University of California, Berkeley
Master of Information and Data Science (MIDS)
W205 – Fundamentals of Data Engineering

Week 3 – Linux CLI Refresher and GitHub git CLI Refresher

Agenda for Today's Class

- Attendance and Participation
- Announcements
- Schedule and Due Dates
- Work / Life / School Balance
- Asynch High Level Review in a Nutshell
- Breakouts
- Summary

Attendance and Participation

Please record your attendance and participation for today's class:

GitHub => ucb_mids_w205_repo => README.md => Attendance and Participation

Announcements

- Upcoming holidays and/or breaks
- Makeup classes for holidays
- Upcoming events
- Student evaluations
- Etc.

Schedule and Due Dates

Take a quick look at the next couple of weeks' due dates:

GitHub => ucb_mids_w205_repo => README.md => Schedule and Due Dates

Work / Life / School Balance
Open Discussion

Student feedback

- About 5 minutes
- How are things going related to work / life / school balance?
- How is w205 going? Difficulty? Time?
- Impact of any natural and/or man-made disasters
- Etc.

Asynch High Level Review in a Nutshell

Each week we will spend about 15 minutes reviewing the most important high level concepts from the asynch

OS

- OS = Operating System
- MS Windows, Mac OS, Linux, iPhone iOS, Android, etc.
- Layers
  - Kernel – innermost
  - GUI – outermost layer
  - Linux can peel off layers, runs much faster, other OS’s cannot
- Linux can run lean and mean without GUI
  - Workarounds – small web server, remote connections from desktop apps

Major Linux Distros

- Red Hat Branch
  - Corporate, professional servers
  - Large
  - CentOS, Amazon Linux, Oracle Linux, Fedora (experimental)
- Debian Branch
  - Desktop, IoT
  - Small, lean
  - Ubuntu, Lubuntu, Raspberry Pi OS

Linux Shells

- Shell = program that creates a Linux CLI
- BASH
  - Bourne-again shell
  - Most widely used
- Bourne, C, Korn, T, Z, etc.

Shell Scripts

- Shell scripts allow us to programmatically run Linux CLI commands
- Variables, if, loops, etc.
- crontab – allows us to schedule shell scripts
- BASH – commonly used
- Python – has be used for shell scripting

Source Code Control

- Multiple programmers work on same code without overwriting each others work
- Versioning of software, often support numerous versions at the same time
- Allows us to separate development from QA from UAT from Production

Breakouts

GitHub => ucb_mids_w205_repo => breakouts

(time permitting, we may not get to all of them)

Summary

Instructor will give a brief (about 2 minute) summary of today’s class.