Fall 2019

CS 643-101: Cloud Computing

Cristian Borcea

CS 643 101: Cloud Computing, Fall '19

Wednesdays, 6-8:50pm, CKB 330

Instructor

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

Announcements

- **08/26** The weekly topics and readings as well as the research topics that you will present in class are posted below. I also posted the slides for the first lecture. I emailed you the credentials to access these materials.

Short description

The course presents a top-down view of cloud computing, from applications and administration to programming and infrastructure. Its main focus is on parallel programming techniques for cloud computing and large scale distributed systems which form the cloud infrastructure. The topics include: overview of cloud computing, cloud systems, parallel processing in the cloud, distributed storage systems, virtualization, security in the cloud, and multicore operating systems. Students will study state-of-the-art solutions for cloud computing developed by Google, Amazon, Microsoft, VMWare, etc. Students will also apply what they learn in one programming assignments and one project executed over Amazon Web Services.

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 the local infrastructure.
- Compare the advantages and disadvantages of various cloud computing platforms.
- Deploy applications over commercial cloud computing infrastructures such as Amazon Web Services, Microsoft Azure, and Google Cloud.
- Program data intensive parallel applications in the cloud.
- Analyze the performance, scalability, and availability of the underlying cloud technologies and software.
- Identify security and privacy issues in cloud computing.
- Solve a real-world problem using cloud computing through group collaboration.

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." In the near future, we may store all 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 as well as how to build and administer cloud systems. By studying real-world systems developed in industry during the past few years, 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 recent papers/articles covering a specific topic. Students are required to read the papers before the class and participate in the discussions in class.

Research Presentations

Students will present, in groups of two to four, one research topic during the semester. These topics cover very recent developments in cloud computing. The presentations (using power point slides) will take place in class, and extra-credit will be assigned for active participation in discussions.

Programming Assignment

There will be one individual programming assignment consisting of creating an AMI for Hadoop and implementing short Hadoop programs on the Amazon Web Services platform.

Project

Students will choose their project topic and work in teams of three or four to design, implement, and evaluate cloud applications using Hadoop on the Amazon Web Services platform. General requirements for all projects will be discussed in class after the midterm.

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 4 times during the semester to prepare students with the type of questions encountered in exams. The solutions will be discussed in class.

Grading

- Midterm exam - 25%
- Final exam - 25%
- Project - 15%
- Research presentation - 15%
- Programming Assignment - 10%

Homework - 10%

Schedule

Week	Lecture topics	Required readings	Research topics
1	Course overview. Introduction to Cloud Computing. Slides.	• M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the Clouds: A Berkeley View of Cloud Computing, Technical Report No. UCB/EECS-2009-28.	
2	Cloud Platforms I. Amazon Web Services. Google AppEngine. Homework 1 handed out.	• Amazon Web Services Documentation. • Google AppEngine Documentation.	
3	Cloud Platforms II. Windows Azure. Discussion of homework 1 solutions. Programming assignment handed out.	• Microsoft Azure Documentation.	
4	Parallel Programming in the Cloud I. Google's MapReduce. Apache's Hadoop.	• J. Dean and S. Ghemawat, MapReduce: Simplified Data Processing in Large Clusters, Communications of the ACM, 2008. • Hadoop Documentation	*1. Serverless Computing and Cloud Functions* • 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	Parallel Programming in the Cloud II. Yahoo's Pig Latin. Homework 2 handed out. Discussion of project requirements.	• C. Olston, B. Reed, U. Srivastava, R. Kumar, and A. Tomkins, Pig Latin: A Not-So-Foreign Language for Data Processing, ACM SIGMOD 2008.	*2. Apache Spark* • https://spark.apache.org/ *3. Apache Storm* • https://storm.apache.org/
6	Parallel	• M. Isard, M. Budiu, Y. Yu,	

	Programming in the Cloud III. Microsoft's Dryad and DryadLINQ. Programming assignment due. Discussion of homework 2 solutions.	A. Birrell, and D. Fetterly, Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks, ACM EuroSys 2007. • Y. Yu, M. Isard, D. Fetterly, M. Budiu, U. Erlingsson, P. K. Gunda, and J. Currey DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing using a High-Level Language, Usenix OSDI 2008.	*4. Apache Zookeeper* • https://zookeeper.apache.org/
7	Midterm (material covered in the first 6 weeks, including research topics). Discussion of midterm solutions.		
8	Distributed Storage Systems I. Google's GFS and BigTable. Presentation of project outlines (who, what, why, how).	• 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.	
9	Distributed Storage Systems II. Amazon's Dynamo. Homework 3 handed out.	• 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.	*5. Databases in the Cloud* • https://aws.amazon.com/products/databases/ • https://azure.microsoft.com/en-us/product-categories/databases/ • https://cloud.google.com/products/databases/
10	Virtualization I. VMWare virtualization. Containers. Discussion of homework 3 solutions.	• J. Sugerman, G. Venkitachalam, and B-H. Lin, Virtualizing I/O Devices on VMWare Workstation's Hosted Virtual Machine Monitor, Usenix	*6. Machine Learning in the Cloud* • https://cloud.google.com/ml-engine/

		2001. • O. Agesen, A. Garthwaite, J. Sheldon, P. Subrahmanyam, The evolution of an x86 virtual machine monitor, ACM SIGOPS Operating Systems Review, 2010.	• https://www.tensorflow.org/ • https://aws.amazon.com/machine-learning/ • https://azure.microsoft.com/en-us/overview/ai-platform/
11	Virtualization II. Xen virtualization; Virtual machine migration.	• 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.	*7. Analytics in the Cloud (non-ML and non-DB)* • https://azure.microsoft.com/en-us/product-categories/analytics/ • https://aws.amazon.com/big-data/datalakes-and-analytics/ • https://cloud.google.com/solutions/big-data/ *8. Kubernetes* • 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/
12	Cloud Security. Homework 4 handed out.	• T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds, ACM CCS 2009.	*9. Mobile Cloud* • https://aws.amazon.com/amplify/ • https://azure.microsoft.com/en-us/product-categories/mobile/ • https://firebase.google.com/ *10. Cloud IoT and Edge Computing* • 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/
13	Multicore Operating Systems. Discussion of homework 4 solutions.	• A. Baumann, P. Barham, P. Dagand, T. Harris, E. Isaacs, S. Peter, T. Roscoe, A. Schupbach, and A. Singhania, The Multikernel: A new OS architecture for scalable multicore systems, ACM SOSP 2009.	*11. Security and Privacy in the Cloud* • https://aws.amazon.com/security/ • https://aws.amazon.com/privacy/ • https://cloud.google.com/security/ • https://azure.microsoft.com/en-us/overview/trusted-cloud/ *12. DevOps in the Cloud* • https://azure.microsoft.com/en-us/product-

			categories/devops/ • https://aws.amazon.com/devops/ • https://cloud.google.com/devops/
14	Final project presentations.		
15	Final exam (material covered in weeks 8-13, including the research topics).		

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.