

- **What is File Structure and what are the main roles of file structure?**

- A **File Structure** is a combination of representations for data in files and of operations for accessing the data , Minimize number of trips to the disk AND Grouping related information to get everything with only one trip to the disk

- **Data processing involves..... , and**

- Storage of data , Organization of data , Access to data

- **Compare between main memory and secondary memory?**

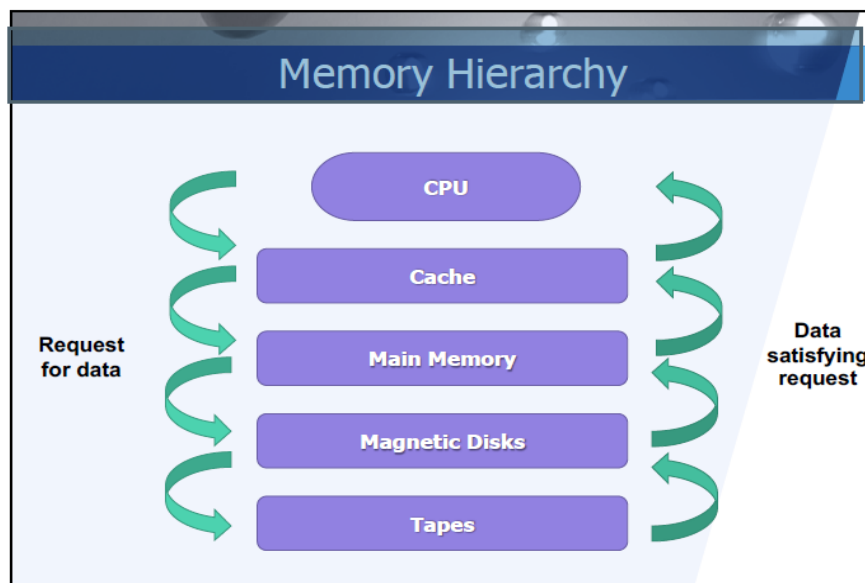
- Main Memory :**

- Fast (since electronic)
 - Small (since expensive)
 - Volatile (information is lost when power failure occurs)

- Secondary Storage:**

- Slow (since electronic and mechanical)
 - Large (since cheap)
 - Stable, persistent (information is preserved longer)

- **Describe the memory hierarchy ?**



- **What are the main characteristics of a good file structure design?**

- Fast access to great capacity
 - Reduce the number of disk accesses
 - collecting data into buffers, blocks or buckets
 - Manage growth by splitting these collections

- **You should know the rules of naming files ?**

Every file has a name and might also have a file extension. When you save a file, you must provide a valid file name that adheres to specific rules, referred to as file-naming conventions.

Each operating system has a unique set of file-naming conventions.

WINDOS

Case sensitive	No
Maximum length of file name	File name, path, and extension cannot exceed 255 characters
Spaces allowed	Yes
Numbers allowed	Yes
Characters not allowed	* \ : < > " / ?
File names not allowed	Aux, Com1, Com2, Com3, Com4, Con, Lpt1, Lpt2, Lpt3, Prn, Nul

MAC

Case sensitive	No
Maximum length of file name	File name, path, and extension cannot exceed 255 characters
Spaces allowed	Yes
Numbers allowed	Yes
Characters not allowed	: (the colon)

- **Define file extension with examples?**

A **file extension** (sometimes referred to as a file name extension) is an optional file identifier that is separated from the main file name by a period, as in Paint.exe.

File extensions provide clues to a file's contents. For example .exe files (Windows) and .app files (Mac OS) contain computer programs.

- **Define file format with examples ?**

The term **file format** refers to the organization and layout of data that is stored in a file.

The format of a file usually includes a header, data, and possibly an end-of-file marker.

For example, graphics data can be stored in file formats such as BMP, GIF, JPEG, or PNG.

