CS3223 Tutorial 2:   Hash-based Indexing

Questions to be discussed: 1, 2, 3 & 4

1. Consider the Linear Hashing index shown below

level = 0, $N_0 = 1$, next=0

0	2*		

Linear Hashing index with one data bucket

where each data bucket can store 3 data entries. Show the Linear Hashing index after inserting the following sequence of data entries:

$$3*, \ 5*, \ 7*, \ 11*, \ 17*, \ 19*, \ 23*, \ 0*, \ 4*, \ 29*, \ 33*, \ 13*, \ 6*$$

Assume that a bucket split is triggered whenever some bucket overflows.

2. (Exercise 11.9, R&G) Consider the snapshot of the Linear Hashing index shown below.

level = 0, $N_0 = 4$, next=0

00	64*	44*		
01	9*	25*	5*	
10	10*			
11	31*	15*	7*	3*

Assume that a bucket split is triggered whenever some bucket overflows.

   (a) What is the maximum number of data entries that can be inserted (given the best possible distribution of keys) before you have to split a bucket? Explain very briefly.

   (b) Show the file after inserting a *single* record whose insertion causes a bucket split.

   (c) Consider the original Linear Hashing index shown.
   (i) What is the minimum number of record insertions that will cause a split of all four buckets? Explain very briefly.
   (ii) What is the value of `next` after making these insertions?
   (iii) What can you say about the number of pages in the fourth bucket shown after this series of record insertions?

3. Consider the Extendible Hashing index shown below

Extendible Hashing index with one data bucket

where each data bucket can store 3 data entries. Show the Extendible Hashing index after inserting the following sequence of data entries:

$$3*, \ 5*, \ 7*, \ 11*, \ 17*, \ 19*, \ 23*, \ 0*, \ 4*, \ 29*, \ 33*, \ 13 * \ 6*$$

Note that k* denotes a data entry having a hashed value of k.

4. (Adapted from Exercise 11.1, R&G) Consider the Extendible Hashing index shown below which was created by inserting a sequence of 13 data entries without any deletions. Which bucket could be the last bucket that was split by the insertions?

Bucket	Base 10	Base 2
B	1	0000001
E	4	0000100
B	5	0000101
D	7	0000111
C	10	0001010
E	12	0001100
D	15	0001111
A	16	0010000
E	20	0010100
B	21	0010101
E	36	0100100
D	51	0110011
A	64	1000000