Data System Questions (Q.pdf)

(a) Record Size = (30 + 9 + 9 + 40 + 10 + 8 + 1 + 4 + 4 + 1) B

Name | SSN | Department-code | Address | Phone | Birth-Date | Sex | Job code | Salary | DELETE-MARKER

= 116 Bytes

(b) Unspanned organization.

$$\text{Blocking factor } (bfr) = \left\lfloor \frac{\text{block size}}{\text{Record size}} \right\rfloor = \left\lfloor \frac{512}{116} \right\rfloor = 4$$

(How many records will fit in a block)

(floor due to unspanned).

$$\text{Number of file blocks} = \left\lceil \frac{30{,}000}{bfr} \right\rceil = \left\lceil \frac{30{,}000}{4} \right\rceil = 7500$$

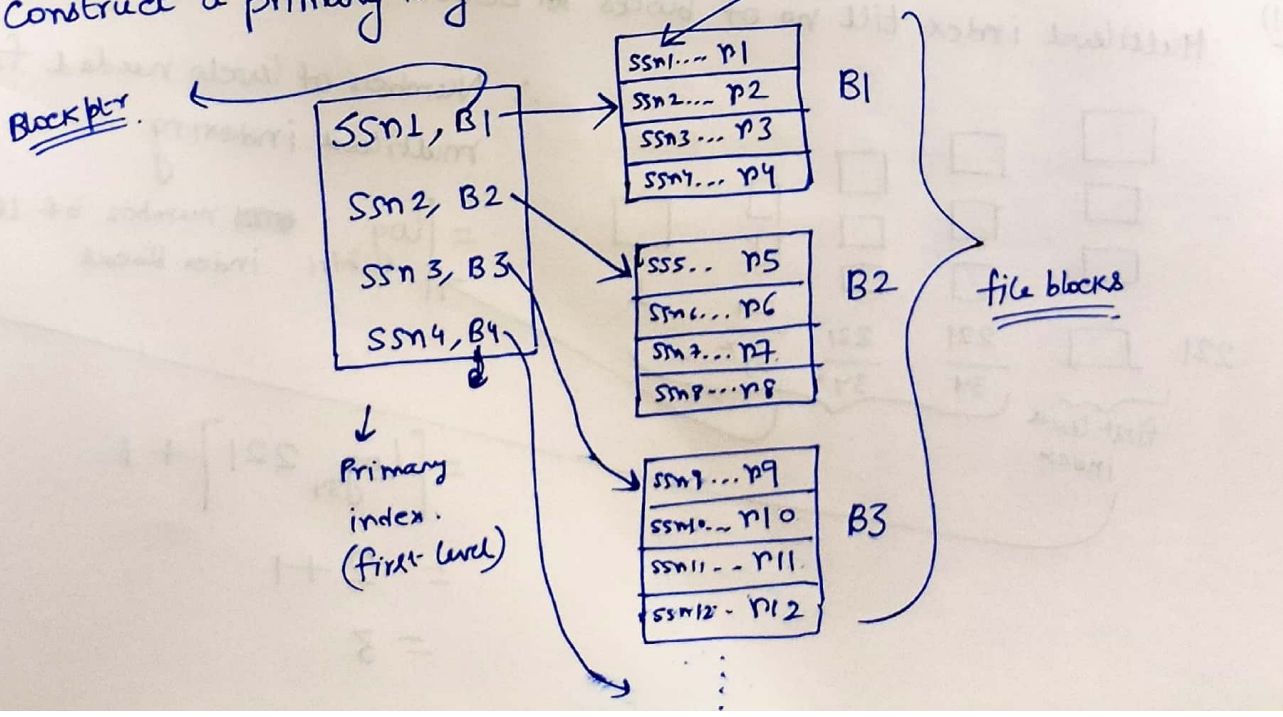
total no. of records / records/block

(c) File is ordered by key field Ssn.

Construct a primary key index on Ssn

ordered on key field SSN.

$r_i = \{SSN, name, dept, \ldots\}$



Length of Index Row = (Size of Ssn + Size of Block ptr)

$= (9+6)B$

$= 15B$

i) Index blocking factor $(bfr_i) = \left\lfloor \frac{\text{Block Size}}{\text{Length of Index Row}} \right\rfloor = \left\lfloor \frac{512}{15} \right\rfloor = 34$

ii) Number of first level index entries in primary index

= number of file blocks

= 7500 (obtained from (b))

Number of first level index blocks

$= \left\lceil \frac{\#\text{ of first level index entries}}{\text{Index blocking factor } (bfr_i)} \right\rceil$

$= \left\lceil \frac{7500}{34} \right\rceil = 221$

iii) Multilevel index till no of blocks in becomes 1.

