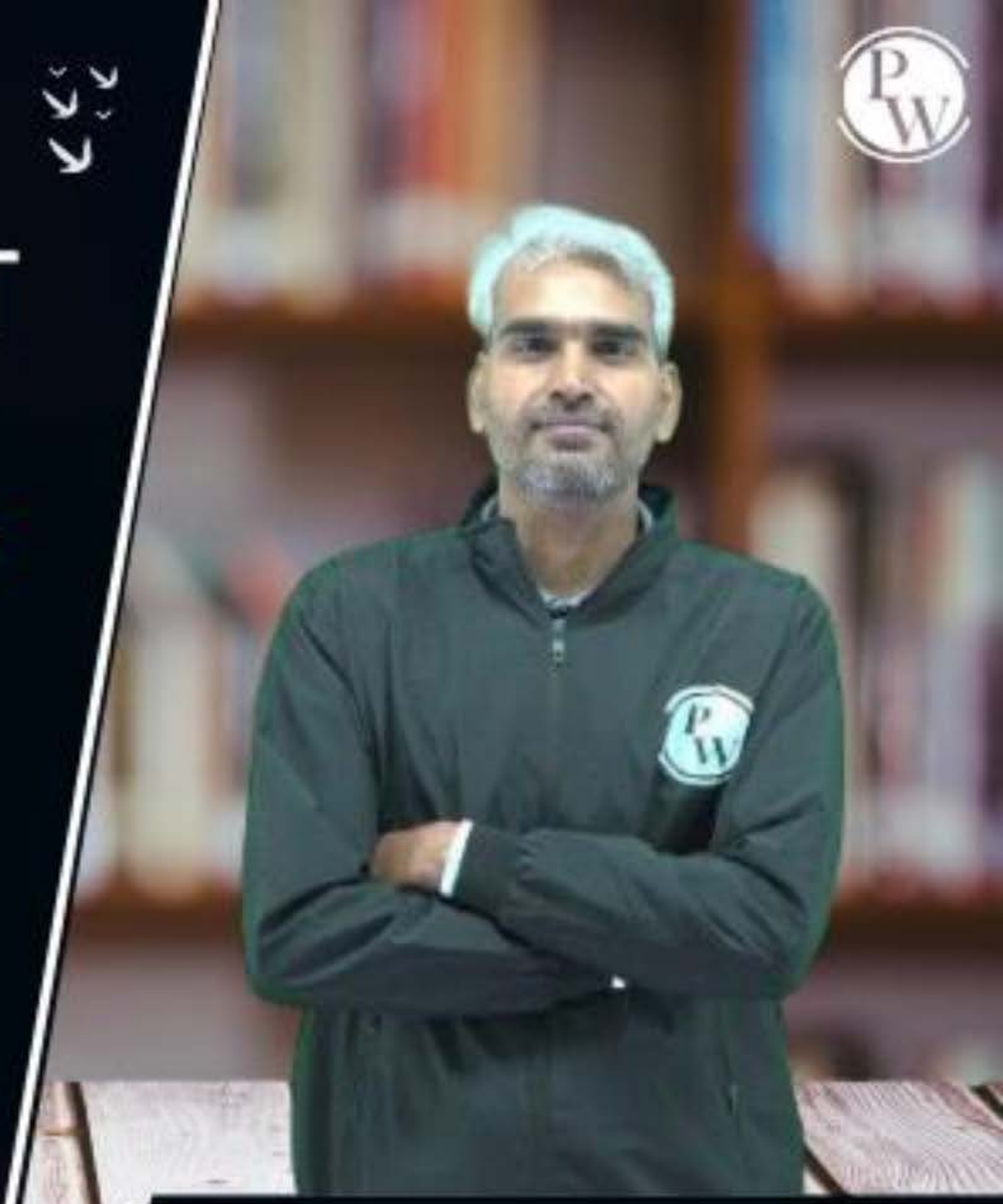
Computer Science & IT

Database Management System

File organization and indexing

Lecture No. 05



By- Vishal Sir

Recap of Previous Lecture

Topic Structure of B+ tree

Slide

Topics to be Covered

Topic Practice questions

Slide

Topic : NOTE

"w.r.t. B+ tree"

