Lecture 0: Course Logistics

CSCI4180 (Fall 2013)

Introduction to Cloud Computing

About the Instructor

- Patrick P. C. Lee (http://www.cse.cuhk.edu.hk/~pclee)
  - B. Eng. in Information Eng., CUHK, 2001
  - M. Phil. in Computer Sci. & Eng., CUHK, 2003
  - Ph. D. in Computer Sci., Columbia, 2008
  - Postdoc in Computer Sci., UMass Amherst, 2009
  - Asst. Prof., Computer Sci. & Eng., CUHK, 2009 - now
- Research interests:
  - Applied topics in storage systems, networks, distributed systems, security and dependability. Focus on system prototyping and implementation.

Course Information

- Course website:
  - http://www.cse.cuhk.edu.hk/csci4180
- TAs
  - QIN Chuan (official TA), SHB 118, cqin@cse
  - XU Min (official TA), SHB 118, xum@cse
  - CHAN Chun Wing Jeremy (supporting TA), SHB 118, cwchan@cse
  - LI Yan Kit Wilson (supporting TA), SHB 118, liyk@cse

Course Prerequisites

- CENG3150/CSCI3150, or equivalent
  - If you haven't taken it or have failed it, then you are advised NOT to take this course
- Comfortable with Java programming
  - We will provide crash courses in the first two weeks of tutorials
- Comfortable with Linux

Course Newsgroup

- Facebook group:
  - http://www.facebook.com/groups/200985016728843/
- I will make announcements in class, on course website, and Facebook group
- Please post your assignment questions to the Facebook group

Course Materials

- Reference books:
  - Tom White, **"Hadoop: The Definitive Guide"**, Second Edition, O'Reilly Media. Reserved in CUHK library
  - Jimmy Lin and Chris Dyer, **"Data-Intensive Text Processing with MapReduce"**, Morgan and Claypool, 2010. You can find an online copy on the CUHK library website.
- Some required research papers are posted online
- ***It's important to read!!***

Course Assessment

- 3 programming assignments (40%)
  - Group assignments of **2-3** people
  - Programming on a real cloud platform based on OpenStack
    - http://www.openstack.org
- Final exam (60%)

Windows Azure

- Trial on real production cloud: **Windows Azure**
  - Run applications across a global network of Microsoft-managed data centers
- Pronunciation:
  - Azure → /ˈæʒə(r)/
  - Measure → /ˈmeʒə(r)/
- Supported by Windows Azure Educator Grant

Academic Honesty

- In short, don't cheat!
- Don't copy code or solutions from your classmates or third-party sources, and don't let others copy yours.
- Cases will be reported to the school
- Details:
  - CUHK: http://www.cuhk.edu.hk/policy/academichonesty/
  - Faculty of engineering: http://www.cse.cuhk.edu.hk/v5/other/A5_BookletN3.pdf
- Ask me if you are unsure

Course Objectives

- **Goals**:
  - Understand the essentials of cloud computing
  - Learn the applied methodologies of using cloud computing for solving practical engineering problems

10 Questions

- What is (and is not) cloud computing?
- How does Google store big data in a scalable, reliable way?
- How does Google perform analytics?
- How do we write elegant programs for big data processing (like Google)?
- How is PageRank implemented with cloud computing?
- How does Yahoo! coordinate thousands of machines?
- How does Dropbox make profit?
- How does Amazon find data in different geographic regions?
- How does Facebook manage your photos?
- How do I allocate resources of a single computer for 100 people?

Centered around two main areas: ***Computation*** *and* ***Storage***

Topics to Cover

- Fundamentals of Cloud Computing
- Big data computation
  - Overview of Hadoop: MapReduce and HDFS
  - MapReduce Programming
  - MapReduce Algorithm Design
  - MapReduce Applications (e.g., PageRank)
- Hadoop data management
  - BigTable, HBase
  - Zookeeper

Topics to Cover

- Cloud storage:
  - Deduplication
  - Dropbox
  - Amazon
  - Azure
  - Facebook
- Virtualization

Student/Faculty Expectations

- Goal: to enhance teaching and learning qualities

- http://www.erg.cuhk.edu.hk/Student-Faculty-Expectations