Introduction to High Performance Computing

Orientation

MacOS, Linux

Search for “Terminal”

OR

Find “Terminal” in *Applications/Utilities/*

Windows

Search for “Git Bash”

OR

Find “Git Bash” in your Start Menu/Git/

http://wintere.github.io/introhpc-archive/

What is This Workshop Series?

- A workshop series to orient researchers to the Talapas cluster and its software ecosystem
- Start with the command line!
- Lessons build on previously introduced concepts
- This is part of an ongoing relationship!

Schedule

Bash and the Command Line
Feb. 4th

Scripting on Talapas
Feb. 6th

Talapas Essentials
Feb. 11th

Scheduling Jobs on Talapas
Feb. 13th

JupyterLab for Researchers
Feb. 14th

R on Talapas, Wrapup
Feb. 20th

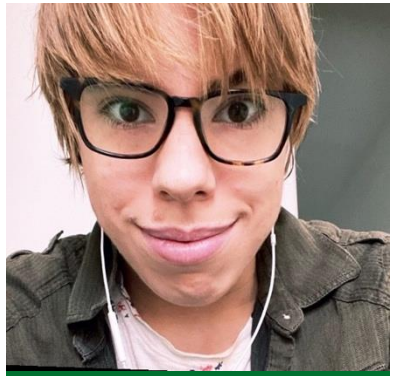
Sessions after today **require Talapas access**.

If you registered before the start of the workshop, you've been added to Talapas.

Workshop Series Goals

- Learn Bash to manage code and data on remote filesystems
- Write Bash scripts for SLURM jobs
- Configure Talapas modules, PIRGS, and permissions when writing jobs
- Cultivate best practices for efficiency, reproducibility, and security

Meet Your Instructors



ERIN WINTER

Data Services

BRUCE EDELMAN

RACS

FREDDY LOPEZ

RACS

WILLIAM WINTER

RACS

JOLINDA SMITH

LCNI

Rules and Procedures

- No hybrid attendance, no video recordings.
- Need to miss a session? Don’t worry, detailed notes will be posted after class finishes
- Don’t be afraid to signal to an assistant for help. We want everyone to follow along!

I’m already proficient in Bash. Do I have to attend this session?

No. If you’d like to skip ahead, please see today’s assistant to make sure your Talapas account is configured for future activities.

To use the software and programming tools available on Talapas at scale…

Rust

R

Python +
PyTorch

conda

FSL

C, gcc

Matlab

JupyterLab

DCM2Niix

FMRIPrep

Julia

CUDA

You need to know Bash.

A Computer

A “Supercomputer”

Supercomputer (Cluster)

Highly connected computers that divide and conquer computationally demanding tasks

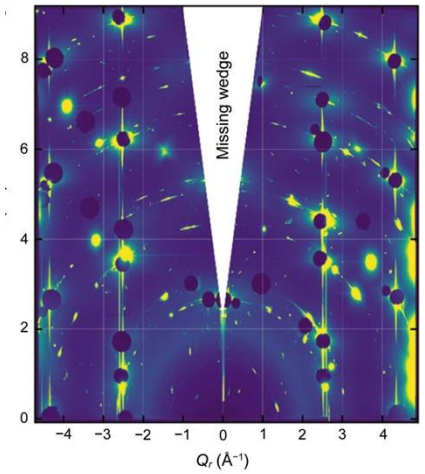
Why Parellel Computing?

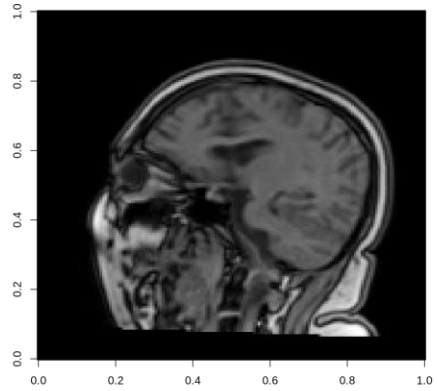
When the software you run for your research is…

