THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

DATS 6450 – CS Foundations

Week 2

Class Overview

- Tech Stack Setup
  - Slack
  - Git
  - Terminal Setup (Windows/Mac)
  - GitHub
  - DataCamp
- Computer Systems and Operating System Overview

Introductions

- Michael Arango
  - Education
    - BS Mathematics & Economics, Covenant College
    - MS Data Science, GWU
  - Current Role
    - Data Scientist, Deloitte

Why Slack?

- So. many. emails.
  - Place to answer emails once for everyone
  - Refer back to much easier than email
- Virtual office hours
- Code Snippets
- Students have a platform to help each other
- Quickly send out revisions or resources
  - “I’ll post this link/code on BlackBoard later.”
- It’s used in industry

Setting up Slack

- Sign up using the invite link below:
  - https://bit.ly/2LRV8J6

Downloading the Slack Desktop App

- Windows Users:
  - https://get.slack.help/hc/en-us/articles/209038037-Slack-for-Windows
- Mac Users
  - https://get.slack.help/hc/en-us/articles/207677868-Slack-for-Mac

Getting Started with Slack

- What is Slack?
  - https://get.slack.help/hc/en-us/articles/115004071768-What-is-Slack
- How to:
  - Channels
  - Messages
  - Search
  - Formatting Messages
    - https://get.slack.help/hc/en-us/articles/202288908-Format-your-messages

Installing Git

- Windows
  - https://git-scm.com/download/win
- Mac
  - https://git-scm.com/download/mac
- Linux
  - https://git-scm.com/download/linux

Installing iTerm (Mac Only)

- https://www.iterm2.com

First time Git Setup

- Open up a terminal (iterm2 for Mac & Git Bash for Windows) and type the following:

```
$ git config --global user.name "John Doe"
$ git config --global user.email johndoe@example.com
```

- You can check your config is successful:

```
$ git config --list
user.name=John Doe
user.email=johndoe@example.com
```

GitHub

- https://github.com

DataCamp

- https://www.datacamp.com/enterprise/cs-foundations-for-data-science/members#invite

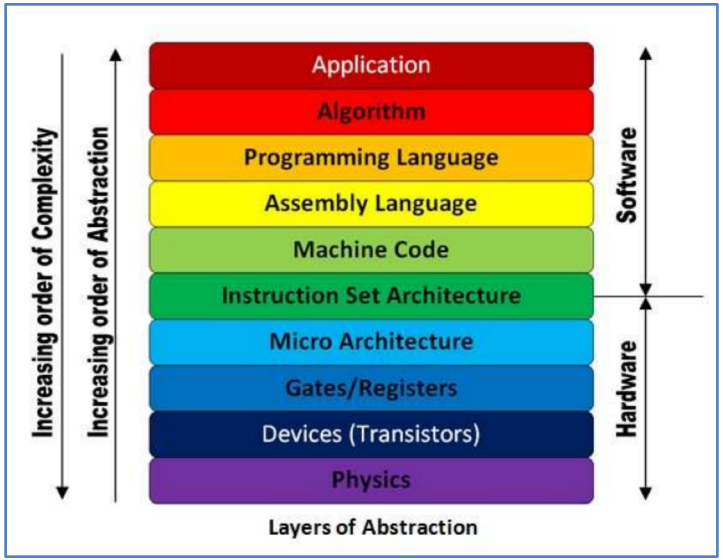
Computer Architecture

Intro to Computer Architecture

- When it comes to computer architecture, the purpose (for this course) is to show you how the computer works from a high level
- The hope is that by knowing more about the design of the underlying system, you can be more effective as a programmer
- This should (in theory) enable you to:
  - Write programs that are more reliable and efficient
  - Troubleshoot your code and debug errors
  - Understand what (physically) is happening when you log-in to a computer remotely and navigate its internal filesystems

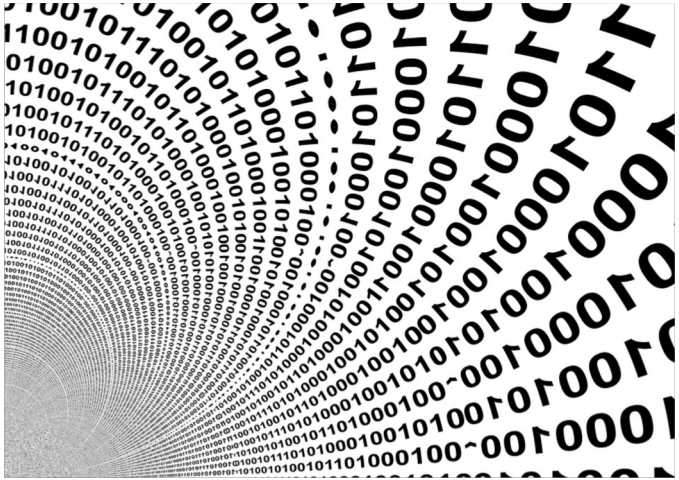
Intro to Computer Architecture

- When programming a computer, we interact with abstractions of abstractions of an abstractions
- At the end of the day it’s all 1’s an 0’s