- If only one order is given in the question, then we will use the same order for both leaf node as well as for internal node, but it will be handled differently

Imp:-

Consider order of node of B+ tree is = n

{only one order is given}

Korth (w.r.t. Internal node)

* Maximum number of Child Pointer an internal node of B+ tree can have = n
* Maximum number of Keys an internal node of B+ tree can have = $(n-1)$
* Minimum number of Child Pointer a non-root internal node of B+ tree must have = $\left\lceil \frac{n}{2} \right\rceil$
* Minimum number of Keys a non-root internal node of B+ tree must have = $\left(\left\lceil \frac{n}{2} \right\rceil - 1\right)$

for root, Min. no. of Keys = 1
Min. no. of Child ptr = 2

Navathe (w.r.t. Internal node)

* Maximum number of Child Pointer an internal node of B+ tree can have = n
* Maximum number of Keys an internal node of B+ tree can have = $(n-1)$
* Minimum number of Child Pointer a non-root internal node of B+ tree must have = $\left\lceil \frac{n}{2} \right\rceil$
* Minimum number of Keys a non-root internal node of B+ tree must have = $\left(\left\lceil \frac{n}{2} \right\rceil - 1\right)$

for root, Min. no. of Keys = 1
Min. no. of Child ptr = 2

Imp:-

Consider order of node of B+ tree is $= n$

{only one order is given}

Korth (w.r.t. leaf node)

Navathe (w.r.t. leaf node)

* Maximum number of key a leaf node of B+ tree can have $= (n-1)$

* Maximum number of key a leaf node of B+ tree can have $= (n-1)$

* Minimum number of keys a non-root leaf node must have $= \left\lceil \frac{(n-1)}{2} \right\rceil$

* Minimum number of keys a non-root leaf node must have $= \left\lceil \frac{n}{2} \right\rceil$

Today's Topic

H.W.

#Q.[redacted] Consider a file of 10,00 records, and disk blocks of 1000 bytes. There are two alternate options to create a single level index file:

Option 1: Index file is created on an $\underbrace{\text{unordered key field}}_{\text{Dense index}}$, where key field is 12 bytes long.

Option 2: Index file is created on an $\underbrace{\text{ordered nonkey field}}_{\text{Clustering (sparse)}}$, where non-key field is 20 bytes long.

Block pointer size is 10 bytes long and $\boxed{\text{unspanned organization}}$ is used.

& Record Pointer size is 15 bytes long.

Let 'X' is the blocking factor of index block using option 1 and 'Y' is the blocking factor of index block using option 2 then |X – Y| is $|37-33| = 4$.

Option 1: $Bf = \left\lfloor \frac{B.S.}{\text{Entry size}} \right\rfloor = \left\lfloor \frac{1000}{\text{Key size} + \text{Rec. Pointer}} \right\rfloor = \left\lfloor \frac{1000}{12+15} \right\rfloor = \left\lfloor \frac{1000}{27} \right\rfloor = 37 = X$

Option 2: $\text{——} \;\; '' \;\; \text{——} = \left\lfloor \frac{1000}{\text{Key size} + \text{Block ptr size}} \right\rfloor = \left\lfloor \frac{1000}{20+10} \right\rfloor = \left\lfloor \frac{1000}{30} \right\rfloor = 33 = Y$

H.W.

#Q. [redacted] Consider an unordered file of 10,0000 records with a record size of 100 bytes stored on blocks of 1KB with an unspanned record organization. We will assume that no system related information is stored within a block. How many blocks would be needed to store this file? ________

$$\#\text{ Blocks required to store the file} = \left\lceil \frac{\text{No. of records}}{\text{No. of records per block}} \right\rceil = \left\lceil \frac{\text{No. of records}}{\text{B.F. w.r.t. database block}} \right\rceil$$

$$= \left\lceil \frac{10,0000}{\left\lfloor \frac{\text{Block size}}{\text{Record size}} \right\rfloor} \right\rceil = \left\lceil \frac{100000}{\left\lfloor \frac{1024}{100} \right\rfloor} \right\rceil = \left\lceil \frac{100000}{10} \right\rceil = 10,000$$

Ans

H.W:

#Q.[redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a primary index on Eid.

