Introduction to Database Systems (CS3700)
Problem Set
Topic: File Storage & Indexing

1. A file has r = 16,500 STUDENT records, each record has the following fields: ID# (3 bytes), NAME (4 bytes), ADDRESS (14 bytes), and PROGRAM (6 bytes). This file is stored on the disk with block size B = 1024 bytes. Find the number of disk blocks required to store this file on the disk.

2. In Question 1, suppose each of the records has the following fields: ID# (3 bytes), NAME (4 bytes), and X, where size of X is fixed, but unknown. It is known that size of X lies between 12 and 35 bytes (inclusive). Find the minimum and maximum disk blocks required to store this file.

3. Consider the following sequence of records to be inserted into a dynamic hash file organized using the linear hashing method: 5, 7, 12, 11, 9, 13, 14, 20 and 23. Bucket capacity is 2. Consider the hash function $h_i(x) = x \bmod 2^i M$, where $M = 2$, initially. Show how file structure evolves as records are inserted.

4. Consider a file of 82,768 records. Each record is 75 bytes long and it includes a key field of size 4 bytes. The file is stored in a file system with block size 1024 bytes, and the size of a block pointer is 6 bytes. Find the maximum number of block accesses required to obtain a record given the value of the key field for the following settings: a) without using primary index b) using primary index, and c) using multi-level primary index. Also, find the blocks required by the primary index and each level of the multi-level primary index structure.

5. For file of data records mentioned in Question 4, suppose there is a candidate key field of length 18 bytes. Assume that a record pointer takes 7 bytes. Design a multi-level secondary index structure for accessing the records using this field. For each level, specify the number of index entries and blocks. How many block accesses are required to get the record with this index?

6. Consider that extendible hashing is being used on a database file that contains records with the following search key values: 1, 2, 5, 3, 11, 7, 19, 23, 29, 31. Construct the extendible hash file directory structure for this file and show how it evolves if each bucket can hold three records and hash function is $h(x) = x \bmod 7$.