

Systems Development for Computational Science CS107/AC207/E-207

Lecture 0

Thursday, September 3rd 2020

Course Policies
Unix & Linux

Teaching Staff

- Instructor: Dr. David Sondak
- Lecturer in Computational Science
- Research in fluid mechanics, machine learning, scientific computing / software
- Hobbies include: dogs, hiking, soccer, reading



Office: <https://harvard.zoom.us/my/dsondak>

Office Hours: Mondays, 10:30 AM - 11:30 AM (Boston time)

Email: dsondak@seas.harvard.edu

Teaching Staff and Guests

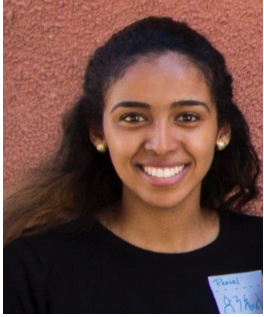


Dr. Andrew Kirby

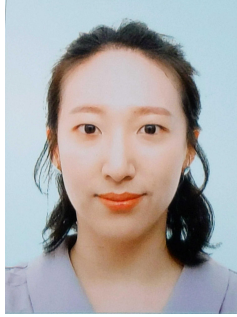


Dr. Mohammad Souri

Teaching Fellows



Peniel Argaw



Hayoun Oh



Simon Warchol



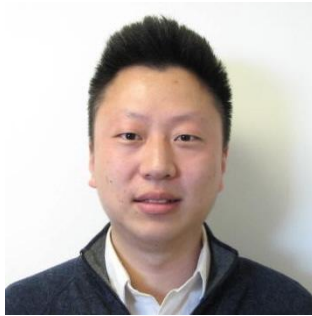
Lihong Zhang



Dovran Amanov



Oluwatosin (Tosin) Alliyu



Johnathan Jiang

Course Policies

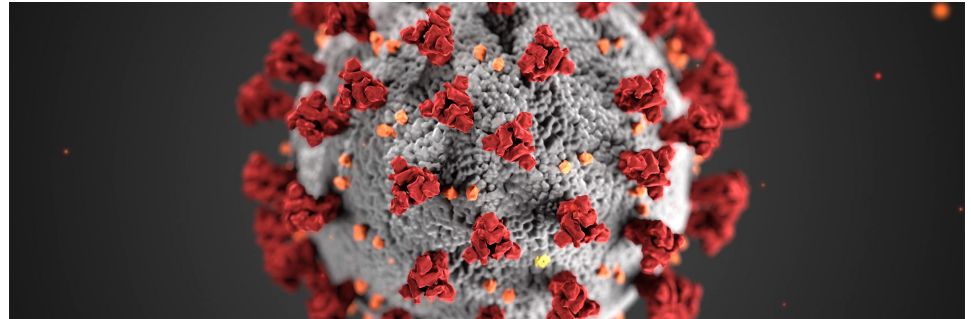
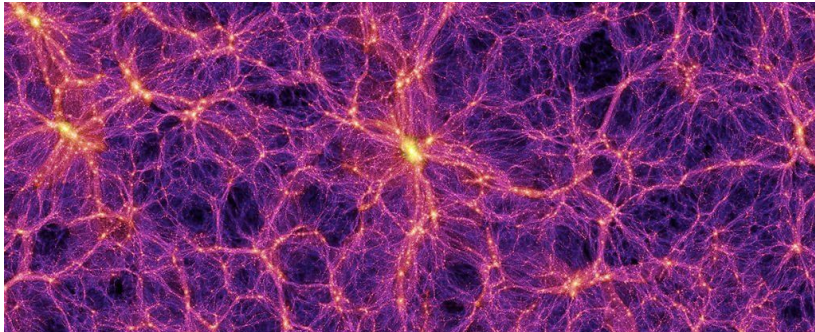
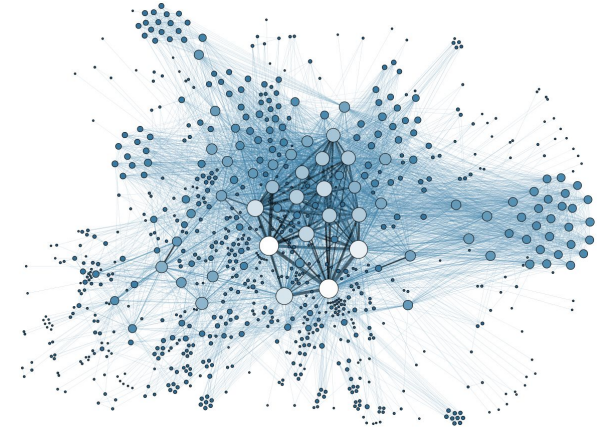
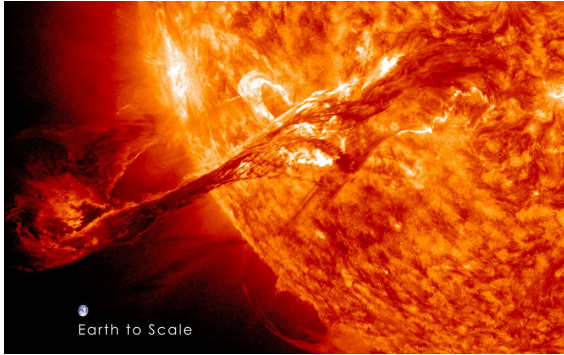
- ✓ Sign up for Piazza (<https://piazza.com/class/kc57xuuysdm64b>)
- ✓ Sign up for Deepnote with course invitation (<https://deepnote.com/join-team?token=91n852m91>)
- ✓ Sign up for Github
 - ✓ Add the teaching staff Github ID (`cs107-sys-dev`) as a collaborator on your course repo when you create it.
- ✓ Understand roles of main course sites:
 - ✓ Main Site: <https://harvard-iacs.github.io/2020-CS107/>
 - ✓ Canvas - Grades and lecture recordings
 - ✓ Piazza - All course announcements and discussions
 - ✓ Github - All assignment submissions (homework, project, pair programming exercises)
 - ✓ Deepnote - Live pair programming exercises
- ✓ Course helplines:
 - ✓ Homework re-grade requests: cs107.sys.dev@gmail.com

Breakout Room Discussion

1. What's the best thing you did this summer?
2. Course policy questions?

Unix and Linux

Motivation



Why Unix / Linux?

- The top 500 most power supercomputers in the world use the Linux family of OS
- Top 500 list
 - <https://www.top500.org/>
- Top 500: Statistics
 - <https://www.top500.org/statistics/list/>