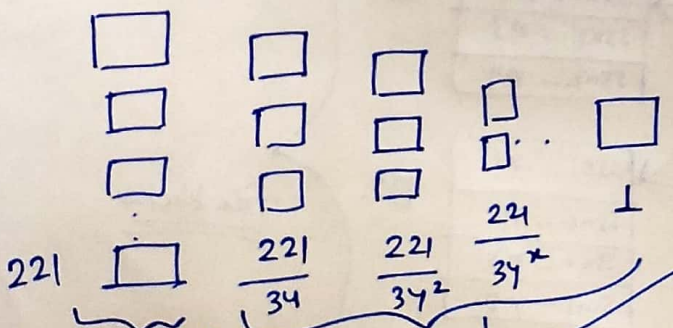
∴ Number of levels needed for multilevel indexing

$= \lceil \log_{bfr_i} (\text{number of 1st level index blocks}) \rceil + 1$

$= \lceil \log_{34} 221 \rceil + 1$

$= 2 + 1$

$= 3$



iv) First level requires = 221 blocks

2nd " " $= \left\lceil \frac{221}{34} \right\rceil = 7$ blocks

3rd " " $= \left\lceil \frac{7}{34} \right\rceil = 1$ block

∴ Total number of blocks required by multilevel index

$= (221 + 7 + 1) = 229$

v) Number of block accesses required to search for and retrieve a record from the file – given its Ssn value – using the primary index:-

Assuming no multilevel indexing.

Number of block accesses $= \lceil \log_2 \#\text{ first level blocks} \rceil + 1$

