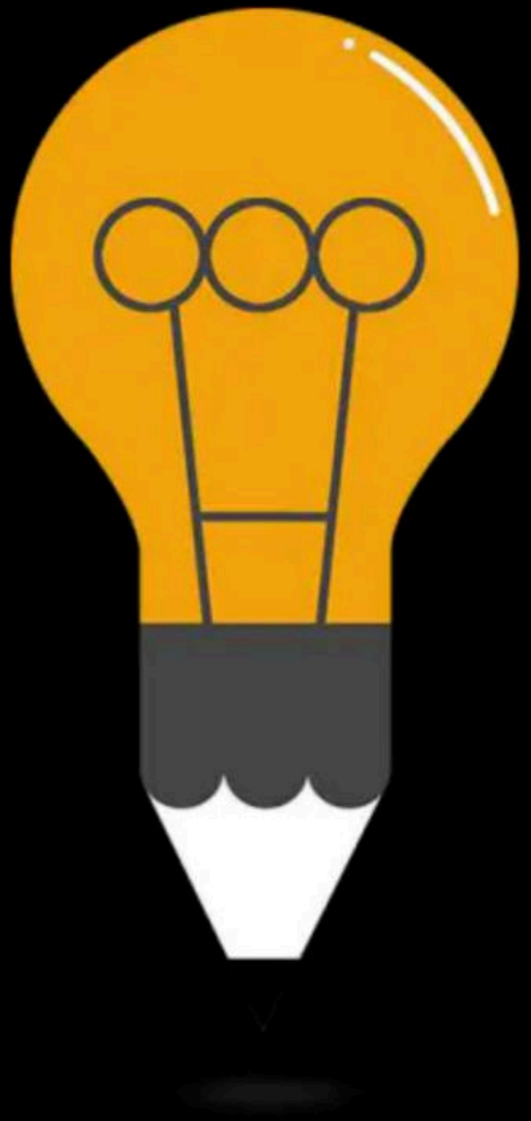
unacademy

# File Organization and Indexing: Part II

Complete Course on Database Management System

**Vishvadeep Gothi** • Lesson 30 • July 26, 2023

unacademy



# DBMS

# Indexing

By: Vishvadeep Gothi

unacademy

# Memory Structure

- Main Memory
- Secondary Memory

unacademy

# Disk Blocks and Record Storages

unacademy

# File Organization

- Sorted File Organization
- Heap File Organization

# Spanned vs Unspanned File Organization

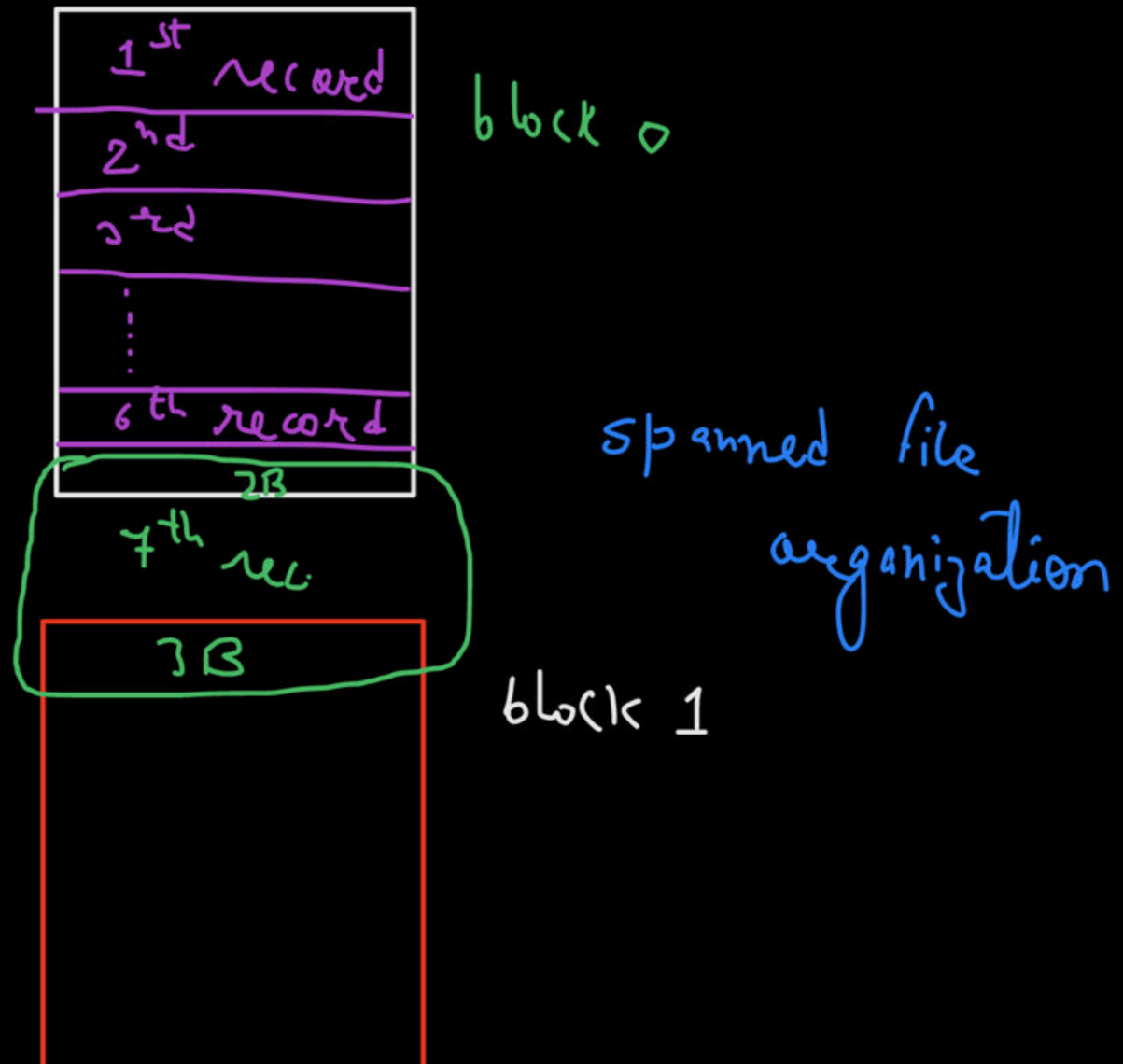
Ex:- Disk block ⇒ 32 bytes

Record size = 5 bytes

---

Assuming table has 100 records.

no. of blocks to store db table



Skanned :-

Total size of db file = 5 * 100 = 500 bytes

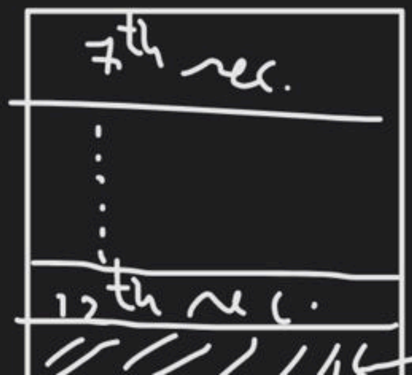
