

Contents

- 1 Introduction to File Structures
- 2 History of File Structures Design
- 3 File Basics
- 4 File Management

7

7

Introduction to File Structures

8

8

File Structure

- ❖ A **File Structure** is a combination of representations for data in files and of operations for accessing the data.
- ❖ A File Structure allows applications to read, write and modify data.
- ❖ It might also support finding the data that matches some search criteria or reading through the data in some particular order.



9

9

Data Processing

- ❖ Data processing from a computer science perspective involves:
 - Storage of data
 - Organization of data
 - Access to data
- ❖ This will be built on your knowledge of:



10

10

Data Structures vs. File Structures

- ❖ **Both involve:**
 - Representation of Data
 - Operations for accessing data

- ❖ **Difference:**
 - Data Structures deal with data in **main memory**
 - File Structures deal with data in **secondary storage device (File)**



Main Storage (Memory)



Secondary Storage

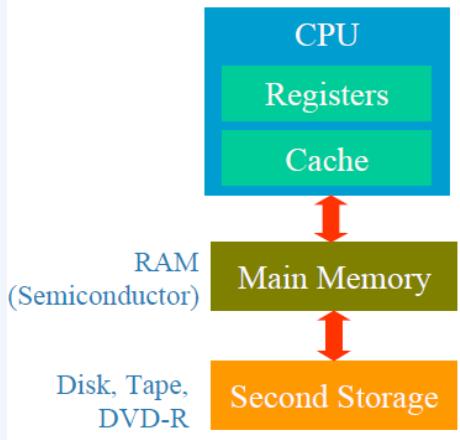
Data Structures

File Structures

11

11

Computer Architecture



The diagram illustrates the computer architecture hierarchy. At the top is the **CPU**, which contains **Registers** and **Cache**. A double-headed red arrow connects the CPU to **Main Memory** (represented by a green box). Another double-headed red arrow connects Main Memory to **Second Storage** (represented by an orange box). To the left of Main Memory is **RAM (Semiconductor)**, and below Second Storage are **Disk, Tape, DVD-R**.

Differences

<ul style="list-style-type: none"> — Fast — Small — Expensive — Volatile
<ul style="list-style-type: none"> — Slow — Large — Cheap — Stable

12

12

Main Memory vs. Secondary Storage

- ❖ **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)



13

13

How Fast ...?

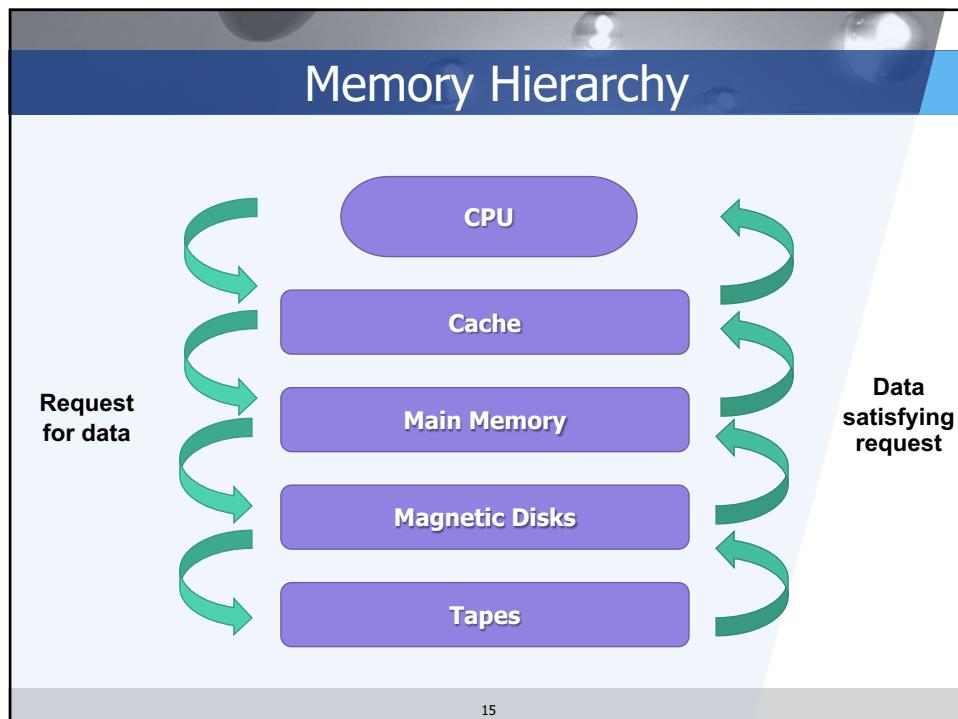
- ❖ Typical times for getting information
 - **Main memory:** ~120 nanoseconds = 120×10^{-9}
 - **Magnetic Disks:** ~30 milliseconds = 30×10^{-6}