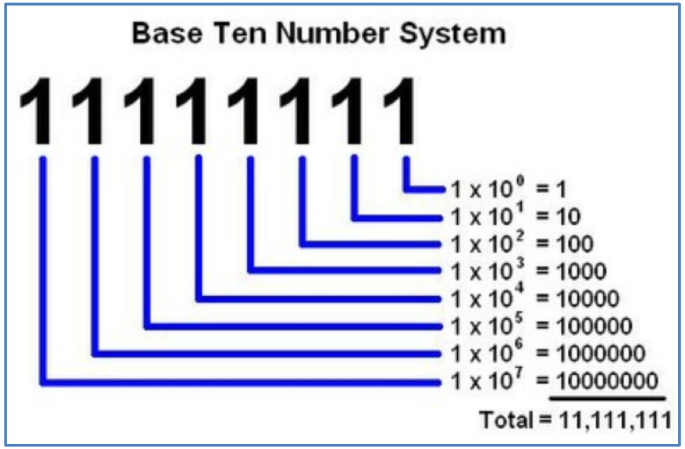
Binary

- “There are 10 kinds of people in the world, those than understand binary, and those who don’t”
- Whatever is stored on a computer, has to ultimately be represented as a finite collection of bits
- This is true whether it’s reals, ints, chars, strings, programs, images, videos, etc.



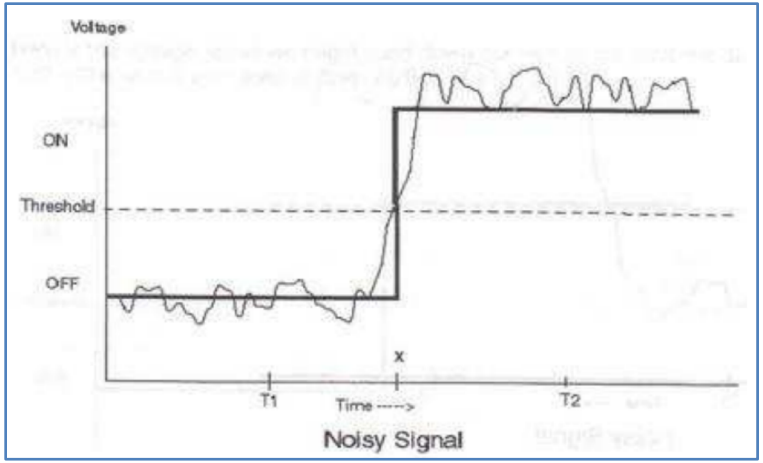
Why Not Base 10 (Decimal) System?

- Hard to store in memory
- Hard to transmit (need high precision to encode 10 different signals on a single wire)
- Messy to implement addition/multiplication/etc.



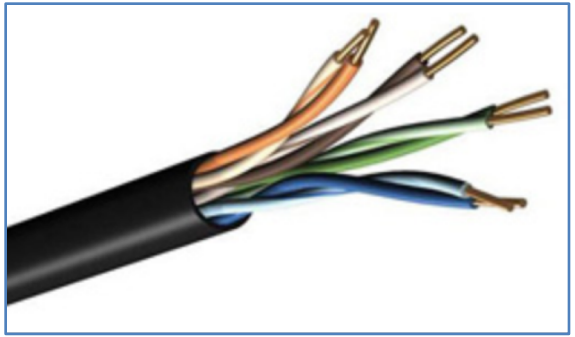
Binary to the rescue

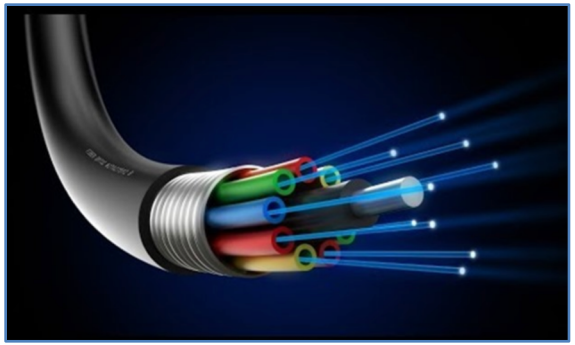
- Easy to store bits with bistable elements
- Easy to transmit signals (streams) on noisey and inaccurate wires



Binary Data

- Data streams are typically voltage changes (on copper wires)
  - Electrons flowing through a conductor
- Recently, with the advent of fiber-optics, data streams can be represented as light pulses (on/off)
  - Photons travelling through glass





Binary to the rescue

- So computers use binary numbers, and therefore use **binary digits** in place of decimal digits.
- The word **bit** is a shortening of the words "Binary digIT." Whereas decimal digits have 10 possible values ranging from 0 to 9, bits have only two possible values: 0 and 1.
- Therefore, a binary number is composed of only 0s and 1s, like this: 1011.
- How do you figure out what the value of the binary number 1011 is?

(1 * 2^3) + (0 * 2^2) + (1 * 2^1) + (1 * 2^0) =

8 + 0 + 2 + 1 = 11

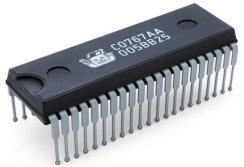
Bits/Bytes/…

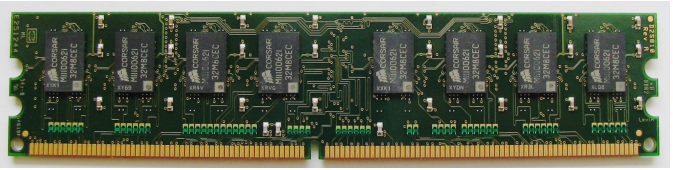
Unit	Value	Size
bit (b)	0 or 1	1/8 of a byte
byte (B)	8 bits	1 byte
kilobyte (KB)	1000^1 bytes	1,000 bytes
megabyte (MB)	1000^2 bytes	1,000,000 bytes
gigabyte (GB)	1000^3 bytes	1,000,000,000 bytes
terabyte (TB)	1000^4 bytes	1,000,000,000,000 bytes
petabyte (PB)	1000^5 bytes	1,000,000,000,000,000 bytes
exabyte (EB)	1000^6 bytes	1,000,000,000,000,000,000 bytes
zettabyte (ZB)	1000^7 bytes	1,000,000,000,000,000,000,000 bytes
yottabyte (YB)	1000^8 bytes	1,000,000,000,000,000,000,000,000 bytes