no. of blocks required to store entire Table $= \left\lceil \frac{500\ B}{32\ B} \right\rceil$

$= 16$ blocks

## Unspanned :-

A record is stored completed in only one block always.


block 6




2 bytes wastage

$$\text{no. of records per block} = \frac{\text{72 B}}{\text{5B}} = 6$$

$$\text{no. of blocks to store table} = \left\lceil \frac{100}{6} \right\rceil = 17 \text{ blocks}$$

Ques Consider a db table with 500 records with each record size 60 bytes. The disk block size is 2048 bytes. No. of blocks required to store table in

① spanned file organ

② unspanned — " —

① spanned :-

Table size = 500 * 60 Bytes

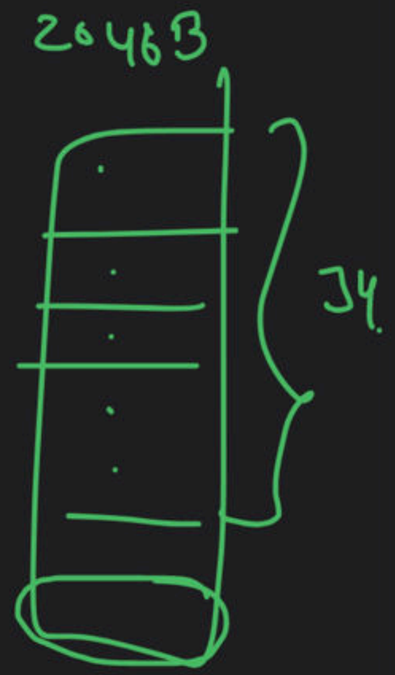
= 30000 bytes

no. of blocks = $\left\lceil \frac{30000 \text{ bytes}}{2048 \text{ bytes}} \right\rceil$

