Summer 2020

CS 643-850: Cloud Computing

Cristian Borcea

CS643850-Cloud Computing

Jump to Today

CS 643 850 Summer 2020 - Syllabus

Instructor

- Cristian Borcea
- borcea@njit.edu
- https://cs.njit.edu/~borcea (Links to an external site.)
- Live office hours: Wednesdays, 6pm-7pm, Webex: https://njit.webex.com/meet/borcea (Links to an external site.)
  - You can also join by phone: 1-650-479-3207 *Call-in toll number (US/Canada).* Access code: 928 115 913

Short Description

The course presents a comprehensive view of cloud computing, from platforms and services to programming and infrastructure. The topics include: cloud computing platforms, with examples from Amazon Web Services (AWS), Google Cloud, and Microsoft Azure; cloud services for data analytics, machine learning, mobile computing, IoT, edge computing, security and privacy, and devops; programming frameworks for parallel computing in the cloud; distributed storage in the cloud; and virtualization and containerization. The course includes homework, programming assignments, and research paper presentations. The programming assignments will be done in AWS.

Learning Outcomes

Upon the successful completion of this course, the student should be able to:

- Analyze the trade-offs between deploying applications in the cloud and over local infrastructure
- Compare the advantages and disadvantages of different types of cloud platforms
- Deploy applications over commercial cloud platforms
- Program data intensive parallel applications in the cloud
- Analyze the performance, scalability, and availability of the cloud systems and applications
- Identify security and privacy issues in the cloud
- Collaborate to present state-of-the-art cloud research

Why Take This Course?

Cloud computing represents a major paradigm shift in computing from the era of personal computers to the era of computing as utility. Most major Internet services are already deployed in the "the cloud." We store most of our data in "the cloud" and execute most applications from "the cloud." This course is aimed at all graduate students (both M.S. and Ph.D. students) who want to learn how to design and program cloud services and applications as well as how to build and administer cloud systems. By studying real-world systems developed in industry, students will acquire cutting-edge knowledge that may be a major advantage when searching for a job.

Lectures and Readings

There is no book required for this class. Each lecture is based on research papers and/or online documentation covering a specific topic (i.e., readings). Each lecture is posted as a module, which includes the lecture slides, the voice-over video recordings of each sub-module of the lecture, and the assigned readings. The readings should be used as reference material to clarify and add details to lectures.

Exams

There will be two exams: a midterm, and a final exam. Both exams are closed book (i.e., papers, notes). The final exam

will cover only the material taught after the midterm. Both exams will take place online using ProctorU and Canvas. In case of missing an exam, a make-up may be taken only after providing written documentation to the Dean of Students.

Homework

Homework will be assigned 5 times during the semester to prepare students with the type of questions encountered in exams. The solutions will be discussed during the live office hours.

Programming Assignments

There will be two individual programming assignments. The first is to build an image recognition pipeline in Amazon AWS, using two EC2 instances, S3, SQS, and Rekognition. The assignment must be done in Java on Amazon Linux VMs. You will learn how to use the AWS cloud platform and how to develop an AWS application that uses existing cloud services. The second is to build a machine learning prediction model in Spark/MLlib over AWS. The model must be trained in parallel on multiple EC2 instances. The assignment must be implemented in Java, Scala, or Python on Ubuntu Linux. You will learn how to develop parallel machine learning applications in the AWS cloud platform.

Research Paper Presentations

Students will present, in groups of two or three, one research paper during the semester. These papers cover state-of-the-art research in cloud computing. The video recordings of voice over slides will be uploaded in Canvas for the whole class to watch. The presentations must be uploaded in the week to which the papers are assigned. Extra-credit is available for asking good questions about the presentations.

Grading

- Midterm exam - 25%
- Final exam - 25%
- Programming Assignment 1 - 10%
- Programming Assignment 2 - 20%
- Research presentation - 10%
- Homework - 10%

Schedule

- **Module 1 [May 18-22]: Course overview. Introduction to Cloud Computing.**
  - Form groups for research paper presentations by 23 May
- **Module 2 [May 23-28]: Cloud Platforms I: Infrastructure as a Service (IaaS), AWS.**
  - Homework 1 handed out on 23 May
  - Programming assignment 1 handed out on 23 May
  - Research papers assigned to groups by 25 May
- **Module 3 [May 29 - June 2]: Cloud Platforms II: Platform as a Service (PaaS), Google App Engine, Windows Azure.**
  - Homework 1 due on 29 May
- **Module 4 [June 3-8]: Cloud Platforms III: Serverless Computing; Function as a Service (FaaS).**
  - Homework 2 handed out on 3 June
- **Module 5 [June 9-14]: Cloud Services I: Data Analytics and Machine Learning.**
  - Homework 2 due on 9 June
  - Research paper presentation 1 due on 10 June
- **Module 6 [June 15-20]: Cloud Services II: Mobile, IoT, and Edge Computing.**
  - Homework 3 handed out on 15 June
  - Research paper presentation 2 due on 17 June
- **Module 7 [June 21-26]: Cloud Services III: Security and Privacy, Devops.**
  - Programming assignment 1 due on 22 June
  - Homework 3 due on 23 June

- **Module 8 [June 27-July 1]: Parallel Programming in the Cloud I: Google's MapReduce, Apache's Hadoop, Yahoo's Pig Latin.**
  - Midterm exam is on 29 June (covers modules 1-7)
- **Module 9 [July 2-7]: Parallel Programming in the Cloud II: Apache's Spark, Storm and Zookeper.**
  - Programming assignment 2 handed out on 2 July
- **Module 10 [July 8-13]: Cloud Storage Systems I: Google's GFS and BigTable.**
  - Homework 4 handed out on 8 July
  - Research paper presentation 3 due on 10 July
- **Module 11 [July 14-19]: Cloud Storage Systems II: Amazon's Dynamo and Other Cloud Databases.**
  - Homework 4 due on 14 July
- **Module 12 [July 20-25]: Virtualization I: VMWare, XEN, Live VM Migration.**
  - Homework 5 handed out on 20 July
  - Research paper presentation 4 due on 22 July
- **Module 13 [July 26-August 3]: Virtualization II: Containerization, Docker, Kubernetes.**
  - Homework 5 due on 26 July
  - Programming assignment 2 due on 29 July
  - Final exam is on 3 August

Academic Integrity

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at this link: University Policy on Academic Integrity (Links to an external site.).

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu.

Modifications to Syllabus

The students will be consulted and must agree to any modifications or deviations from the syllabus throughout the course of the semester.