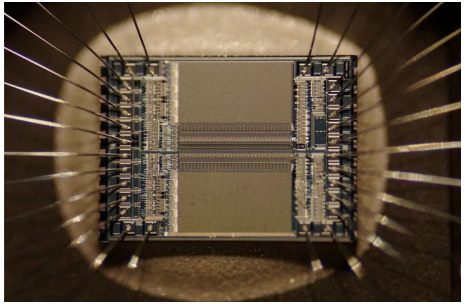
- Internet Spees
  - www.Speedtest.net

Computer Hardware

- Basic element
  - Solid-state transistor (i.e., electrical switch)
  - Building block of integrated circuits (ICs)
- Why IC's?
  - High Performance, High Reliability, Low Cost, Low Power
  - Easily mass-produced
- Modern IC's
  - Use tiny components that can be etched onto the surface of silicon chips
  - Easily mass-produced
- Moore's Law

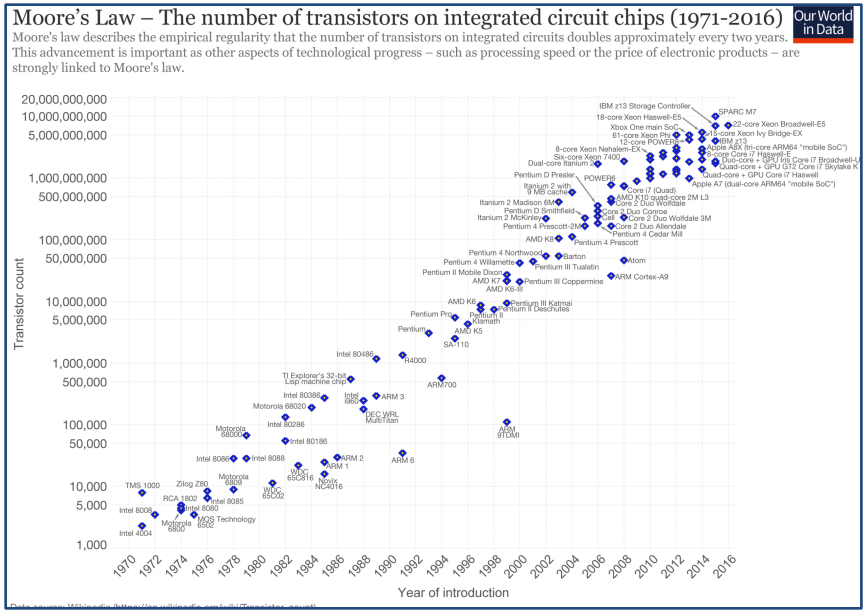




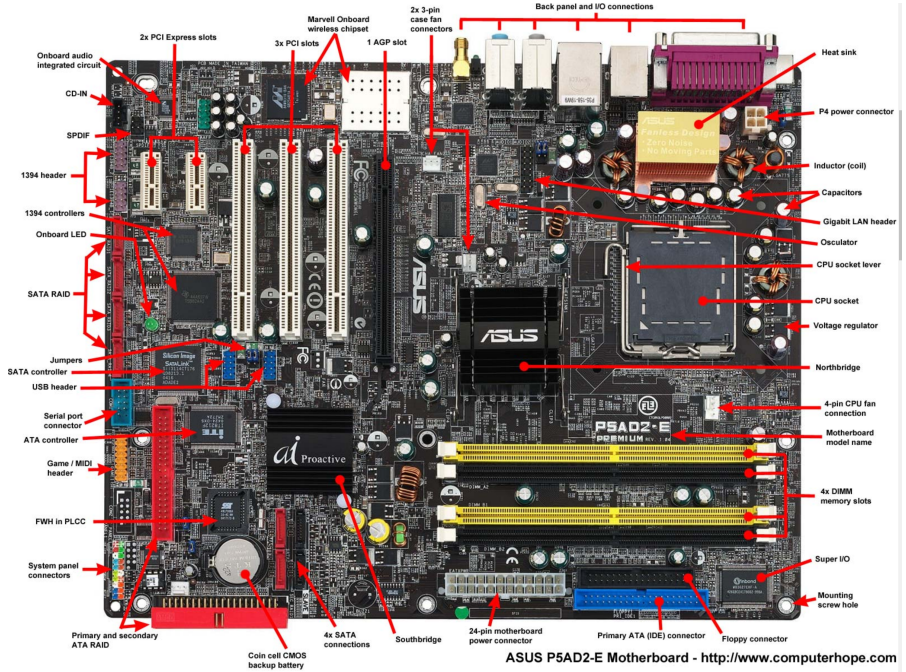


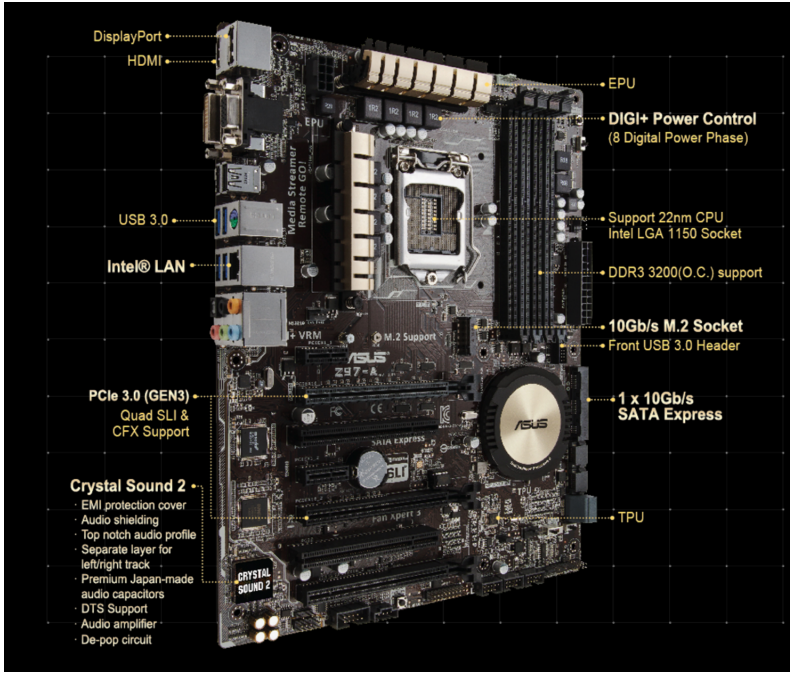
Moore’s Law

- Transistors get 2x smaller about every 2 years
  - sometimes listed as about 18 months
- Can fit twice as many transistors per chip
- Due to better chip etching technology
  (But a cutting edge chip factory costs more than 1 billion dollars)
- Moore's law (Gordon Moore, Intel co-founder) states that the density of transistors on a chip doubles about every 2 years or so
- It is not a scientific law, just a broad prediction that seems to keep working.



Computer Hardware

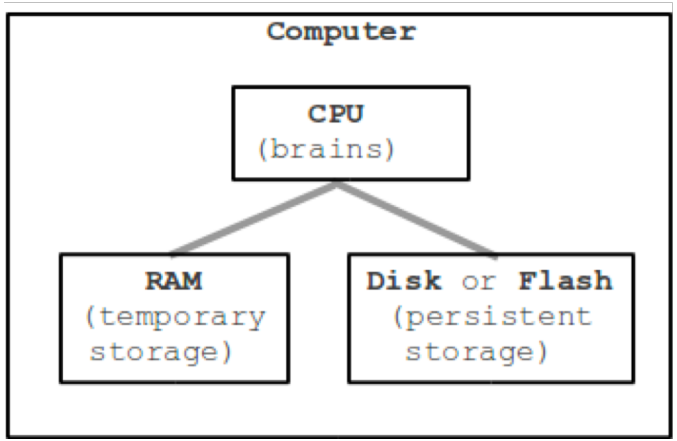




CPU/RAM

CPU - Central Processing Unit

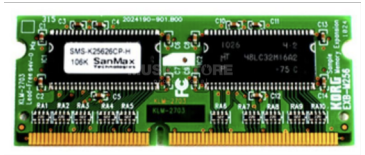
- Acts like a brain: follows the instructions in the code
- "general processing" - images, networking, math .. all on the CPU
