

File MANAGEMENT

File System Architecture

- Bits are grouped together to represent data in computer system. Bits are grouped basically into two patterns; bytes and characters.
- Bytes are made up of 8 bits. Size of character differs in each computer system. Usually size of a character is 2 bytes and there are 2^8 or 256 possible character represented in a character set.
- File management is organised into a number of levels. Figure 9.3 shows the logical structure of file.
- The various elements of file structure are :
 1. Field
 2. Record
 3. File
 4. Database

Field

- A field is the basic element of data. It is usually a group of characters. An individual field contains a single value, such as a student's name, age, marks etc. It is characterized by its length and data type (e.g. ASCII string, decimal).
- Depending on the file design, fields may be fixed length or variable length.
- In case of variable length, a field often consists of two or three subfields; the actual value to be stored, the name of the field, and, in some cases, the length of the field.

Record

- A record is a collection of related fields that can be treated as a unit by some application program. For example, a student record would contains fields like name, age, marks, address, date of birth and so on.
- A field that uniquely identifies a record is know as record key or key field.
- Record can be of fixed length or variable length depending upon the application program and file structure.

File

- A file is a collection of records.
- A file is treated as a single entity by users and applications and may be referenced by name.
- Size of file may be fixed or variable.
- Files have file names and can be created, deleted and moved from one directory to another.
- Access control restrictions usually apply at the file level. That is, in a shared system, users and programs are granted or denied access to entire files.
- Some file systems are structured only in terms of fields, not records. In such a case, a file is a collection of fields.

Database

- It is the highest level in file system and is defined as a collection of files.
- Data elements of database have relationship between them that is explicitly designed *i.e.* designed manually by the user.
- A database may contain all of the information related to an organization or project such as a business or a scientific study.

Aman	21	...	15 Jul 1991
Rohit	19	...	23 Sep 1993
Charanjeet	25	...	8 Apr 1986
.			
.			
.			
Harshan	20	...	24 Dec 1992

