COMP 3311: Database Management Systems

Tutorial 5
Storage and File Structure

Exercise 1. Which file organization, heap, sequential or hash, would you choose for a file where the most frequent operations are as follows?

a) Search for records based on a range of field values.

b) Perform inserts and scans, where the order of records does not matter.

c) Search for a record based on a particular field value.

Exercise 2. A file has 10,000 student records of fixed-length. Each record has the following fields: studentId (9 bytes), name (30 bytes), address (40 bytes), phone (8 bytes), birthdate (8 bytes), gender (1 byte) and degreeProgram (3 bytes). An additional byte is used as a deletion marker.

a) What is the size of a record in bytes

b) What is the blocking factor bf_r if the page size is 4096 bytes?

c) How many pages are required to store the file:
   i. if a sequential file organization is used?

   ii. if a heap file organization is used?

   iii. if a hash file organization is used (assuming 100% page occupancy)?

   iv. if a hash file organization is used (assuming 100% page occupancy)?

   v. if a hash file organization is used (assuming 80% page occupancy)?

d) Consider the query “Find a student record given a particular student id”. Assuming that a record with the student id exists in the file, calculate the cost, in page I/Os, to answer this query:
   i. if a sequential file organization is used?

   ii. if a heap file organization is used?

   iii. if a hash file organization is used?

Name: _____ **Student#:** _____ **Date:** _____

COMP 3311: Database Management Systems

Tutorial 5
Storage and File Structure

Exercise 3: Assume that a school keeps a file with the records of its students:

Student(studentId: 4 bytes, name: 10 bytes, deptId: 4 bytes)

where deptId is the department id to which a student belongs. There are 10,000 student records and 50 departments. A page is 128 bytes. The data file is sorted sequentially on studentId.

a) What is the size of a record in bytes?

b) How many records can fit on each page?

c) How many pages are needed to store these student records?

d) Given this data file, what is the cost, in page I/Os, to find a particular student by studentId?

You must upload this completed exercise sheet to Canvas by **11 p.m. today.**