New Jersey Institute of Technology

Digital Commons @ NJIT

Computer Science Syllabi

NJIT Syllabi

Spring 2020

CS 643-102: Cloud Computing

Cristian Borcea

Follow this and additional works at: https://digitalcommons.njit.edu/cs-syllabi

Recommended Citation

Borcea, Cristian, "CS 643-102: Cloud Computing" (2020). *Computer Science Syllabi*. 41.
https://digitalcommons.njit.edu/cs-syllabi/41

CS 643 102: Cloud Computing, Spring '20

Wednesdays, 6-8:50pm, CKB 215

Instructor

Cristian Borcea
Office: GITC 4201A
Phone: 973 596-2866
Office Hours: Mondays 4pm-6pm and by appointment

borcea@njit.edu

Announcement

- **03/19** Due to the coronavirus crisis, the course will move online in Canvas (please check your email for more updates). We will use video recordings for lectures, and we will have shorter meetings at 6pm on Wednesdays to give you a short overviw of the lectures, discuss homework/assignments (including solutions), and answer any questions that you may have. For consistency, I will keep updating the lecture slides and assignments/homework in this page as well, but everything will also be posted in Canvas.

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.

Prerequisites

CS 656 or instructor's permission. If you didn't take CS 656, but you would like to take this class, you should come and talk with me about your background. Basic Unix/Linux skills and good programming skills are necessary for the assignment and the project.

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). The lectures will be posted a few days before each class, and students are advised to familiarize with them before class in order to be able to actively participate in class discussions. 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. 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 in class.

Programming Assignments

There will be two individual programming assignments. The first will consists of building, training, and deploying a machine learning (ML) model in Amazon Web Services (AWS). Its goal is to help you learn how to use the AWS platform and how to deploy applications over this platform. The second assignment will consists of building an application using Apache's Spark, MLlib, and HDFS, and running the application in parallel on multiple EC2 instances in AWS. Its goal is to learn how to program and deploy parallel applications in the cloud.

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 presentations (using PowerPoint slides) will be done online: 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%

Weekly Schedule

Week	Lecture topics	Readings	Research papers
1	Course overview. Introduction to Cloud Computing. Slides.		
2	Cloud Platforms I: Infrastructure as a Service (IaaS), AWS. Slides. Homework 1 handed out.	- AWS Documentation.	
3	Cloud Platforms II: Platform as a Service (PaaS), Google App Engine, Windows Azure. Homework 1 due. Slides. Assignment of research papers to groups finalized.	- Google AppEngine Documentation. - Microsoft Azure Documentation.	
4	Cloud Platforms III: Serverless Computing; Function as a Service (FaaS). Slides. Homework 2 handed out. Programming Assignment 1 handed out.	- Cloud Programming Simplified: A Berkeley View on Serverless Computing, UC Berkeley Technical Report, 2019. - https://aws.amazon.com/lambda/ - https://cloud.google.com/functions/ - https://azure.microsoft.com/en-us/services/functions/ - https://aws.amazon.com/step-functions/ - https://aws.amazon.com/rds/aurora/serverless/	
5	Cloud Services I: Data Analytics and Machine Learning. Slides. Homework 2 due. First set of research presentations due.	- https://cloud.google.com/ai-platform/ - https://www.tensorflow.org/ - https://aws.amazon.com/machine-learning/ - https://azure.microsoft.com/en-us/overview/ai-platform/ - https://azure.microsoft.com/en-us/product-categories/analytics/ - https://aws.amazon.com/big-data/datalakes-and-analytics/ - https://cloud.google.com/solutions	- Narrowing the Gap Between Serverless and its State with Storage Functions. ACM SoCC, 2019. - Pocket: Elastic Ephemeral Storage for Serverless Analytics. Usenix