Layered Architecture

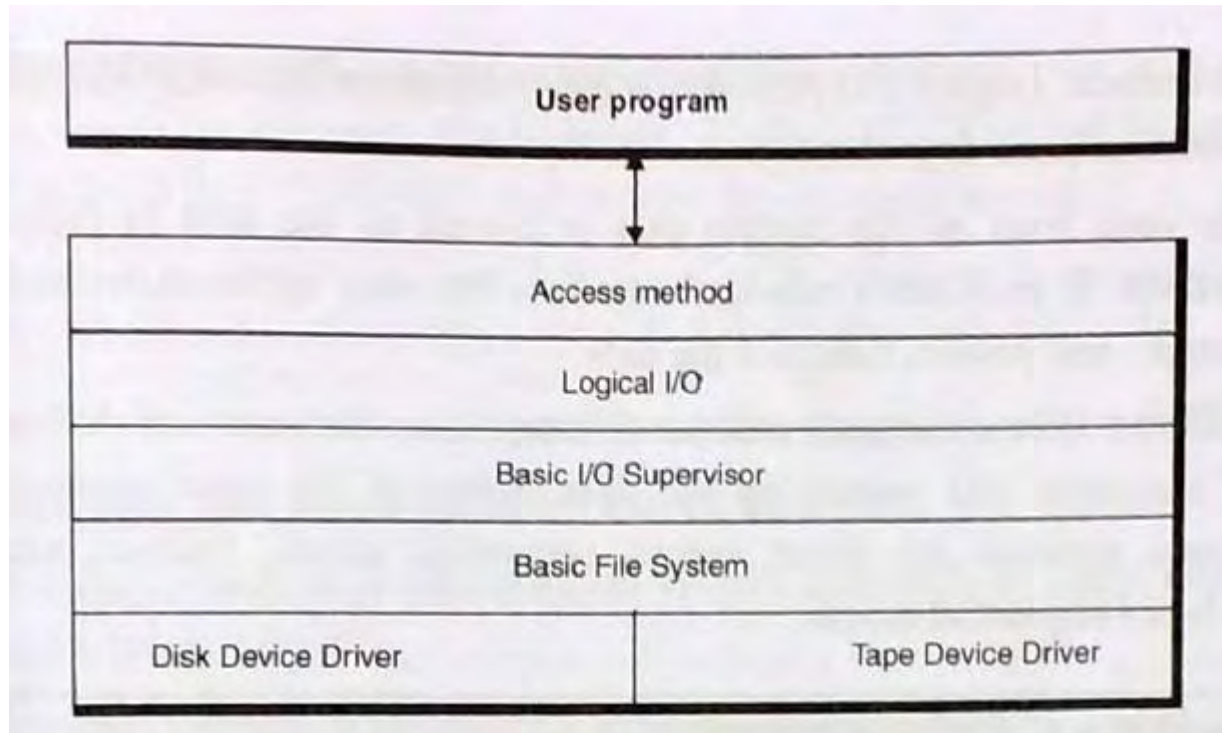
- At the lowest level, **device drivers** communicate directly with peripheral devices or their controllers or channels.
- A device driver is responsible for starting I/O operations on a device and processing the completion of an I/O requests. For file operations, typical device controlled are disk and tape drives. Device drivers are usually considered to be part of the operating system.
- The next level is known as the **basic file system** or the **physical I/O** level. It is the primary interface with the environment outside of the computer system. It deals with the blocks of data that are exchanged with disk or tape system. Thus, it is concerned with the placement of these blocks on the secondary storage device and on the buffering of those blocks in main memory. It does not understand the content of the data or the structure of the files involved. The basic file system is often considered part of the operating system.
- **The basic I/O supervisor** is responsible for all file I/O initiation and termination. At this level control structures are maintained that deal with device I/O, scheduling and file status.
The basic I/O supervisor selects the device on which file I/O is to be performed, based on the particular file selected.
It is also concerned with scheduling disk and tape accesses to optimize performance.
I/O buffers are assigned and secondary memory is allocated at this level. The basic I/O supervisor is part of the operating system.
- **Logical I/O** enables users and applications to access records. The basic file system deals with blocks of data whereas logical I/O module deals with the

file records. Logical I/O provides a general purpose record I/O capability and maintains basic data about files.

- The next level of file system that is closest to the user is called **access method**. It provides a standard interface between applications and the file systems and devices that hold the data.

Different access methods reflect different file structures and different ways of accessing and processing the data. Some of the most commonly used access methods are direct access, sequential access, indexed access and indexed sequential access.

Layered Architecture



Logical File System

- The logical file system is concerned with mapping the structure of the logical records onto the linear byte string view of a file provided by the physical file system.
- The logical file system provides facilities for accessing data stored in files. It is known as *access methods*.

9.4.1 Access Methods

Files are used to store data. The information present in the file can be accessed by various methods.

Thus, the way of retrieving data from a file is known as **access method**.

Different systems use different access methods. The various access methods used are:

1. Sequential access
2. Direct access
3. Indexed access
4. Indexed sequential access

1. Sequential Access

- It is the simplest and most commonly used access method.
- In sequential access, information in the file is accessed in the order it is stored in the file *i.e.* one record after the other.
- Thus, the various records are read sequentially one after the other in an order, starting at the beginning to the end of the file.

- The various records cannot be read randomly out of order *i.e.* we cannot skip any record in between. For example, reading of 34th record followed by 5th record and then 1st record is not possible in sequential access.
- In this way, the file pointer is automatically advanced in the read operation which tracks the I/O location.
- In case of write operation, the new content is appended to the end of file.
- Sequential file can be reset to the starting of the file.
- In some system, it is also possible to shift n records of some integer value backward or forward.
- Sequential access is convenient when the storage medium is magnetic tape, rather than a disk.

Advantages of Sequential Access Method

1. It is one of the simplest method of file access and easy to implement.
2. There is no need for any storage space identification.
3. It uses disk and memory efficiently.
4. It also allows data to be stored on many types of media, in a device independent manner.
5. Errors in the files remain localized.
6. It is also economical and easier to organise and maintain.

