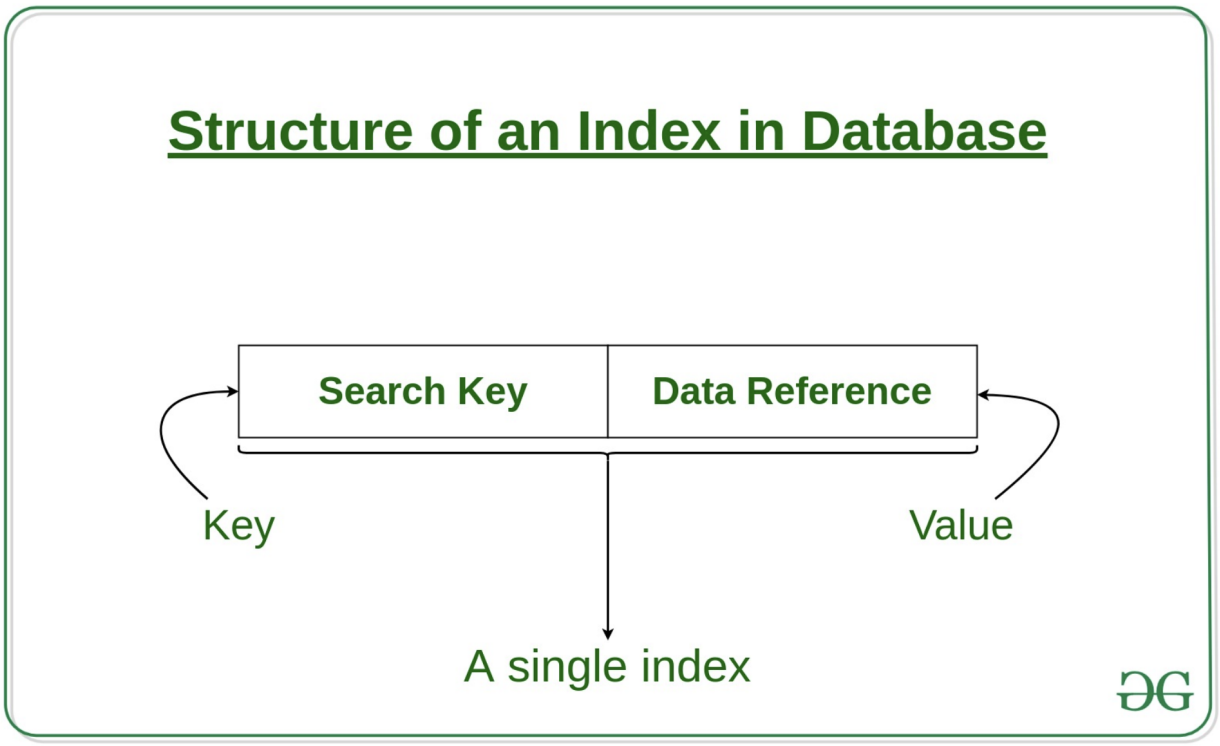
4: Indexing

What is Indexing?