- ❖ An analogy keeping same time proportion as above
 - Looking at the [index of a book](#): 20 seconds versus
 - [Going to the library](#): 58 days



14

14



15

Main Goal of This Course

- ❖ Minimize number of trips to the disk in order to get desired information (Ideally get what we need in one disk access or get it with as few disk access as possible).
- ❖ Grouping related information so that we are likely to get everything we need with only one trip to the disk (e.g. name, address, phone number, account balance).

Locality of Reference in Time and Space

In order to achieve these goals, we need good file structure design

16

16

Good File Structure Design

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



17

17

History of File Structures Design



18

18

History of File Structures Design

1. In the beginning... it was the **tape**
 - Sequential access
 - Access cost proportional to size of file
 - [Analogy to sequential access to array data structure]

2. **Disks** became more common
 - Direct access
 - [Analogy to access to position in array]
 - **Indexes** were invented
 - list of keys and points stored in small file
 - allows direct access to a large primary file

Great if index fits into main memory.

As file grows we have the same problem we had with a large primary file

19

19

History of File Structures Design

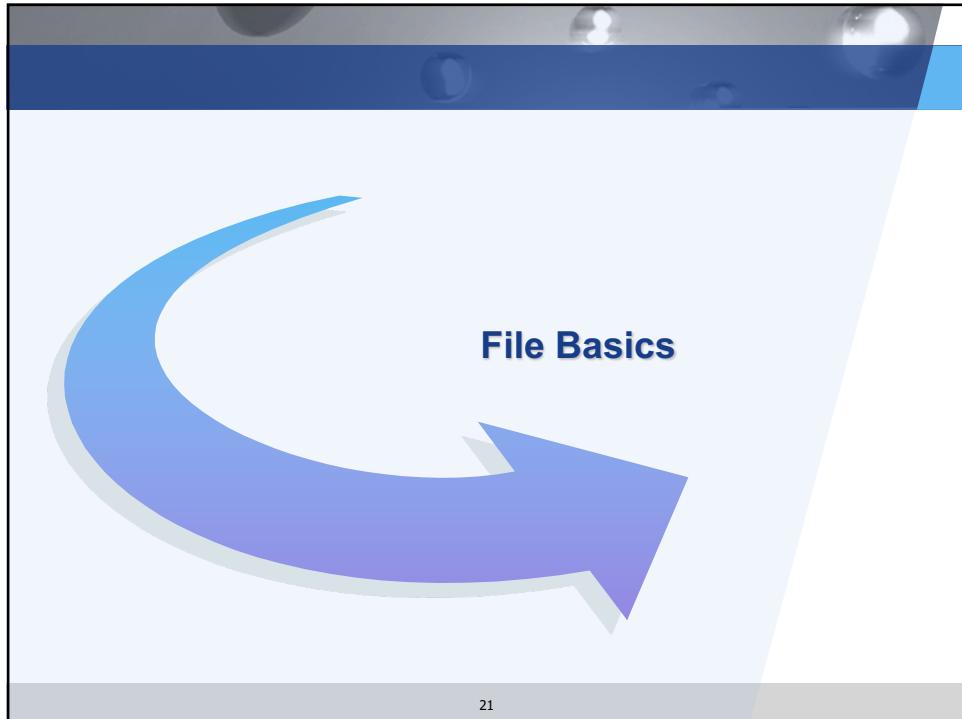
3. Tree structures emerged for main memory (1960's)
 - Binary search trees (BST's)
 - Balanced, self adjusting BST's: e.g. AVL trees (1963)