Disadvantages of Sequential access method

1. Searching a record is a time consuming affair as it allows only sequential access. To search a record at n^{th} location entire file needs to be processed.
2. New records can only be added to the end of a file.
3. It requires the transactions to be sorted in a particular sequence before processing.
4. There is high data redundancy.
5. It is not possible to handle the random enquiries.

2. Direct Access

- In direct access method it is possible to access the records of a file in any order (*i.e.* out of order). For example, if we are reading block 13, we can read block 46 after this and then block 20.

- The various records can be read or write randomly. There is no restriction on the order of reading or writing for a direct access file. Hence it is also known as **random access method** or relative.
- In this way the records can be accessed by key, rather than by position.
- This method is also known as **relative access method** and is typically used in systems that use direct access storage devices (DASD) like disks rather than magnetic tapes.
- In direct access method, the various records or blocks of a file are numbered for reference purpose.
- In order to perform read or write operation, we specify block number where read or write operation is to be performed. Thus, an instruction **read n** will read block number n.
- Thus a user provides a block number to an operating system. This number is known as **relative block number**. It is an index to starting of the file. So, in a file, the first relative block number is 0, the next relative block number is 1 and so on.
- Direct access method is important for many applications, for example, database systems. In a banking application, a customer may want to look up his current balance. This can be done by locating this customer's record using his account number as a key, rather than sequentially reading the records for thousand of the customers before this customer's record is located and read.

Advantages of direct or random access method

1. It results in faster access of desired record. As a result retrieval process is not so time consuming as in direct access.
2. It does not require records to be sorted before processing.
3. It allows faster updating of several files.
4. It also allows the random record deletion and insertion.
5. It is best suited for online transaction processing system like online reservation system.

isadvantages of direct or random access method

1. This method requires backup facility as records are directly updated.
2. It requires expensive direct access storage devices such as hard disk to store records.

3. It is less efficient as compared to sequential file organisation in terms of usage of storage space.
4. Data may be accidentally erased or overwritten unless special precautions are taken.
5. It is not possible to access such a file sequentially.
6. Only one key is used.

3. Indexed Access

- This method uses the combination of sequential and random access techniques.
- In this method, an index is created for the file.
- This index contains pointer for various blocks of a file, just like an index in the back of a book.
- Here various records are stored randomly on a DASD such as disks using a primary key. The data is accessed either sequentially or randomly using index.
- If we want to find a record of a file, first the index is searched and then the pointer from index is used to access that file. In this way, a required record is found.
- When the size of a file is large, its index itself becomes too large. In such a situation a further an index is created for the index file. The primary index or main index contains the pointers to secondary index files, which would point to the actual data items. Thus another index is created for an index.

Advantage of indexed access method

1. This method allows records to be processed both sequentially and randomly in a efficient manner.
2. Accessing of records is faster and less time consuming but requires index table to be properly organised.

Disdvantage of indexed access method

1. It requires a lot of storage space because of the presence of index.
2. As compared to other access method, indexed access is less efficient in terms of usage of storage space.
3. The indexed files have to be reorganised from time to time to get rid of

deleted records and improve performance that gets gradually decreased with addition of new records.

4. It is also expensive as it requires special software and extra storage for index.

Indexed Sequential Access

- This method of file access is an advancement over sequential access.
- In this method index of records is maintained in a sequential file.
- This index provide the facility of searching records randomly.
- Index to a file is a simple sequential file that contains index as its record.
- Each entry in index file contains two fields : key field and a pointer pointing to some record in the main file.
- To find a specific field in main file, index is searched for the highest key value that is equivalent to the desired value. The pointer related to key field starts searching the record at the location it indicates.
- The main advantage of this methods is that its reduces the searching time to a great extent keeping the sequential nature of file.
- To increase the efficiency, multiple level of index are possible in this method. Lower level of index is considered as the sequential file and its higher level is considered as its index file.