Started with 🔃

> https://www.youtube.com/watch?v=SxHX1T53n_A

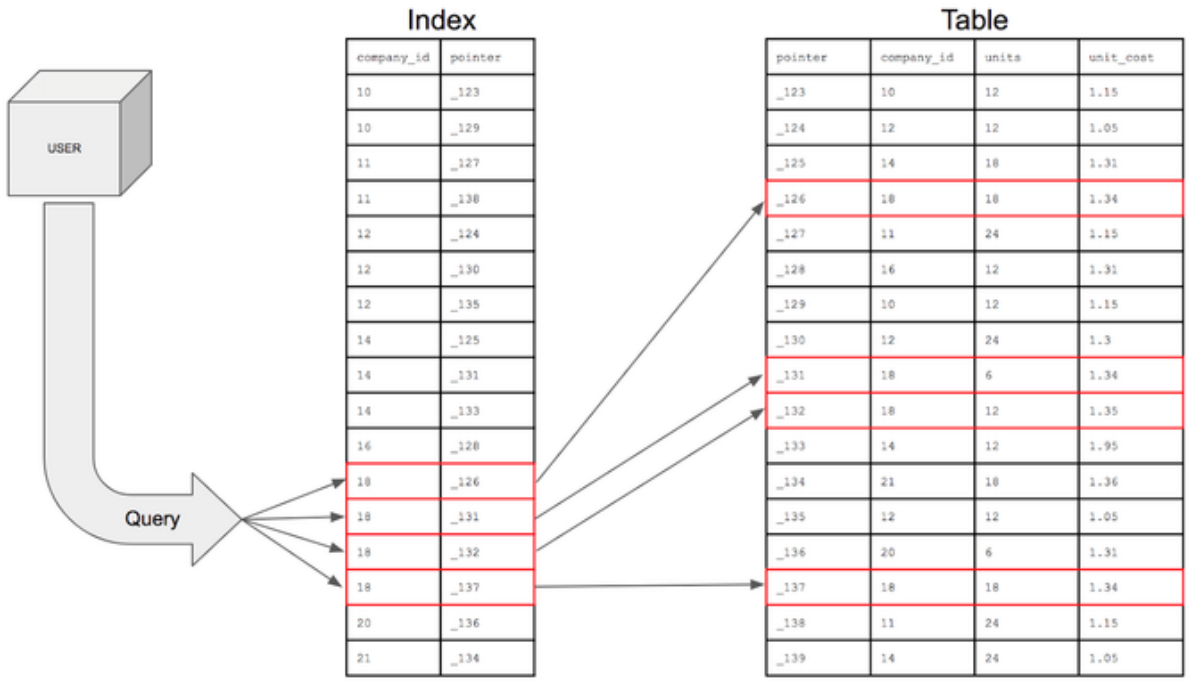
- **What?:** Indexing is done to make **search faster** in the main database file (regardless data is sorted or unsorted)
- **Is Indexing necessary?** No, its optional. It is not a primary means to access the record, its a secondary means. You can access the data sequentially also.
- Searching for the right block takes more time than searching for the actual record within the block. Hence, we solve primarily for reducing the time to find the block.
- The index file does not store all the data, only 2 things:
  - The anchor or the key term eg. Photosynthesis
  - The block address where it can be found
- **AN INDEX FILE IS ALWAYS SORTED. SO WE CAN EASILY APPLY BINARY SEARCH.**
- The index file is now meta-data.
- The larger the memory, the larger will be the search time. Hence, we want the index file to have lesser memory than the main database file where the record is found. (The index has to be smaller than the book).
- Indexing doesn't affect the main database in any way.
- **Sparse indexing:** When not all records in all the blocks are not a part of the index file. Just a few records from each block are used. Some records from each block are chosen.
- **Dense indexing:** When all records from each block are chosen for the index file.



What are types of Indexing?

https://www.youtube.com/watch?v=O-Mbn6VI1zc

An index file can further be indexed (Done when even the index file is considerably large). It’s called multilevel indexing.



1. **Primary indexing:** When the main file is sorted. An example of sparse indexing.
2. **Clustered indexing:** The main file is sorted but we use non-key records/non-key attributes as anchors in the index file. Non-key attributes can be repeated. (eg. Roll no of students are key attributes but non-key attributes are age, gender, and address which can be repeated).
3. **Secondary indexing:** When the main file is not sorted.

What are the disadvantages of Indexing?

https://stackoverflow.com/questions/29842622/why-and-where-to-use-indexes-pros-and-cons

While it (mostly) speeds up a select, it slows down inserts, updates and deletes because the database engine does not have to write the data only, but the index, too. An index need space on hard disk (and much more important) in RAM. An index that can not be held in RAM is pretty useless. An index on a column with only a few different values doesn't speed up selects, because it can not sort out much rows (for example a column "gender", which usually has only two different values - male, female).