

Date:

Name 1:

Name 2:

Assume we have a relation  $R(\underline{a}, b)$ .

- This relation contains 1 million tuples. Its primary key is  $a$ .
- Each block can hold at most 20 tuples.
- The values of  $b$  are distributed between the values of 1 and 1000., and any value of  $b$  is equally likely to appear (uniform distribution).
- There are three indexes on  $R$ . A sparse index on  $R(a)$ , and dense indexes on  $R(a)$  and  $R(b)$ .
- Each index block can hold at most 150 index records. However, on average, only 100 index records are placed in each block.
- Values of  $a$  vary from 1 to 1,000,000

Compute the cost (number of blocks read) of the following queries.

a)  $\sigma_{a=5}R$

b)  $\sigma_{a > 10 \text{ and } a \leq 100} R$

c)  $\sigma_{b=5}R$

Let us first calculate # of blocks in heap:

$$B(R) = \frac{10^6}{20}$$

## Sparse Index on R.

We will have  $B(R)$  records in index.

$$h = \left\lceil \log_{100} B(R) \right\rceil = \left\lceil \frac{\log_{10} \frac{1}{2} \cdot 10^5}{\log_{100}} \right\rceil = \left\lceil \frac{4.7}{2} \right\rceil$$

$\uparrow$  average value

$$= 3$$

Dense Index on R (same for index on a or b)  
# index records in index  $B(R)$

$$h = \left\lceil \log_{100} |R| \right\rceil = \left\lceil \log_{100} 10^6 \right\rceil = 3$$

Both indexes have same height

a)  $\sigma_{a=5} R$      a is primary key

Values of  $a \in [1, 10^6]$

# values =  $10^6$

$\Rightarrow a=5$  exists (every value between 1 and  $10^6$  exists).

For sparse and dense

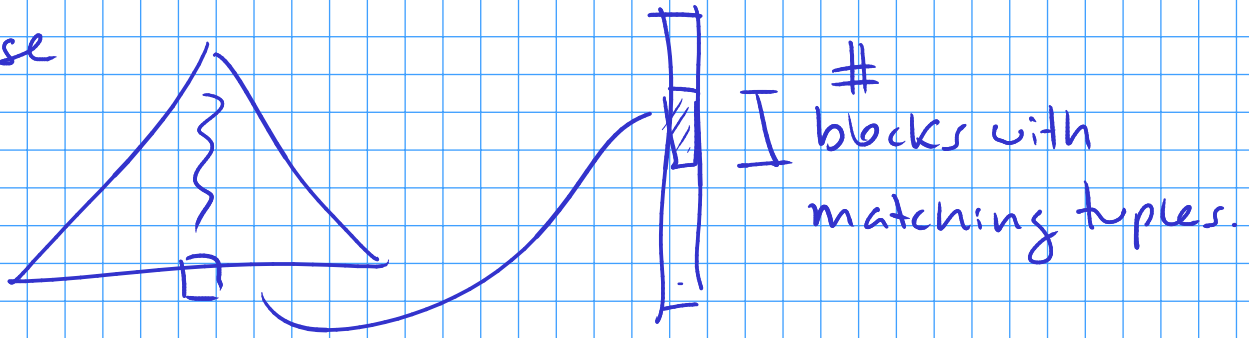
# matching tuples = 1.

$\Rightarrow h+1 = 4$ .

b)  $\sigma_{a > 10 \text{ and } a \leq 100} R$

Sparse

$h$



# blocks in heap with matching tuples.

$$= \left\lceil \frac{\text{\# matching tuples}}{\text{\# tuples per block in heap}} \right\rceil$$

