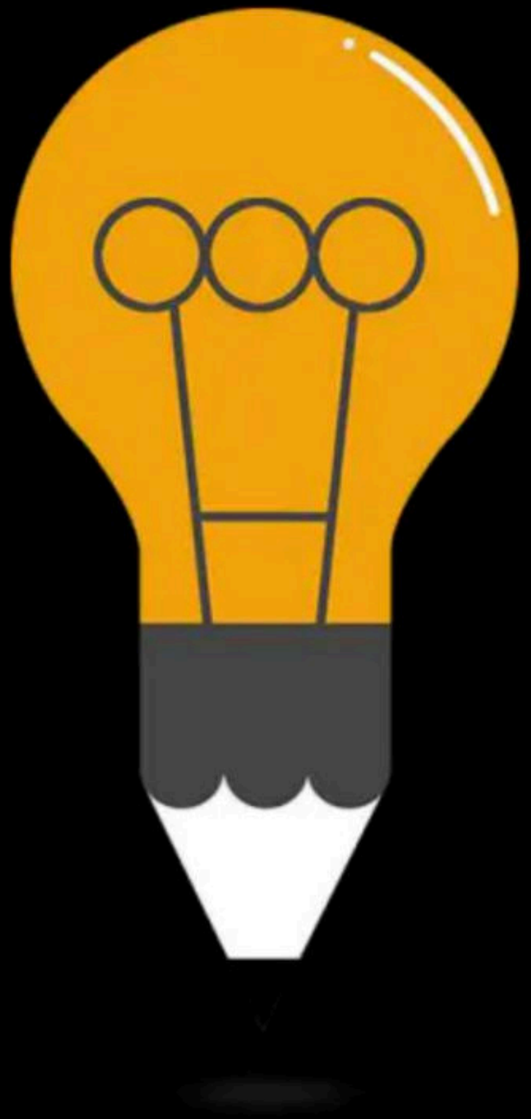
unacademy

File Organization and Indexing: Part III

Complete Course on Database Management System

Vishvadeep Gothi • Lesson 31 • July 27, 2023

unacademy



DBMS

Indexing 3

By: Vishvadeep Gothi

Spanned vs Unspanned File Organization

unacademy

Why Indexing

unacademy

Index File

unacademy

Indexing Techniques

- Clustered Indexing
- Non-Clustered Indexing

unacademy

Clustered Indexing

- Data order and index order are same

unacademy

Non-Clustered Indexing

- Data order and index order are not same

unacademy

Dense Vs Sparse Index

1. Dense: Index record is for each database record
2. Sparse (non-dense): Index record is for a few database records only

unacademy

Indexing Techniques

- Primary Indexing
- Clustering Indexing
- Secondary key Indexing
- Secondary non-key Indexing

unacademy

Indexing Techniques

- Primary Indexing
- Clustering Indexing
- Secondary key Indexing
- Secondary non-key Indexing

unacademy

Primary Indexing

- Indexing done on primary key or any super key
- Data must be ordered on index
- Its always sparse index

Clustering Indexing

- Indexing done on non-key field
- Data must be ordered on index field

} Here indexing is done for each unique value of non-key field.

- It can be dense or sparse

dense → when non-key field has all unique values

sparse → when non-key field has duplicate values

unacademy

Secondary key Indexing

- Indexing done on primary key or any super key
- Data must not be ordered on index field
- It is dense index

unacademy

Secondary Non-key Indexing

- Indexing done on non-key field
- Data must not be ordered on index field
- It is sparse index

Primary Index

Rno.	Block Pointer
1	B1
5	B2

Rno.	Block Pointer
9	B3
13	B4

1	A
2	B
3	C
4	D

5	A
6	B
7	E
8	W

9	V
10	D
11	I
12	A

13	C
14	S
15	H
16	A

unacademy

Primary Index

1	A
2	B
3	C
4	D

5	A
6	B
7	E
8	W

9	V
10	D
11	I
12	A

Select * from Students

where rno=1

unacademy

Primary Index

1	A
2	B
3	C
4	D

5	A
6	B
7	E
8	W

9	V
10	D
11	I
12	A

Select * from Students

where rno=12

unacademy

Question

Consider a database file of 65536 records, each record of size 64 bytes. Key field is 10 bytes and block pointer size is 22 bytes. Assume that block size is 256 bytes.

1. The number of blocks required to store file?
2. The number of blocks required to store the index file for primary indexing?