"By default it is Sparse"

The record size of the file is ___115___ bytes

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a primary index on Eid.

Sparse

The number of disk blocks required to store the file assuming an unspanned organization is used 7500

$$\#\text{Disk blocks for database file} = \left\lceil \frac{\text{No. of records}}{\text{B.f. w.r.t. database block}} \right\rceil = \left\lceil \frac{30{,}000}{\left\lfloor \frac{\text{Block size}}{\text{Record size}} \right\rfloor} \right\rceil = \left\lceil \frac{30000}{\left\lfloor \frac{512}{115} \right\rfloor} \right\rceil = \left\lceil \frac{30000}{4} \right\rceil = 7500$$

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a $\underbrace{\text{primary index}}_{\text{Sparse}}$ on Eid.

The blocking factor of index block is ______ 34 = Ans

$$= \left\lfloor \frac{\text{Block size}}{\text{Index file Entry size}} \right\rfloor = \left\lfloor \frac{512}{\text{Key size} + \text{Block pointer size}} \right\rfloor = \left\lfloor \frac{512}{9+6} \right\rfloor = \left\lfloor \frac{512}{15} \right\rfloor$$

$$= 34$$

H.W:

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is Unordered by the key field Eid and we want to construct a secondary index on Eid.

(always dense)

The blocking factor of index block is____ $\boxed{32 = \text{Ans}}$

$$B.f = \left\lfloor \frac{\text{Block size}}{\text{Entry size}} \right\rfloor = \left\lfloor \frac{512}{\text{Key size} + \text{Rec. pointer size}} \right\rfloor = \left\lfloor \frac{512}{9+7} \right\rfloor = 32$$

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a primary index on Eid.

{sparse}

Total number of entries in first level index is____ = No. of disk blocks required for database file

= 7500

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is Unordered by the key field Eid and we want to construct a secondary index on Eid.

(always dense)

Total no. of Entries in 1st level index is = No. of records in Database file = 30,000

Ans

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a primary index on Eid.

Total number of disk blocks required to store first level index is 221 = Ans

$$\text{Sparse} = \left\lceil \frac{\text{No. of Entries}}{\text{Bf. w.r.t. sparse index}} \right\rceil = \left\lceil \frac{7500}{34} \right\rceil$$

$$= 221$$

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is Unordered by the key field Eid and we want to construct a secondary index on Eid.

Total number of disk blocks required to store first level index is $\boxed{\text{Ans} = 938}$

$$\#\text{Disk block} = \left\lceil \frac{\text{No. of Entries}}{\text{B.f. w.r.t. dense index}} \right\rceil = \left\lceil \frac{30{,}000}{32} \right\rceil = 938$$

