

Tutorial CS3402
File Organization & Indexing

1. Consider the extendible hashing scheme. Recall that we consider the binary representation of the hash values for assigning items to buckets. Suppose that the *global depth* is $d=2$ and that each bucket can hold 2 items. Initially, there are two (empty) buckets b_1 and b_2 that have a *local depth* $d'=1$, where
- b_1 is the bucket for records whose binary hash values start with 0, and
 - b_2 is the bucket for records whose binary hash values start with 1.

Below is the initial state of the directory:

Directory entry ($d=2$)	Pointer to Bucket
00	b_1
01	b_1
10	b_2
11	b_2

- a) What is the content of buckets b_1 and b_2 after inserting items with hash values 0100, 0000, 1000, 1100? Are there any changes to the pointers in the directory?
- b) Describe the steps that take place if we next insert an item with hash value 1110.
- c) After performing the insertions mentioned in a) and b), we next insert items with hash values 1101 and 1111. Describe how these insertions are handled. What is the final state of the directory and the contents of the buckets?

2. Suppose that you have formatted your disk with a block size of 1024 bytes and assume that we have 40,000 STAFF records of fixed length. A block pointer is $P=6$ bytes long, and a record pointer is $Pr = 7$ bytes long. Each STAFF record has the following fields: Name (20 bytes), Ssn (9 bytes), Department (9 bytes), Address (40 bytes), Phone (10 bytes), Salary (8 bytes), Position_code (4 bytes), Email (32 bytes), and Job_description (200 bytes). An additional byte is used as a deletion marker. Suppose that the file is ordered by the key field Ssn and we want to construct a primary index on Ssn.
- Calculate the blocking factor (bfr) and the number of file blocks b needed to store the STAFF records. Assume that records are stored unspanned. How much space remains unused per block?
 - Calculate the index blocking factor bfr_i .
 - Assume that we only have a single-level index. Calculate the number of index entries and the number of index blocks.
 - Now suppose that we want to make the index a multilevel index. How many levels are needed and what is the total number of blocks required by the multilevel index?
 - Consider the multilevel index from question (d). What is the number of block accesses needed to search for and retrieve a record from the file given its Ssn value?