TOP500 Release

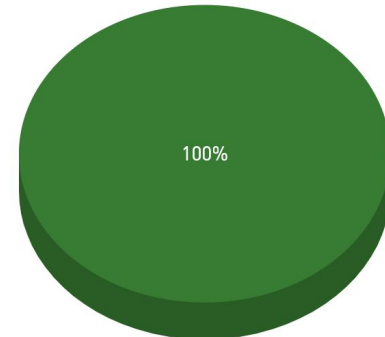
June 2020

Category

Operating system Family

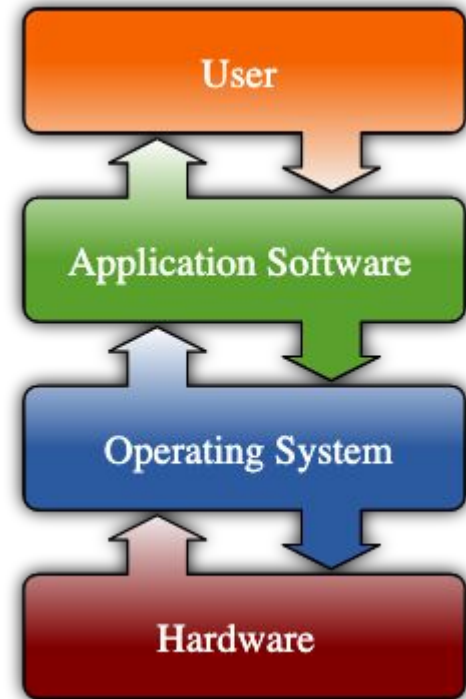
Submit

Operating system Family System Share

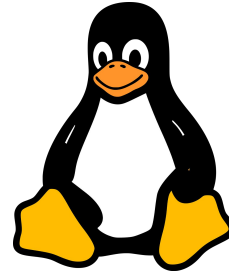


What is Unix?

- Unix is a multi-user, preemptive, multitasking, **operating system (OS)**
- It provides several facilities
 - Management of hardware resources
 - Directories and file systems
 - Loading, execution, and suspension of programs
- There are many versions of Unix
 - Solaris
 - AIX
 - BSD
 - Linux (NOT Unix, but pretty close)
 - ...



What is Linux?

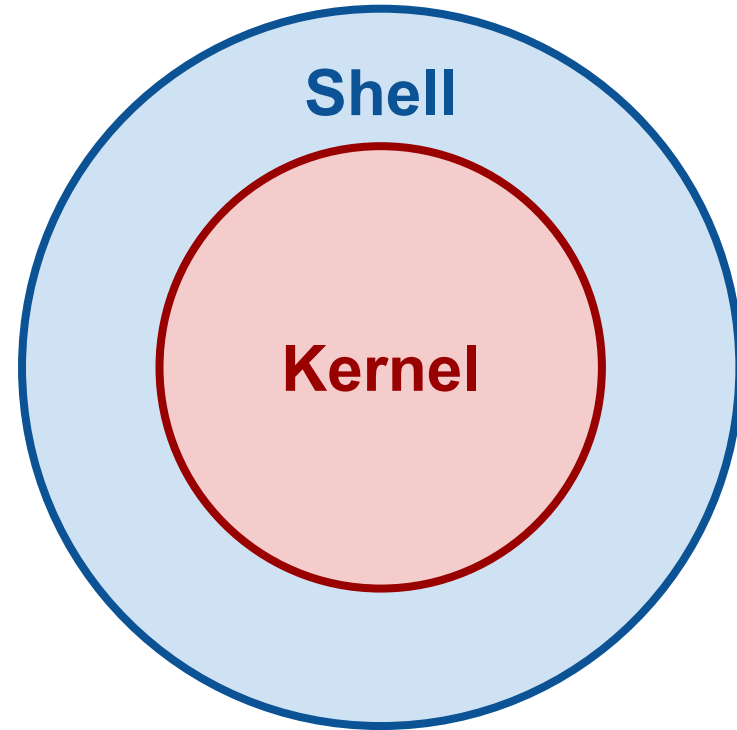


- Linux is a clone of Unix
 - Written by Linus Torvalds
- First version dates to September 1991
- Linux has been further developed by people around the world
 - <https://github.com/torvalds/linux/graphs/contributors>
- Developed under the [The GNU General Public License v3.0](#)
 - Source code for Linux is freely available: <https://github.com/torvalds/linux>



