CompSci 401: Cloud Computing

Introduction

Prof. Ítalo Cunha

昆山杜克大学
DUKE KUNSHAN UNIVERSITY

Introduction

- Ítalo Cunha <cunha@dcc.ufmg.br>
  - B.Sc. from UFMG, Brazil (2004)
  - M.Sc. from UFMG, Brazil (2007)
  - Ph.D. from UPMC Sorbonne, France (2011)
  - Professor at UFMG since 2012
  - Post-doc at Columbia University in 2019

Introduction

- Ítalo Cunha <cunha@dcc.ufmg.br>
  - Interest Areas
    - Computer Networks
    - Distributed Systems
    - Operating Systems
    - Cybersecurity
  - Research
    - Internet routing (traffic engineering, monitoring, and modeling)
    - Network security (monitoring, characterizing, and tracking malicious activities)
    - Network performance (congestion controls and transport optimizations)

Course Goals

- Understand the cloud computing paradigm in general
  - Underlying technologies
  - Core concepts
  - Availability, performance, and cost trade-offs
- Learn how to build applications in the cloud
  - Containerization and virtualization
  - Serverless, function-based, and stream computing
  - Microservices
  - Automation, continuous integration, and continuous delivery
- Get experience with data analysis in the cloud

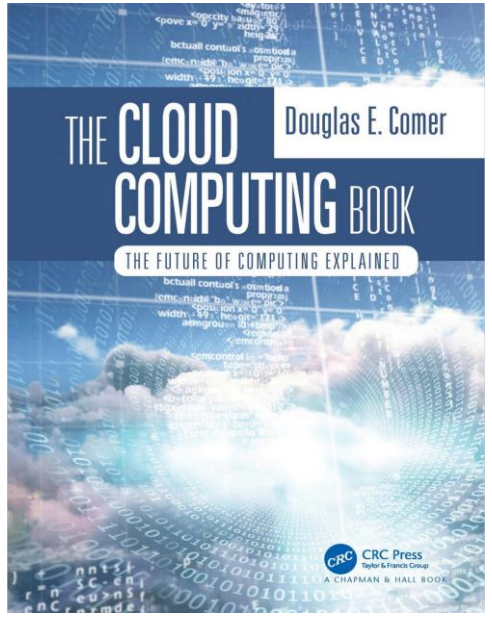
How you may apply the concepts in this course

- Developers
  - Many applications today run *totally* on the cloud
  - *Most* applications have some functionality that runs on the cloud
- Data scientists
  - Many data science processes run or can benefit from the cloud
  - A data science task can be approach in different ways, need to choose
- Managers and software architects
  - Choosing how and where a solution should be built is key

Course Requirements

- Important
  - Experience with coding algorithms and programs (CompSci 201)
    - We will use Java, Python, and Bash
- Beneficial
  - Familiarity with computer architecture concepts (CompSci 205)
  - Familiarity with manipulating large datasets (CompSci 301)
  - Familiarity with computer network concepts (CompSci 311)

Planned Activities



- Lectures covering concepts
  - *The Cloud Computing Book: The Future of Computing Explained.* Douglas Comer. Chapman and Hall/CRC, 1st ed.
  - Classes and pre-recorded videos
- Reading of conference papers covering select topics in depth
  - Students will choose a topic and record a short video about it
- Programming assignments
  - Hands-on practice with cloud computing technologies
- Midterm and final exams
  - Focusing on concepts (lecture content)
  - But also covering practical knowledge (assignments)

Course Structure

- Week 1: Foundations, motivation, use cases, services, infrastructure
- Week 2: Virtualization
- Week 3: Cloud and distributed storage
- Week 4: Automation, monitoring, and orchestration
- Week 5: Programming paradigms
- Week 6: DevOps, security, privacy
- Week 7: Data analytics in the cloud

Assignments

- Assignment 1: Cloud computing paradigms
  - MapReduce
  - Spark
- Assignment 2: Building cloud-native applications
  - Containerization
  - Continuous Integration
  - Autoscaling
- Assignment 3: Stream processing
  - Serverless functions
  - Data handling
- Assignment 1b: Data analytics in the cloud

Assignments

- We will use open-source frameworks in our assignments
- But students can try out commercial offerings for the technologies used in the assignments from cloud providers
  - Alibaba Cloud
  - Amazon AWS
  - Google Cloud
  - IBM Cloud
  - Microsoft Azure
  - Oracle Cloud
- All cloud providers have *free tier* services or *academic accounts* you can use to try out the technologies in the assignments

Grading

- Assignments: 45%
- Midterm: 15%
- Final exam: 30%
- Video: 10%

Suggestions Are Welcome

- Please e-mail me your suggestions as early as possible

Books

- **The Cloud Computing Book: The Future of Computing Explained**
  Douglas Comer. Chapman and Hall/CRC, 1st ed
  Available from the university library.
- *Cloud Computing for Science and Engineering*
  Ian Foster and Dennis B. Gannon. MIT Press, 1st ed
  https://cloud4scieng.org/chapters
- *Learning Spark*
  J. S. Damji, B. Wenig, T. Das, and D.Lee. O'Reilly, 2nd ed
  https://databricks.com/p/ebook/learning-spark-from-oreilly
- *Mining of Massive Datasets*
  J. Leskovec, A.Rajaraman, and J. Ullman. Cambridge Univ. Press, 3rd ed
  http://www.mmds.org