Database Management System

Files, File Organization and File Structures :—

(1) A file is a collection of bytes stored as an individual entity. All data on disk is stored as a file with an assigned file name, that is unique with in the directory it resides in.

(2) To the computer, a file is nothing more than a series of bytes. The structure of a file is known to the software that manipulates it.

(3) A file contains data that is needed for information processing. These data is about entities.

(4) An entity is anything about which information can be stored.

(5) An attribute is a characteristic of interest about an entity. The values of the attributes describe a particular entity.

(6) An instance of the entity is represented by a set of specific values for each of the attributes.

(7) Each attribute of an entity is represented in storage by a data item. A data item is the elementary unit in data storage. Data items are usually grouped together to describe an entity.

(8) The data representation in storage of each instance of an entity, is commonly called as a record. A collection of related records, is called a file.

Note: In physical storage, a record has physical storage location or address associated with it.

Creation and maintainence of stored data is one of the primary functions of information processing.

Files and the relationships among them, is the main key point of any file-processing environment.

There are mainly 3 types of files available —.

1. Master files —
2. Transaction files —.
3. Report files —.

Master files → This is the file of relatively permanent information about entities. These files are used as a source of reference data for transaction processing and accumulate information based on the transaction data.

Transaction files → This is the collection of records describing the activities or transactions by organization. It is created as a result of processing transaction & preparing transaction documents. Transaction files are also used to update the details in the master file.

Report files → This is the file created by the extracting data to prepare a report.

~~What are the operations that can be executed for the processing of records in a file?~~

There are mainly two kinds of file operations —:

1. Retrival operations
2. Update operations.

* Retrival operations do not change the contents of the file; it only locates records in the file matching certain specific criteria.
* Update operations on the other hand, change the file, by modifying the records, deleting the records and inserting new records.

In both update and retrieval operations — one or more records have to be located for retrieval, modification or deletion based on a selection condition or ~~Sear~~ search criteria.

There are several representative operations that are used in most systems.

1. Find (locate) → The goal of this operation, is to locate the record or records that satisfy the search criteria. The block that contains the record is transferred to the main memory and the records are searched.

2. Read → Read is sometimes referred to as Get. In this operation, the contents of the records are copied from the memory to a program variable or work area.

③ ReadNext (or GetNext) → Searches for the next record that matches the selection condition and when found the contents of the record are copied to the program variable or work area.

④ Modify → Also known as update. This comman modifies the field values of the current record and then writes the modified record back to the disk.

⑤ Insert → Inserting a new record to the file.

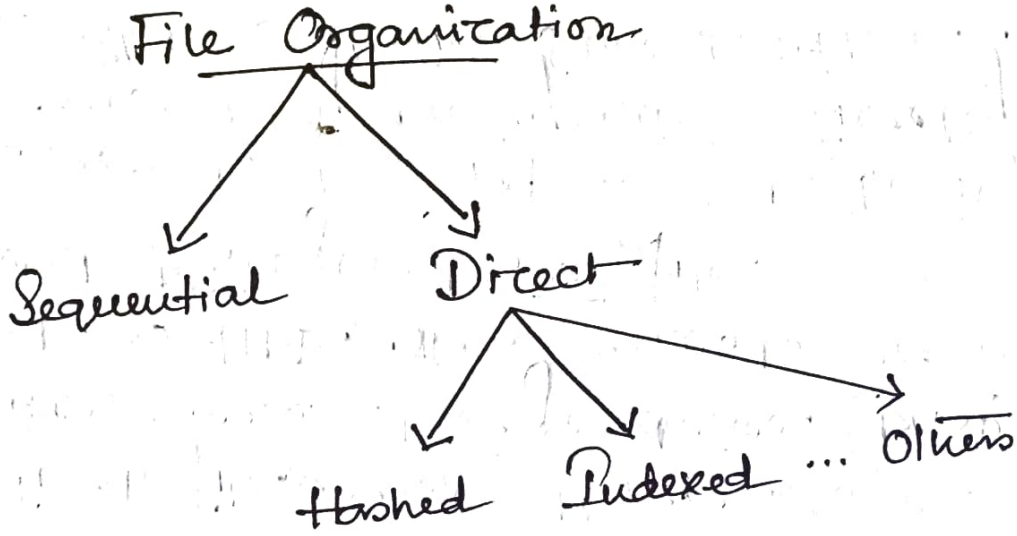
⑥ Delete → Deletes the current record and updates the file on the disk to reflect the deletion.

What are the operations that are required for the processing of records in files?

There are five operations are required for the processing of records in files —.

① File Creation
② Record Location (finding the record)
③ Record Creation
④ Record Deletion
⑤ Record Modification.

Each file organization is more efficient in some operations than others.



What is Sequential file organization?

In Sequential file organization, records are stored in some pre-defined sequence, that means one after another. and one feild reffered as Primary Key usually determins the sequence or order.

NOTE :- Primary key → A candidate key, selected to uniquely identify all other attribute values. in any given row & it can not contain any null entries.

What are the advantages and dis-advantages of Sequential file organization?

Advantages :-

① Magnetic tape, the least expensive method of Secondary storag, can be used.

② It contains a fast and efficient method, for huge amount of data.

P.T.O

③ It is simple in design. It requires no much effort to store the data.

④ This method is used for report generation or statistical calculations.

Dis-advantages :—

① Can't move to a particular record.

② Sorted file method (for sorted values) takes more time and space for sorting the records.

What do you mean by Direct file organization?

In a direct file organization, data may be organized in such a way that they are scattered throughout that disk and what may appear in a random order. This form of organization supports direct access or random access.

① Direct file access is also known as Random file access or relative file organization.

② In direct file, ~~access~~ all records are stored in direct access storage device (DASD), such as HDD, where records are normally placed throughout the file.

③ The records does not need to be in sequence because they are updated directly and rewritten back in same location.

④ This file organization is useful for immediate access to large amount of information. It is used in accessing large databases.

⑤ It is also called as hashing.

Explain – what do you mean by direct file processing?

Direct access systems do not search the entire file, rather they move directly (or nearly directly) to the needed record. To be able to do this, they must have some way of determining where the record is stored. This is the principal challange of direct processing. Several different strategies are used to find a record, including relative addressing, hashing and indexing.

What is relative addressing?

The simplest method of finding a record is called relative addressing, in which a record's primary key is associated with a specific physical location (storage). The contents of the record are stored in this address.

When the record need to be retrived, the user enters the key and the disk operating system associates this number with the appropriate location on the disk.

Wasted space is often considered one of the biggest problem with relative addressing.

#1 What is hashing?

Hashing (also known as Randomizing) is a method of determining the physical location of a record. In this method, the record key (such as emp id, emp no, post no etc.) is processed mathematically and another number is computed that represents the location where the record will be stored.