**FILE ORGANIZATION :**

1. **Sequential**
2. **Indexed**
3. **Direct**

**Indexing** is a data structure technique which allows you to quickly retrieve records from a database file.

An Index is a small table having only two columns. The first column comprises a copy of the primary or candidate key of a table. Its second column contains a set of [pointers](https://www.guru99.com/c-pointers.html) for holding the address of the disk block where that specific key value stored.

An index -

* Takes a search key as input
* Efficiently returns a collection of matching records.

|  |
| --- |
| Block 0 |
| B 1 |
| B2 |
| B3 |
| .. |

Why indexing? RAM Hard disk

CPU

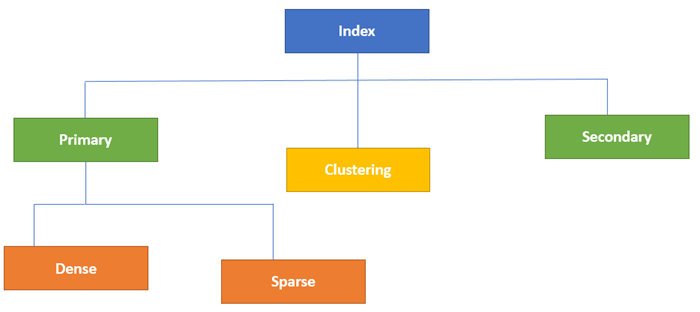
|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

-CPU : MIPS

-I/o cost

**Types of indexing:**

|  |  |  |
| --- | --- | --- |
| **Ordered file** | **Primary index** | **Clustered index** |
| **Unordered file** | **Secondary index** | **Secondary index** |
|  | **Key** | **Non key** |

* 

1. **Primary Index**

* Primary Index is an ordered file which is fixed length size with two fields.
* The first field is the same a primary key and second, filed is pointed to that specific data block.
* In the primary Index, there is always one to one relationship between the entries in the index table.

Index table

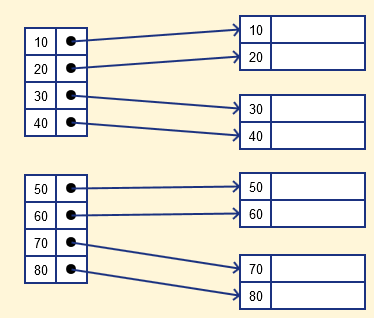
|  |  |
| --- | --- |
| pointer | Key value |
|  |  |
|  |  |

The primary Indexing in DBMS is also further divided into two types.

* Dense Index
* Sparse Index

**a)Dense Index**

* In a dense index, a record is created for every search key valued in the database.
* This helps you to search faster but needs more space to store index records.
* In this Indexing, method records contain search key value and points to the real record on the disk.

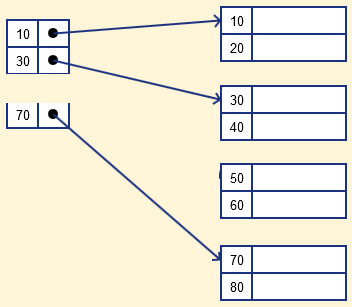
[[](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)

**[b) Sparse Index](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)**

* [It is an index record that appears for only some of the values in the file. Sparse Index helps you to resolve the issues of dense Indexing in DBMS. In this method of indexing technique, a range of index columns stores the same data block address, and when data needs to be retrieved, the block address will be fetched.](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)

[However, sparse Index stores index records for only some search-key values. It needs less space, less maintenance overhead for insertion, and deletions but It is slower compared to the dense Index for locating records.](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)

[Below is an database index Example of Sparse Index](https://cdn.guru99.com/images/1/070119_0833_IndexinginD2.png)

[](https://cdn.guru99.com/images/1/070119_0833_IndexinginD3.png)

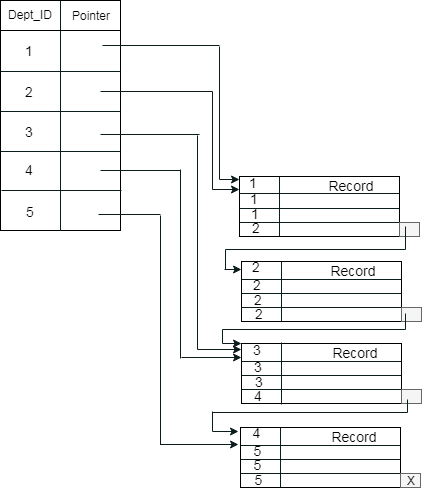
**3)Clustering Index**

* A clustered index can be defined as an ordered data file. Sometimes the index is created on non-primary key columns which may not be unique for each record.
* In this case, to identify the record faster, we will group two or more columns to get the unique value and create index out of them. This method is called a clustering index.
* The records which have similar characteristics are grouped, and indexes are created for these group.

**Characteristic of Clustered Index**

* Default and sorted data storage
* Use just one or more than one columns for an index
* Helps you to store Data and index together
* Operations
* Clustered index scan and index seek
* Key Lookup
* **Sparse in nature**

**Example**: suppose a company contains several employees in each department. Suppose we use a clustering index, where all employees which belong to the same Dept\_ID are considered within a single cluster, and index pointers point to the cluster as a whole. Here Dept\_Id is a non-unique key.



**Complexity: log base 2 N+1+1, where N is number of blocks in index table**

**Secondary Index**

* **Unordered – a) key**

**b) non key**

* In secondary indexing, to reduce the size of mapping, another level of indexing is introduced. In this method, the huge range for the columns is selected initially so that the mapping size of the first level becomes small. Then each range is further divided into smaller ranges. The mapping of the first level is stored in the primary memory, so that address fetch is faster. The mapping of the second level and actual data are stored in the secondary memory (hard disk).

Diagram, engineering drawing

Description automatically generated

Diagram

Description automatically generated