$= \lceil \log_2 221 \rceil + 1$

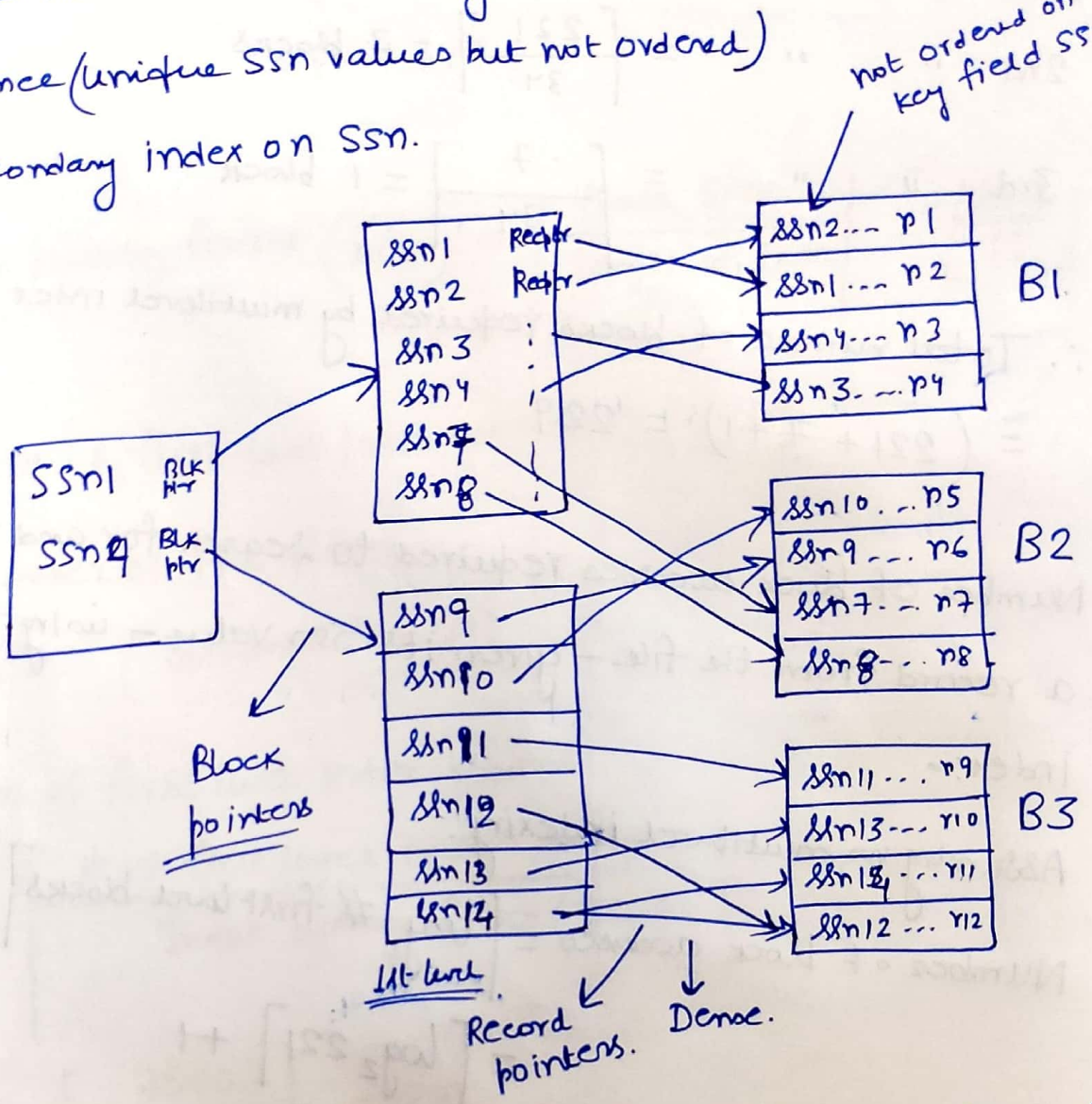
$= 8 + 1$

$= 9$

(d) File is not ordered on the key field Ssn.

Hence (unique Ssn values but not ordered)

Secondary index on Ssn.



Length of Index Row (in 1st Level) = (Length of Ssn + Length of Rec.ptr)

= (9+7)B

= 16B

* Length of Index Row (in Multilevel 2nd level onwards.) = (Length of Ssn + Length of Block ptr)

= (9+6)B

= 15B

i) Index blocking factor (for 1st level) $(bfr_i) = \left\lfloor \frac{\text{Block Size}}{\text{Length of Index Row}} \right\rfloor = \left\lfloor \frac{512}{16} \right\rfloor$

$= 32$

Blocking factor for multilevel (2nd level onwards) $= \left\lfloor \frac{512}{15} \right\rfloor = 34$

ii) Number of first level index entries

= Number of total records (∵ we have to include each and every record here since not sorted)

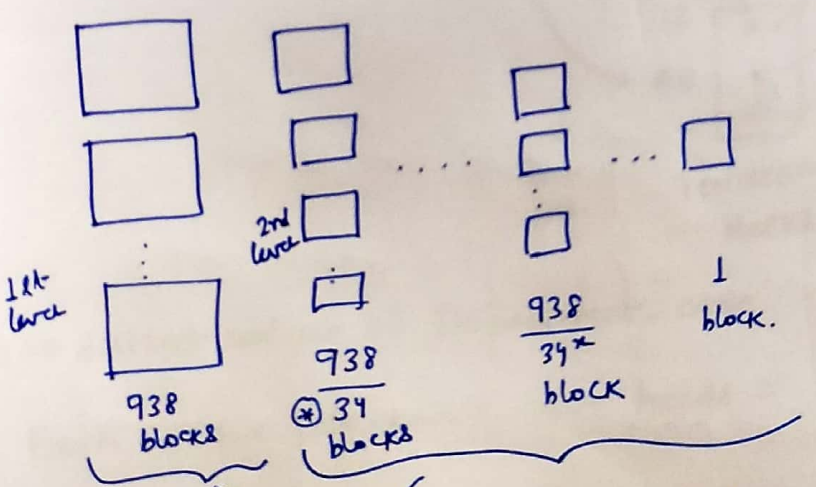
= 30,000

Number of blocks for 1st level Index $(n_I) = \left\lceil \frac{\#\text{ of first level entries}}{bfr_i} \right\rceil$

$= \left\lceil \frac{30,000}{32} \right\rceil$

$= 938$

iii)



∴ # of ~~blocks~~ levels needed for multilevel index

$= \lceil \log_{bfr_i \text{ for multilevel index}} (n_I) \rceil + 1$

$= \lceil \log_{34} (938) \rceil + 1 = 2 + 1 = 3$

iv) No. of blocks at first level = 938

No. ——— 2nd " = $\left\lceil \frac{938}{34} \right\rceil$ = 28

bfr for multilevel

——— 3rd " = $\left\lceil \frac{28}{34} \right\rceil$ = 1

∴ Total no. of blocks = (938 + 8 + 1) = 967

v) No. of block accesses required (Assuming no multilevel indexing)

= $\lceil \log_2 (\text{blocks at first level}) \rceil + 1$

= $\lceil \log_2 938 \rceil + 1$

= 10 + 1

= 11.

