Homework4

Deadline: Nov 1st, 2015

Q1. B+ Structure

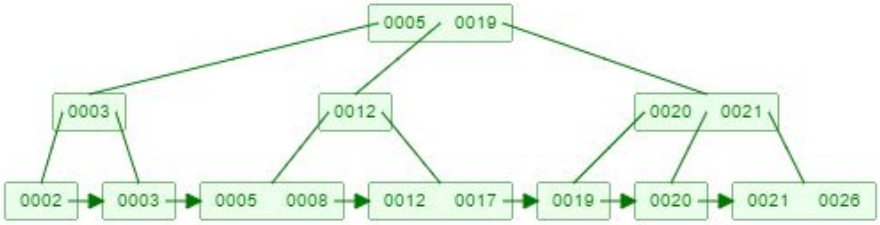
1. How many levels you need for representing 100000 entries using B+ structure with 25-entries per block.

Assumptions for questions 2,3,4:

- The branching factor of the given B+ tree is 3 (meaning the maximum number of entries in a node is 2)
- Value of keys ranges from 1 to 32

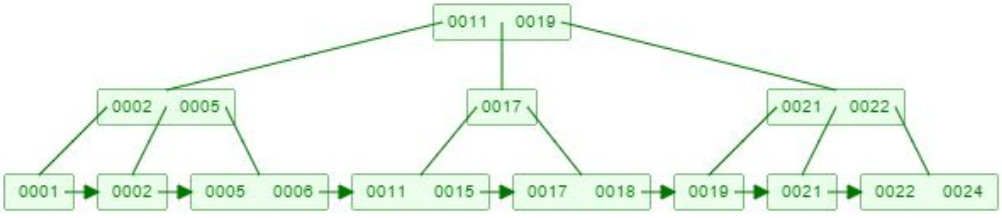
2. Given the B+ tree in Fig. 1, draw the B+ tree after each insertion of following keys (so five drawings in total) : 1, 4, 18, 13, 7

Fig 1:

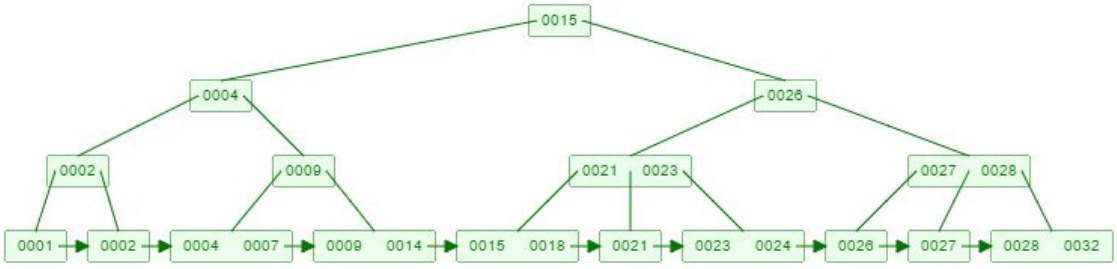


3. Given the B+ tree in Fig. 2, find all the keys which can increase the height of the tree from 3 to 4 with only one insert operation (Please see the assumption section for the range of values).

Fig 2:



4. In this question, you will learn how to delete keys from B+. You can follow the tutorial in the following link: http://www.cburch.com/cs/340/reading/btree/
For the given B+ tree in Fig. 3, perform deletion of the key values 24, 32 and 28, 27, 26. Draw the B+ tree after each deletion.
Fig 3:

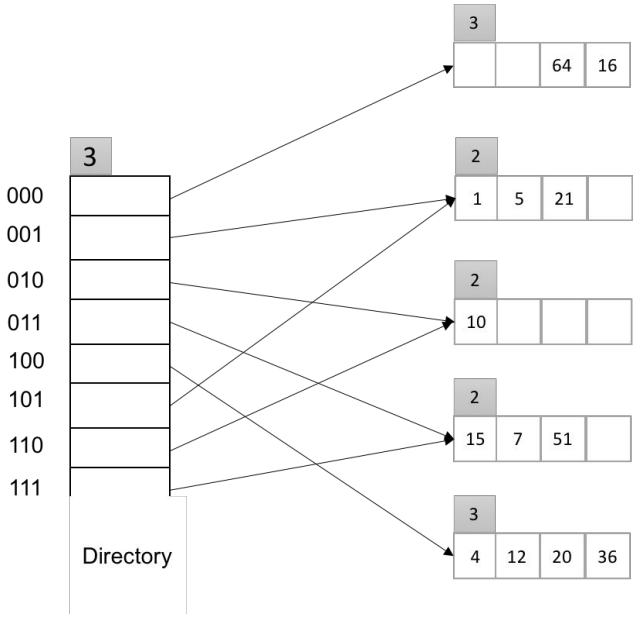


Q2. Extendable Hashing

Fig 4. is an extendible hashing structure with 4-slot buckets. The hash function in this example is simply the binary representation of numbers.

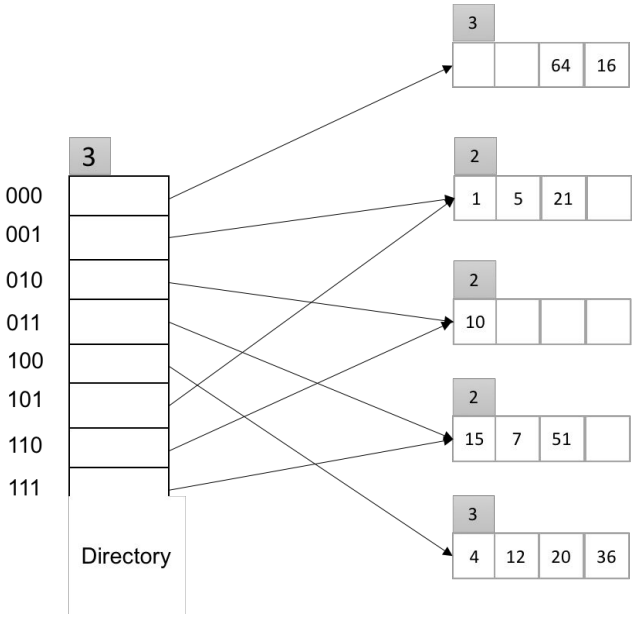
1. Insert the following keys into the Fig. 4 and show the structure after each insertion: 32, 13, 28

Fig 4:



2. For Fig. 5, show two deletions that would lead to decrease of global depth by one. Draw the hashing structure after each deletion.

Fig. 5:



Q3. Optimizing queries by using indexes

Given the following tables, which store information of users and ads of an online advertisement website.

- Users(username, password, name, dob, gender, email, last_logout)
- Ads(ad_id, username, title, price, description, post_date, last_edited, img_url, category_id, region_id, status)

For the following queries, you need to list all the attributes that are needed to optimize each query. You also need to point out a suitable indexing technique for each optimization (remember that choosing only one technique). We consider clustered B+, unclustered B+, hashing techniques in this question.

1. We want to draw a histogram of year-of-birth for users that were born in or after 1970. Write a query that shows years starting 1970 and the number of users born in that year.
2. How many ads were posted by the username 'lhartj'?
3. How many ads were posted by the username ‘lhartj’ and price less than 500$?

Note:

- DateTime format in this database is UNIX time.
- You may assume a uniform distribution of values if that affects your answer.