Ans

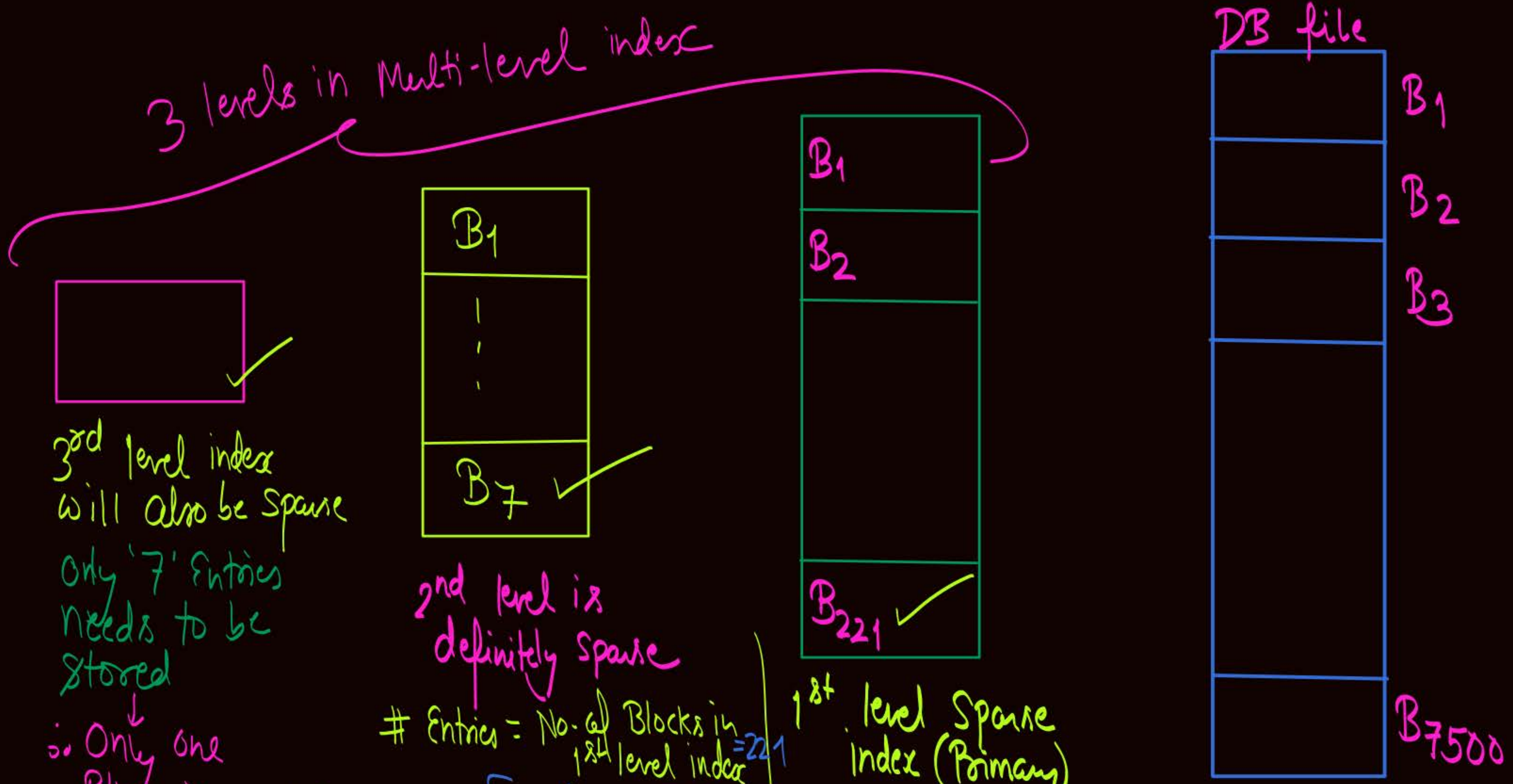
HW:

#Q[redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a primary index on Eid.

Sparse 1st level index

If we decide to create a multi-level index, then total number of levels needed in multi-level index is '3' = Ans

Soluⁿ on next slide



3rd level index will also be Sparse

Only '7' Entries needs to be stored

↓

∴ Only one Block is sufficient

↓

Hence it is last level

2nd level is definitely Sparse

Entries = No. of Blocks in 1st level index = 221

$$\#\text{Disk blocks required} = \left\lceil \frac{\text{No. of Entries}}{\text{B.f w.r.t. Sparse index}} \right\rceil = \left\lceil \frac{221}{34} \right\rceil = 7$$

HW:

