Exercises

(Course: Database Management Systems)

Chapter 1

Disk Storage, Basic File Structures and Hashing

1. Exercise 17.27 in the text book ("Fundamentals of Database Systems- 6th Edition", Elmasri et al.)

Consider a disk with the following characteristics (these are not parameters of any particular disk unit): block size B = 512 bytes; interblock gap size G = 128 bytes; number of blocks per track = 20; number of tracks per surface = 400. A disk pack consists of 15 double-sided disks.

a) What is the total capacity of a track, and what is its useful capacity (excluding interblock gaps)?
b) How many cylinders are there?
c) What are the total capacity and the useful capacity of a cylinder?
d) What are the total capacity and the useful capacity of a disk pack?
e) Suppose that the disk drive rotates the disk pack at a speed of 2400 rpm (revolutions per minute); what are the transfer rate (tr) in bytes/msec and the block transfer time (btt) in msec? What is the average rotational delay (rd) in msec? What is the bulk transfer rate? (See Appendix B.)
f) Suppose that the average seek time is 30 msec. How much time does it take (on the average) in msec to locate and transfer a single block, given its block address?
g) Calculate the average time it would take to transfer 20 random blocks, and compare this with the time it would take to transfer 20 consecutive blocks using double buffering to save seek time and rotational delay.

2. Exercise 17.28 in the text book ("Fundamentals of Database Systems- 6th Edition", Elmasri et al.)

A file has r = 20,000 STUDENT records of *fixed length*. Each record has the following fields: Name (30 bytes), Ssn (9 bytes), Address (40 bytes), PHONE (10 bytes), Birth_date (8 bytes), Sex (1 byte), Major_dept_code (4 bytes), Minor_dept_code (4 bytes), Class_code (4 bytes, integer), and Degree_program (3 bytes). An additional byte is used as a deletion marker. The file is stored on the disk whose parameters are given in Exercise 17.27.

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) Calculate the average time it takes to find a record by doing a linear search on the file if (i) the file blocks are stored contiguously, and double buffering is used; (ii) the file blocks are not stored contiguously.
d) Assume that the file is ordered by Ssn; by doing a binary search, calculate the time it takes to search for a record given its Ssn value.

3. Exercise 17.31 in the text book ("Fundamentals of Database Systems- 6th Edition", Elmasri et al.)

A PARTS file with Part# as the hash key includes records with the following Part# values: 2369, 3760, 4692, 4871, 5659, 1821, 1074, 7115, 1620, 2428, 3943, 4750, 6975, 4981, and 9208. The file uses eight buckets, numbered 0 to 7. Each bucket is one disk block and holds two records. Load these records into the file in the given order, using the hash function h(K) = K mod 8. Calculate the average number of block accesses for a random retrieval on Part#.

4. Exercise 17.32 in the text book ("Fundamentals of Database Systems- 6th Edition", Elmasri et al.)

Load the records of Exercise 17.31 into expandable hash files based on *extendible hashing*. Show the structure of the directory at each step, and the global and local depths. Use the hash function h(K) = K mod 128.

5. Exercise 17.33 in the text book ("Fundamentals of Database Systems- 6th Edition", Elmasri et al.)

Load the records of Exercise 17.31 into an expandable hash file, using *linear hashing*. Start with a single disk block, using the hash function h_0 = K mod 20, and show how the file grows and how the hash functions change as the records are inserted. Assume that blocks are split whenever an overflow occurs, and show the value of n at each stage.