= 15

(2) Unspanned :-

$$\text{no. of records per block} = \left\lfloor \frac{2048\,B}{60\,B} \right\rfloor = 34$$

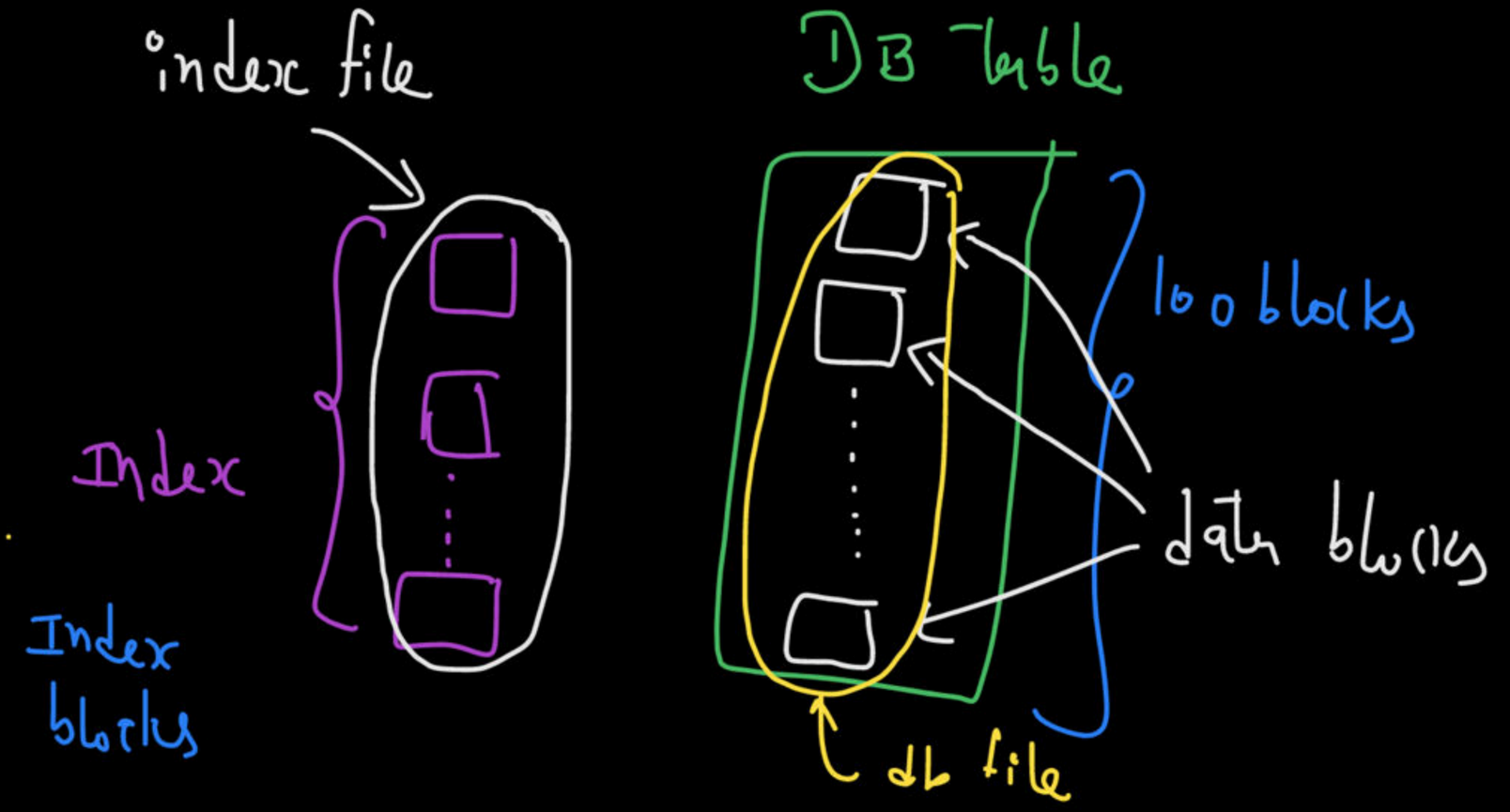



$$\text{no. of blocks} = \left\lceil \frac{500}{34} \right\rceil = 15$$

unacademy

# Why Indexing

used to access records in less time.