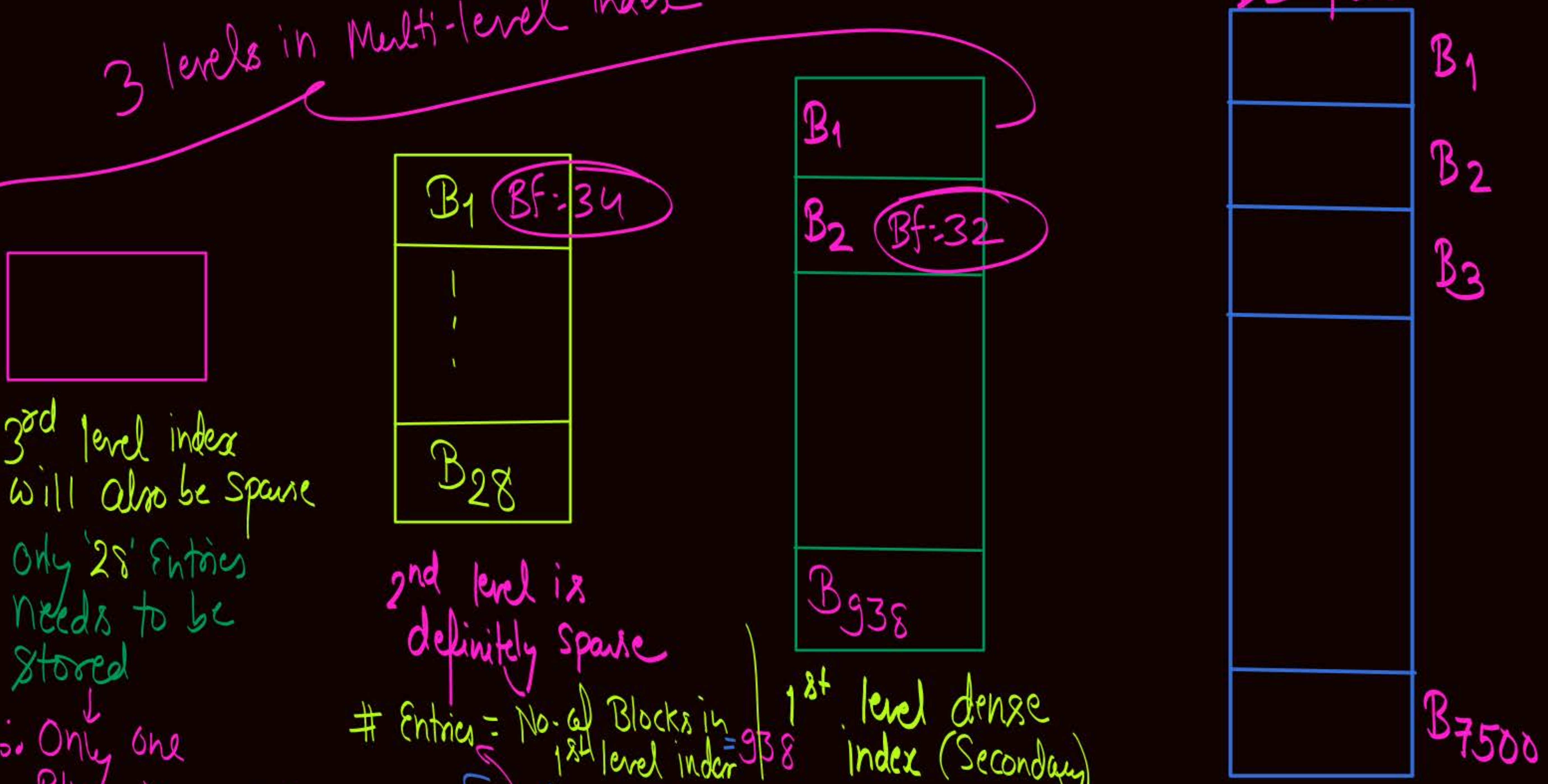
#Q[redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is Unordered by the key field Eid and we want to construct a Secondary index on Eid.

(First level index is dense)

If we decide to create a multi-level index, then total number of levels needed in multi-level index is "3"

Soln on next slide

3 levels in Multi-level index



3rd level index will also be sparse

Only '28' Entries needs to be stored

∴ Only one Block is sufficient

Hence it is last level

2nd level is definitely sparse

Entries = No. of Blocks in 1st level index = 938

1st level dense index (Secondary)

$$\#\text{Disk blocks required} = \left\lceil \frac{\text{No. of Entries}}{\text{Bf w.r.t. Sparse index}} \right\rceil = \left\lceil \frac{938}{34} \right\rceil = 28$$

H.W.

#Q. Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is ordered by the key field Eid and we want to construct a $\underbrace{\text{primary index}}_{\text{Sparse}}$ on Eid.