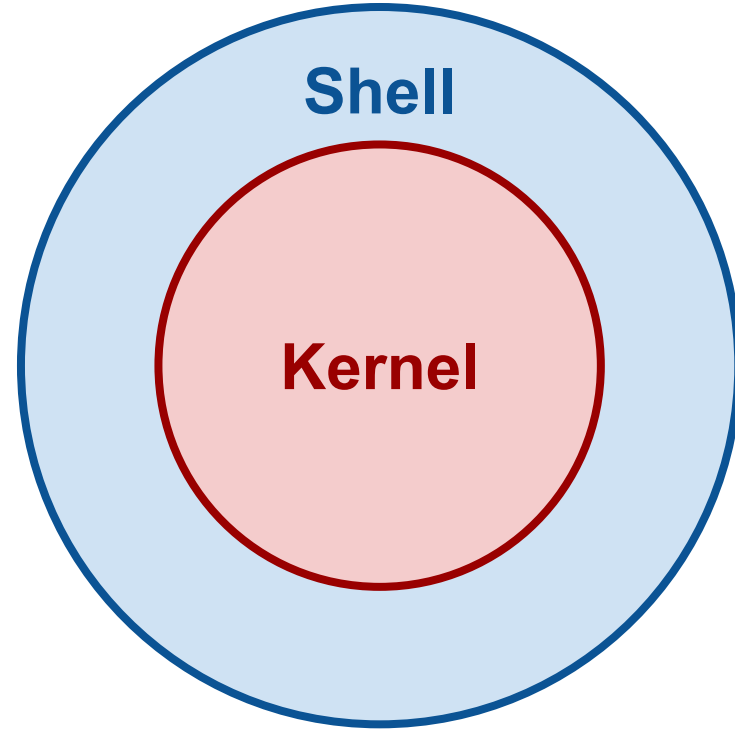
How Does Unix Work? (I)

- Unix has a *kernel* and one or more *shells*
- The kernel is the core of the OS
- It receives tasks from the shell and executes them
- **Users interact with the shell!**



How Does Unix Work? (II)

- Everything in Unix is a *process* or a *file*
- A process
 - Is an executing program
 - An executing program has a unique process ID (PID)
 - May be short or run indefinitely
- A file
 - Is a collection of data
 - Is created by users
- The Unix kernel is responsible for organizing processes and interacting with files

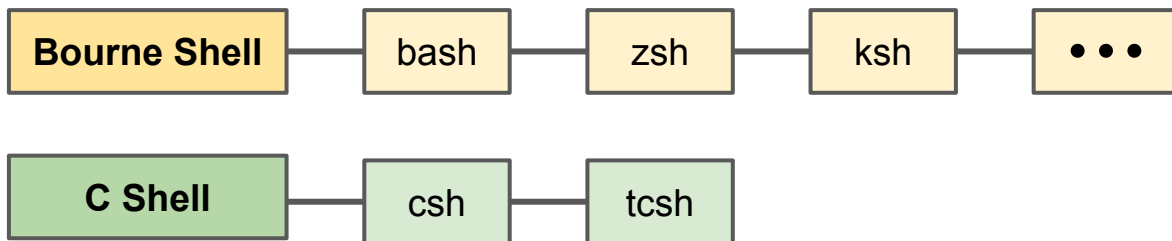


The Shell

- The Unix interface is called the shell
- The shell basically does four things repeatedly:
 - a. Display prompt
 - b. Read command
 - c. Process command
 - d. Execute command

How to Interact with Unix

- The user interacts with Unix via a shell
- Different kinds of shells
 - Graphical (e.g. X-Windows)
 - Text-based (command-line)
 - e.g. `bash` and `tcsh`
- Two main shell families



- Use `ssh` (secure shell) to remotely access a shell session
 - (More on this later)

Some Common Unix Terminology

- Unix has the notion of *accounts*, which include:
 - A username / password
 - `userid / groupid`
 - Home directory
 - A shell preference
- `userids` are called *UIDs*
- Unix has the notion of *groups*:
 - A Unix group can share files and active processes
 - Each account is assigned a primary group
 - The `groupid` corresponds to this primary group
- `groupids` are called *GIDs*

Unix Files and Directories

- A file is a basic unit of storage
- Every file must have a name
- Unix is case-sensitive
- A directory is a special kind of file
 - Directories hold information about other files
- We often think of a directory as a container that holds other files
 - e.g. folders for Mac or Windows users

