File - Organization

Data base is collection of files, each file is a collection of records, each record is a sequence of fields.

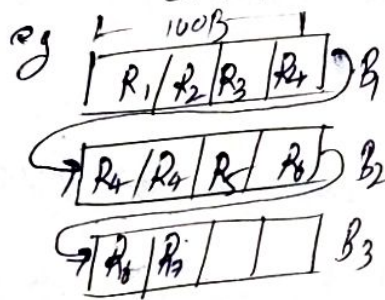| Database | → | Files | → | Records | → | fields |

\# **Blocking factor** - Avg Number of records per block.

\# <u>strategies</u> for storing file of records in to a Blocks

① spanned strategy   ② Unspanned strategy

\# ① **spanned strategy :-** It allow, Partial Part of record can be stored in a Block.

eg.

	100B			
R₁	R₂	R₃	R₄	B

| → R₄ | R₄ | R₅ | R₆ | B₂ |

| → R₆ | R₇ | | | B₃ |

Adv ① No wastage of memory

DS ① Block access increases

NOTE :- It is suitable for variable length records

\# ② **Un-spanned :-** No record can be stored in more than 1 Block.

Adv Block access is reduced

Dis wastage of memory

Note : It is suitable for fixed length record.

\# Organization of Records in a file :-

① **Ordered file Organization :-** All records of file are ordered on some search key value.

Searching :- Binary search.

If you have B data block to access a record. The average \# of block access = $\lceil log_2 B \rceil$

Adv :- Searching is efficient

Diff :- Insertion is Inefficient

② Unordered file organization

All file of records are inserted at where ever the place is available usually at the end of the file.

Searching: Linear Search

Note:- If you have B data block to access record the average no. of block access = $B/2$,

$$\text{worst case} = B$$

Adv:- Insertion is easy

Dis:- Searching is inefficient

Index:- Index are used to improve the searching efficiency. Index is a record consist of two field.

| Key | Block | , Pointer to a Block where key is available
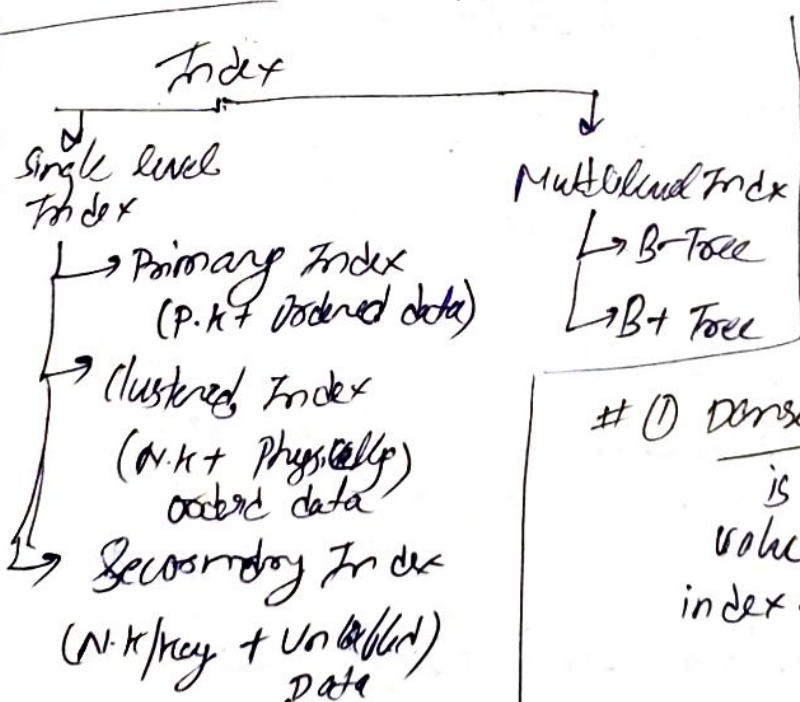
Note:-
① Index is a ordered file
② Searching Binary.
- To access a record using index, the average # of block access $= \lceil \log_2 B \rceil + 1$

Index

- Single level Index
  → Primary Index (P.K + Ordered data)
  → Clustered Index (N.K + Physically ordered data)
  → Secondary Index (N.K/Key + Unordered) Data

- Multilevel Index
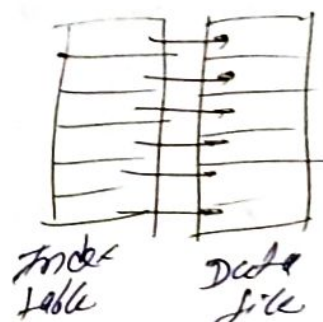 → B-Tree
  → B+ Tree

Classification of Index

① Dense Index
② Sparse Index

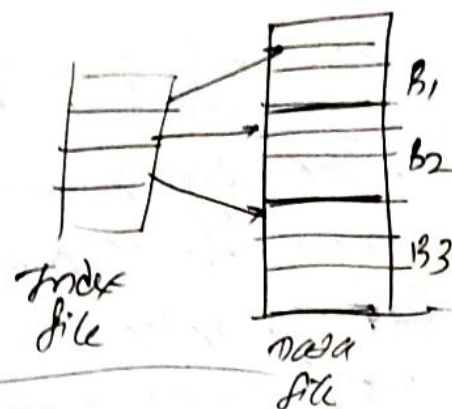# ① Dense Index:- If an index entry is created for every search key value that index is called Dense index.

Index Table   Data file

 # ② Sparse Index :-

- If an Index entry is created only for some search key/value it is called sparse Index.

Index file

Data file

 # Primary Index :- A primary Index is an ordered file. whose records are of fixed size with two field, first field is same. as Primary key of data file & second field is a Pointer to data block where they key is available.
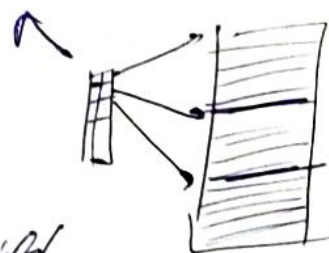
Note :-

① Index entry is created for first record of each block called anchor record.

② The no. of Index entries equal to the no. of blocks is data file.

③ The avg. no. of block access using index = $\lceil \log B \rceil + 1$

④ The type of Index is called sparse.

 # Cluster Index (Nk + Physical Ordered)

Clustered Index is an ordered file with two field the first field is same as the clustering field is called non-key and second field is called Block pointer. cluster Index is created on Data file whose file records are physically ordered on a non-key field.

• Index entry is created for each distinct value of clustering field. The block pointer points to the first Block in which the key is available. Too . Type of Index is dense

④

Secondary Index: (N.k/k + unordered data)

Secondary index Provides a Secondary
means of access a file for which some
Primary index already exists. Secondary index
may be on a non-key or C.k.

* Index entry is created for each record in a data file

 # of Index entries = No. of records in a data file

 # Type of Index is dense.