4. A tree structure suitable for files was invented:
 - **B trees** (1979) and **B+ trees**
 - good for accessing millions of records with 3 or 4 disk accesses.

5. What about getting info with a single request?
 - **Hashing Tables** (Theory developed over 60's and 70's but still a research topic)
 - Good when files do not change too much in time.
 - **Expandable, dynamic hashing** (late 70's and 80's)
 - One or two disk accesses even if file grows dramatically

20

20



21

Computer File

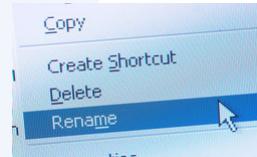
- ❖ A **computer file**, or simply a file, is defined as a named collection of data that exists on a storage medium, such as a hard disk, CD, DVD, or USB flash drive.
- ❖ A file can contain a group of records, a document, a photo, music, a video, an e-mail message, or a computer program.

	Name	Date modified	Type
vacation.png	20/04/2012 1:40 AM	PNG File	
pic-23353.png	20/04/2012 1:31 AM	PNG File	
pic-34366.png	20/04/2012 1:32 AM	PNG File	
pic-48642.png	20/04/2012 1:40 AM	PNG File	
pic-54735.png	20/04/2012 1:32 AM	PNG File	
pic-63858.png	20/04/2012 1:34 AM	PNG File	
pic-65687.png	20/04/2012 1:34 AM	PNG File	
pic-95421.png	20/04/2012 1:34 AM	PNG File	

22

Rules for 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.



23

23

Rules for Naming Files



Microsoft Windows

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 OS

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)

24

24

Rules for Naming Files

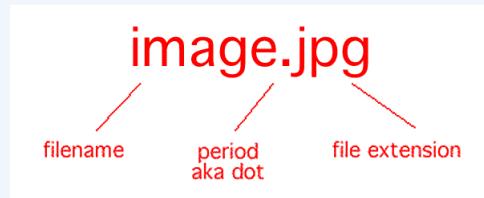
- ❖ Some operating systems also contain a list of reserved words that are used as commands or special identifiers. You cannot use these words alone as a file name.
 - ❖ You can also use spaces in file names. That's a different rule than for e-mail addresses, where spaces are not allowed.