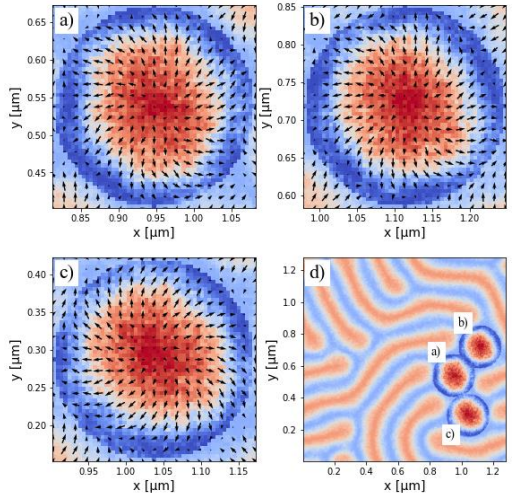
+Running out memory

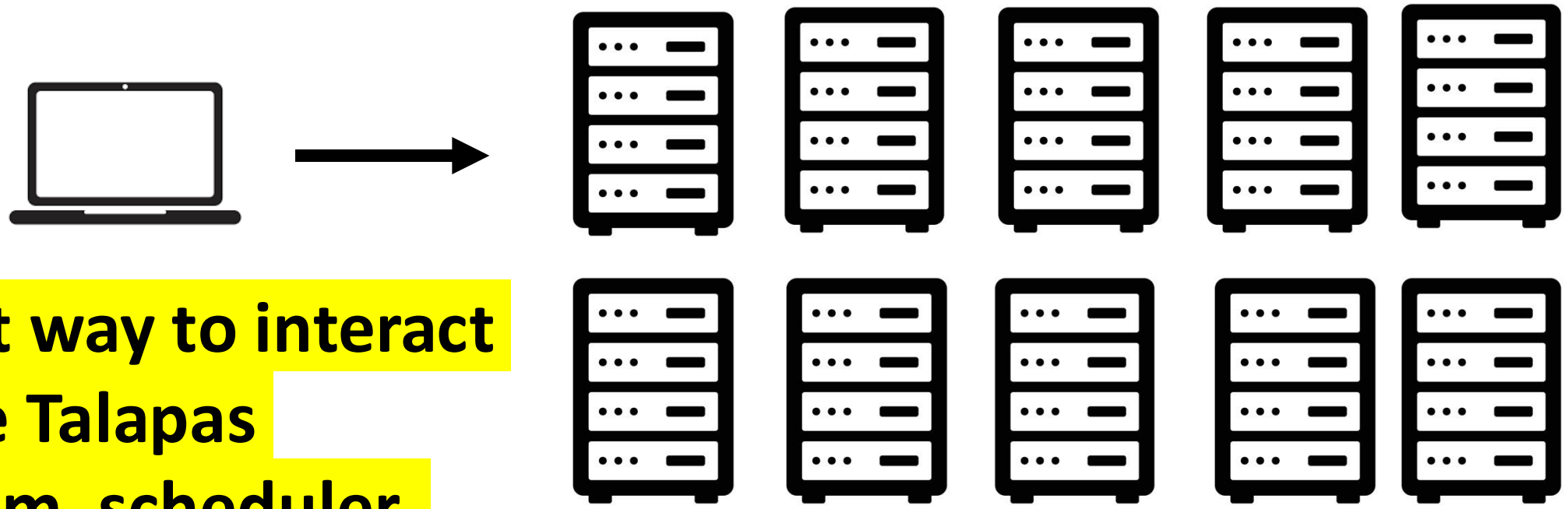
+Takes too much time to run one job at a time (serially)

+Requires communication between concurrent components









The best way to interact with the Talapas filesystem, scheduler, and environment module system is through Bash.

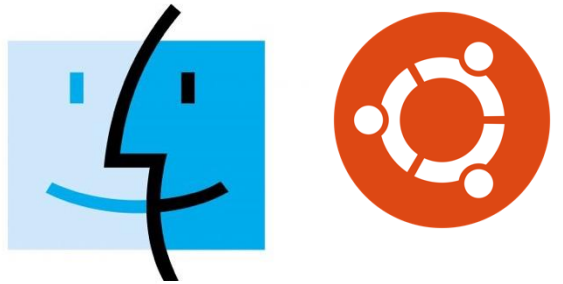
What is Bash?

- Shells (aka the command line) are textual interfaces that execute operating system commands
  + Shells are also programming languages
- Bash or the “Bourne-Again Shell” is a popular UNIX shell
  + Fast, simple, and long-standing!
- Most Bash commands work on descendants like zsh (MacOS)

What is UNIX Anyway?

- UNIX is a family of operating systems. The most popular UNIX-like operating systems today are MacOS and Linux.
- Talapas is a RedHat Linux cluster.
- Windows is not UNIX. The emulator Git Bash can “talk” to Windows and accept Bash commands.