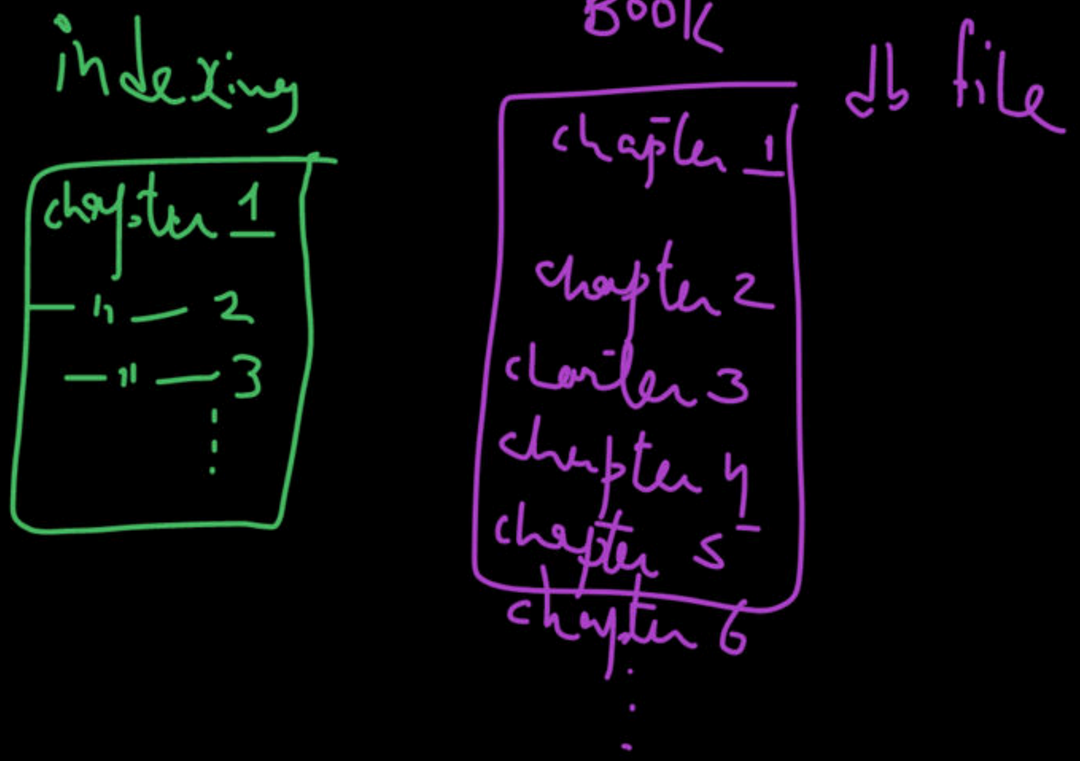
unacademy

# Index File

unacademy

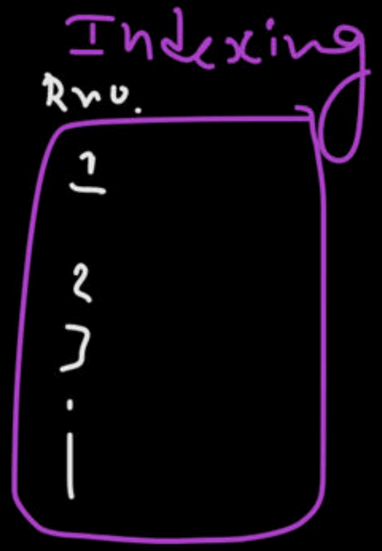