25

25

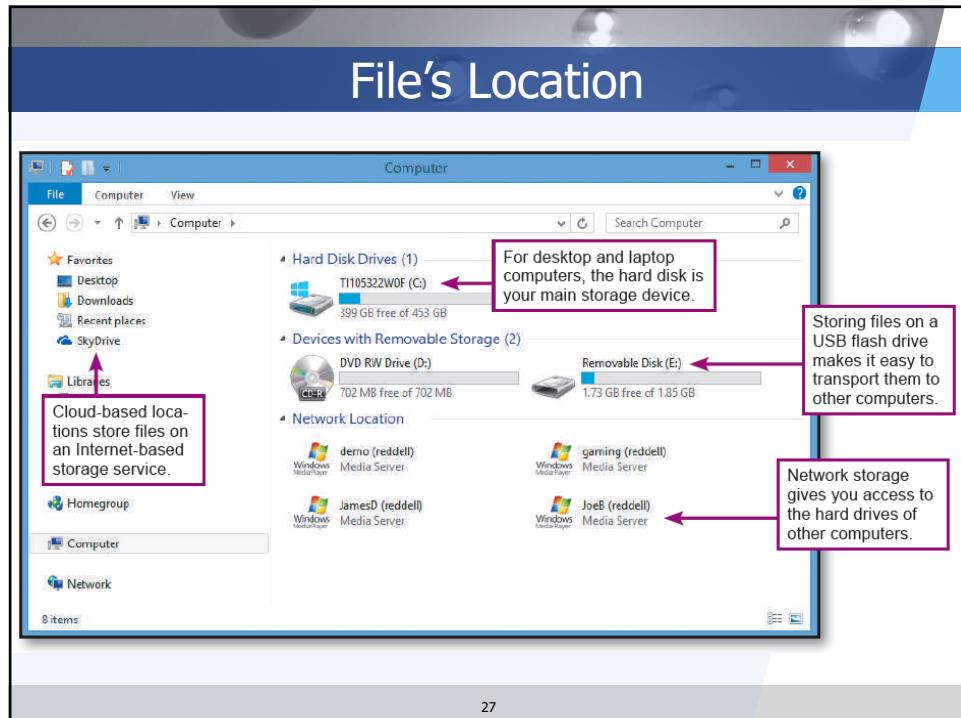
File Extension

- ❖ 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**.



26

26



27

File's Location

- ❖ To determine a file's location, you must first specify the device where the file is stored.
- ❖ You can store files on a hard drive, removable storage, a network computer, or cloud-based storage.
- ❖ When working with Windows, each local storage device is identified by a device letter. The main hard disk drive is referred to as drive C:
- ❖ Macs do not use drive letters. Every storage device has a name. The main hard disk is called Macintosh HD, for example.

28

28

File's Location

- ❖ A **disk partition** is a section of a hard disk drive that is treated as a separate storage unit.
- ❖ Every storage device has a directory containing a list of its files.
- ❖ The main directory is referred to as the **root directory**. On a PC, the root directory is identified by the device letter followed by a backslash (C:\).
- ❖ A root directory can be subdivided into smaller lists. Each list is called a **subdirectory**

29

29

File's Location

- ❖ A computer file's location is defined by a **file path** (sometimes called a **file specification**), which on a PC includes the **drive letter**, **folder(s)**, **file name**, and **extension**.
- ❖ Suppose that you have stored an MP3 file called Marley One Love in the Reggae folder on your hard disk.



30

30

File Format

- 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.
- A 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.

BMP File Format	GIF File Format
File header	File header
Bitmap header	Logical screen descriptor
Color palette	Global color table
Image data	Local image descriptor
	Local color table
	Image data
	End-of-file character

31

31

File Format

- Music files are stored differently than text files or graphics files; but even within a single category of data, there are many file formats.
- For example, graphics data can be stored in file formats such as BMP, GIF, JPEG, or PNG.
- Although a file extension is a good indicator of a file's format, it does not really define the format.

**Don't
FORGET!**

32

32

File Format: Executable File Extensions

- ❖ Windows software program consists of at least one executable file with an .exe file extension. It might also include a number of support programs with extensions such as .dll, .vbx, and .ocx.

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)

33

33

File Format: Data File Extensions

- ❖ The list of data file formats is long.

Type of File	Extensions
Text	.txt .dat .rtf .docx (Microsoft Word) .doc (Microsoft Word 2003) .odt (OpenDocument text) .wpd (WordPerfect) .pages (iWork)
Sound	.wav .mid .mp3 .m4p .aac
Graphics	.bmp .tif .wmf .gif .jpg .png .eps .ai (Adobe Illustrator)
Animation/video	.flc .swf .avi .mpg .mp4 .mov (QuickTime) .wmv (Windows Media Player)
Web page	.htm .html .asp .vrml .php
Spreadsheet	.xlsx (Microsoft Excel) .xls (Microsoft Excel 2003) .ods (OpenDocument spreadsheet) .numbers (iWork)
Database	.accdb (Microsoft Access) .odb (OpenDocument database)
Miscellaneous	.pdf (Adobe Acrobat) .pptx (Microsoft PowerPoint) .qxp (QuarkXPress) .odp (OpenDocument presentations) .zip (WinZip) .pub (Microsoft Publisher)

34

34

Why can't I open some files

- ❖ When a **file doesn't open**, one of **three things** probably went Wrong:
 - The **file** might have been **damaged** by a **transmission** or **disk error**.
 - Someone might have **accidentally changed the file extension**.
 - Some **file formats** exist in **several variations**, and your software might **not have the capability** to open a particular variation of the format.