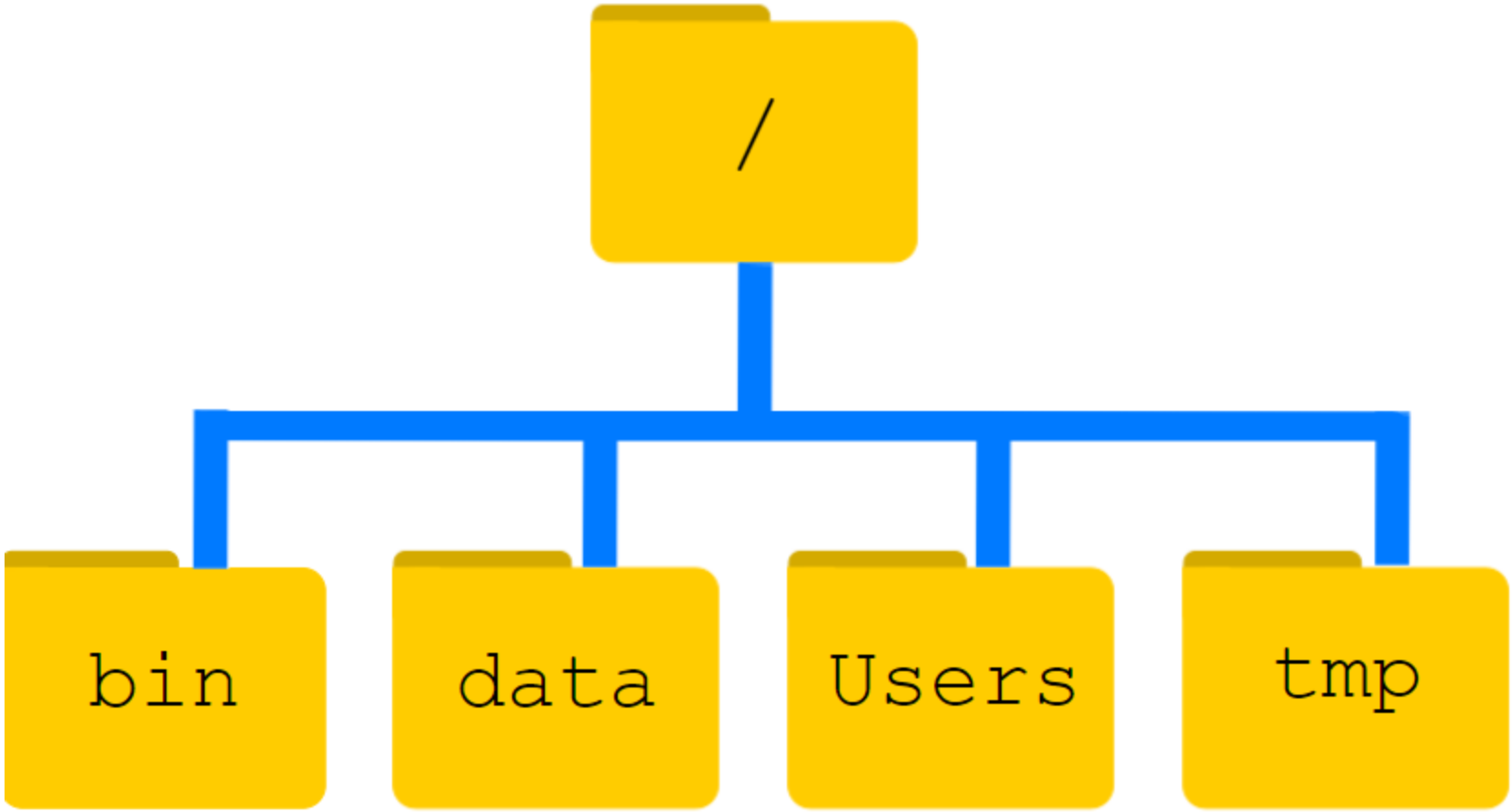
UNIX

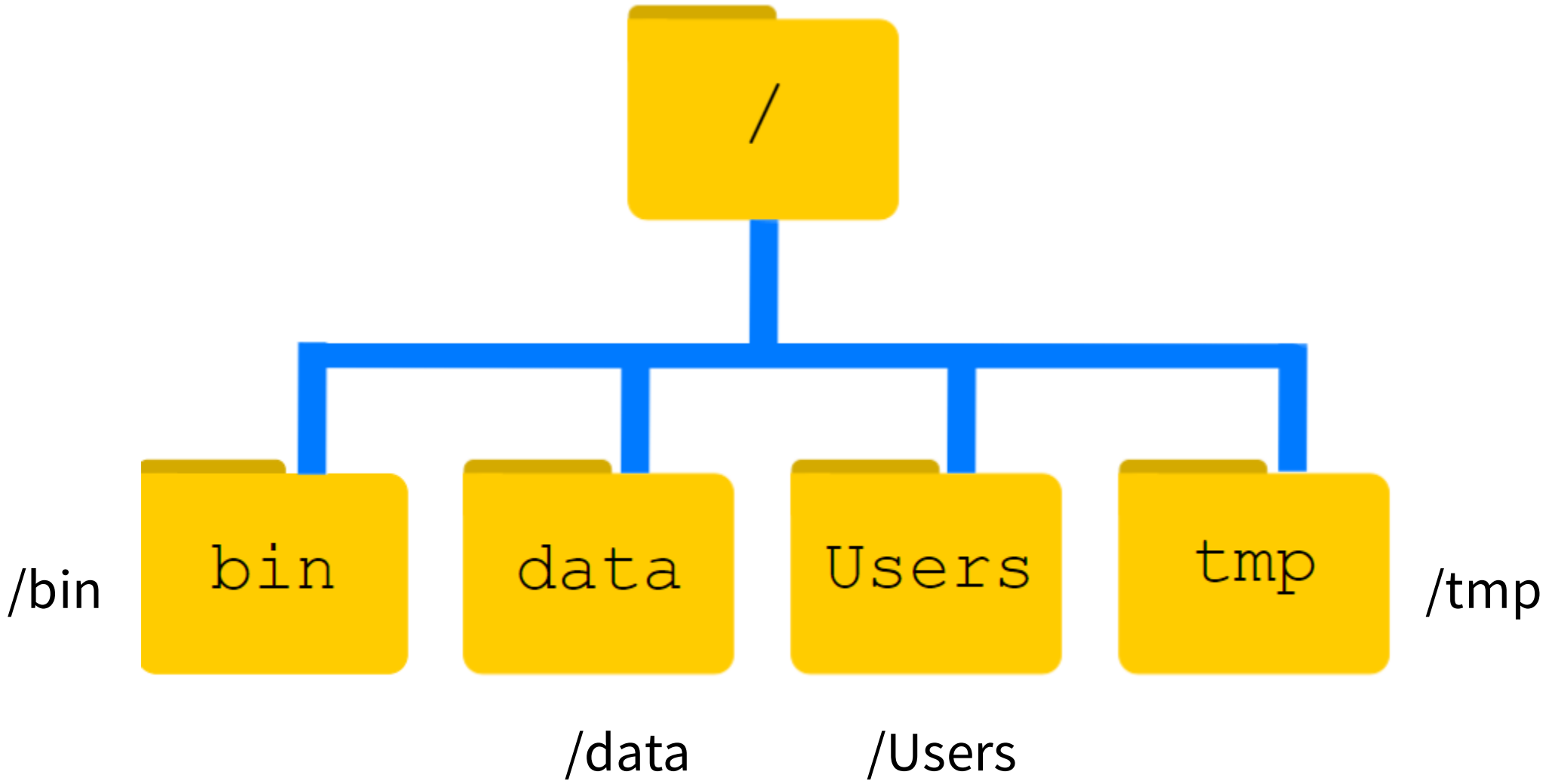


