Relation R(A,B)

• R has 7,000 records

• 70 records per page/block

• B+ tree index on attribute A (n=4)

• The values that the attribute A takes in relation R are integers that are uniformly distributed in the range 1 – 200.

1. Assuming that the index on A is unclustered, estimate the cost needed to compute the query

**σA=18(R)**

1. What would be the cost estimate if the index were clustered?



Solution for (a)

We will use *A5*

br= noof blocks of relation R=7000/70=700 block

B+ on attribute A

n=4

Height(B+)=logn/2(keys)= log4/2(200)= log2(200)=7.6=8

Numbers in attribute A: 1, 2 , 3, 4, ……200

1 is a category

2 is a category

3 is a category

.

.

200 is a category

frequency of each category= 7000/200=35 records

category 18: A=18=35 records

* + - *Cost* = *HTi* + *number of records retrieved*
    - *Cost = 8 + 35=43 blocks*

Solution for (b)

We will use A4

* **A4** (primary index on nonkey, equality): Retrieve multiple records
  + Records will be on consecutive pages
  + *Cost* = *HTi* + number of pages containing retrieved records

Cost= 8+ number of pages containing retrieved records??

Category 18 has 35 records. We know that each block has 70 records. This means the 35 records can be stored in one block/page.

So

Cost= 8+1=9 blocks