step as follows:

- Number of first-level index entries (N_1) = Number of distinct DEPARTMENTCODE values = 1000 entries.
- Number of first-level index blocks (b_1) = ceiling (N_1/bfr_i) = ceiling (1000/34) = 30 blocks.
- Number of second-level index entries (N_2) = number of first-level index blocks (b_1) = 30 entries.
- Number of second-level index blocks (b_2) = ceiling (N_2/bfr_i) = ceiling (30/34) = 1 block.
- Since the second level has one block, it is the top index level. Hence, the index has x = 2 levels.

We can also calculate the number of levels directly as follows:

- Number of levels x = ceiling ($\log_{bfri}$ N_i) = ceiling ($\log_{34}$ 1000) = 2 levels.

b. Total number of blocks for the index $b_i = b_1 + b_2 = 30 + 1 = 31$ blocks.

c. Number of block accesses to search for the first block in the cluster of blocks = x + 1 = 2 + 1 = 3 blocks.

   There are 1000 distinct values of DEPARTMENTCODE, so the average number of records for each value is: (N/1000) = (30000/1000) = 30 records.

   The 30 records are clustered in ceiling (30/bfr) = ceiling (30/4) = 8 blocks.

   Hence, total block accesses needed on average to retrieve all the records with a given DEPARTMENTCODE = x + 8 = 2 + 8 = 10 blocks.