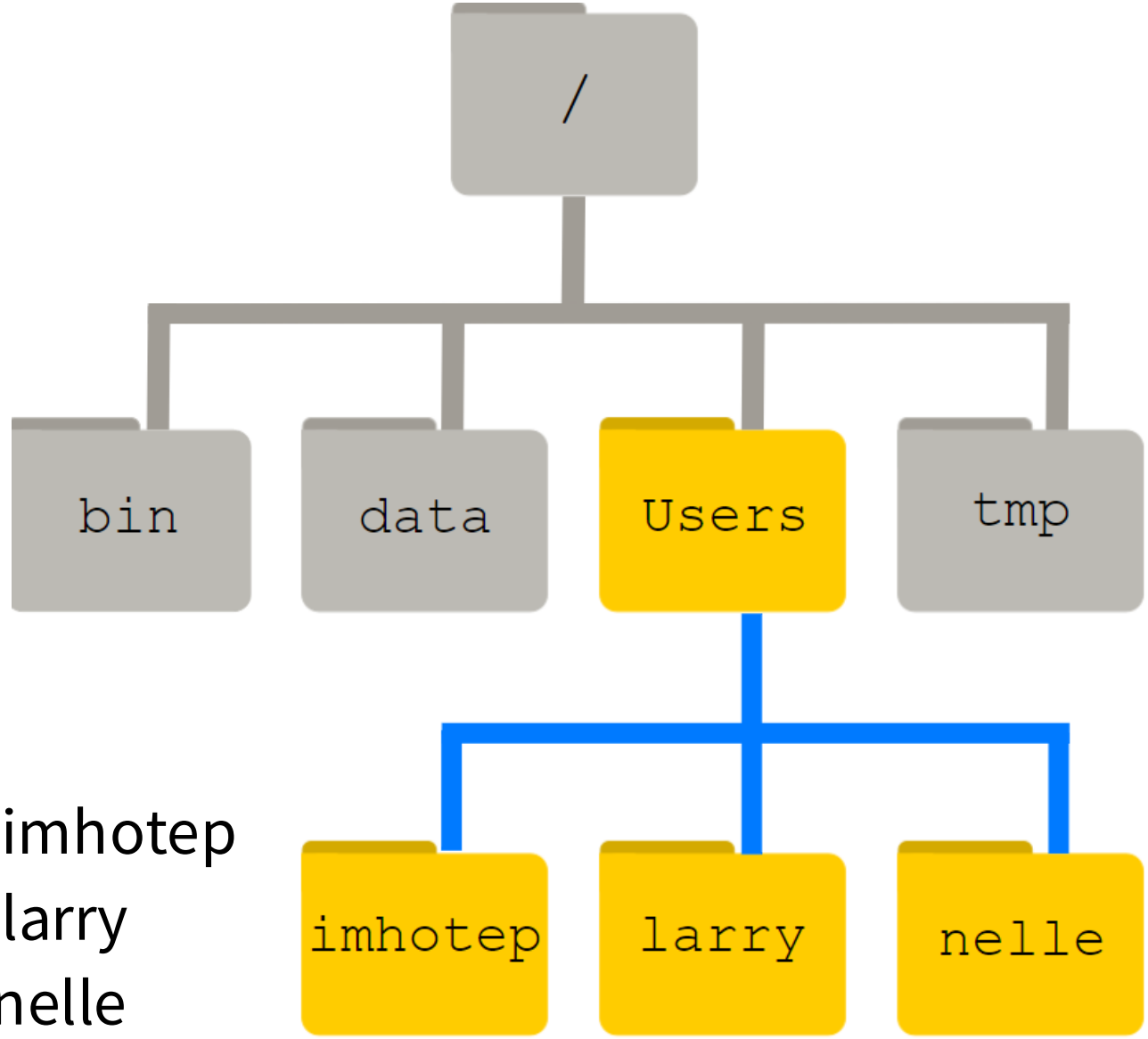
NOT UNIX (DOS)