---

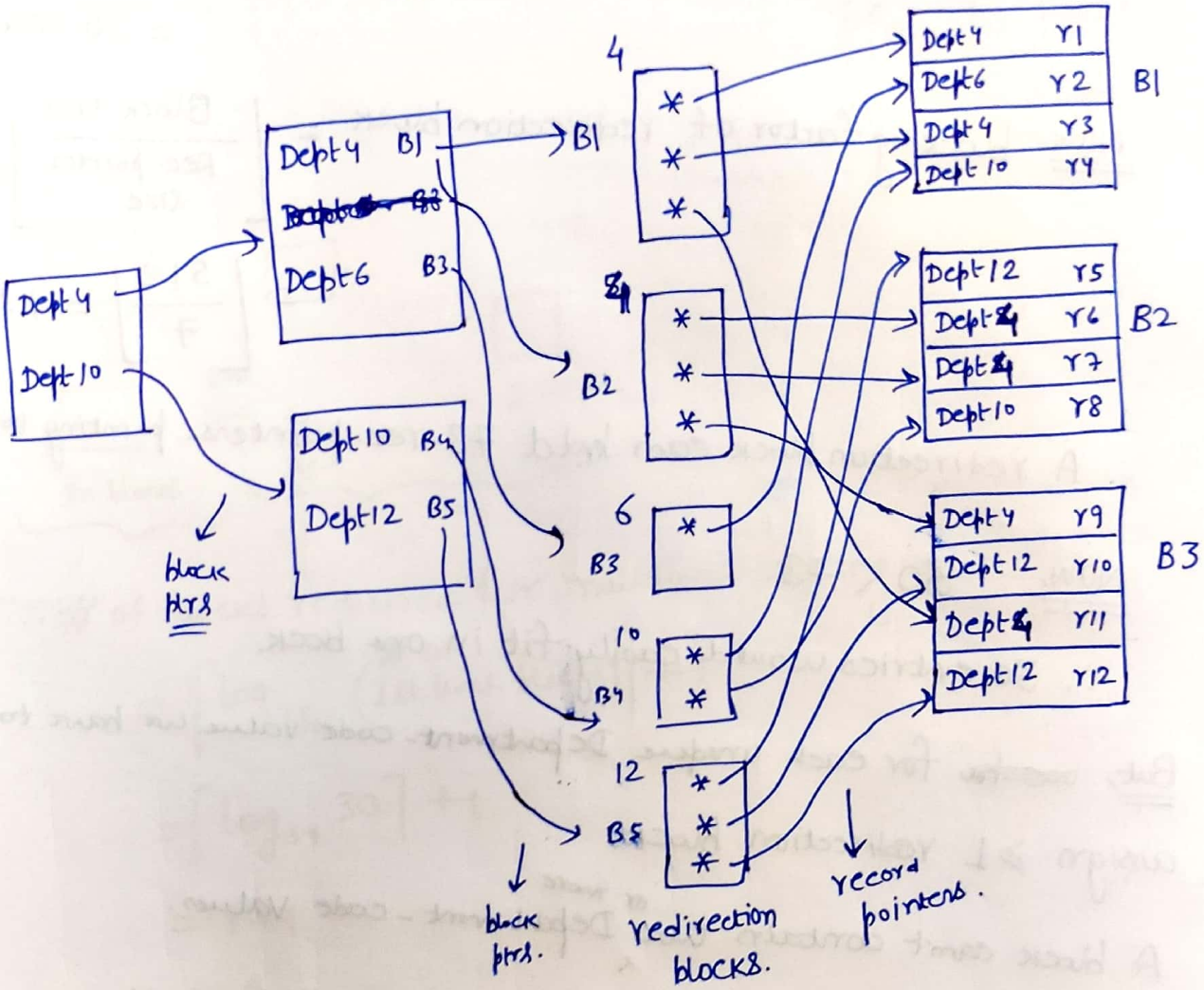
vi)

(e) Department_code → non key field

File is not ordered by Department_code.

Hence (duplicate + unordered)

Construct a secondary index on Department_code



1000 distinct values of Department_code.

∴ Each unique Department_code holds = $\frac{30,000}{1000}$ = 30 rec. values.

Size of Index Row = (Size of Department_code + Block-ptr)

= (9+6) = 15B

i) Blocking factor (b_{fri}) = $\left\lfloor \frac{\text{Block Size}}{\text{Size of Index Row}} \right\rfloor = \left\lfloor \frac{512}{15} \right\rfloor = 34$

ii) Redirection level.

Total no. of rows = # of total records = 30,000

~~We will store~~ 1000 distinct values of Department-code

∴ Each unique Department-code value holds = $\frac{30,000}{1000}$

= 30 records.

NOW, blocking factor of redirection block = $\left\lfloor \frac{\text{Block size}}{\text{Rec. pointer size}} \right\rfloor$

= $\left\lfloor \frac{512}{7} \right\rfloor$ = 73.

∴ A redirection block can hold 73 rec. pointers. pointing to file blocks.

