Unit 4

File Management and Data processing

Files :- files are the most basic unit of data that users can store on a disk. A file is a common storage unit in a computer. All programs and data are contained in a file, and the computer reads and writ files. In every program, image, video, song and document are stored in a file.

One can create, open, move, save and delete files. It is possible to move a file from one ~~folder~~ folder to another. We can also download files from other network and internet.

Types of files :- Text file, programs files, images files, music files etc. we know the file types by the extension of file name. g:word documents, they are having extension .doc

Name of files :- ~~There are different types of files, depending~~ on type of information they

Name of files :- The name can have up to 255 characters it contains letters, numbers, blank spaces, special characters like dashes, underlines etc. but there is a group that can't be used (" ; ? , \ , > , < , |)

File Extensions files :- are identified by a "short" "extension" at the end of their name. for example ABC.JPG is JPEG file image, ABC.DOC is a microsoft word document file and ABC.EXE is an executable Application of windows.

Folders :- A folder is a collection of multiple files. A folder holds one or more files, and it can be empty with just a name. Folders provide a method for organizing files similar to a mania file folder containing paper documents in a file cabinet. Folders are of great help in organising files. For ex:- a person can store all photos in a folder name photos, while he can store videos in another similar named folder. He can then place all such folder in folder called My documents. Folders can store other folders called sub folders. Folders were also called "directories" in operating system

Difference between File and Folder

File	Folder
1) File store data, whether text, music, film	1) Folder store files and other folders.
2) ~~Folders~~ Files usually ~~take no~~ have a size ranging from a few bytes to kilobytes to gigabytes in case of files containing music and video content. (as word files)	2) Folders usually take no space in the harddrives.
3)	

File Access and Storage Methods.

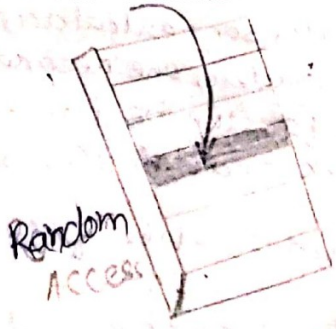
An access method defines the technique that is used to store and retrieve data. An access method is a function of a mainframe operating system that enables the access of data on a disk, tape or other external devices. Access methods have their own data structures to organize data, system-provided programs (macros) to define data sets, and utility program to process data set.

In computing, an access method is a program or hardware mechanism that moves data between the computer and an outlying device such as hard disk or a display terminal.

The term is sometimes used to refer to the mechanics of placing or locating specific data at a particular place on a storage medium and then writing the data or reading it. It is also used to describe the way the data is located within a larger unit of data such as data set or file.

There are two types of access method.

(1) Random access
(2) sequential access.



Random (Direct) Access.

Following are some of key features of direct access method

(*) A file is made up of fixed-length logical records that allows programs to read and write records rapidly in no particular order. Thus, we may read block 14, then read block 53, then write block 7. There are no restrictions on the order of reading or writing for a direct-access file.

→ The direct-access method is based on a disk model of a file, since disks allow random access to any file block

→ Direct-access files are of great use for immediate access of large amounts of information. Database are often of this type.

→ For the direct-access method, the file operations must be modified to include the block numbers as a ~~parm~~ parameter.

→ The block number provided by the user to the OS is normally a relative block number.

→ A relative block number is an index relative to the beginning of the file.

→ Thus, a first relative block of the file is 0, the next is 1, and so on, even though the actual absolute disk address of the block may be 14703 for the first block and 3192 for the second.

→ The use of relative block numbers allow the OS to decide where the file should be placed (called the allocation problem) and helps to prevent the user from accessing portions of the file system that may not be part of her file.

Sequential Access :-

Following are some of the key features of sequential access method.

→ The simplest access method is sequential access. Information in the file is processed in order, one record after other.

→ This mode of access is by far the beginning current position most common. for ex :- editors & compilers usually access in this fashion.

→ Reads and writes make up the bulk of the operations on a file.

→ A read operation read next portion of file and

automatically advance a file pointer, which tracks the I/O location.

→ Similarly, the write operation write next appends to the end of the file and advances to the end of the newly written material (the new end of file).

Indexed Sequential Access Method (ISAM)

An indexed sequential method is static, hierarchical, disk based index structure that enable both (single-dimensional) range and membership queries on an ordered data file. ISAM initially stores records sequentially and permits both sequential and random processing. The features that provide the flexibility are indexes to locate a correct cylinder and track and keys to locate a record on a track.

The records of the data files are stored in sequential order according to some data attributes. Since ISAM is static, it doesn't change it structure if records are added or deleted from the data file. Should new records be inserted into the data file, they are stored. Each index defines a different ordering of the records. An employee database may have several indexes, based on the information being sought. for ex:- A name index may order employees alphabetically by last name, while a department index may order employees by their department. A key is specified in each index. for an alphabetical index of employee name, the last name field would be the key.

ISAM (other non sequential file organization methods) differs from sequential organization is that the record keys in an indexed file must be unique, this is a system requirement not just a programming practice. Indexed file is typically a master file. Also, there is a clear difference between updating a sequential file and updating an indexed file. when you update a sequential file, you rewrite the entire file, this practice leaves the original file as a convenient ~~file~~ backup in case the job must be rerun. when you update an indexed file, the system rewrite records in the file directly in place thereby not providing automatically backup. To create backup you ~~y~~ periodically copy file to another ~~dif~~ device ISAM helps to locate records in file, second the system store new added records in special ~~reserved~~ reserved overflow areas.

