CSE441 DATABASE SYSTEMS

ASSIGNMENT 2

PART 1 (60%)

Implement B+ tree to handle the following type of queries

1. INSERT x - insert x into the B+tree
2. FIND x - print YES if x is already inserted , else NO
3. COUNT x - print number of occurrences of x in B+tree
4. RANGE x y - print number of elements in range x to y (both x and y included)

Constraints

1. -10^9 <= x <= 10^9 and -10^9 <= y <= 10^9.
2. Number of queries will be less than 10^6.

Marks will be based on the efficiency of the code.

Input: Filename,M,B (explained in Note section)

1. Inputs must be taken as command line arguments
2. Each line in the filename consists one of the above mentioned query.

Output:

Print output of each command in a separate line

PART 2 (40%)

Implement Linear Hashing (paper explained in class) to handle duplicate elimination(explained in Output section).

Input: Filename,M,B (explained in Note section)

1. Inputs must be taken as command line arguments
2. Each line in the filename consists one of single integer(x) (-10^9 <= x <= 10^9).

Output:

After reading every line (call it record), If record is not inserted into data structure, print it and insert it into data structure.

Note:

1. M denotes the number of buffers and B denotes the buffer size (M>=2 and M*B<=10^6)
2. Out of the M buffers, M-1 Buffers will be used as input buffers (which will hold the records from the input file), 1 buffer will be used as output buffer (holds the distinct

records). If the output buffer gets filled, it should be flushed to the output. If the input buffers get empty, next chunk of records should be read from the input file.

Upload Format :

1. Create a folder with your roll number.
2. Put all the code files & ReadMe.txt into the folder created in 1.
3. zip the folder and name the archive as rollnumber_assign2.zip

Deadline :

6th March 2017 09:00PM.

Any kind of plagiarism will be severely punished.

You are not allowed to use inbuilt data structures to handle these queries.