NOW, 30 < 73.

∴ 30 entries would easily fit in one block.

But, ~~we have~~ for each unique Department-code value we have to assign ≥ 1 redirection block.

A block can't contain two (or more) Department-code values.

So, 30 will be occupied, rest (73-30) = 43 will be empty

∴ Number of redirection blocks required = $\left\lceil \frac{30,000}{30} \right\rceil$

= 1,000.

Shortcut: $\frac{\text{No. of entries per unique Department-code}}{\text{Blocking factor of redirection blocks}} = \left\lfloor \frac{30}{73} \right\rfloor$

iii) Number of first level index entries

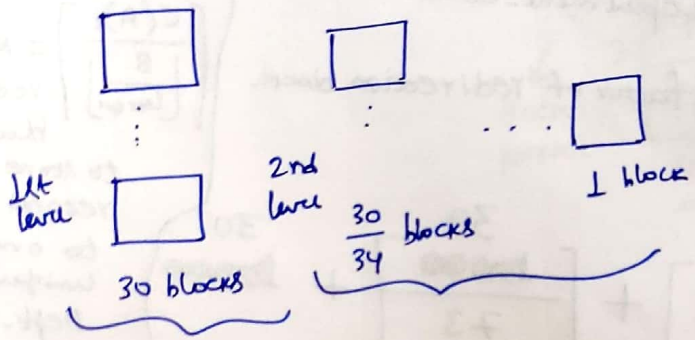
= # of distinct values of Department_code

= 1000

∴ Number of first level index blocks = $\left\lceil \frac{1000}{bfr_i} \right\rceil = \left\lceil \frac{1000}{34} \right\rceil$

= 30

iv)



∴ # of blocks required for multilevel indexing

$= \lceil \log_{bfr_i} (\#\text{ of 1st level blocks}) \rceil + 1$

$= \lceil \log_{34} 30 \rceil + 1$

$= 1 + 1$

$= 2.$

v) # of blocks in 1st level = 30

# of blocks — 2nd " $= \left\lceil \frac{30}{34} \right\rceil = 1$

# of blocks required for redirection = 1000

∴ Total no of blocks required = $(30 + 1 + 1000)$

= 1031

vi) Block access needed to search for and retrieve all records in the file that have a specific Department-code value using the index.

$$(NBA)_{SI_{Non\text{-}key}} = \left\lceil \log_2 (\#\text{ of 1st-level blocks}) \right\rceil + \left\lceil \frac{C(A)i}{\left\lfloor \frac{B}{l_{rec/ptr}} \right\rfloor} \right\rceil + C(A)_i$$

(NBA) ↙ Number of block access

$C(A)i \Rightarrow$ No of records per unique Department-code

$\left\lfloor \frac{B}{l_{rec/ptr}} \right\rfloor \Rightarrow$ Blocking factor of redirection blocks.

$\left\lceil \frac{C(A)i}{\left\lfloor \frac{B}{l_{rec/ptr}} \right\rfloor} \right\rceil$ = No of redirection blocks req. to store all records corresp. to one unique Dept-code value

