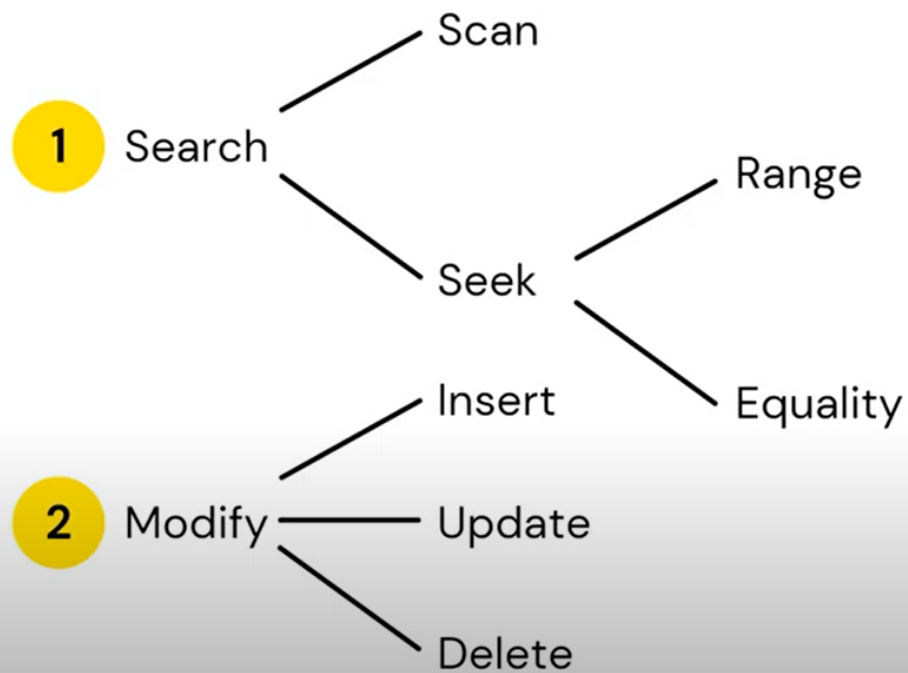
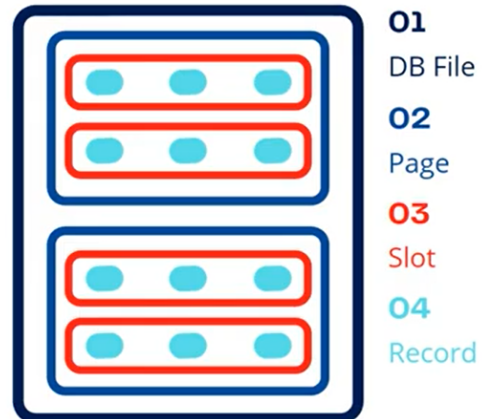


File Organization

File: A collection of pages, each containing a collection of records.

Must support:

- 1 Insert/delete/modify record.
- 2 Read a particular record.
- 3 Scan all records.



Operations we can do with database are as above. Definition of search and seek are as below.

1. Search - retrieving all records
2. Seek - searching for something

File Organization Types

1

Heap File Organization

2

Sequential File Organization

3

Hashing File Organization

Files in databases can be organised according to above methods.

1. Heap file organisation - Records are stored randomly, There is no order or anything. So this is good for insert queries. But not good for other queries.
2. Sequential file organization - Records are sorted. So this is good for range seeks.
3. Hashing file organization - Records are going through a function, and storing in a place according what that function returns. For an example if you want to store 34, you are going to send 34 through a function which will calculate its remainder ($34\%2$). The function will return 0. So you're going to store 34 in the section where everyone who are remainder 0. This is good for equality seeks because if you want to retrieve 34, you 34 is available in 0 remainder section.

INDEXES

Indexes

- An index on a file speeds up selections on the search key fields for the index.
- Any subset of the fields of a relation can be the search key for an index on the relation.
- Search key is not the same as key (minimal set of fields that uniquely identify a record in a relation).

Characteristics

- Indexes provide fast access
- Indexes takes space -Need to be careful in creating only useful indexes
- May slow-down certain inserts/updates/ deletes (maintain indexes)

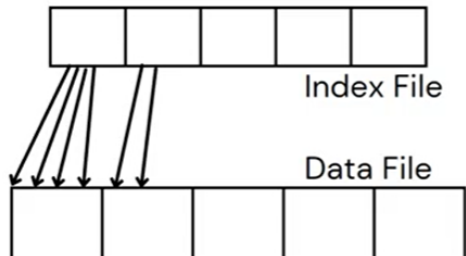
Alternatives for Data Entry in Index



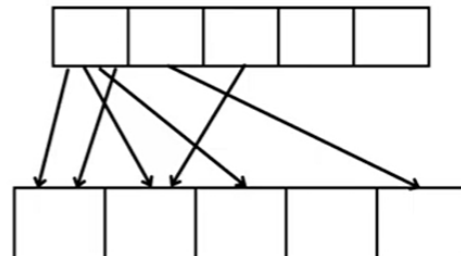
1. Alternative 1 - records and indexes are in the same file and linked together.
2. Alternative 2 - records and indexes are stored in separate files. For an example, lets just say you want to find 27. Indexes are sorted unlike record file, so you easily can find 27 in index file. With the index record you have stored recordId (record id = page number + slot number) of the real record. So you can find the page and slot where the real record is available.
3. Alternative 3 - Same as alternative 2 but duplication values are not stored in the index file.

Properties of Indexes

1 Clustered vs Unclustered Index



Can have at most one clustered index per table



Cost of retrieving data records through index varies greatly based on whether index is clustered or not!

One table can contain more than one index. But if you want clustered indexes, one table can only contain one clustered index.

Normally when you want to retrieve a record, you take the page/block where the record is stored to the memory. Lets just say you want to do a range selection. If your index is clustered, you can take the whole page/record into memory which all the values related to the range are available. But if it is unclustered, you will have to take different different pages/blocks into memory in order to access the records.