Ques) Primary indexing

no. of records = 50000

block size = 1024 bytes

record size = 50 bytes

key field = 10 bytes

block pointer = 35 bytes

	Spanned	Unspanned
no. of blocks to store data file	2442	2500
no. of blocks to store index file	108	114

Channel :- unacademy

db file size = 500000 * 50 B

= 25000000 B

no. of blocks for db file = $\left\lceil \frac{25000000\ B}{1024\ B} \right\rceil$ = 2442

---

no. of index records = 2442

index record size = 10 + 35 = 45 B

} ⇒ index file size = 2442 * 45 B

= 109890 B

no. of blocks for index file = $\left\lceil \frac{109890\ B}{1024\ B} \right\rceil$ = 108

Unspanned :-

$$\text{no. of records per block} = \left\lfloor \frac{1024\,B}{50\,B} \right\rfloor = 20$$

$$\text{no. of blocks to store 50000 records} = \frac{50000}{20} = 2500$$

---

no. of index records = 2500, index record size = 10 + 35 = 45 B

$$\text{no. of index records per block} = \left\lfloor \frac{1024\,B}{45\,B} \right\rfloor = 22$$

$$\text{no. of blocks for 2500 index records} = \left\lceil \frac{2500}{22} \right\rceil = 114$$

Clustering Index

indexing done for each unique value of non-key field

S_Name	Block Pointer
A	B1
B	B2
B	B3
E	B4

S_no	S_name	
1	A	B1
2	A	
3	A	
4	B	
5	B	B2
6	B	
7	B	
8	B	
9	B	B3
10	B	
11	C	
12	D	
13	E	B4
14	F	
15	F	
16	G	

Clustering Index

S_Name	Block Pointer
A	B1
B	B1
C	B3
D	B3
E	B4
F	B4
G	B4

1	A	B1
2	A	
3	A	
4	B	
5	B	B2
6	B	
7	B	
8	B	
9	B	B3
10	B	
11	C	
12	D	
13	E	B4
14	F	
15	F	
16	G	

Question

2^{30}

DB File size 1G records

Record size = 64bytes

Block size = 4096bytes

Index field = 10 bytes

Block pointer size = 22 bytes

Number of distinct values in index file = 16384

Indexing is done on non-key, data is ordered on non-key and indexing is done for each distinct non-key value ⇒ clustering indexing

db file size $= 2^{30} * 64B$

no. of blocks to store db file $= \frac{2^{30} * 64B}{4096B} = \frac{2^{36}}{2^{12}} = 2^{24}$

---

no. of records in index file $= 16384 = 2^{14}$

index record size $= 10B + 22B = 32B$

index file size $= 2^{14} * 32B = 2^{19}B$

no. of blocks to store index file $= \frac{2^{19}B}{4096B} = 2^{7}$

Ques)

no. of records in db file = 2^{24}

record size = 70 bytes

block size = 2048 bytes

unique values in index = 2^{22}

non-key index field = 12 Bytes

block pointer = 35 bytes

clustering indexing.

	spanned	unspanned
no. of blocks for db file		
no. of blocks for index file		

Spanned :-

db file size = $2^{24} * 70B$

no. of blocks for db file = $\left\lceil \frac{2^{24} * 70B}{2048B} \right\rceil = 573440$

---

index rec. size = 12 + 35 = 47B

no. of index records = 2^{22}

index file size = $2^{22} * 47B$

no. of blocks for index file = $\left\lceil \frac{2^{22} * 47B}{2048B} \right\rceil = 96256$

Unspanned :-

no. of db records per block $= \left\lfloor \frac{2048\,B}{70\,B} \right\rfloor = 29$

no. of blocks to store 2^{24} records of db file $= \left\lceil \frac{2^{24}}{29} \right\rceil$

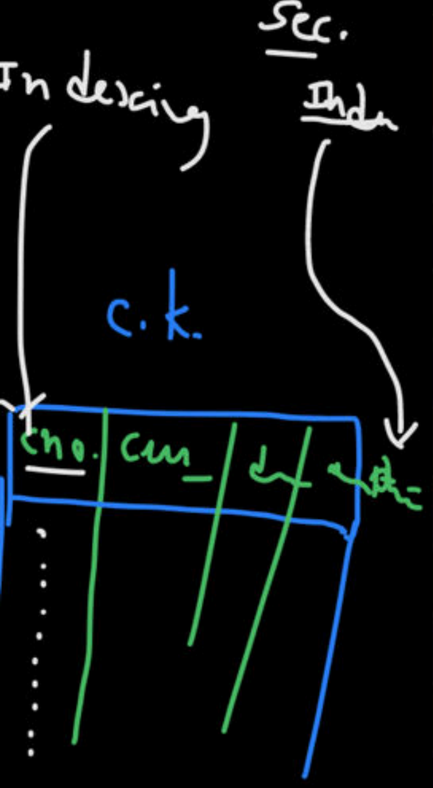
$= 578525$

---

no. of index records per block $= \left\lfloor \frac{2048\,B}{47\,B} \right\rfloor = 43$

no. of blocks to store 2^{22} records of index file $= \left\lceil \frac{2^{22}}{43} \right\rceil = 97542$