If we decide to create a multi-level index, then total number disk blocks required to store multi-level index is____

$$\left(\underset{\substack{1^{st}\\ \text{level}}}{221} + \underset{\substack{2^{nd}\\ \text{level}}}{7} + \underset{\substack{3^{rd}\\ \text{level}}}{1}\right) = 229$$

H.W.

#Q. [redacted] Consider a disk with block size B = 512 bytes. A block pointer is PB = 6 bytes long, and a record pointer is PR = 7 bytes long. A file has 30,000 EMPLOYEE records of fixed length. Each record has the following fields: Ename (30 bytes), Eid (9 bytes), Did(9 bytes), Location (40 bytes), Contact (9 bytes), DOB(8 bytes), Gender (1 byte), Role(4 bytes), and Salary (4 bytes) Other than the specified fields, an additional byte is used as a deletion marker in each record. Suppose the file is Unordered by the key field Eid and we want to construct a secondary index on Eid.

If we decide to create a multi-level index, then total number disk blocks required to store multi-level index is_____

$$\left(\underset{1^{st}\text{ level}}{938} + \underset{2^{nd}\text{ level}}{28} + \underset{3^{rd}\text{ level}}{1}\right) = 967$$

H.W.

#Q. [redacted] Consider a B tree in which the maximum numbers of keys in a node is '9', The minimum numbers of keys a non-root may have is ___4___ Ans

let order = p

then, Max no. of keys $= (p-1) = 9 \Rightarrow \boxed{p = 10}$

Min no. of keys

a non-root node must have $= \left(\left\lceil \frac{p}{2} \right\rceil - 1\right) = \left(\left\lceil \frac{10}{2} \right\rceil - 1\right) = 4$

H.W.

#Q. Let ‘X’ denotes the order of an internal node of B+tree, and ‘Y’ denotes order of a leaf node of B+tree.

If disk block size is 1024 bytes, search key is 15 bytes long, block pointer is 10 bytes long and record pointer is 10 byte long then maximum value of X+Y is ___ $41 + 40 = 81$ Ans

$$\text{Max of } (X+Y) = x_{max} + y_{max}$$ = Max order for internal node + Max order for leaf node

$$[(x * \text{Block pointer size}) + (x-1) \text{ key size}] \leq \text{Block size}$$

$$[(x * 10) + (x-1) * 15] \leq 1024$$

$$25x \leq 1039$$

$$\boxed{x_{max} = 41}$$

$$[y * (\text{key size} + \text{R.P. size}) + 1 * \text{Block ptr size}] \leq \text{Block size}$$

$$y * (15 + 10) + 1 * 10 \leq 1024$$

$$25y \leq 1014$$

$$\boxed{y_{max} = 40}$$

HW:

#Q. [redacted] Let 'X' denotes the order of an internal node of B[redacted]tree, and 'Y' denotes order of a leaf node of B[redacted]tree.

If disk block size is 1024 bytes, search key is 15 bytes long, block pointer is 10 bytes long and record pointer is 10 byte long then maximum value of X+Y is $X_{max} + Y_{max} = 29 + 29 = 58$ Ans

For B tree,

Order of leaf node = Order of Non-leaf node

i.e. $X = Y$

$$\left[\left(X * \begin{matrix}\text{Block Pointer}\\ \text{size}\end{matrix} \right) + (X-1)\left(\begin{matrix}\text{Key}\\ \text{size}\end{matrix} + \begin{matrix}\text{Record Pointer}\\ \text{size}\end{matrix} \right) \right] \le \begin{matrix}\text{Block}\\ \text{size}\end{matrix}$$

$$\left[(X * 10) + (X-1)(15+10) \right] \le 1024$$

$$35X \le 1049$$

$$\boxed{X_{max} = 29} \quad \therefore \quad \boxed{Y_{max} = 29}$$

H.W.