Data capture :-

Data capture is the process of identification and extraction of data from a scanned document, often to be sent to a workflow for routing and action as part of business method. (letter, invoices, email, fax etc)

OCR (Optical character Recognition)

OCR technology has the ability to successfully capture machine produced characters in preset zone are full pages. Depending upon the capabilities of the particular OCR product, this can be used to capture low or high volumes of data.

ICR (Intelligent character Recognition)

ICR is the computer translation of hand printed and written characters. Data is entered from hand printed forms through a scanner, and the image captured data is then analysed and translated by ICR software.

Bar code Recognition :->

Dependent upon the type of barcode that is used the amount of metadata that can be included in high as it level of recognition. The application of single or multiple bar codes to particular document types such as POD (proof of delivery), membership forms, applicati forms, gift aid etc can dramatically increase the effectiveness of a business process.

Intelligent Document Recognition. (IDR)

The level of capability is dependent upon the individual product. These applications are used to capture metadat from documents that is rules based.

ex:- the product will identify postcodes, logos, key words VAT registration numbers and, through an ongoing learning process, capture information and multiple document types.

Data storage

Data storage is the holding of data in an electro magnet form for access by a computer processor.

There are two main kinds of storage

Primary storage is data that is held in RAM and oth

memory devices that are built into computers. secondary storage is data that is stored on external storage devices such as hard disk, tapes, CD etc.

following are some main devices for data storage

Hard disk- often called disk drive, hard drive or hard disk drive. A hard disk drive is a device used to store large amount of digital information in computer and related equipments like ipod and games consoles such as the X box 360 and PS3 and provides relatively quick access to large amounts of data. Hard disk drives are used to store operating system, software and working data.

Floppy disk :- is a type of magnetic disk memory which consists of a flexible disk with a magnetic coating. Almost all floppy disks for personal computers now have a capacity of 1.44 megabytes. floppy disk are readily protable, and are very popular for transferring software from one PC to another. Therefore they are slow compared to hard disk and lack storage capacity.

Tape storage :- Tape is used as an external storage medium. It consists of a loop of flexible celluloid like material that can store data in the form of electromagnetic charges. A tape drive is a device that position, write froms and reads to the tape. A tape cartridge is a protectively encased tape that is portable.

Optical disk :- An optical disc is a storage medium that can be written to and read using low powered laser beam. A laser reads these dots, and the data is converted to an electrical signal, finally converted into the original data.

Compact Disc Recordable "CD-R" :- has become a universal data storage medium worldwide. CD's tend to be used for large files which are too big for a floppy disc to hold such as music and general animation. CDs are becoming increasingly popular for music recording and file storage or transfer between personal computers. CDR discs are write once media. This means data can't be erased or re-recorded upon once used.

Memory sticks / pendrive USB flash drives :- are typically removable and rewritable, much smaller than a floppy disk. storage capacities typically ranges from 64 MB to 64 GB. USB flash drives are portable

storage devices than that of floppy disk. They are more compact, operate faster, hold more data, have a more durable design and operate more reliably due to their lack of moving ~~data~~ parts.

Flash memory cards :- A memory card or flash memory card is a solid state electronic flash memory data storage device used in digital cameras, handheld and mobile computers, telephones, music players, video game. Now a days PC have inbuilt memory card slots. memory stick, compact Flash. SD etc.

Data Processing and Retrieval.

Data Processing

Data must be processed in order to convert it into information. For this purpose, different operations may be performed on a data.

Defn :- A sequence of operations on data to convert it into useful information.

The data processing can be ~~done~~ accomplished through following methods.

(1) Manual Data Processing
(2) Mechanical Data Processing
(3) Electronic Data Processing.

(1) Manual Data Processing :- data is processed manually without using any machine or tool to get required result. In manual data processing, all the calculations and logical operations are performed manually on data. The method of data processing is very slow and errors may occur in the output. ~~Ex~~ In an educational institute for ~~ex~~ marksheets, fee receipts and other financial calculations are performed by hand.

(2) Mechanical Data Processing :- data is processed by using different devices like typewriters, mechanical printers or other mechanical devices. This method of data processing is faster and more accurate than manual data processing. Examination boards and printing press use mechanical data processing devices frequently.

(3) Electronic Data Processing :- is a modern technique of processing of data. The data is processed through computer. Data and set of instructions are given to the computer as input and computer automatically processes the data according to the given set

of instructions. The computer is also known as electronic data processing machine. This method of processing data is very fast and accurate. eg:- In banks customer a/c are maintained through computer

Data Retrieval

Data is one of the most important assets of any business. Data recovery refers to the whole process of salvaging this lost data that is corrupted, failed, damaged or inaccessible. Data recovery is the process of restoring data that has been lost, deleted, corrupted or made inaccessible for any reason.

Lost files can occur because of the below possibilities

(1) File ~~was~~ was mistakenly deleted.
(2) File was corrupt and deleted by scandisk
(3) Another program deleted the file
(4) file is password protected.

Following are some of the data recovery method.

(1) Physical damage to storage devices :→

Different failures can cause physical damage to your storage media. This could range from breaking of tapes to the metallic substrate or dye layers of CD ROM's being scratched. While physical loss occur, some form of data loss is likely to follow. Although it may not be possible for end users to repair physical damage, data recovery techniques may involve replacing parts in the harddisk or recovering every readable bit from the surface through a specialized disk imaging procedure

(2) Media errors and corrupt partitions and file systems :-

In some cases, media error or damage to the file system or partition table can make the data on a harddisk to be unreadable. Specialized data recovery software like Test Disk can be used to repair the damaged file system or partition table. This make it possible to recover all the original data.

(3) Online data recovery :→

This is another popular method of data recovery used to restore deleted or lost data. It is method of data recovery that is performed over the internet without having the computer or drive in possession while using this method your data is regularly backed up online and therefore if any data is lost we are able to recover the last backed up version of data.