

Linux Shell

DS 5110: Big Data Systems

Spring 2025

Lecture 2

Yue Cheng



UNIVERSITY
of
VIRGINIA

Some material taken/derived from:

• Wisconsin CS 544 by Tyler Caraza-Harter.

@ 2025 released for use under a [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Learning objectives

- Setting up an EC2 VM instance via AWS Academy
- Navigate a Linux file system
- Operating within a Linux shell
- Automate repeated tasks

What is a Shell?

A shell program

Loop:

```
print(SOME PROMPT)
cmd = get_input()
run(cmd)
```



If you can type it, you can automate it

A shell program

Loop:

```
print(SOME PROMPT)
cmd = get_input()
run(cmd)
```

Script.sh

Shell commands

...
...



You can run a shell inside of a shell

The image shows a terminal window titled "bash" with a window icon in the top-left corner and a system menu icon in the top-right corner. The terminal content is as follows:

```
Yue ~ $ echo hello world
hello world
Yue ~ $ pwd
/Users/yue
Yue ~ $ bash

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh
-s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.

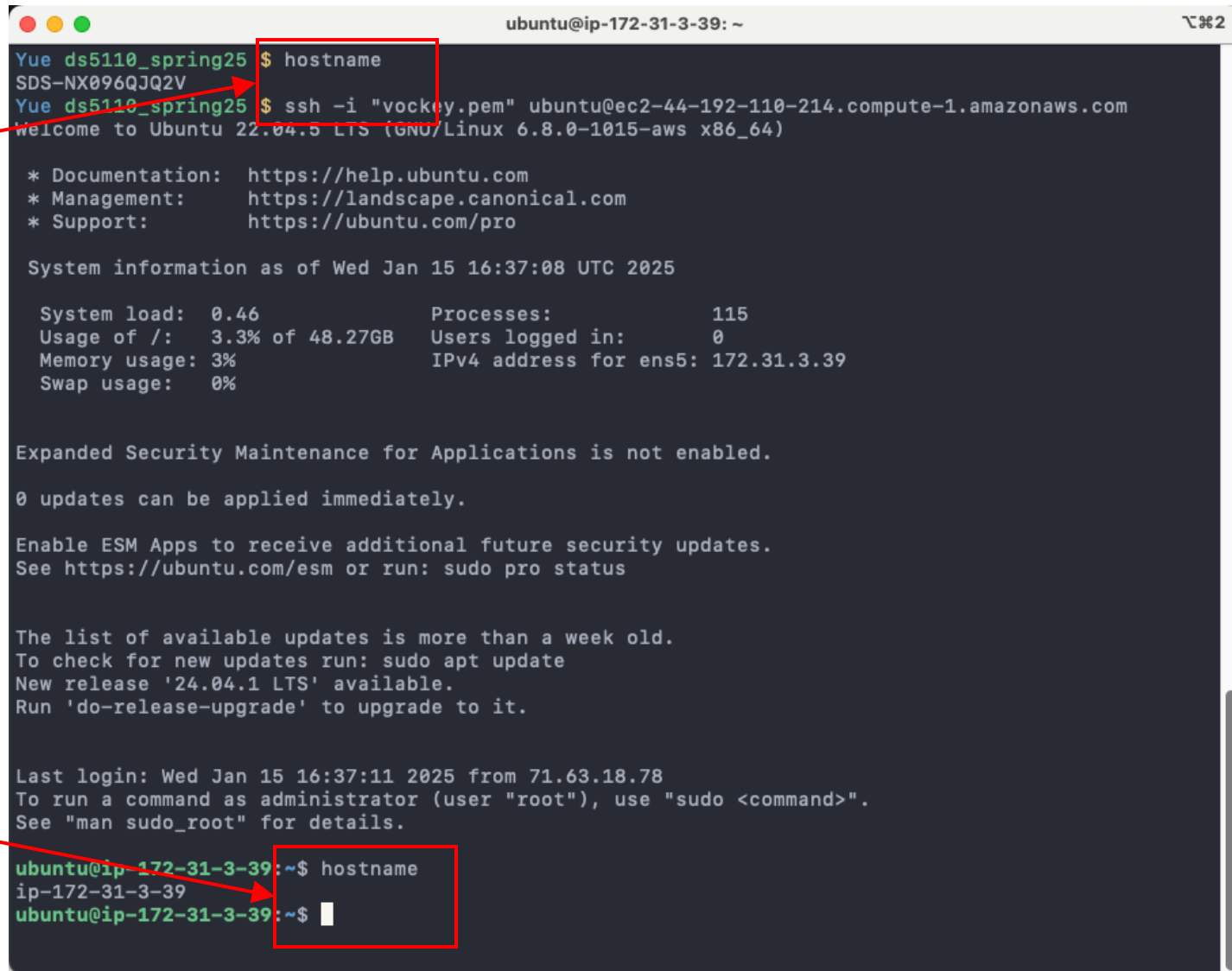
bash-3.2$
```

Annotations on the left side of the terminal:

- zsh prompt**: Three red arrows point to the prompts "Yue ~ \$" on the first, second, and third lines of the terminal.
- Start running bash**: A blue arrow points to the "bash" command on the third line.
- bash prompt**: A red arrow points to the "bash-3.2\$" prompt on the final line.

