Using CloudLab as a Scalable Platform for Teaching Cluster Computing

LINH B NGO
WEST CHESTER UNIVERSITY

JEFF DENTON
CLEMSON UNIVERSITY

Motivation

- Provide student with hands-on experience in working with *actual* computing cluster
- Physical resources
- Virtualization through local resources
  - Shared computing resources
    - Not scalable
    - Difficulty to maintain cloud infrastructure
  - Local VMs on students' laptop
    - Highly dependent on individual resources
    - Not an affordable approach

CloudLab

- www.cloudlab.us
- NSF-Funded Cloud-Based Infrastructure for Next Generation Computing Research
- Enable dynamic creation of profiles describing computing infrastructures using Python-based libraries
- Computing experiments can be instantiated on virtual machines or bare-metals using these profiles

Educational Setting

- Upper undergraduate (could be shared with graduate)
- Prerequisites
  - Networking
  - Operating System
- Recommended (but not required)
  - Linux Administration

Course Design

- System-oriented
- Narration-based
  - Cluster of Computers
    - Beowulf Cluster
    - Introduction to Message Passing (MPI)
    - Project Part 1: Deployment of Cluster of Computers
  - Shared and Parallel File Systems
    - Project Part 2: Incorporating NFS and OrangeFS
  - Parallel Programming Models
    - Project Part 3: Measuring performance improvement on complex MPI programs
  - Scheduler
    - Project Part 4: Incorporating scheduler (OpenPBS)

Course Design

- CloudLab Settings
  - Baseline OpenStack profile from CloudLab
  - Students develop offline *image templates* for login, storage, scheduler, and compute nodes using VirtualBox
  - *Image templates* are uploaded to persistent online storage (DropBox) to be incorporated into the default OpenStack profile for automatic deployment
- Small Individual Assignments
  - Prior to Project
    - Installation of a Linux baseline compute node (Improving Linux skillset) and running it in OpenStack
  - Interleaving with Project
    - Short MPI programming assignments

Assessment Technique

- Theoretical Knowledge Assessment
  - Multiple-choice quizzes weekly
  - Major exams (also multiple-choice)
- Practical Knowledge Assessment
  - In-class project discussion session where instructor can assess team (and individual) performance
  - The nature of the project requires the participation of all members
  - Skills necessary for project are also included in quizzes and exams

Measures of Success

- Difficult to follow-up, as most students take this class at their senior-level and already have a job lined up
- Anecdotal success stories
  - Recruitment of students into research labs
  - Enable new job opportunities
  - Increased usage of CloudLab as students moved on to other research groups
- Highly positive evaluation from students, even though the course is considered to be more difficult and time-consuming than the average CS courses at the same level

Lesson Learned

- Turn-around time for experiment deployment is still a noticeable issue
- Linux technical skills are highly desirable
  - Good programmers, bad system tinkerers
  - Linux administrative skills are considered “IT career”???
- The course would benefit students the most when taken earlier, yet students were not adequately prepared to take this course early
- Students will make the simplest error imaginable

Reproducibility

- Reproducible across institutions
  - Clemson to West Chester
- CloudLab's Github integration helps ease the transition process

THANK YOU!

QUESTIONS?

Linh B Ngo (LNGO@WCUPA.EDU)
Jeff Denton (DENTON@CLEMSON.EDU)

https://github.com/linhbngo/Distributed-and-Cluster-Computing