#Q. Consider the B tree of order 10. Assume that the number of levels in the tree is four including root. The maximum number of record pointers that can be stored in B tree is $(9000+900+90+9)$
$= 9999$ Ans

order = 10

∴ Max. no. of key in a node = Max. no. of record pointer in a node = (10 − 1) = 9

level 1 = Root ⇒ Max No. of nodes = 1 ⇒ ∴ Max no. of record pointers at level 1 = $1 * 9 = 9$

level 2 = Max. No. of Nodes = Max. no. of Child pointer at previous level = $1 * 10$ ⇒ ∴ Max no. of record pointers at level 2 = $10 * 9 = 90$

level 3 = Max no. of nodes = Max no. of Child pointer at previous level = $10 * 10$ ⇒ ∴ Max. no. of RP = $100 * 9 = 900$

level 4 = Max no. of nodes = $100 * 10$ → ∴ Max No. of RP = $1000 * 9 = 9000$

H.W.

#Q. [redacted] Consider the B tree of order 10. Assume that the number of levels in the tree is four including root. The minimum number of record pointers that can be stored in B tree is $\underline{(1+8+40+200)} = 249$

Min no. of nodes → Min no. of Rec. point

level 1 = Root ⇒ Min. no. of nodes = 1 ⇒ Min. no. of Child ptr = 2 ⟹ Min. no. of record pointers = $1 \times 1 = 1$

level 2 = Min no. of nodes = Min no. of Child Pointer at previous level = 2 = Min. no. of Child pointer $= 2 \times \left\lceil \frac{10}{2} \right\rceil = 10$ = of Min. no. of Record pointers $= 2 \times \left(\left\lceil \frac{10}{2} \right\rceil - 1 \right) = 2 \times 4 = 8$

level 3 = Min. no. of nodes = Min no. of Child ptr at previous level = 10 = Min no. of Child ptr $= 10 \times \left\lceil \frac{10}{2} \right\rceil = 50$ = Min no. of record pointers $= 10 \times 4 = 40$

level 4 = —"— = —"— = 50 = Min. no. of Rec. pointer $= 50 \times 4 = 200$

H.W:

#Q. $\boxed{\text{Consider the B+tree of order 10.}}$ Assume that the number of levels in the tree is four including root. The $\boxed{\text{maximum number of record pointers}}$ that can be stored in B+tree is

"only one order is given"

let, order $n = 10$

In B+ tree record pointers are available only at leaf level

$\therefore$ Max no. of child pointer a non-leaf node can have $= n = 10$

Max. no. of keys a non-leaf node can have $= (n-1) = 9$

Max no. of record pointers a non-leaf node can have $= 0$

Max no. of keys a leaf node can have $=$ Max no. of record pointers a leaf node can have $= (n-1) = (10-1) = 9$

$\therefore$ Max. no. of Records pointers

$=$ Max no. of nodes at leaf level $\times$ Max no. record pointers in each node at leaf level

level - 1 Root with max 10 Child pointer

level - 2 Max # nodes = 10, each with max 10 Child pointer
∴ total max Child pointer = 10×10 = 100

level - 3 :- Max # nodes = 100, each with max 10 Child pointer
∴ total max Child pointer = 100×10 = 1000

level - 4 : Max no of nodes = 1000
(leaf level)

∴ Max no. of record Pointer = 1000 × Max no. of record Pointer in each node = 1000 × 9 = 9000

Ans

Hint:

#Q. [redacted] Consider the B+tree of order 10. Assume that the number of levels in the tree is four including root. The minimum number of record pointers that can be stored in B+tree is 250 Ans

= Min no. of nodes at leaf level × Min no. of Record pointers in a node at leaf level

(level-1) Root → One node with Min. '2' child pointer

level-2 → '2' nodes each with min $\lceil \frac{10}{2} \rceil = 5$ child ptr

i.e. $2 \times 5 = 10$ child ptr

level-3 → 10 nodes each with min '5' child ptr.

i.e. with Min $10 \times 5 = 50$ child ptr

level 4 → 50 nodes each with minimum '5' record pointer.

× $\lceil \frac{n}{2} \rceil$ or $\lceil \frac{n-1}{2} \rceil$

× $\lceil \frac{10}{2} \rceil$ or $\lceil \frac{10-1}{2} \rceil$

× 5 or 5

∴ Min No. of Record pointers $= 50 \times 5 = 250$

2 mins Summary

Topic Practice questions

Slide

THANK - YOU

Slide