- Performs computations, e.g. add two numbers
- vs. RAM and persistent storage which just store data
- "gigahertz" = 1 billion operations per second

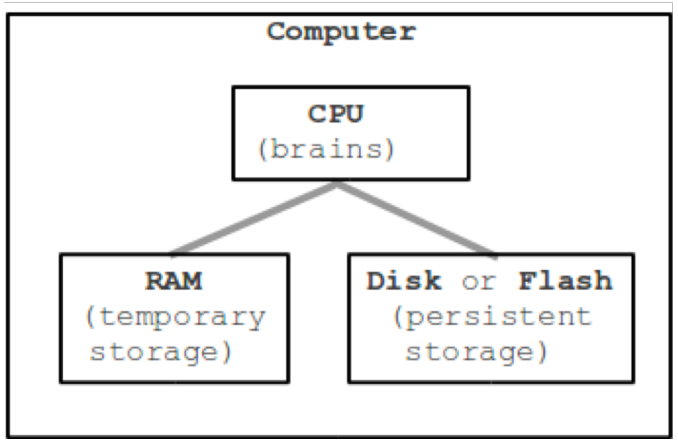


RAM

RAM – Random Access Memory

- Acts like a whiteboard
- Temporary, working storage bytes
- RAM stores both code and data (temporarily)
- e.g. open an image in Photoshop
  - image data loaded into the bytes of RAM
- "persistent"
- - RAM is not persistent. State is gone when power turned off
  - e.g. You're working on a doc, then power goes out and you lose your work (vs. "Save")





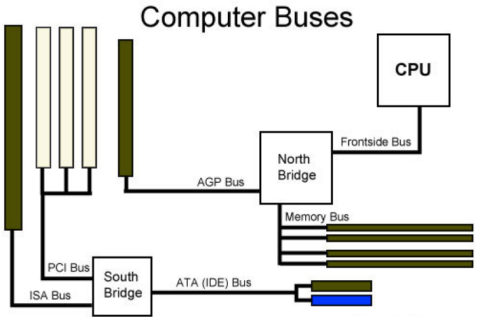
BIOS

- BIOS (Basic Input/Output System)
  - ROM chip on motherboards that allows you to access and setup your computer system at the most basic level

- The four main functions of a PC BIOS
  - POST - Test the computer hardware and make sure no errors exist before loading the operating system.
  - Bootstrap Loader - Locate the OS. If a capable operating system is located, the BIOS will pass control to it.
  - BIOS drivers - Low-level drivers that give the computer basic operational control over your computer's hardware.
  - BIOS or CMOS Setup - Configuration program that allows you to configure hardware settings including system settings such as computer passwords, time, and date.