SSH: Secure shell

Running on
my laptop



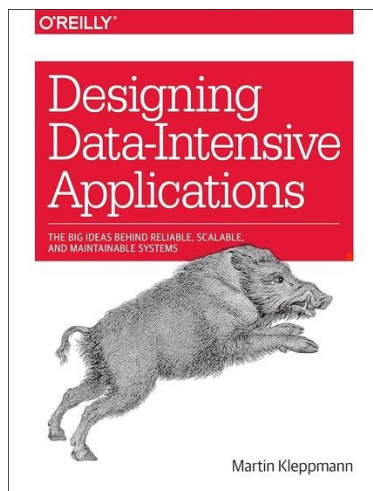
```
ubuntu@ip-172-31-3-39: ~  
Yue ds5110_spring25 $ hostname  
SDS-NX096QJQ2V  
Yue ds5110_spring25 $ ssh -i "vockey.pem" ubuntu@ec2-44-192-110-214.compute-1.amazonaws.com  
Welcome to Ubuntu 22.04.5 LTS (GNU/Linux 6.8.0-1015-aws x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:       https://ubuntu.com/pro  
  
System information as of Wed Jan 15 16:37:08 UTC 2025  
  
System load:  0.46          Processes:            115  
Usage of /:   3.3% of 48.27GB Users logged in:     0  
Memory usage: 3%          IPv4 address for ens5: 172.31.3.39  
Swap usage:  0%  
  
Expanded Security Maintenance for Applications is not enabled.  
  
0 updates can be applied immediately.  
  
Enable ESM Apps to receive additional future security updates.  
See https://ubuntu.com/esm or run: sudo pro status  
  
The list of available updates is more than a week old.  
To check for new updates run: sudo apt update  
New release '24.04.1 LTS' available.  
Run 'do-release-upgrade' to upgrade to it.  
  
Last login: Wed Jan 15 16:37:11 2025 from 71.63.18.78  
To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo_root" for details.  
ubuntu@ip-172-31-3-39:~$ hostname  
ip-172-31-3-39  
ubuntu@ip-172-31-3-39:~$
```

Running on
my EC2
VM

Linux pipe

Unix philosophy

- “Make each program do one thing well. To do a new job, build afresh rather than complicate old programs by adding new ‘features’.”
- “Expect the output of every program to become the input of another, as yet unknown, program. Don’t clutter output with extraneous information. Avoid stringently columnar or binary input formats. Don’t insist on interactive input.”




* Designing Data-Intensive Applications (“Batch Processing with Unix Tools” of Chapter 10)

Linux pipe

Simple Log Analysis

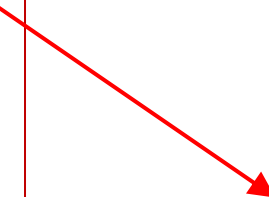
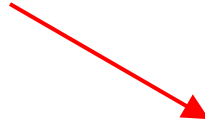
Various tools can take these log files and produce pretty reports about your website traffic, but for the sake of exercise, let's build our own, using basic Unix tools. For example, say you want to find the five most popular pages on your website. You can do this in a Unix shell as follows:ⁱ

```
cat /var/log/nginx/access.log | ❶  
awk '{print $7}' | ❷  
sort | ❸  
uniq -c | ❹  
sort -r -n | ❺  
head -n 5 | ❻
```

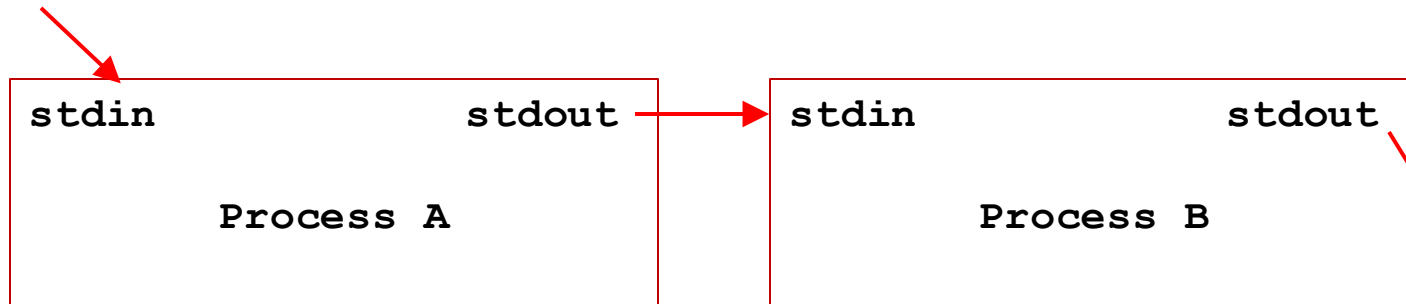


The pipe | connects output of one process to input of the next.

Standard input and output (I/O)



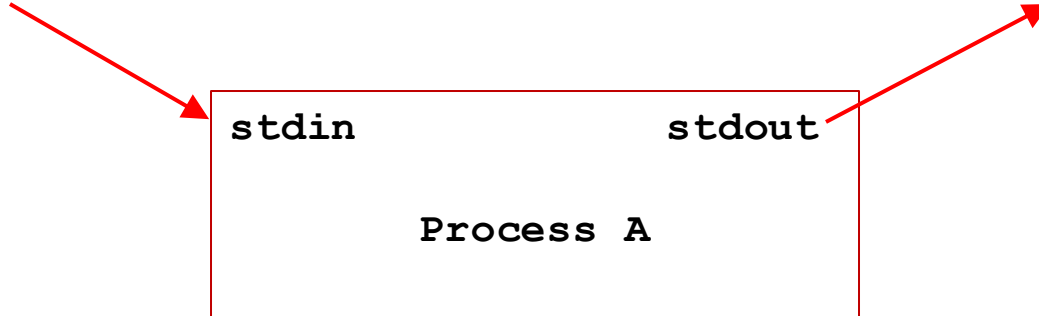
stdout => stdin



Command
A | B



Redirection



`output.txt`

Command
`A > output.txt`



Announcement

- A0 is out
 - Due by 11am ET, Thursday, Jan 30

Demos ...