File Organization: Organization of records in file, Data Dictionary Storage.
Indexing and Hashing: Basic Concepts, Ordered Indices, B+ Tree Index files, B tree index files – Static Hashing – Dynamic Hashing – Comparison of Indexing with Hashing.

File Organization

Introduction

- The **File** is a collection of records.
- A **file** is organized logically as a sequence of records.
- A file is named a collection of related information that is recorded on secondary storage such as magnetic disks, magnetic tapes, and optical disks.
- Each file is also logically partitioned into fixed-length storage units called **blocks**, which are the units of both storage allocation and data transfer.
- Most databases use block sizes of 4 to 8 kilobytes by default, but many databases allow the block size to be specified when a database instance is created.
- A block may contain several records; the exact set of records that a block contains is determined by the form of physical data organization being used.

Organization of Records in Files

File organization contains various methods and each methods have pros and cons.
The programmer decides the best-suited file organization method according to his requirement.
Types of file organization are as follows:

- o   Sequential file organization
- o   Heap file organization
- o   Hash file organization
- o   B+ file organization
- o   Indexed sequential access method (ISAM)
- o   Cluster file organization

Data Dictionary Storage

- Relational schemas and other metadata about relations are stored in a structure called the **data dictionary** or **system catalog**.
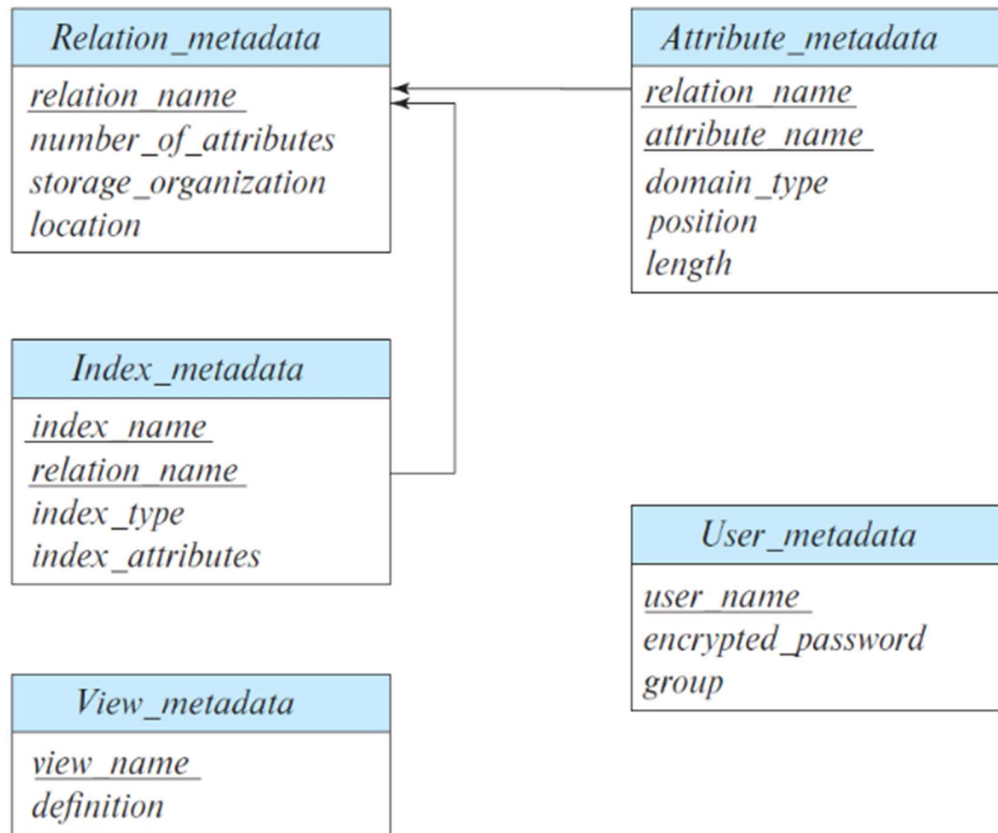- "Data about data" are referred to as **metadata**

Figure Relational schema representing part of the system metadata.

- Among the types of information that the system must store are these:
  - ➢ Names of the relations
  - ➢ Names of the attributes of each relation
  - ➢ Domains and lengths of attributes
  - ➢ Names of views defined on the database, and definitions of those views
  - ➢ Integrity constraints (e.g., key constraints).

- Many systems keep the following data on users of the system:
  - ➢ Names of users, the default schemas of the users, and passwords or other information to authenticate users
  - ➢ Information about authorizations for each user.

- The data dictionary may also note the storage organization (heap, sequential, hash, etc.) of relations, and the location where each relation is stored:
  - ➢ If relations are stored in operating system files, the dictionary would note the names of the file (or files) containing each relation.
  - ➢ If the database stores all relations in a single file, the dictionary may note the blocks containing records of each relation in a data structure such as a linked list.

Indexing and Hashing

Basic Concepts

- An index is a type of **data structure**. It is used to locate and access the data in a database table quickly.
- An index for a file in a database system works in much the same way as the index in this textbook.
- Indexing is used to optimize the performance of a database by minimizing the number of disk accesses required when a query is processed.
- Indexing mechanisms used to speed up access to desired data.
- For example, index of book, author catalog in library etc.
- There are two basic kinds of indices:
  1. **Ordered indices:** Based on a sorted ordering of the values.
  2. **Hash indices**: Based on a uniform distribution of values across a range of buckets.
- We shall consider several techniques for ordered indexing. Each technique must be evaluated on the basis of these factors:
  - ✓ **Access types:** The types of access that are supported efficiently.
  - ✓ **Access time:** The time it takes to find a particular data item, or set of items.
  - ✓ **Insertion time:** The time it takes to insert a new data item.
  - ✓ **Deletion time:** The time it takes to delete a data item.
  - ✓ **Space overhead:** The additional space occupied by an index structure.

- The structure of an index in the database management system (DBMS) is

Search key	Data Reference

Fig: Structure of Index

- An attribute or set of attributes used to look up records in a file is called a **search key**.
- The data reference contains a group of pointers that point to the disk block where the value of a specific key can be found.