BUS

- When referring to a computer, the **bus** also known as the **address bus**, **data bus**, or **local bus**, is a data connection between two or more devices connected to the computer.
- The bus contains multiple wires (signal lines) that contain addressing information that describes the memory location of where the data is being sent or where it is being retrieved.
- Each wire in the bus carries a single bit of information, which means the more wires a bus has the more information it can address. For example, a computer with a 32-bit address bus can address 4 GB of memory, and a computer with a 36-bit bus can address 64 GB of memory.

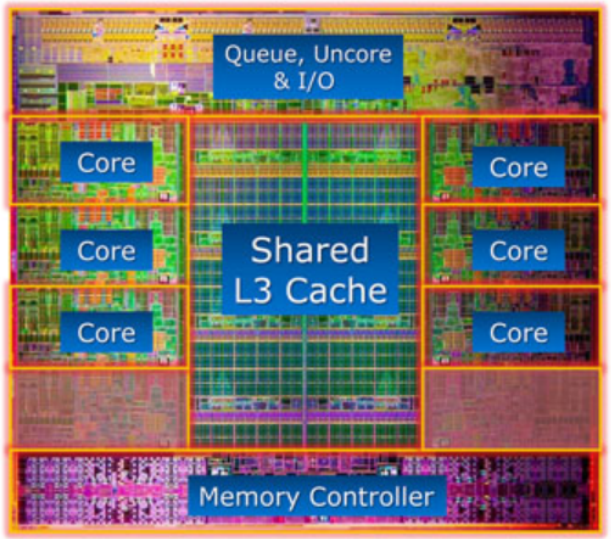


Cache

Cache is a high-speed access area that can be a reserved section of main memory or on a storage device. The two main types of cache are memory cache and disk cache.

Memory cache is a portion of the high-speed static RAM (SRAM) and is effective because most programs access the same data or instructions repeatedly

Intel® Core™ i7-3960X Processor Die Detail



Internet Cache

- Cache, Internet browser cache, or temporary Internet files with an Internet browser, is used to improve how fast data loads while browsing the Internet.
- In most cases, each time you open a web page, the page and all its files are sent to the browser's temporary cache on the hard drive.
- If the web page and its resources have not changed since the last time you viewed it, the browser loads the data from cache rather than downloading the files again.

Hard Drive Formats

- **NTFS**: The NT File System (NTFS) is the file system that modern Windows versions use by default.
- **HFS+:** The Hierarchical File System (HFS+) is the file system modern macOS versions use by default.
- **APFS**: The proprietary Apple file system developed as a replacement for HFS+, with a focus on flash drives, SSDs, and encryption. APFS was released with iOS 10.3 and macOS 10.13, and will become the mandatory file system for those operating systems.
- **FAT32**: The File Allocation Table 32 (FAT32) was the standard Windows file system before NTFS.
- **exFAT**: The extended File Allocation Table (exFAT) builds on FAT32 and offers a lightweight system without all the overhead of NTFS.
- **EXT 2, 3, & 4**: The extended file system (EXT) was the first file system created specifically for the Linux kernel.

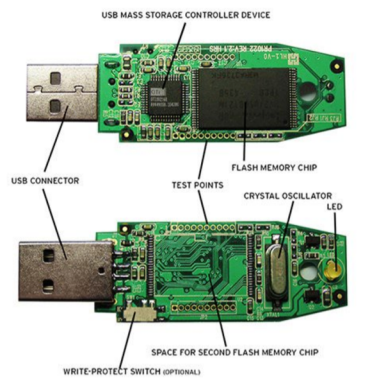
Persistent Storage: Hard Drive, Flash Drive

- Persistent storage of bytes
- "Persistent" means preserved even when not powered
- e.g. Hard drive - stores bytes as a magnetic pattern on a spinning disk
  - aka "hard disk"
  - High pitch spinning sound you may have heard
- Hard drives have been the main, persistent storage tech for a long time
- BUT now flash is getting more popular.

