What is File? File is a collection of records related to each other. The file size is limited by the size of memory and storage medium.

There are two important features of file:

1. File Activity

2. File Volatility

File activity specifies percent of actual records which proceed in a single run.

File volatility addresses the properties of record changes. It helps to increase the efficiency of disk design than tape.

File Organization File organization ensures that records are available for processing. It is used to determine an efficient file organization for each base relation. For example, if we want to retrieve employee records in alphabetical order of name. Sorting the file by employee name is a good file organization. However, if we want to retrieve all employees whose marks are in a certain range, a file is ordered by employee name would not be a good file organization.

Types of File Organization

There are three types of organizing the file:

1. Sequential access file organization

2. Direct access file organization

3. Indexed sequential access file organization

1. Sequential access file organization

• Storing and sorting in contiguous block within files on tape or disk is called as sequential access file organization.

• In sequential access file organization, all records are stored in a sequential order. The records are arranged in the ascending or descending order of a key field.

• Sequential file search starts from the beginning of the file and the records can be added at the end of the file.

• In sequential file, it is not possible to add a record in the middle of the file without rewriting the file.

Advantages of sequential file

• It is simple to program and easy to design.

• Sequential file is best use if storage space.

Disadvantages of sequential file

• Sequential file is time consuming process.

• It has high data redundancy.

• Random searching is not possible.

2. Direct access file organization

• Direct access file is also known as random access or relative file organization.

• In direct access file, all records are stored in direct access storage device (DASD), such as hard disk. The records are randomly placed throughout the file.

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

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

• It is also called as hashing.

Advantages of direct access file organization

• Direct access file helps in online transaction processing system (OLTP) like online railway reservation system.

• In direct access file, sorting of the records are not required.

• It accesses the desired records immediately.

• It updates several files quickly.

• It has better control over record allocation.

Disadvantages of direct access file organization

• Direct access file does not provide back up facility.

• It is expensive.

• It has less storage space as compared to sequential file.

3. Indexed sequential access file organization

• Indexed sequential access file combines both sequential file and direct access file organization.

• In indexed sequential access file, records are stored randomly on a direct access device such as magnetic disk by a primary key.

• This file have multiple keys. These keys can be alphanumeric in which the records are ordered is called primary key.

• The data can be access either sequentially or randomly using the index. The index is stored in a file and read into memory when the file is opened.

Advantages of Indexed sequential access file organization

• In indexed sequential access file, sequential file and random file access is possible.

• It accesses the records very fast if the index table is properly organized.

• The records can be inserted in the middle of the file.

• It provides quick access for sequential and direct processing.

• It reduces the degree of the sequential search.

Disadvantages of Indexed sequential access file organization

• Indexed sequential access file requires unique keys and periodic reorganization.

• Indexed sequential access file takes longer time to search the index for the data access or retrieval.

• It requires more storage space.

• It is expensive because it requires special software.

• It is less efficient in the use of storage space as compared to other file organizations.

File Organization

o The File is a collection of records. Using the primary key, we can access the records. The type and frequency of access can be determined by the type of file organization which was used for a given set of records.

o File organization is a logical relationship among various records. This method defines how file records are mapped onto disk blocks.

o File organization is used to describe the way in which the records are stored in terms of blocks, and the blocks are placed on the storage medium.

o The first approach to map the database to the file is to use the several files and store only one fixed length record in any given file. An alternative approach is to structure our files so that we can contain multiple lengths for records.

o Files of fixed length records are easier to implement than the files of variable length records.

Objective of file organization

o It contains an optimal selection of records, i.e., records can be selected as fast as possible.

o To perform insert, delete or update transaction on the records should be quick and easy.

o The duplicate records cannot be induced as a result of insert, update or delete.

o For the minimal cost of storage, records should be stored efficiently.