Secondary Key Index



Rno.	Record Pointer
1	B1, R0
2	B2, R1
3	B2, R2
4	B1, R1
5	B4, R2
6	B4, R3
7	B3, R0
8	B2, R3
9	B1, R2
10	B4, R1
11	B1, R3
12	B2, R2
13	B3, R3
14	B3, R1
15	B4, R0
16	B2, R0

Rno	Name	
1	A	0
4	A	1
9	A	2
11	B	3
16	B	0
2	B	1
12	B	2
8	B	3
7	B	0
14	B	1
3	C	2
13	D	3
15	E	0
10	F	1
5	F	2
6	G	3

B1
B2
B3
B4

Bi, Rj

Block i
Record j

unacademy

Question

DB File size 2^{40} bytes

Record size = 128bytes

Block size = 4096bytes

Index key field = 10 bytes

Record pointer size = 22 bytes

unspanned file organization

Indexing is done on key; data is unordered on key and indexing is done for each key value

1. Number of blocks required to store database file $= 2^{28}$
2. Number of blocks required to store index file $= 2^{26}$

no. of db records per block $= \frac{4096\,B}{128\,B} = 2^5 = 32$

no. of db records $= \frac{2^{40}\,B}{2^7\,B} = 2^{33}$

no. of blocks for db file $= \frac{2^{33}}{2^5} = 2^{28}$ $= \frac{2^{40}\,B}{2^{12}\,B} = 2^{28}$

no. of index records $= 2^{33}$

index rec. size $= 10 + 22 = 32B$

no. of index records per block $= \frac{4096B}{32B} = 2^7$

no. of blocks for index file $= \frac{2^{33}}{2^7} = 2^{26}$

Secondary Non-Key Index

Rno.	Block Pointer of Record pointers
A	
B	
C	
D	

1	D
2	A

B1

3	B
4	C

B2

5	C
6	A

B3

7	B
8	C

B4

unacademy

Recap

Ordering	Key or Non Key	Type
Ordered	Key	
Ordered	Non-Key	
Unordered	Key	
Unordered	Non-Key	

unacademy

Question

DB File size 10^8 records

Record size = 400bytes

Block size = 4096bytes

Index key field = 16 bytes

index pointer size = 4 bytes

unspanned file organization

Indexing is done on key; data is unordered on key and indexing is done for each key value

1. Number of blocks required to store database file
2. Number of blocks required to store index file

unacademy

B-Tree

- Tree based indexing
- Dynamic Indexing technique
- Based on insertion and deletion, the tree automatically adjusted
- Self balancing search tree

unacademy

Binary Search Tree

unacademy

B-Tree

An order-p B-tree:

1. Every node other than root should have atleast $\left\lceil \frac{p}{2} - 1 \right\rceil$ nodes
2. In every node there are atmost (p-1) keys and n tree pointers
3. Root can have minimum 1 node
4. All leaves appear on the same level

unacademy

B-Tree Node Structure

- Key
- Record Pointer
- Tree Pointer

unacademy

Insertion in B-Tree

- B-tree of order-3
- Insert keys 1, 2, 3, 4, 5, 6, 7

unacademy

Insertion in B-Tree

- B-tree of order-5
- Insert keys 7, 4, 14, 25, 3, 10, 12, 15, 17, 9, 29, 1, 38, 3, 11

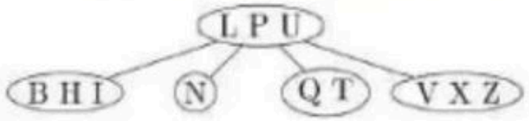
unacademy

Insertion in B-Tree

- B-tree of order-3
- Insert keys 14, 3, 5, 10, 35, 40, 1, 37

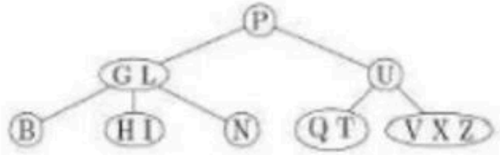
Question GATE-2003

Consider the following $2-3-4$ tree (i.e., B-tree with a minimum degree of two) in which each data item is a letter. The usual alphabetical ordering of letters is used in constructing the tree.

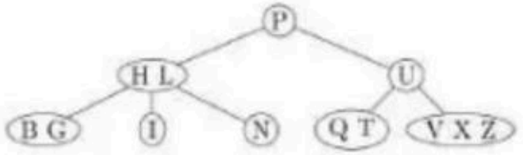


What is the result of inserting G in the above tree?

A.



B.



C.

D. None of the above

unacademy

Insertion in B-Tree

- B-tree of order-4
- Insert keys 15, 5, 8, 22, 10, 1

unacademy

Insertion in B-Tree

More split in left-biasing or right-biasing?

unacademy

Happy Learning.!