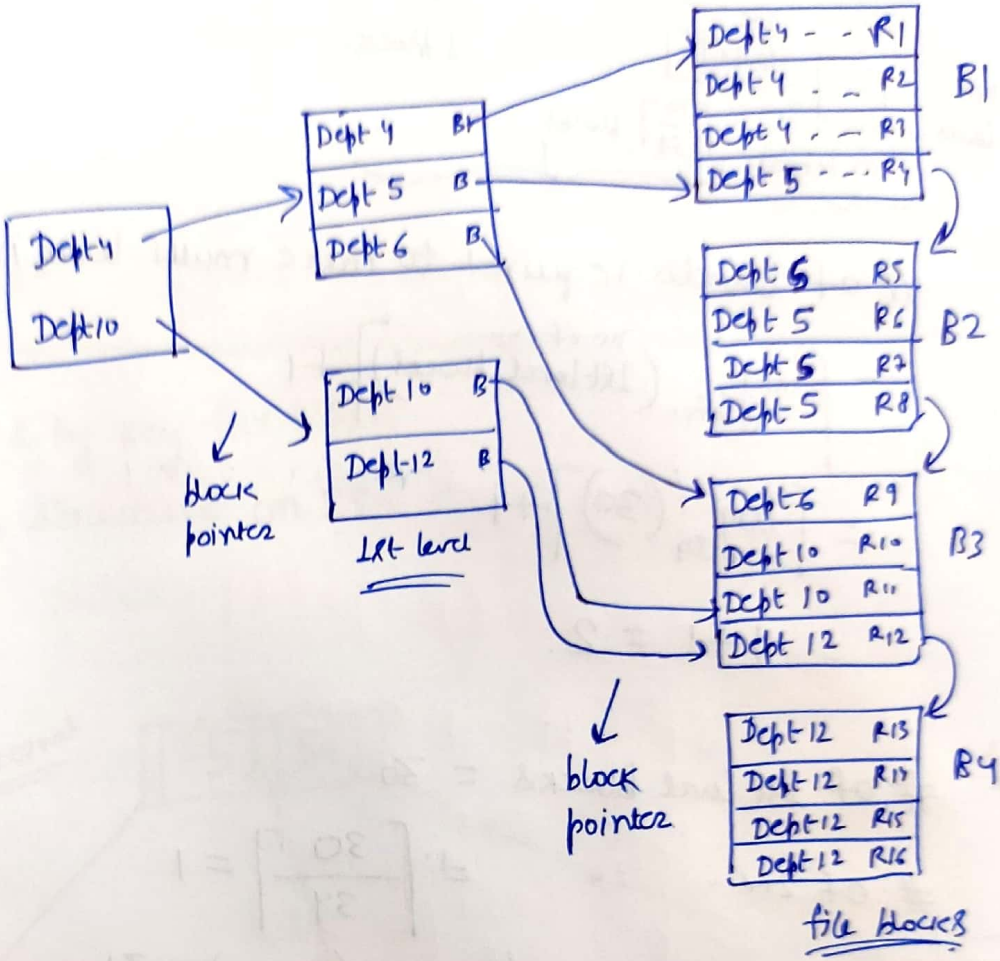
$$\therefore NBA_{SINK} = \left\lceil \log_2 (30) \right\rceil + \left\lceil \frac{30}{73} \right\rceil + 30$$

$$= 5 + 1 + 30$$

$$= 36$$

(f) Non-key field Department-Code.

File is ordered by Non-key ↓ " (Sorted + duplicates).



Index Row Length = (Dept-code + Block-ptr)

= (9+6) B = 15 B.

i) Index blocking factor (bfri) = $\left\lfloor \frac{\text{Block Size}}{\text{Index Row Length}} \right\rfloor = \left\lfloor \frac{512}{15} \right\rfloor = 34$

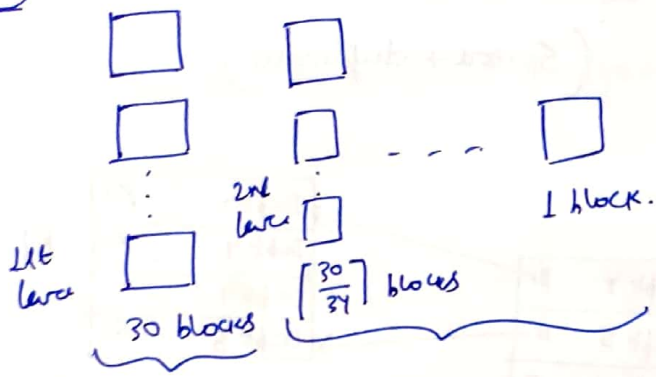
ii) The number of first level index entries = # of distinct values of Dept-code

= 1000

Number of first level index blocks = $\left\lceil \frac{\text{\# of first level index entries}}{\text{bfri}} \right\rceil$

= $\left\lceil \frac{1000}{34} \right\rceil = 30$

iii)



# of levels required to make multi-level indexing

$$= \lceil \log_{bfr_i} (\text{no of 1st level blocks}) \rceil + 1$$

$$= \lceil \log_{34} (30) \rceil + 1$$

$$= 1 + 1 = 2.$$

iv) # of 1st level blocks = 30

# of 2nd " " $= \lceil \frac{30}{34} \rceil = 1$

∴ Total no. of blocks = $(30+1) = 31$

v) Number of block access (NBA) for clustering index

$$= \log_2 (\text{\# of 1st level blocks}) + \left\lceil \frac{c(A)i}{bfr} \right\rceil$$

$c(A)i$ ⇒ Number of records per distinct value of Dept_code

bfr ⇒ blocking factor of file blocks

$\lceil \frac{c(A)i}{bfr} \rceil$ ⇒ those many rows spanned in how many blocks.

NOW, 30,000 total records.

1,000 distinct Department_code values.

∴ Per Dept_code contains = $\frac{30,000}{1000} = 30$ records.