# Indexing Techniques

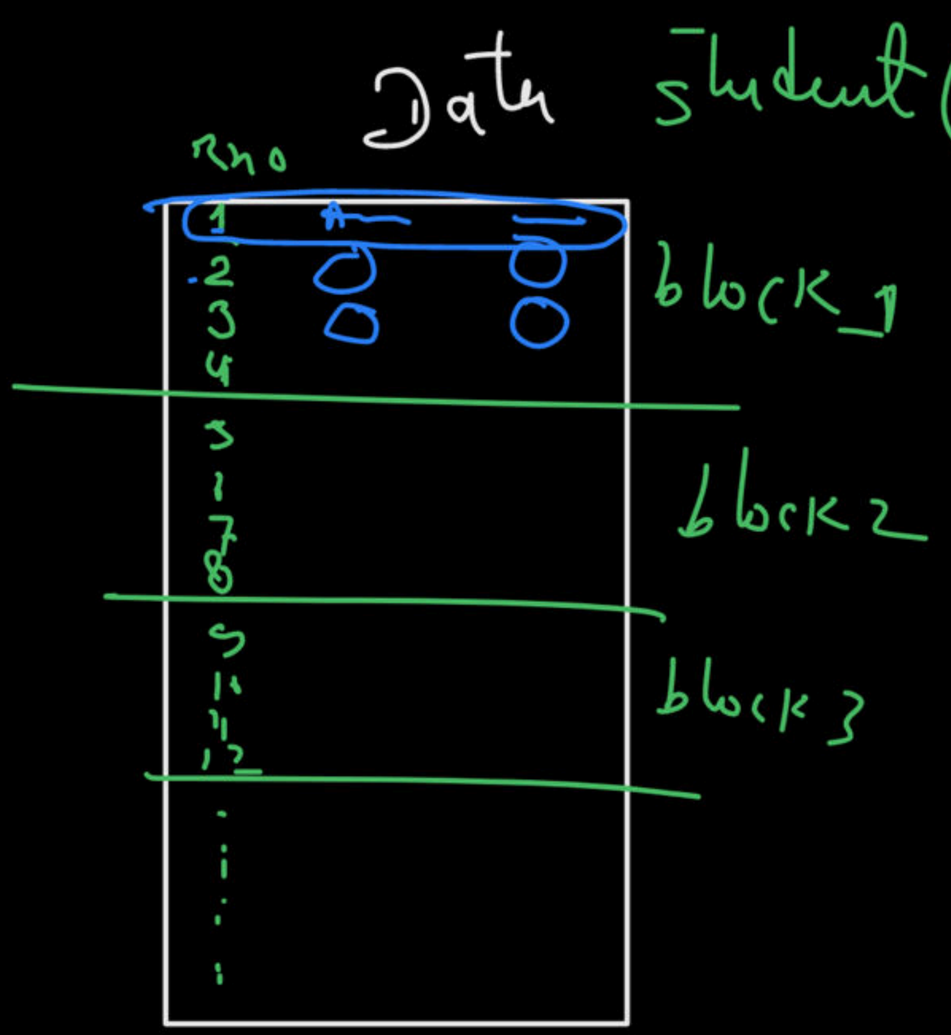
- Clustered Indexing
- Non-Clustered Indexing