# matching tuples : 90 (every value exists once)

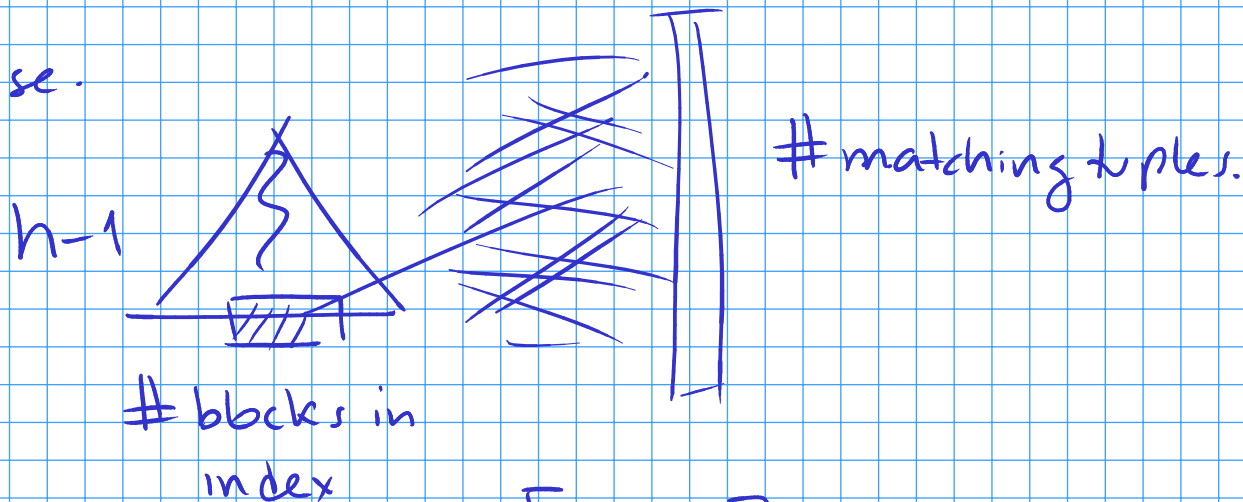
# blocks with matching tuples

$$= \left\lceil \frac{90}{20} \right\rceil = 5$$

Total cost =  $h + \text{\# blocks in heap with matching tuples}$

$$= 3 + 5 = 8$$

Dense.



$$\# \text{ Blocks in index} = \left\lceil \frac{90}{100} \right\rceil = 1.$$

$$\# \text{ matching tuples} = 90$$

average records per index block

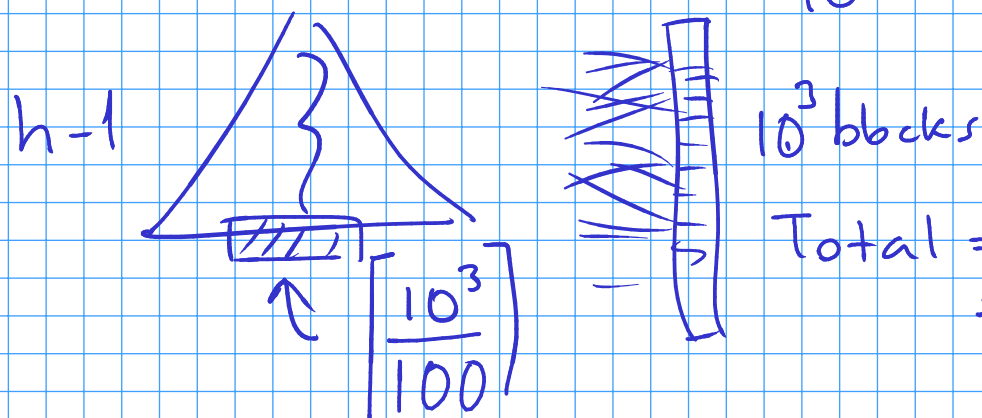
$$\text{Cost} = (\underbrace{3}_{h}-1) + 1 + 90 = 93 \text{ blocks.}$$

$\sigma_{b=s} R$

# matching tuples

There are  $10^6$  tuples, and 1000 different values.  
all equally likely

$$\Rightarrow \# \text{ matching tuples} = \frac{10^6}{10^3} = 10^3$$



$$\text{Total} = 2 + 10 + 10^3 = 1012 \text{ blocks.}$$