- **What .hlp, .com, .avi and ods file formats refer to ?**

Type of File	Description	Extension
Batch file	A sequence of operating system commands executed automatically when the computer boots	.bat
Configuration file	Information about programs the computer uses to allocate the resources necessary to run them	.cfg .sys .mif .bin .ini
Help	The information displayed by on-screen Help	.hlp
Temporary file	A sort of scratch pad that contains data while a file is open, but is discarded when you close the file	.tmp
Support program	Program instructions executed along with the main .exe file for a program	.ocx .vbx .vbs .dll
Program	The main executable files for a computer program	.exe .com .app (Mac OS)

- **Define file header?**

file header is a section of data at the beginning of a file that contains information about a file, such as the date it was created, the date it was last updated, its size, and its file type.

- **What is file management, and what are the advices or tips that make the file management easier and efficient?**

File management encompasses any procedure that helps you organize your computer-based files so that you can find and use them more efficiently

File Management Tips :

- Use descriptive names
- Maintain file extensions.
- Group similar files.
- Organize your folders from the top down.
- Consider using default folders.
- Use Public folders for files you want to share.
- Do not mix data files and program.

Don't store files in the root directory.
Access files from the hard disk.
Follow copyright rules.
Delete or archive files you no longer need.
Back up!

- **What are the types of secondary storage devices with examples?**

Direct Access Storage Devices (DASDs): Magnetic Disks (hard disks): High capacity, low cost .

Optical Disks (CD-ROM, DVD-ROM): Read-only or write-once, holds large data, inexpensive.

Serial Devices: Magnetic Tapes: Used for sequential access.

- **What are the types of storage technologies, with examples of each?**

There are three common storage technologies for personal computers:

Magnetic: hard drives ,,Optical: CDs, DVDs ,,Solid State: SSDs

- **Describe the structure of hard disk in details?**

Disk Structure:

- Disk blocks: Storage units
- Tracks: Concentric circles on the disk
- Platters: Stacked disks
- Cylinder: Aligned tracks across platters
- Sectors: Smallest addressable units
- Disk heads: Read/write mechanisms
- Disk Controller: Manages operations
- Seek Time: Time for head positioning
- Rotational Delay: Time for sector alignment

- **You should study well how to calculate disk capacity as given in lecture ٢, slides(٣٠,٣١ and ٣٢)**

- **Define cluster and extent?**

A cluster is a fixed number of contiguous sectors.

An extent is a group of contiguous clusters.

- **When to choose large cluster size and small cluster size?**

- **Large clusters** are used for large, sequential files (e.g., bank updates).
- **Small clusters** are used for small or randomly accessed files (e.g., airline reservations).

- **What is the cost of disk Access?**

Seek time: Time to move the access arm to the correct track.

Rotational delay: Time for the disk to rotate so the desired sector is under the read/write head.

Transfer time: Time required to move data to/from the disk surface.

Average Total Time = Average Seek Time + Average Rotational Delay + Transfer Time.

- **Write short notes about Optical Storage Technology ?**

Optical storage uses light and dark spots on the disc surface to represent data, with dark spots (pits) and light areas (lands).

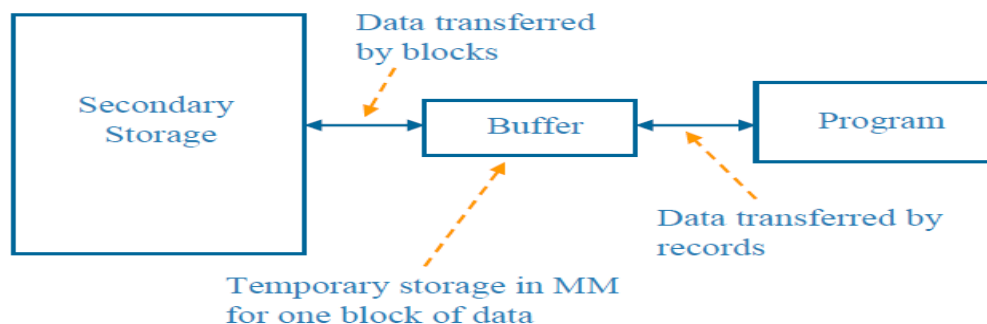
Optical technologies are categorized into three types: Read-only (ROM): Cannot be written to, data is pre-recorded.