# Clustered Indexing

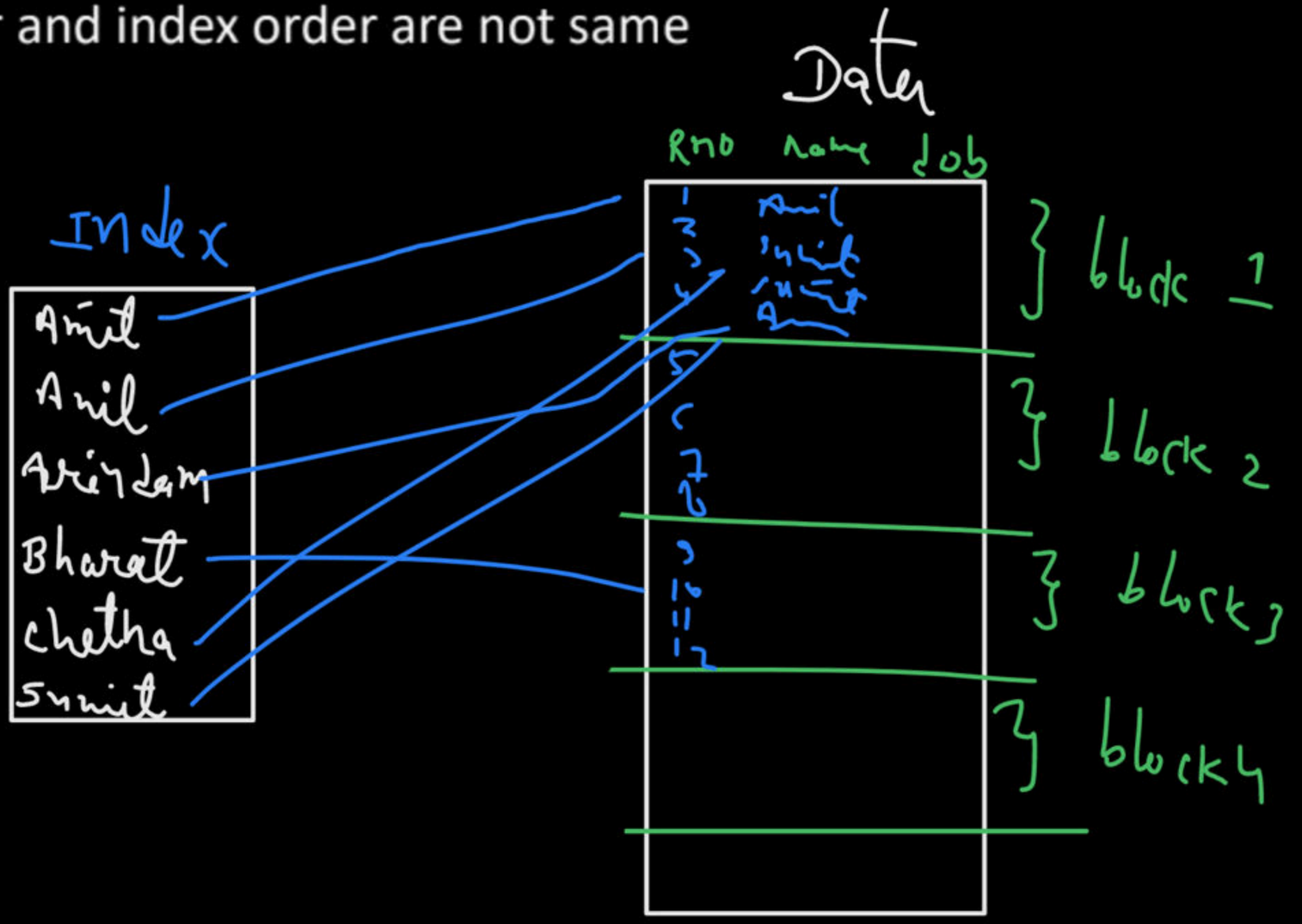
- Data order and index order are same



# 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

unacademy

# Clustering Indexing

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

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

unacademy

# Primary Indexing

ex:- student

| Rno | name |
|---|---|
| 1 | |
| 2 | |
| ⋮ | |
| 16 | |

block size ⇒ 4 records per block

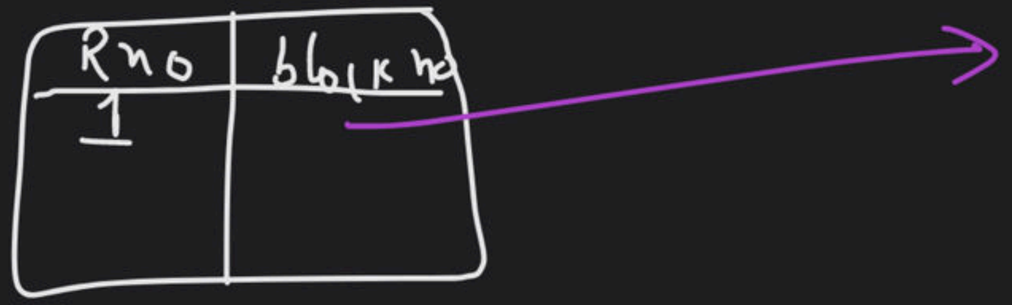
for 16 records ⇒ 4 blocks

---

2 indexes per block

Index

| Rno | block no |
|---|---|
| 1 | |



# Primary Index

| Rno. | Block Pointer |
|---|---|
| 1 | B1 |
| 5 | B2 |

| Rno. | Block Pointer |
|---|---|
| 9 | B3 |
| 13 | B4 |

Anchor records

| | |
|---|---|
| 1 | A |
| 2 | B |
| 3 | C |