		/big-data/	OSDI, 2018. • Shuffling, Fast and Slow: Scalable Analytics on Serverless Infrastructure. Usenix NSDI, 2019.
6	Cloud Services II: Mobile, IoT, and Edge Computing. Homework 3 handed out. Slides.	• https://aws.amazon.com/amplify/ • https://azure.microsoft.com/en-us/product-categories/mobile/ • https://firebase.google.com/ • https://aws.amazon.com/greengrass/ • https://aws.amazon.com/lambda/edge/ • https://cloud.google.com/iot-core/ • https://cloud.google.com/edge-tpu/ • https://azure.microsoft.com/en-us/overview/iot/ • https://azure.microsoft.com/en-us/services/iot-edge/	
7	Cloud Services III: Security and Privacy, Devops. Homework 3 due. Slides. Second set of research presentations due.	• https://aws.amazon.com/security/ • https://aws.amazon.com/privacy/ • https://cloud.google.com/security/ • https://azure.microsoft.com/en-us/overview/trusted-cloud/ • https://azure.microsoft.com/en-us/product-categories/devops/ • https://aws.amazon.com/devops/ • https://cloud.google.com/devops/	• Cirrus: a Serverless Framework for End-to-end MLWorkflows. ACM SoCC, 2019. • Gandiva: Introspective Cluster Scheduling for Deep Learning. Usenix OSDI, 2018.
8	Midterm (material covered in weeks 1-7). Discussion of midterm solutions. Programming assignment 1 due.		
9	Parallel Programming in the Cloud I: Google's MapReduce, Apache's Hadoop, Yahoo's Pig Latin. Slides.	• J. Dean and S. Ghemawat, MapReduce: Simplified Data Processing in Large Clusters, Communications of the ACM, 2008. • Hadoop Documentation • C. Olston, B. Reed, U. Srivastava, R. Kumar, and A. Tomkins, Pig Latin: A Not-So-Foreign Language for Data Processing, ACM SIGMOD 2008.	

10	Parallel Programming in the Cloud II: Apache's Spark, Storm and Zookeper. Second programming assignment handed out. Third set of research presentations due. Slides.	- Apache Spark - Apache Storm - Apache Zookeper	- BigDL: A Distributed Deep Learning Framework for Big Data. ACM SoCC, 2019. - Ray: A Distributed Framework for Emerging AI Applications, Usenix OSDI, 2018. - Cartel: A System for Collaborative Transfer Learning at the Edge. ACM SoCC, 2019.
11	Cloud Storage Systems I: Google's GFS and BigTable. Homework 4 handed out. Slides.	S. Ghemawat, H. Gobioff, and S-T. Leung, The Google File System , ACM SOSP 2003. F. Chang, J. Dean, S. Ghemawat, W. Hsieh, D. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. Gruber, Bigtable: A Distributed Storage System for Structured Data, Usenix OSDI 2006.	
12	Cloud Storage Systems II: Amazon's Dynamo and Other Cloud Databases. Homework 4 due. Fourth set of research presentations due. Slides.	G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, amd W. Vogels, Dynamo: Amazon's Highly Available Key-value Store, ACM SOSP 2007. - https://aws.amazon.com/products/databases/ - https://azure.microsoft.com/en-us/product-categories/databases/ - https://cloud.google.com/products/databases/	- Why Does Your Data Leak? Uncovering the Data Leakage in Cloud from Mobile Apps. IEEE S&P, 2019. - Houdini's Escape: Breaking the Resource Rein of Linux Control Groups. ACM CCS, 2019.
13	Virtualization I: VMWare, XEN, Live VM Migration. Homework 5 handed out. Slides.	J. Sugerman, G. Venkitachalam, and B-H. Lin, Virtualizing I/O Devices on VMWare Workstation's Hosted Virtual Machine Monitor, Usenix 2001. O. Agesen, A. Garthwaite, J. Sheldon, P. Subrahmanyam, The evolution of an x86 virtual machine monitor, ACM SIGOPS Operating Systems Review, 2010.	

		P. Barham, B. Dragovic, K. Fraser, S. Hand, T, Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, Xen and the Art of Virtualization, ACM SOSP 2003. C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, Live Migration of Virtual Machines, Usenix NSDI 2005.	
14	Virtualization II: Containerization, Docker, Kubernetes. Homework 5 due. Programming assignment 2 due.	- https://docs.docker.com/ - https://aws.amazon.com /kubernetes/ - https://aws.amazon.com/fargate/ - https://aws.amazon.com/ecs/ - https://cloud.google.com /kubernetes-engine/ - https://azure.microsoft.com/en-us/services/kubernetes-service/	
15	Final exam (material covered in weeks 9-14).		

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.

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.