Unix Filesystem Essentials

The shortest valid path in a UNIX filesystem in “/”. It’s a special character designating the **root** of the filesystem tree.

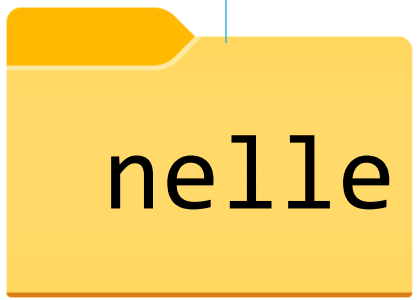



/
bin
data
Users
tmp
/bin
/tmp
/data
/Users

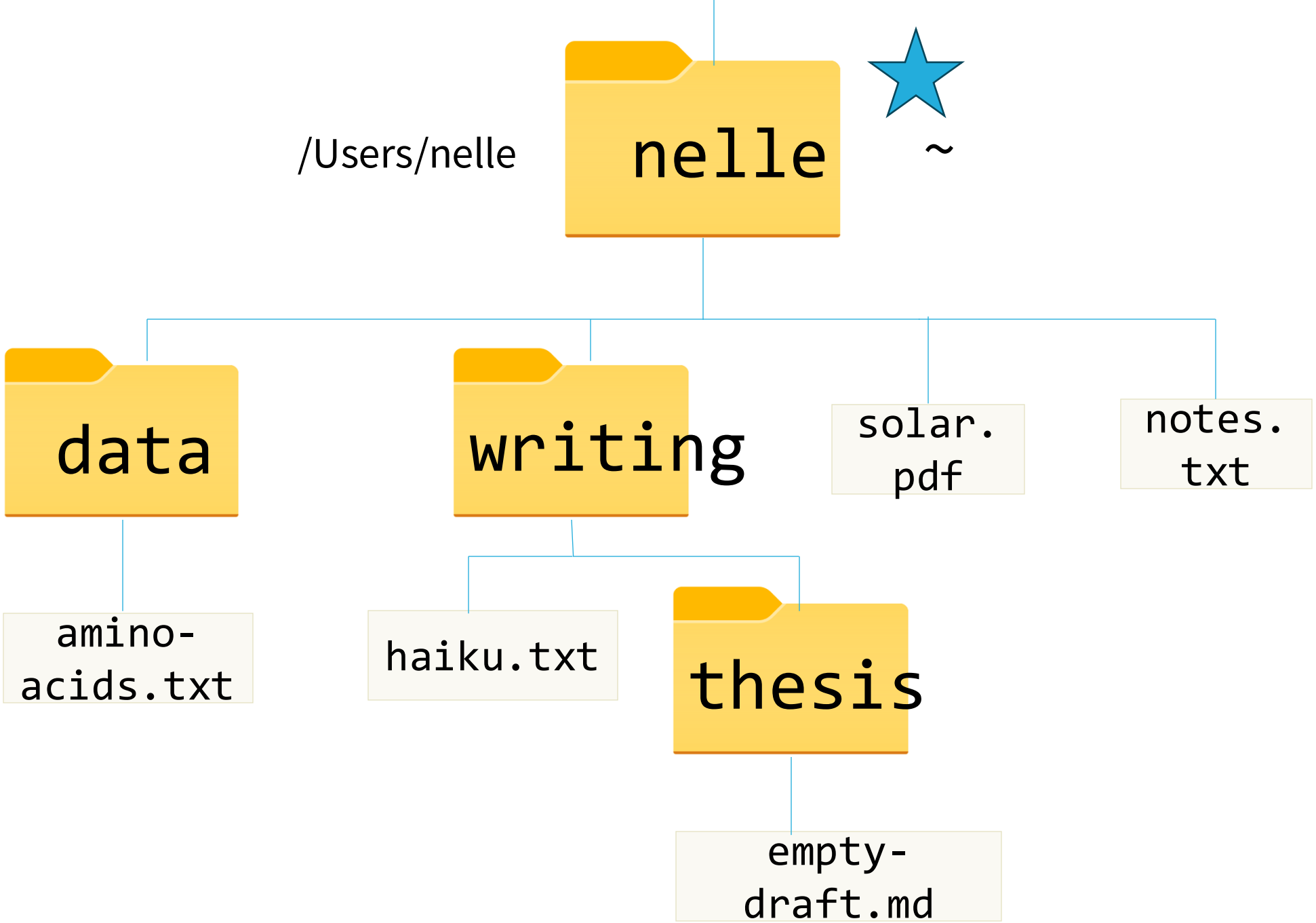


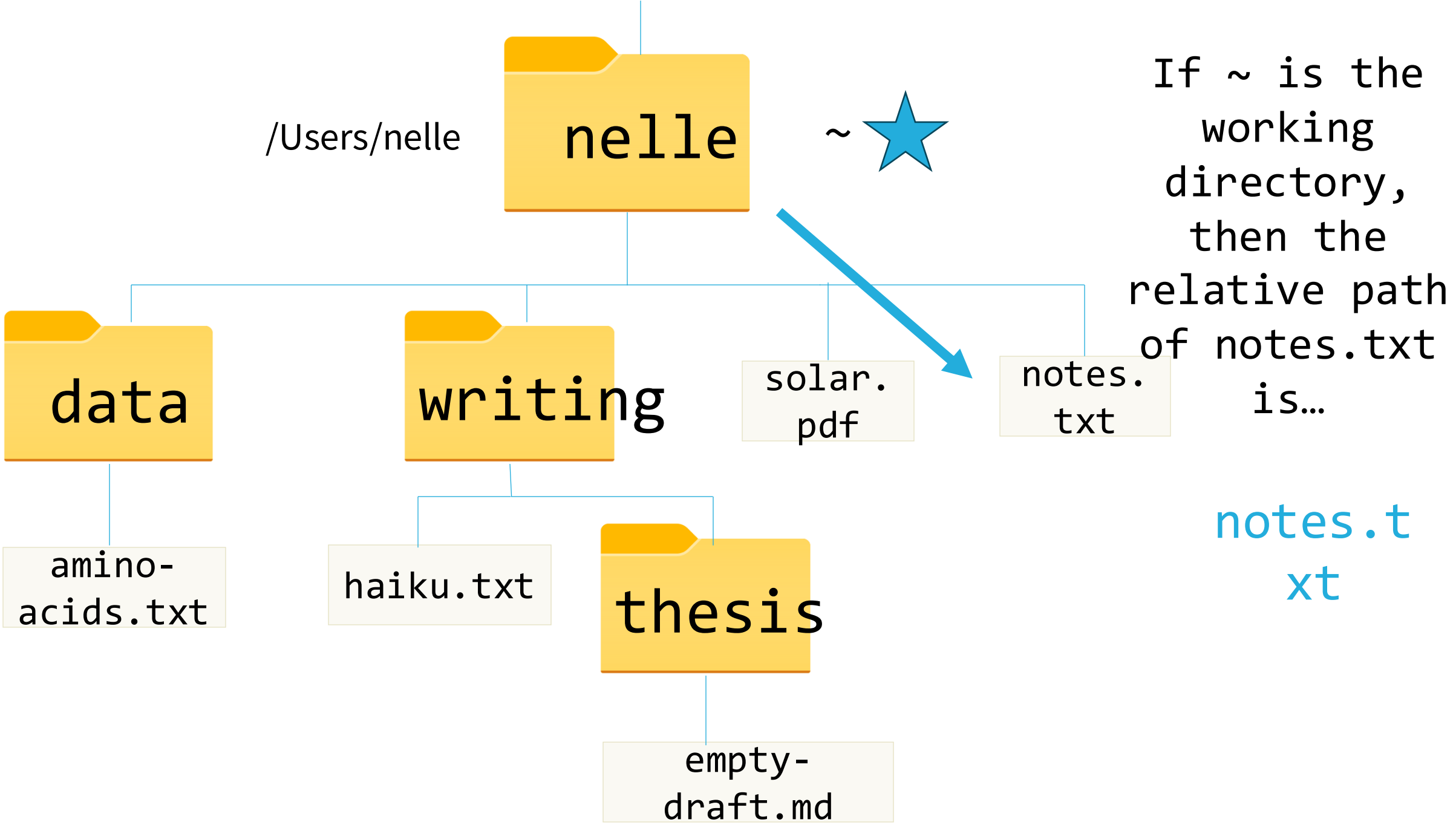
/Users/imhotep
/Users/larry
/Users/nelle