Comments on the Unix Filesystem

- The filesystem is a hierarchical system of files and directories
- The top level in the hierarchy is call the `root`
- The full *pathname* of a file includes the filename and all directories up to the `root`
 - `/Users/dsondak/Teaching/Harvard/CS107/`
- **Absolute and relative pathnames:**
 - Absolute pathnames start at the root
 - Relative pathnames are specified in relation to the current working directory (`cwd` for short)
 - If I'm already in `/Users/dsondak/Teaching` then it makes sense to talk about the directory `Harvard/CS107/`

Special Directory Names

- There is a special relative pathname for the current working directory
 - .
 - That's not a typo; it's just a *dot*!
- There is a special relative pathname for the parent directory
 - ..
 - Pronounced *dot-dot*
- There is a special symbol for the home directory
 - ~
 - Just a tilde
- These commands will become second (first?) nature to you

Basic Unix Commands

UNIX / LINUX CHEAT SHEET



FILE SYSTEM

ls — list items in current directory
ls -l — list items in current directory and show in long format to see permissions, size, and modification date
ls -a — list all items in current directory, including hidden files
ls -F — list all items in current directory and show directories with a slash and executables with a star
ls dir — list all items in directory dir
cd dir — change directory to dir
cd .. — go up one directory
cd / — go to the root directory
cd ~ — go to your home directory
cd - — go to the last directory you were just in
pwd — show present working directory
mkdir dir — make directory dir
rm file — remove file
rm -r dir — remove directory dir recursively
cp file1 file2 — copy file1 to file2
cp -r dir1 dir2 — copy directory dir1 to dir2 recursively
mv file1 file2 — move (rename) file1 to file2
ln -s file link — create symbolic link to file
touch file — create or update file
cat file — output the contents of file
less file — view file with page navigation
head file — output the first 10 lines of file
tail file — output the last 10 lines of file
tail -f file — output the contents of file as it grows, starting with the last 10 lines
vim file — edit file
alias name 'command' — create an alias for a command

SYSTEM

shutdown — shut down machine
reboot — restart machine
date — show the current date and time
whoami — who you are logged in as
finger user — display information about user
man command — show the manual for command
df — show disk usage
du — show directory space usage
free — show memory and swap usage
whereis app — show possible locations of app
which app — show which app will be run by default

COMPRESSION

tar cf file.tar files — create a tar named file.tar containing files
tar xf file.tar — extract the files from file.tar
tar czf file.tar.gz files — create a tar with Gzip compression
tar xzf file.tar.gz — extract a tar using Gzip
gzip file — compresses file and renames it to file.gz
gzip -d file.gz — decompresses file.gz back to file

PROCESS MANAGEMENT

ps — display your currently active processes
top — display all running processes
kill pid — kill process id pid
kill -9 pid — force kill process id pid

SEARCHING

grep pattern files — search for pattern in files
grep -r pattern dir — search recursively for pattern in dir
grep -rn pattern dir — search recursively for pattern in dir and show the line number found
grep -r pattern dir --include='*.ext' — search recursively for pattern in dir and only search in files with .ext extension
command | grep pattern — search for pattern in the output of command
find file — find all instances of file in real system
locate file — find all instances of file using indexed database built from the updatedb command. Much faster than find
sed -i 's/day/night/g' file — find all occurrences of day in a file and replace them with night - s means substitute and g means global - sed also supports regular expressions

PERMISSIONS

ls -l — list items in current directory and show permissions
chmod ugo file — change permissions of file to ugo - u is the user's permissions, g is the group's permissions, and o is everyone else's permissions. The values of u, g, and o can be any number between 0 and 7.
7 — full permissions
6 — read and write only
5 — read and execute only
4 — read only
3 — write and execute only
2 — write only
1 — execute only
0 — no permissions
chmod 600 file — you can read and write - good for files
chmod 700 file — you can read, write, and execute - good for scripts
chmod 644 file — you can read and write, and everyone else can only read - good for web pages
chmod 755 file — you can read, write, and execute, and everyone else can read and execute - good for programs that you want to share

