INFS2200/7903 – Relational Database Systems

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

Tutorial 4

Question 1 Consider a file which has N = 20000 STUDENT records. Each record has the following fields:

NAME (30 bytes), SSN (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes), BIRTHDATE (8 bytes), SEX (1 byte), MAJORDEPTCODE (4 bytes), MINORDEPTCODE (4 bytes), CLASSCODE (4 bytes), and DEGREEPROGRAM (3 bytes)

Now suppose that only 80% of the STUDENT records have a value for PHONE, 85% for MAJORDEPTCODE, 15% for MINORDEPTCODE, and 90% for DEGREEPROGRAM, and we use a variable-length record file. Each record has a 1-byte field type for each field occurring in the record, plus a 1-byte deletion marker and a 1-byte end-of-record marker. Suppose we use a spanned record organization, where each block is of size 512 bytes and has a 5-byte pointer to the next block (this space is not used for record storage).

A. Calculate the average record length R in bytes.

B. Calculate the number of blocks b needed for the file.

Question 2 Suppose that a disk has the following parameters:

- Seek time (s) = 20 msec
- Rotational delay (rd) = 10 msec
- Block transfer time (btt) = 1 msec
- Block size (B) = 2400 bytes

An EMPLOYEE file has the following fields:

SSN (9 bytes), LASTNAME (20 bytes), FIRSTNAME (20 bytes), MIDDLE INIT (1 byte), BIRTHDATE (10 bytes), ADDRESS (35 bytes), PHONE (12 bytes), SUPERVISORSSN (9 bytes), DEPARTMENT (4 bytes), JOBCODE (4 bytes), deletion marker (1 byte).

The EMPLOYEE file has N = 30000 records, fixed-length format, and unspanned blocking. Write down appropriate formulas and calculate the following values for the above EMPLOYEE file:

A. Calculate the record size R (including the deletion marker), the blocking factor bfr, and the number of disk blocks b needed for the file.

B. Calculate the wasted space in each disk block because of the unspanned organization.

C. Calculate the transfer rate tr.

D. Calculate the average number of block accesses needed to search for an arbitrary record in the file, using linear search.

E. Calculate the average time in msec needed to search for an arbitrary record in the file, using linear search, if the file blocks are stored on consecutive disk blocks.

F. Calculate the average time in msec needed to search for an arbitrary record in the file, using linear search, if the file blocks are not stored on consecutive disk blocks.

G. Assume that the records are ordered via some key field. Calculate the average number of block accesses and the average time needed to search for an arbitrary record in the file, using binary search.

Answers for Question 1 are given below:

A. Assuming that every field has a 1-byte field type, and the fields not mentioned above (NAME, SSN, ADDRESS, BIRTHDATE, SEX, and CLASSCODE) have values in every record, we need the following number of bytes for these fields in each record, plus 1 byte for the deletion marker, and 1 byte for the end-of-record marker:

R fixed = (30+1) + (9+1) + (40+1) + (8+1) + (1+1) + (4+1) + 1 + 1 = 100 bytes.

For the fields (PHONE, MAJORDEPTCODE, MINORDEPTCODE, and DEGREEPROGRAM), the average number of bytes per record is:

R variable = ((9+1)*0.8) + ((4+1)*0.85) + ((4+1)*0.15) + ((3+1)*0.9)
= 8 + 4.25 + 0.75 + 3.6 = 16.6 bytes.

The average record size R = R fixed + R variable = 100 + 16.6 = 116.6 bytes.

B. Using a spanned record organization with a 5-byte pointer at the end of each block, the bytes available in each block are (B - 5) = (512 - 5) = 507 bytes.

Total bytes needed for the whole file = N * R = 20000 * 116.6 = 2332000 bytes

The number of blocks needed for the whole file is b = ceiling ((N * R) / (B - 5)) = ceiling (2332000 / 507) = 4600 blocks.

Answers for Question 2 are given below:

A. R = (9 + 20 + 20 + 1 + 10 + 35 + 12 + 9 + 4 + 4) + 1 = 125 bytes.

bfr = floor (B / R) = floor (2400 / 125) = 19 records per block.

b = ceiling (N / bfr) = ceiling (30000 / 19) = 1579 blocks.

B. Wasted space per block = B - (R * bfr) = 2400 - (125 * 19) = 25 bytes.

C. Transfer rate tr = B / btt = 2400 / 1 = 2400 bytes/msec.

D. For linear search we have the following cases:

1) Search on key field: half the file blocks are searched on average b/2 = 789.5 blocks.

2) Search on non-key field: all file blocks must be searched b = 1579 blocks.

E. If the blocks are stored consecutively, the time to read n consecutive blocks = s + rd + (n * btt)

1) If n = b/2, then time = 20 + 10 + ((1579/2) * 1) = 819.5 msec.

2) If n = b, then time = 20 + 10 + (1579 * 1)) = 1609 msec.

F. If the blocks are scattered over the disk, a seek and a rotation are needed for each block, so the time to search n blocks = n * (s + rd + btt)

1) If n = b/2, then time = (1579/2) * (20+10+1) = 24474.5 msec = 24.475 sec.

2) If n = b, then time = 1579 * (20+10+1) = 48949 msec = 48.949 sec.

G. For binary search:

Average number of block accesses = ceiling ($\log_2 b$) = ceiling ($\log_2 1579$) = 11 blocks.

Time to search for a record is estimated as = ceiling ($\log_2 b$) * (s + rd + btt) = ceiling ($\log_2 1579$) * (20+10+1) = 11 * 31 = 341 msec = 0.341 sec.