Date:
Name 1:
Name 2:

Assume you have a relation R(a,b,c). Suppose the blocks can hold either 10 tuples (heap) or 100 search keys (b+tree index). Nodes of the index are 70% full. The relation contains 1 million records. The values of a are expected to be >=0 and < 10^{6}.

Consider the following two queries: 70 fanout

a) $\sigma_{a=5}R$ 1 matching record

b) $\sigma_{a>=10,000 \text{ and } a < 20,000}R$ ⇒ 10,000 matching rec.

Every value of a between 1 and 10^6

$1/10^6 = 10^{-6}$ selec..

$\frac{10^4}{10^6} = 10^{-2}$ selec.

1. Determine for each query:
   - the number of expected matching records

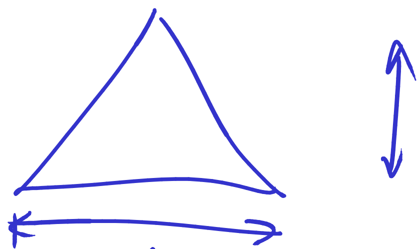
2. Determine, for each of the structures below:
   i) Expected number of leaf blocks of each index
   ii) The expected height of each index
   iii) The average number of disk I/Os needed to answer each query

Assume that nothing is in memory initially, and that the search key is the primary key of the table.

a) The B+tree is dense and the heap is unsorted
b) The B+tree is a sparse.

Redo assuming that a is not a primary key, values of a vary from >=0 and <= 10^5

a) B+ tree dense



h

10^6 index records

70 per block

$$h = \left\lceil \log_{70} \left\lceil \frac{10^6}{\dots} \right\rceil \right\rceil = \left\lceil \frac{\log_{10}(10^6)}{\log_{10} 70} \right\rceil$$

$$h = 4$$

$$\text{Block index} = \frac{10^6}{70} = \frac{1}{7} \cdot 10^5 \text{ blocks.}$$

For $\sigma_{a=3} R \Rightarrow h+1 = 5$

For $\sigma_{a>=10k \text{ and } a<20k} R$

h-1

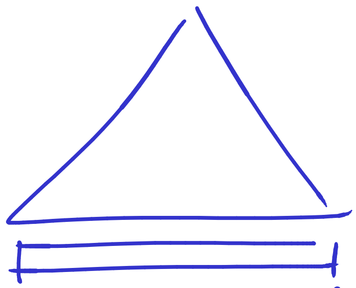
+ matching tuples

+ blocks

$$\text{blocks} = \left\lceil \frac{10^4}{70} \right\rceil = 143$$

$$\text{Cost} = 3 + 143 + 10{,}000 = 10{,}146$$

b) Tree is sparse.

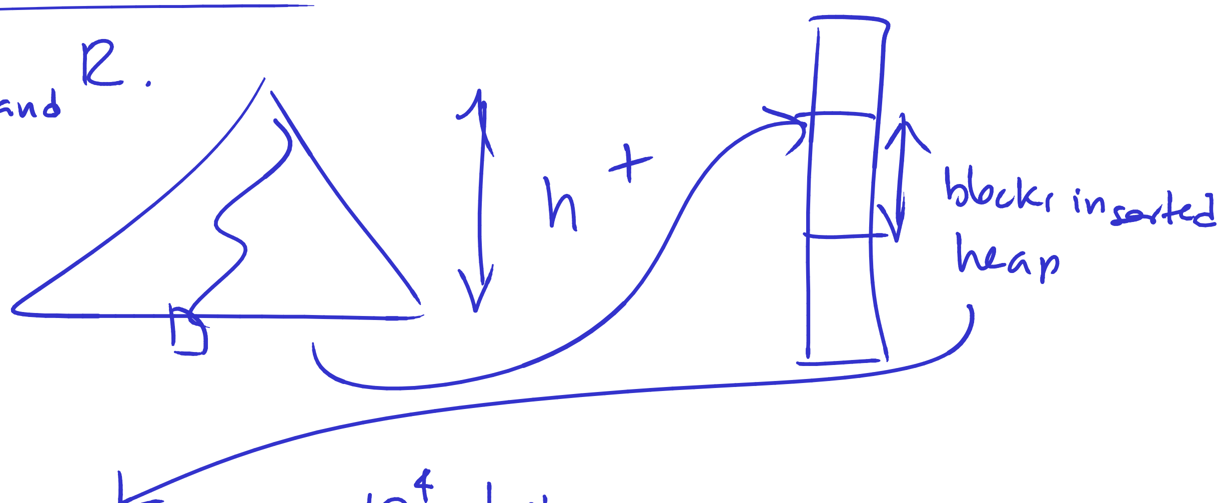


\# index records $= B(R) = 10^5$

$$h = \left\lceil \log_{70} 10^5 \right\rceil = \left\lceil \frac{\log_{10}(10^5)}{\log_{10} 70} \right\rceil = \lceil 2.7 \rceil = 3$$

$$\sigma_{a=5} R \Rightarrow h+1 = 4.$$

$\sigma_{a \geq 10k \text{ and } a < 20k} R.$



$$\text{blocks in heap} = \frac{10^4 \text{ tuples}}{10 \text{ tuples/block}} = 1{,}000 \text{ block.}$$

Cost $= 1003$ blocks.

$\sigma_{a=s} R$ a not primary key.

Indexes will be the same

matching tuples $\underline{\underline{10}}$ (each value is repeated 10 times).

$$Cost = h + 10 = 14$$

(10 index records found in one block of index)

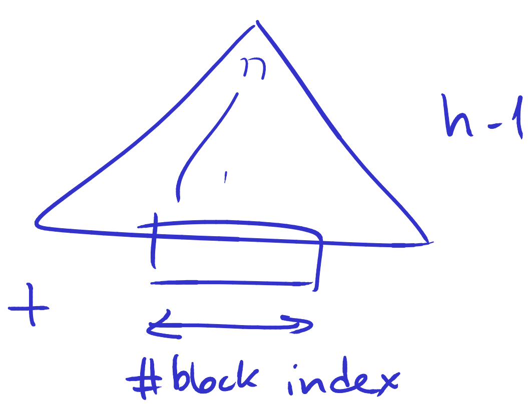
Sparse index 10 tuples can be found in one block of heap (most likely 2 if they don't start in the beginning of block

$$Cost = h + 1 \text{ or } h+2$$

$$= 4 \text{ or } 5.$$

$\sigma_{a>=10k \text{ and } a<20k} R$. $\Rightarrow 10^5$ tuples would match.

Dense:



$$Cost = h-1 + \frac{10^5}{70} + 10^5$$

$$= 3 + \frac{10^5}{70} + 10^5$$

$$\#\text{block index} = \frac{10^5 \text{ tuples}}{70}$$

Sparse index

h + # blocks in heap.

$$\text{Cost} = h + \frac{10^5}{10} = 3. + 10^4 = 10{,}003 \text{ blocks.}$$