| 4 | D |

Block 1

| | |
|---|---|
| 5 | A |
| 6 | B |
| 7 | E |
| 8 | W |

Block 2

| | |
|---|---|
| 9 | V |
| 10 | D |
| 11 | I |
| 12 | A |

Block 3

| | |
|---|---|
| 13 | C |
| 14 | S |
| 15 | H |
| 16 | A |

Block 4

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

⇓

how many blocks are accessed

⇓

1 block access for index block
1 — " — for data block

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 |

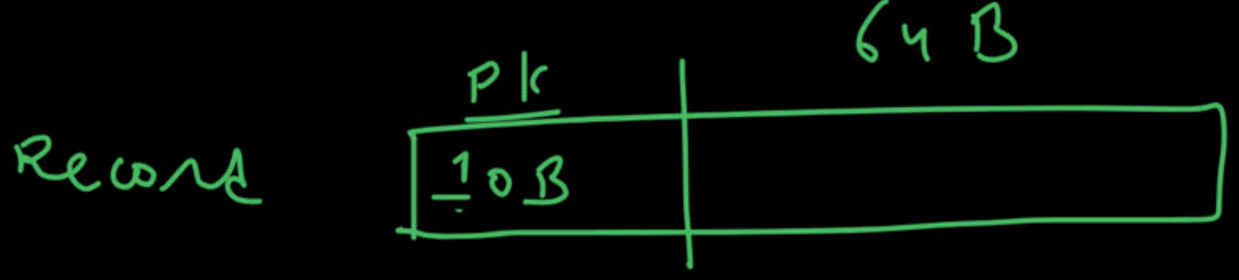
Select * from Students
where rno=12

3 block accessed ⇒ 2 for index
1 for data

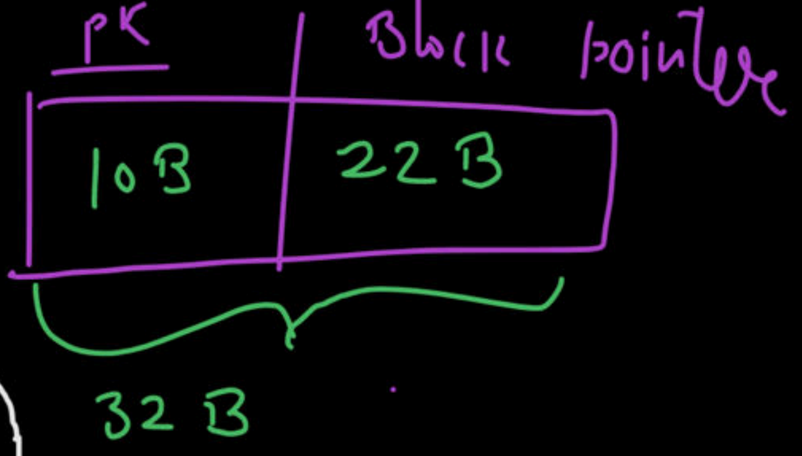
# Question

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? $2^{14}$ blocks
2. The number of blocks required to store the index file for primary indexing? $\Rightarrow 2^{11}$ blocks