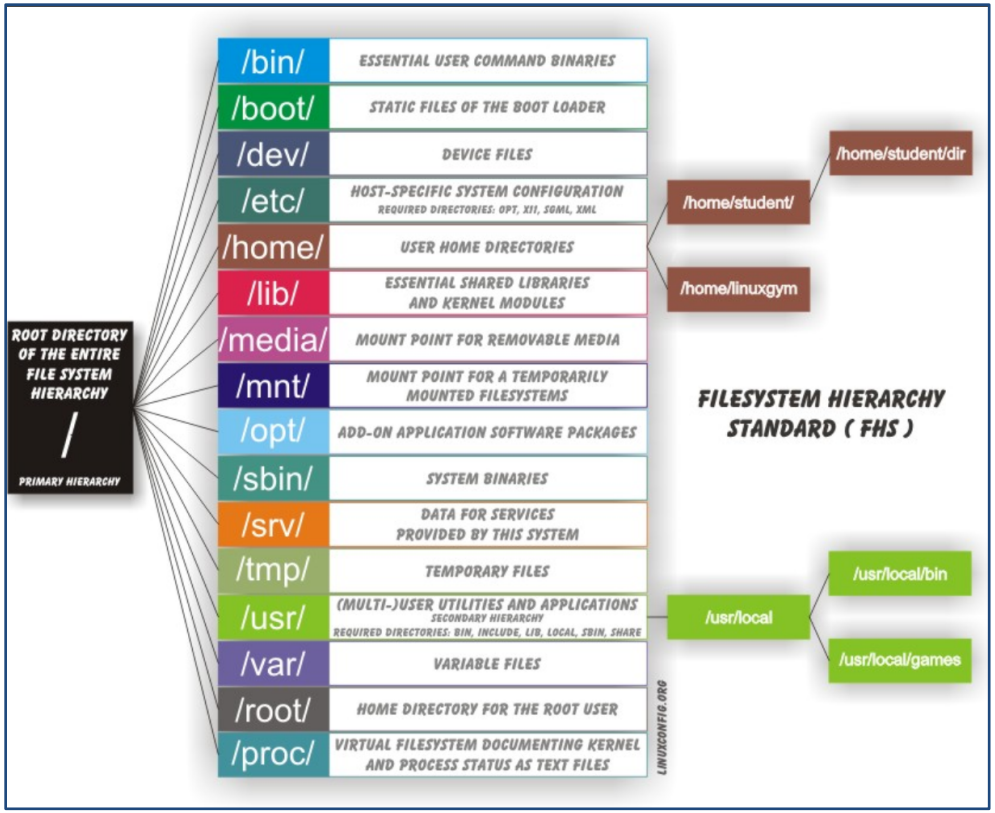
Persistent Storage, Newer Technology: Flash

- "Flash" is a transistor-like persistent storage technology
  "solid state" - no moving parts
- Flash is better than a hard drive in every way but cost - faster, more reliable, less power
- Flash is more expensive per byte
- Formats: usb key, SD card in camera, flash storage built into a phone or tablet or computer
- Flash used to be very expensive, so most computers used hard disks
- Flash is getting cheaper (Moore's law)
- However per-byte, hard drives are still substantially cheaper
- Not to be confused with "Adobe Flash", a proprietary media format

- **Flash does not persist forever. It may not hold the bits past 10 or 20 years. Nobody knows for sure



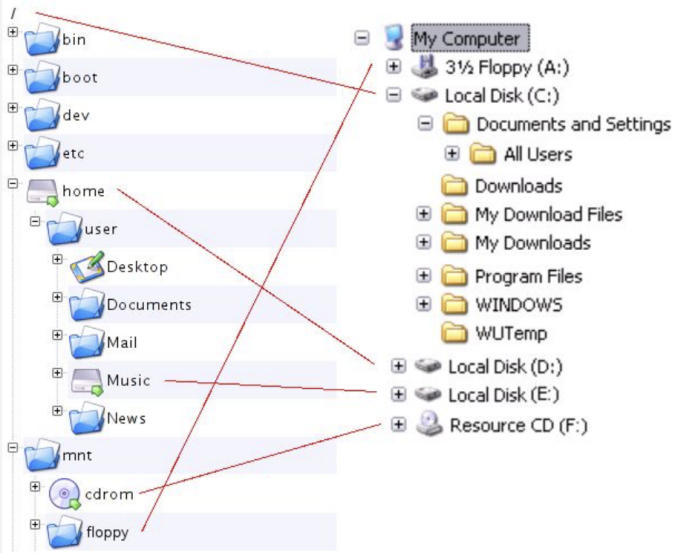
Linux File Systems



Difference between Mac/Windows

WINDOWS vs LINUX FILE SYSTEM

- In Linux there is a **single hierarchical directory structure.** In Windows, there are typically **many partitions with directories** under these partitions.
- In Linux, everything **starts from the root directory,** represented by '/', and then expands into sub-directories. In Windows, it had **various partitions and then directories under those partitions**
- unlike Windows, **Linux is case sensitive**



File Systems: On the left is a typical Linux file system, and on the right is Windows' Explorer. The Linux file system consists of one "tree" with each drive attached to that tree and acting like a folder. The Windows file heirarchy consists of each physical drive having its own seperate file system. The lines between the two point to where these drives appear in each graphic.

Paths Explained

- A file has two key properties:
  - Filename (usually written as one word + extension)
  - Path (specifies the location of a file on the computer.)

