Carnegie Mellon University
15-415 Database Applications
Spring 2012, Faloutsos
Assignment 4: Query Processing
Lead TA: Bin Fu (binf@andrew.cmu.edu)

Solutions

Q1.1

Q1.2

(a)

(b)
Q1.3

(a)

(b)

Note: sometimes it is good to re-distribute multiple keys from sibling, but not quite necessary here.

Q2.1

Depend on whether you use lower-bits or higher-bits on hashing, the solution may change slightly. Notice that we start from one bucket.

000 -> a

001 -> bhi

01 -> df

1 -> ceg

Q2.2

00 -> abh -> i (overflow)

01 -> df

10 -> cg

11 -> e
Q3.1

Number of passes: \( \lceil \log_2 N \rceil + 1 = 21 \)

Time: \( 2N (\lceil \log_2 N \rceil + 1) t = 6.72 \times 10^5 \) seconds.

Q3.2

\( N' = \lfloor N/B \rfloor = 3907 \)

Number of passes: \( \lceil \log_{B^{-1}} N' \rceil + 1 = 3 \)

Time: \( 2N (\lfloor \log_{B^{-1}} N' \rfloor + 1) t = 9.6 \times 10^4 \) seconds.

Q3.3

\( F = \lceil B/b \rceil - 1 = 7 \)

Number of passes: \( \lfloor \log_F N' \rfloor + 1 = 6 \)

Or, if you have \( N' = \lfloor N/2B \rfloor = 1954 \), \( \lfloor \log_F N' \rfloor + 1 = 5 \)

Time to read/write each page now: \( (16 + 31) / 32 = 47/32 \) ms

Time: \( 2N (\lfloor \log_{B^{-1}} N' \rfloor + 1) (47/32) ms = 1.76 \times 10^4 \) seconds (or, \( 1.47 \times 10^4 \) seconds).