COMP2411 Fall 2023 Class Exercise 6

Student Name: _____

Student ID: _____

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 in bytes.

B. Calculate the blocking factor and the number of file blocks assuming an unspanned block 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.
   b. Calculate the number of index entries and the number of index blocks.
   c. Calculate the number of block accesses needed to search for and retrieve all records in the file having a specific SSN value using the primary index.

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.
   b. Calculate the number of index entries and the number of index blocks.
   c. Calculate the number of block accesses needed to search for and retrieve all records in the file having a specific SSN value using the secondary index.

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 reservation (i.e., 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.
   b. Calculate the number of index entries and the number of index blocks.
   c. Calculate the number of file blocks after constructing the clustering index with block reservation.

Answers: *(floor(X) is the largest integer less than or equal to X; ceiling(X) is the smallest integer greater than or equal to X)*

A. Record size (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.

c. I/O cost = $\log_2 b_i$ (binary search on the index) + 1 (data access) = ceiling ($\log_2 221$)+1 = 8+1 = 9 block accesses.

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.

c. I/O cost = $\log_2 b_i$ (binary search on the index) + 1 (data access) = ceiling ($\log_2 883$)+1 = 10+1 = 11 block accesses.

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.

c. Number of records per DEPARTMENTCODE value = 30,000/1000 = 30 records.
Number of file blocks per DEPARTMENTCODE value = ceiling (30/bfr) = ceiling (30/4) = 8 blocks.
Total number of file blocks = 8 * 1000 = 8000 blocks.