/Users/nelle



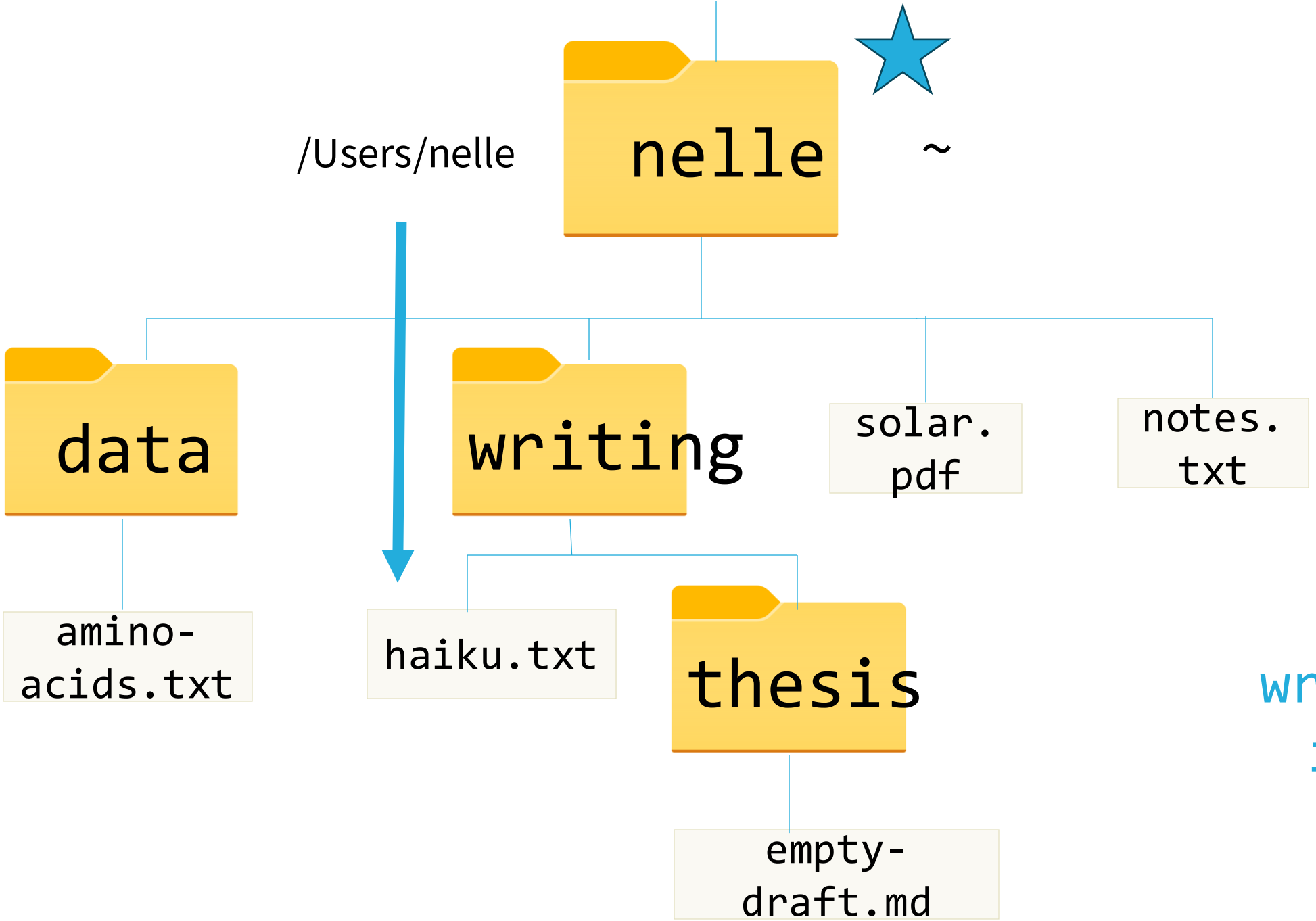
~


/Users/nelle
nelle
~
data
writing
solar.
pdf
notes.
txt
amino-
acids.txt
haiku.txt
thesis
empty-
draft.md