- On Windows, paths are written using backslashes (\) as the separator between folder names. OS X and Linux, however, use the forward slash (/) as their path separator.
  - If you want your programs to work on all operating systems, you will have to write your Python scripts to handle both cases. (import os)

- There are two ways to specify a file path.
  - An absolute path, which always begins with the root folder
  - A relative path, which is relative to the program's current working directory

https://automatetheboringstuff.com/chapter8/

Directory Terminology

- Root Directory: /
  - top-most directory in any UNIX file structure
- Home Directory: ~
  - directory owned by a user
  - default location when user logs in
- Current Directory: .
  - default location for working with files
- Parent Directory: ..
  - directory immediately above the current directory

Directory Terminology

- path: list of names separated by “/”
- Absolute Path
  - Traces a path from root to a file or a directory
  - Always begins with the root (/) directory

  Example: `/home/student/Desktop/assign1.txt`
- Relative Path
  - Traces a path from the current directory
  - No initial forward slash (/)
    - dot (.) refers to current directory
    - two dots (..) refers to one level up in directory hierarchy

  Example: `Desktop/assign1.txt`

Directory Terminology

- `pwd` to show path of current working directory
- `cd` to change the current working directory

- `mkdir` to create a new directory

- `rmdir` to delete an empty directory
  - use "`rm -r`" to remove non-empty directory

Directory Terminology

- most frequently used file system command:

Syntax:

```
ls [options] [path]
```

- common options:
  - `-a` show all files
  - `-l` show long version of listing
  - `-t` show files sorted by time stamp
  - `-S` show files sorted by file size
  - `-r` show files in reverse sorted order

Directory Terminology

```
% ls -al
total 126
drwxr-xr-x 13 ege  csci 1024 Apr 26 15:49 .
drwxr-xr-x 15 root root  512 Apr 24 15:18 ..
-rwxr--r--  1 ege  csci  885 Dec  2 13:07 .login
-rwx------  1 ege  csci  436 Apr 12 11:59 .profile
drwx------  7 ege  csci  512 May 17 14:11 330
drwx------  3 ege  csci  512 Mar 19 13:31 467
drwx------  2 ege  csci  512 Mar 31 10:16 Data
-rw-r--r--  1 ege  csci   80 Feb 27 12:23 quiz.txt
```

Directory Terminology

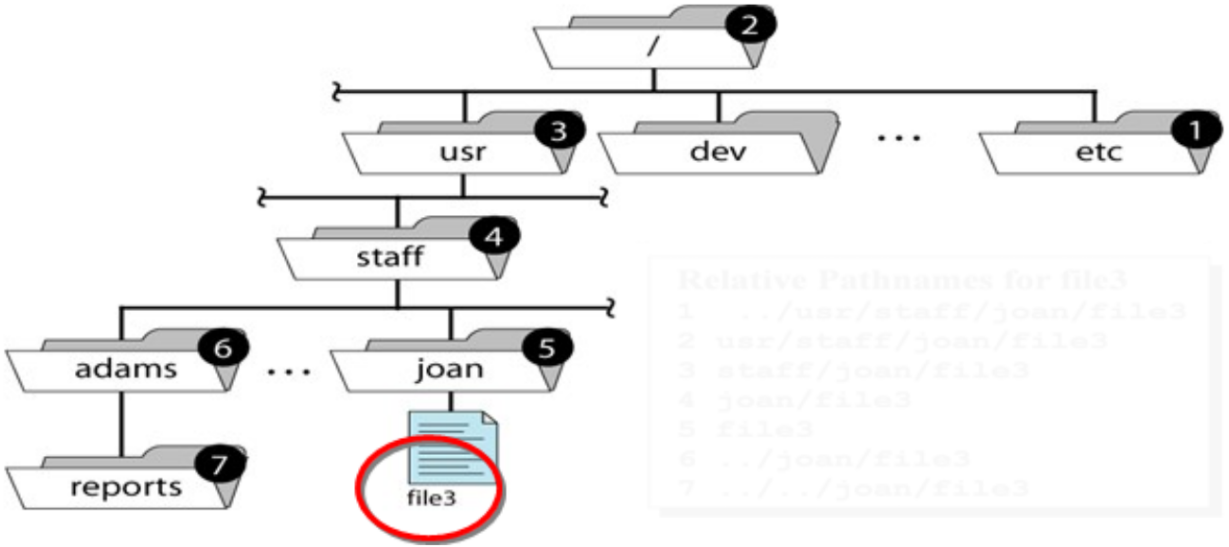
- Copying files or directories

  Syntax: `cp source(s) target`

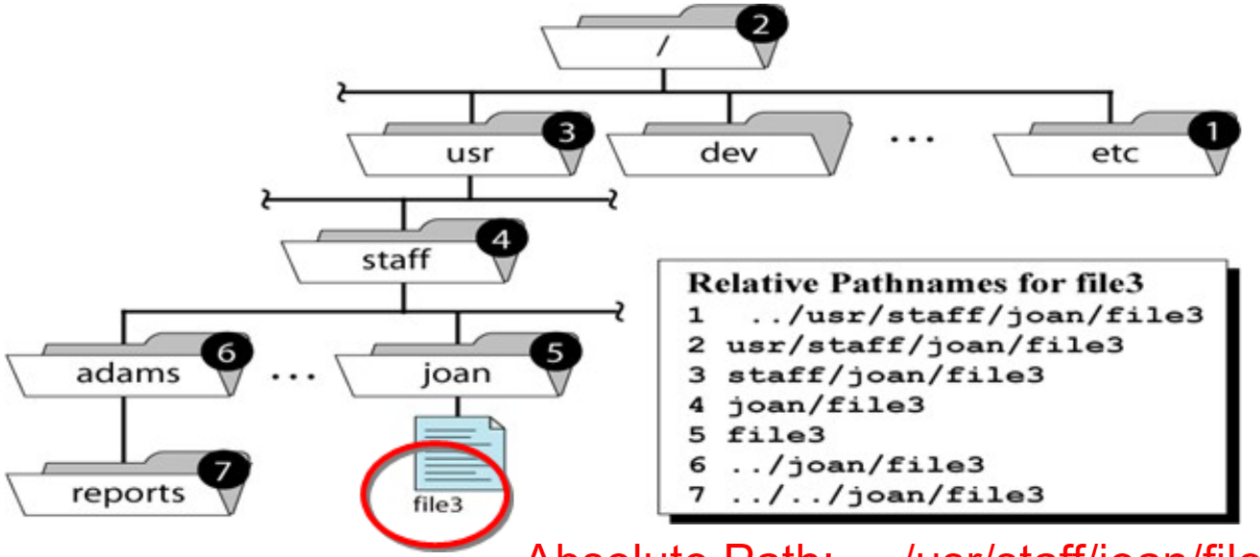
  - source(s) is one or more items to copy
  - target is either name of copied item, or directory
  - commonly used options:
    - `-i` if “target” exists, then prompts for confirmation before overwriting
    - `-r` recursively copy entire directories
    - `-p` preserve access times and permission modes
- Moving files or directories