NOW, $bfr = \left\lfloor \frac{B}{\text{Record size}} \right\rfloor = \left\lfloor \frac{512}{116} \right\rfloor = 4$

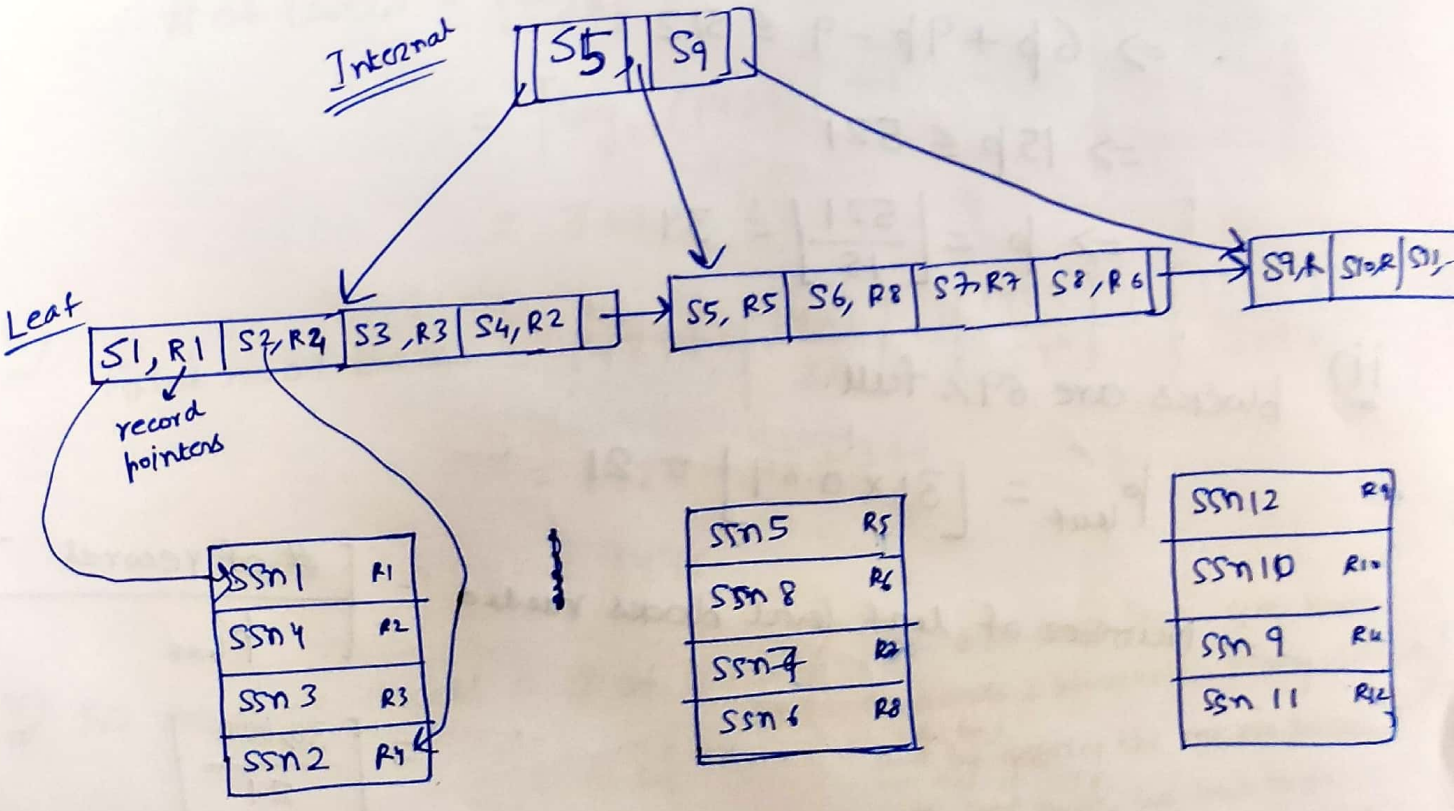
$\therefore NBA = \log_2(30) + \left\lceil \frac{30}{4} \right\rceil$

$= 5 + 8$

$= 13$

(g) File is not ordered by key field ssn.

B+ tree access structure on ssn



i) p_{leaf}

$$p_{leaf}(SSn + \text{record ptr}) + \text{blockptr} \leq \text{Block Size}$$

$$\Rightarrow p_{leaf}(9+7) + 6 \leq 512$$

$$\Rightarrow p_{leaf} * 16 \leq 506$$

$$\Rightarrow p_{leaf} = \left\lfloor \frac{506}{16} \right\rfloor = 31$$

(p)

$$p*(\text{blockptr}) + (p-1)*(SSn) \leq \text{Block Size}$$

$$\Rightarrow p*6 + (p-1)*(9) \leq 512$$

$$\Rightarrow 6p + 9p - 9 \leq 512$$

$$\Rightarrow 15p \leq 521$$

$$\Rightarrow p = \left\lfloor \frac{521}{15} \right\rfloor = 34$$

ii) blocks are 69% full.