35

35

File Management



36

36

File Management

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

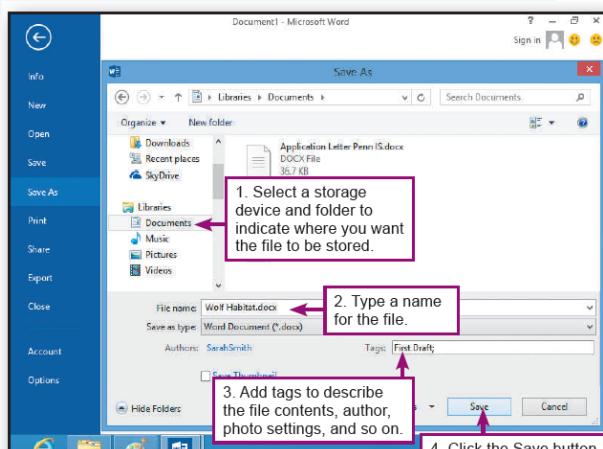


37

37

Application-based File Management

❖ Applications generally provide a way to open files and save them in a specific folder on a designated storage device. Some applications also allow you to delete and rename files.



38

38

Application-based File Management

❖ Creating a new folder while saving a file

1. Select a storage device and folder to indicate where you want the new folder created.

2. Click the New Folder button.

3. Enter a name for the new folder.

39

39

Saving Files on Windows

Save or Save As When saving a file for the first time, you can use Save or Save As.

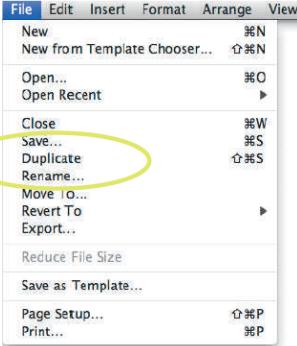
Save When you've revised a file and want to save just the newly revised file with the same name and on the same storage device, use Save.

Save as When you've revised a file and want to save the original version in addition to the newly revised version, use Save As, give the file a different name, and/or select a different storage location.

40

40

Saving Files on Macs



The screenshot shows the 'File' menu of a Mac OS X application. The 'Save...', 'Duplicate', and 'Rename...' options are circled in yellow. The menu also includes 'New', 'Open...', 'Close', 'Move To...', 'Revert To', 'Export...', 'Reduce File Size', 'Save as Template...', 'Page Setup...', and 'Print...'. Descriptions for each option are provided to the right:

- Save** Use this option when saving a file for the first time or after you've revised a file and want to save the revisions.
- Duplicate** Use this option when you want to save a copy of a file, such as when you've revised a document but want to keep the original version in addition to the revised version.
- Rename** Use this option along with the Duplicate option to give your revised version a different name than the original.

41

File Management Metaphors

- ❖ The operating system has a **file management utility**, such as the **Windows File Explorer** or the **Mac OS X Finder**, to handle different file operations.
- ❖ **File management utilities** often use some sort of **storage metaphor** to help you **visualize and mentally organize** the files on your disks.




42

File Management Metaphors

Filing Cabinet

In this metaphor, each storage device corresponds to one of the drawers in a filing cabinet. The drawers hold folders and the folders hold files.

Tree Structure

In this metaphor, a tree represents a storage device.

43

43

File Management Metaphors

Combined Filing Cabinet & Tree Structure

Microsoft programmers combined the filing cabinet metaphor to depict a tree structure in the Windows file management utility

44

44

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!



45