  Syntax: `mv source(s) target`

Directory Terminology

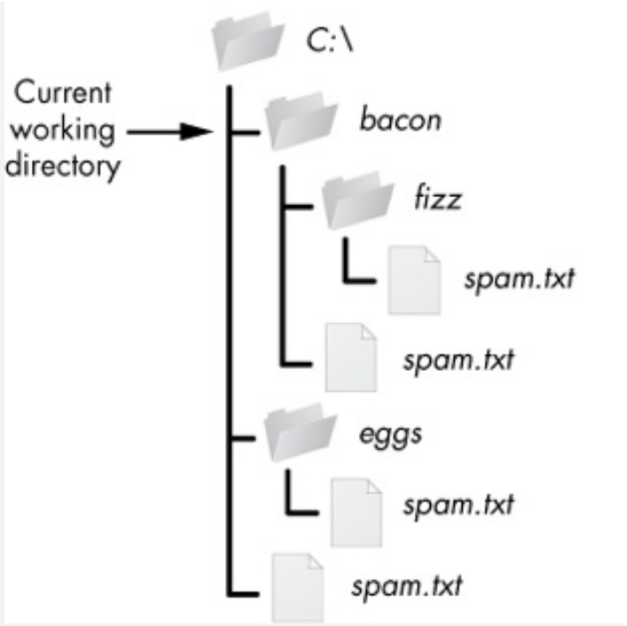


Directory Terminology



Absolute Path: /usr/staff/joan/file3

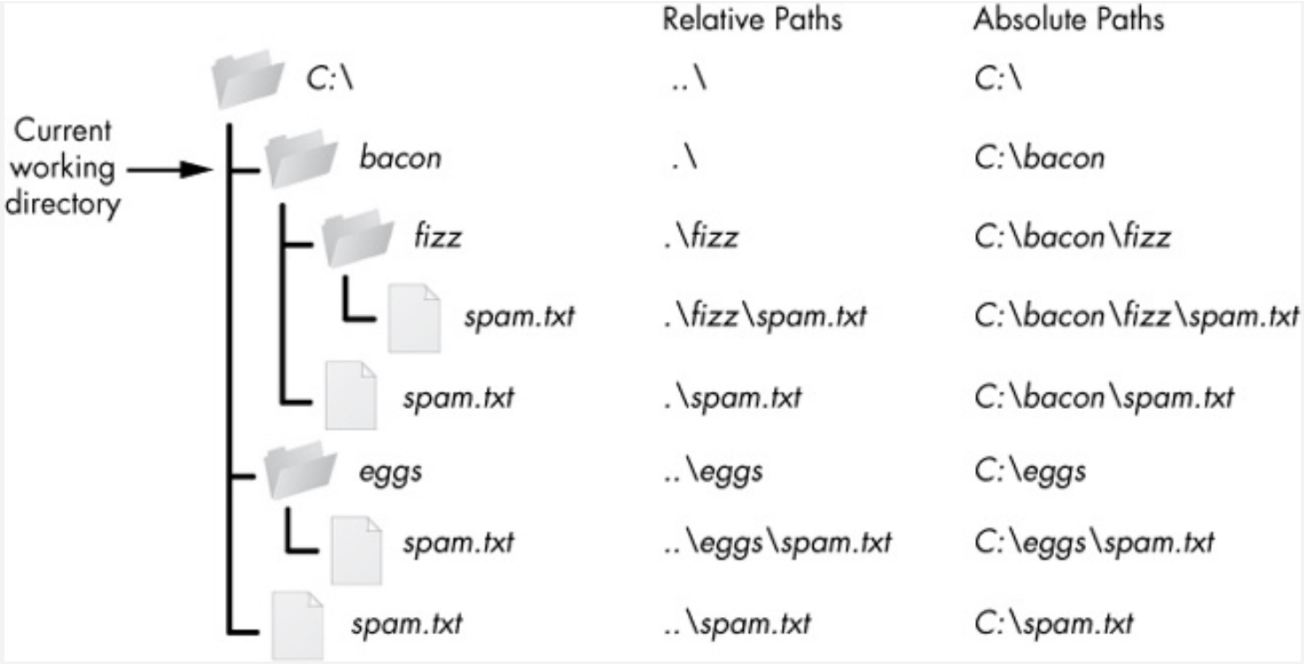
Directory Terminology



Relative Path

Absolute Path

Directory Terminology



Next Class

- Intro to terminal navigation and shell programming