$$p'_{leaf} = \lfloor 31 \times 0.69 \rfloor = 21$$

$\therefore$ Number of leaf-level blocks needed $= \left\lceil \frac{\#\text{ of records}}{p'_{leaf}} \right\rceil$

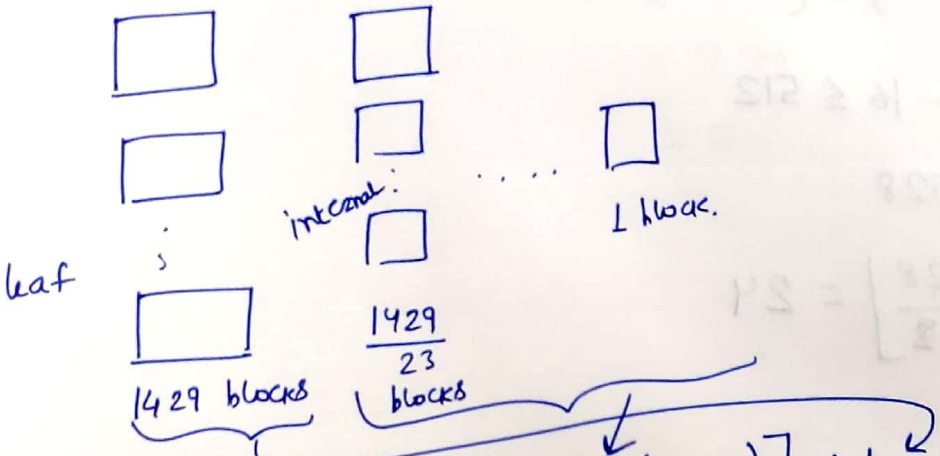
$$= \left\lceil \frac{30,000}{21} \right\rceil$$

$$= 1429.$$

iii) 69% full

$\therefore p'_{leaf} = 0.69 \times p_{leaf} = \lfloor 0.69 \times 31 \rfloor = 21$

$\therefore p' = 0.69 \times p = \lfloor 0.69 \times 34 \rfloor = 23$



$\therefore$ # of levels $= \lceil \log_{p'}(1429) \rceil + 1$

$= \lceil \log_{23}(1429) \rceil + 1$

$= 3 + 1 = 4$

iv) Total no of blocks $= 1429 + \left\lceil \frac{1429}{23} \right\rceil + \left\lceil \frac{1429}{23^2} \right\rceil + \left\lceil \frac{1429}{23^3} \right\rceil$

$= 1429 + 63 + 3 + 1$

$= 1496$

v) NO. of block access = # of levels + 1 (1 for data)

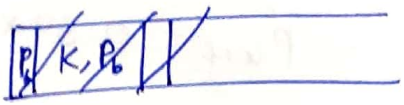
$= 4 + 1$

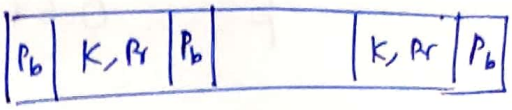
$= 5$

(bec. in each level we have to access 1 block only. Binary search and by getting the record pointer at leaf node, we have to go to the block to fetch the record)

(h) For b-tree $p = p_{leaf} = p_{internal}$

i) $p * L_{blkptr} + (p-1) * (L_{key} + L_{recptr}) \leq B$





$\Rightarrow p * 6 + (p-1) * (9+7) \leq 512$

$\Rightarrow 6p + 16p - 16 \leq 512$

$\Rightarrow 22p \leq 528$

$\Rightarrow p = \left\lfloor \frac{528}{22} \right\rfloor = 24$

$\therefore p = 24$

ii) 69% full

$p' = 0.69 \times p = 0.69 \times 24 = 16$

iii) Number of keys in root level $= (16-1) = 15$

Level	Pointer	Keys
1	16	15
2	16 x 16	240 = 15 x 16
3	16 x 16 x 16	3840 = 15 x 16 x 16
4	16 x 16 x 16 x 16	61,440 = 15 x 16 x 16 x 16

Now, 61,440 >= 30,000 rec ∴ we can stop.

∴ Number of leaf blocks required $= (16 \times 16 \times 16)$

$= 4096$

iii) No of levels = 4

iv) Total no of blocks req = 4096 + 256 + 16 + 1 = 4369

(4096 ↙ last level; 1 ↓ root level)

v) No. of block access req = # of levels + 1

= 4 + 1

= 5

↳ (for each level, 1 block access, by getting rec. ptr at last level, we have to visit blocks to fetch the record).