Recordable (R): Data can be written once, after which it becomes read-only.

Rewritable (RW): Data can be written, erased, and rewritten multiple times

- **What is buffering, describe with a chart system I/O buffer?**

Buffering means working with large chunks of data in main memory so the number of accesses to secondary storage is reduced.



- What are the different buffer strategies in details?

Buffering Strategies:

Single Buffering: Only one buffer is used, leading to potential bottlenecks.

Double Buffering: Two buffers are used, allowing I/O and CPU operations to overlap.

Multiple Buffering: Uses several buffers to handle overlapping operations.

Buffer Pooling: A pool of buffers managed using strategies like Least Recently Used (LRU) for replacement when all buffers are occupied.

- Define field, record and key, then compare between the two types of keys?

Field: a data value, smallest unit of data with logical meaning.

Record: A group of fields that forms a logical unit.

Key: a subset of the fields in a record used to uniquely identify the record

Primary Key: must identify records uniquely ,It is not dataless, Has a canonical form.

Secondary Key : Does not identify records uniquely , It is not dataless, Has a canonical form

- Explain in detail the different methods for organizing fields (example, advantages and disadvantages)?

Methods for organizing fields are:

- Fix the length of fields

Advantages:

- Easy to Read/Store

Disadvantages:

- Waste space with padding

- Begin each field with a length indicator

Advantages:

- Easy to jump ahead to the end of the field

Disadvantages:

- Long fields require more than 1 byte to store length (Max is 255)

- Separate the fields with delimiters

Advantages:

- May waste less space than with length-based

Disadvantages:

- Have to check every byte of field against the delimiter
- Use a “Keyword=Value” Expression to identify fields

Advantages:

- Fields are self describing allows for missing fields

Disadvantages:

- Waste space with keywords (20% or more of the file's space could be taken up by the keywords)
-

○ **Write the five methods of organizing records?**

Methods for organizing records are:

- Make records a predictable number of bytes (fixed length records)
 - Make records a predictable number of fields
 - Begin each record with a length indicator
 - Use an index to keep track of addresses
 - Place a delimiter at the end of each record
-

Describe in detail the methods for record access?

1- **Record Keys** : Key: a subset of the fields in a record used to uniquely identify the record.

Primary Key: A key that uniquely identifies a record.

Secondary Key: Other keys that may be used for search

In general not every field is a key

Keys correspond to fields, or combination of fields, that may be used in a search

2- Sequential Search

Search for a record matching a given key

Look at records sequentially until matching record is found.

The work required to search sequentially for a

record in a file with n records is proportional to n: It takes at most n comparisons; $n/2$ on average.

Time is in $O(n)$ for n records

3- Direct Access

Being able to seek directly to the beginning of the record.

Time is in $O(1)$ for n records.

Possible when we know the Relative Record Number (RRN)

- **What are the main strategies of record deletion?**

Record Deletion and Storage Compaction

Deleting Fixed-Length Records and Reclaiming Space Dynamically

Deleting Variable-Length Records

- **What are the main strategies of placement of new record?**

First-Fit Strategy

Best-Fit Strategy

Worst-Fit Strategy

- **Show an example of fragmentation and defragmentation?**

Slide 29 lec five

- **What is the motivation for data compression?**

Make optimal use of limited storage space

Save time and help to optimize resources

- **Define entropy and write the formula?**

Entropy is the measure of information content in a message

The entropy $H(x)$ of the symbol x is:

$$H(x) = - p(x) \cdot \log_2 p(x)$$

- **List data compression methods?**

lossless methods (text or program) : Run-length , Huffman , Lemple-Zif.

lossy methods (image , audio , video) : JPEG , MP3 , MPEG.