1 index record of primary indexing $\Rightarrow$




No. of index Records = no. of blocks to store db file

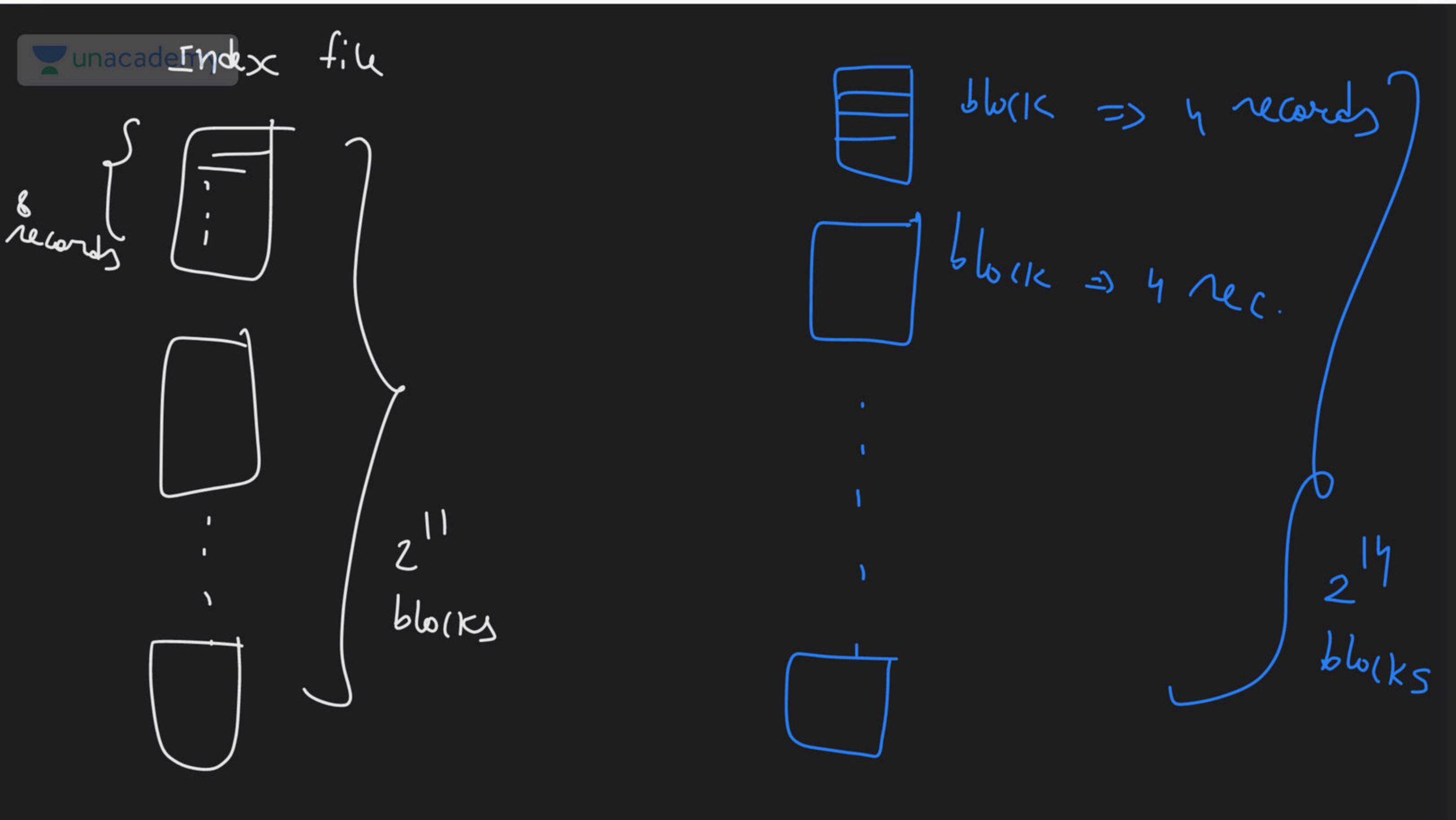
no. of records per block (db) $= \frac{256\ B}{64\ B} = 4$

no. of blocks required to store db file $= \frac{65536}{4} = 16384 = 2^{14}$

---

no. of index records per block $= \frac{256\ B}{32\ B} = 8$

no. of blocks required to store $2^{14}$ index records $= \frac{2^{14}}{8} = 2^{11}$


unacademy
Index file
8 records
$2^{11}$ blocks
block => 4 records
block => 4 rec.
$2^{14}$ blocks

Ques) DB file => $2^{16}$ records

DB record size = 256 Bytes

block size = 16 KBytes

key field => 20 bytes

block pointer = 4 bytes

Primary indexing

① No. of blocks to store DB file = $2^{12}$ Ans

② No. of blocks to store index file = $2^{7}$ Ans.

No. of db records per block = $\frac{16\text{KB}}{256\text{B}} = \frac{2^{14}}{2^8} = 2^6$

No. of blocks to store db file = $\frac{2^{18}}{2^6} = 2^{12}$

---

Index Record size = 20 + 44 = 64 bytes

No. of index records per block = $\frac{16\text{KB}}{64\text{B}} = \frac{2^{14}}{2^6} = 2^8$

No. of blocks to store index file = $\frac{2^{12}}{2^8} = 16$

# Clustering Index

| S_Name | Block Pointer |
|---|---|
| A | B1 |
| B | B2 |
| B | B3 |
| E | 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 | |

# 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 | |

unacademy

# Question

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

# Secondary Key Index

| Rno. | Record Pointer |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |

| | |
|---|---|
| 1 | A |
| 4 | A |
| 9 | A |
| 11 | B |

B1

| | |
|---|---|
| 16 | B |
| 2 | B |
| 12 | B |
| 8 | B |

B2

| | |
|---|---|
| 7 | B |
| 14 | B |
| 3 | C |
| 13 | D |

B3

| | |
|---|---|
| 15 | E |
| 10 | F |
| 5 | F |
| 6 | G |

B4

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. Number of blocks required to store index file

# 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

# Happy Learning.!