NETWORKING

wget file — download a file
curl file — download a file
scp user@host:file dir — secure copy a file from remote server to the dir directory on your machine
scp file user@host:dir — secure copy a file from your machine to the dir directory on a remote server
scp -r user@host:dir dir — secure copy the directory dir from remote server to the directory dir on your machine
ssh user@host — connect to host as user
ssh -p port user@host — connect to host on port as user
ssh-copy-id user@host — add your key to host for user to enable a keyed or passwordless login
ping host — ping host and output results
whois domain — get information for domain
dig domain — get DNS information for domain
dig -x host — reverse lookup host
lsof -i tcp:1337 — list all processes running on port 1337

SHORTCUTS

ctrl+a — move cursor to beginning of line
ctrl+f — move cursor to end of line
alt+f — move cursor forward 1 word
alt+b — move cursor backward 1 word

Absolutely Essential Commands

These commands should be at your fingertips at all times

`ls` — list items in current directory

`ls -l` — list items in current directory and show in long format to see permissions, size, and modification date

`ls -a` — list all items in current directory, including hidden files

`ls -F` — list all items in current directory and show directories with a slash and executables with a star

`ls dir` — list all items in directory dir

`cd dir` — change directory to dir

`cd ..` — go up one directory

`cd /` — go to the root directory

`cd ~` — go to your home directory

`cd -` — go to the last directory you were just in

`pwd` — show present working directory

`mkdir dir` — make directory dir

`rm file` — remove file

`rm -r dir` — remove directory dir recursively

`cp file1 file2` — copy file1 to file2

`cp -r dir1 dir2` — copy directory dir1 to dir2 recursively

`mv file1 file2` — move (rename) file1 to file2

`ln -s file link` — create symbolic link to file

`touch file` — create or update file

`cat file` — output the contents of file

`less file` — view file with page navigation

`head file` — output the first 10 lines of file

`tail file` — output the last 10 lines of file

`tail -f file` — output the contents of file as it grows, starting with the last 10 lines

`vim file` — edit file

`alias name 'command'` — create an alias for a command

Breakout Room Discussion

Don't forget to note your breakout room number!

You should appoint one person as the “speaker” for the group. Appoint the person who got up the earliest this morning!

1. What is one thing you learned about Unix / Linux that you didn't know before?

The `ls` command

- The `ls` command displays the names of files
- Giving it the name of a directory will list all files in that directory
- `ls` commands
 - `ls` - list files in the current directory
 - `ls /` - list files in the root directory
 - `ls .` - list files in the current directory
 - `ls ..` - list files in the parent directory
 - `ls /usr` - list files in the `/usr` directory

Command Line Options

- Modify output format of `ls` with *command line options*
- There are many options for the `ls` command, e.g.:
 - `-l` - *long* format
 - `-a` - *all*; shows hidden files as well as regular files
 - `-F` - include special character to indicate file types
- Note: Hidden files have names that start with `.`
 - `-rw-r--r-- 1 dsondak staff 1687 Jul 2 13:32 .gitignore`
- All of those columns have meanings --- we'll get to them eventually

ls Command Line Options

- How to use the command line options:
 - `ls -a`, `ls -l`, ...
- Two or more options can be used at the same time!
 - `ls -ltra`

General `ls` Command Line

- The general form is:
 - `ls [options] [names]`
 - Note: Options must come first!
 - You can mix any options with any names
 - Examples: `ls -al /usr/bin`
- The brackets around options and names means that something is optional
- You will see this kind of description often in the Unix commands documentation
- Some commands have required parameters
- You can also use variable argument lists:
 - `ls /usr /etc`
 - `ls -l /usr/bin /tmp /etc`
 - This will display many files or directory names

Lecture 0 Recap

- Course intro
- What is Unix? What is Linux? Why do I care?
- Unix terminology
- Intro to navigating on the command line
 - `ls`