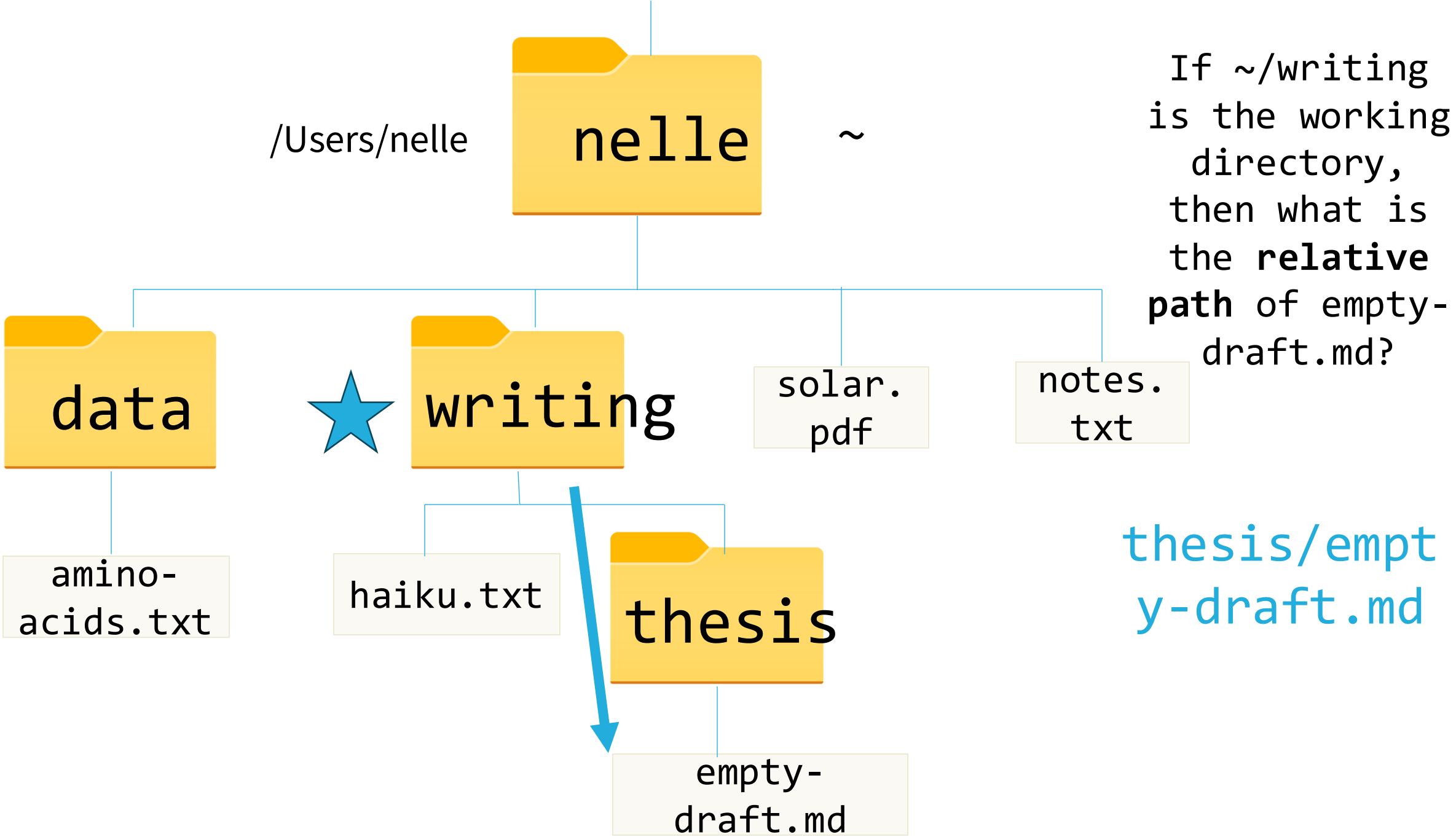
If ~ is the working directory, then the relative path of notes.txt is…

notes.t
xt



If ~ is the working directory, then what is the relative path of haiku.txt?

writing/haiku.txt



If ~/writing is the working directory, then what is the **relative path** of empty-draft.md?

thesis/empty-draft.md

Filepaths: UNIX vs. Windows

UNIX: "/users/emwin/documents/code/input.csv"

WINDOWS: "C:\Users\Home\Erin\My Documents\code\input.csv"

WINDOWS (GIT BASH): "/c/users/home/erin/my documents/code/input.csv"

Which of these are appropriate for use in code you intend to share or use on Talapas?

Avoid Absolute Paths in Code

~~"/users/emwin/documents/code/input.csv"~~

~~"C:\Users\Home\Erin\My Documents\code\input.csv"~~

~~"/c/users/home/erin/my documents/code/input.csv"~~

```
data = pd.read_csv("input.csv")
```

```
input_path = os.path.join("code", "input.csv")
```

Use relative paths and/or libraries like Python's os.path and R's here package whenever possible.

A Closer Look at the Shell

Git Bash (Windows)

zsh (Mac OS)

This is the shell prompt. Do not type the prompt character when entering commands.

The $ or % symbol means that the shell is ready for you to type your input and press Enter!

Username, Hostname, Prompt

emwin@UO-2012173 ~ %

emwin@UO-2012173 ~ $

Pay attention to the hostname! This is the **computer name**, and it will change when you are on Talapas.