

Indexed file organization

Dr. R. Nageshwar Rao

Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology, Vellore.

October 1, 2024

Indexed Sequential Access Method (ISAM)

ISAM is a popular method for indexed file organization.

In ISAM:

- The primary file is stored in a sequential manner based on a primary key.
- There's a static primary index built on the primary key.
- Overflow areas are designated for insertion of new records, which keeps the main file in sequence.
- Periodically, the overflow area can be merged back into the main file.

Indexing

- Indexing involves creating an auxiliary structure (an index) to improve data retrieval times.
- Just like the index in the back of a book, a database index provides pointers to the locations of records.

Structure of Index We can create indices using some columns of the database.

Search Key	Data Reference
.	.
.	.
.	.

- The search key column contains a copy of the table's candidate (or primary) key. The primary key values are saved in sorted order so that the related data can be quickly accessible.
- The data reference column contains a group of pointers that point to the disk block where the value of a specific key can be found.

Single-level Index: A single index table that contains pointers to the actual data records.

Multi-level Index: An index of indexes. This hierarchical approach reduces the number of accesses (disk I/O operations) required to find an entry.

Dense and Sparse Indexes:

- In a dense index, there's an index entry for every search key value in the database.
- In a sparse index, there are fewer index entries. One entry might point to several records.

Primary and Secondary Indexes:

- A primary index is an ordered file whose records are of fixed length with two fields. The first field is the same as the primary key, and the second field is a pointer to the data block. There's a one-to-one relationship between the number of entries in the index and the number of records in the main file.
- A secondary index provides a secondary means of accessing data. For each secondary key value, the index points to all the records with that key value.

Clustered vs. Non-clustered Index:

- In a clustered index, the rows of data in the table are stored on disk in the same order as the index. There can only be one clustered index per table.
- In a non-clustered index, the order of rows does not match the index's order. You can have multiple non-clustered indexes.

Bitmap Index: Used mainly for data warehousing setups, a bitmap index uses bit arrays (bitmaps) and usually involves columns that have a limited number of distinct values.

B-trees and B+ trees:

- Balanced tree structures that ensure logarithmic access time.
- B+ trees are particularly popular in DBMS for their efficiency in disk I/O operations.

Benefits and Drawbacks of Indexing

Benefits of Indexing: Faster search and retrieval times for database operations.

Drawbacks of Indexing:

- Overhead for insert, update, and delete operations, as indexes need to be maintained.
- Additional storage requirements for the index structures.