CS 498: Cloud Computing Applications Syllabus

Instructor: Reza Farivar

Course Outline 2022

Week	Dates	Topics	Estimated time to cover learning material (hours)	Quiz /Exam (hours) *Quizzes are due on Sunday nights*	Programming Assignment (aka. Machine Problem or MP) + *estimated completion time (**AVERAGE**, but you may need more time, so plan accordingly.)*	MP Due Date *Due on the last day of the week, on Sunday night 11:59PM CST*
1		Course Orientation + Cloud Computing Foundations, Cloudonomics, and cloud models: IaaS, PaaS, SaaS	~1:30 hours for course logistics ~3:30 Learning Material	Course Access Quiz (5 mins.) ProctorU Readiness, Orientation (30 mins.) Orientation quiz (15 mins.) Week 1 quiz (1 hours)	1) Java/Python Warmup (~3 hours) *If you are really struggling in this MP, you may want to reconsider taking this course*	
2		Networking, IP, HTTP, REST, VPC	4	Week 2 quiz (1 hour)	2) AWS Load Balancer (~ 6 hours)	1) Java/Python Warmup
3		Serverless Computing	2	Week 3 quiz (1 hour)	3) AWS Lex & Lambda (~ 6 hours)	2) AWS Load Balancer
4		Big Data Programming: MapReduce model, Hadoop, YARN, Spark and HDFS	5	Week 4 quiz (1 hour)	4) Hadoop MapReduce (~ 8 hours)	----------
5		Data Storage Part 1: Cloud-based Storage (Object, filesystem, archival)	3:30	Week 5 quiz (1 hour)	5) Spark MapReduce (~ 6 hours)	3) AWS Lex & Lambda
6		Data Storage Part 2: Cloud-based Databases and Data warehousing (RDBMS, NewSQL, NoSQL)	3	Week 6 quiz (1 hour)	6) AWS RDS & ElastiCache (~6 hours)	4) Hadoop MapReduce
7		Data Storage Part 3: Scalable Data Storage (Caching, HBase, Spark SQL, HIVE, Queues, PubSub systems)	4	Week 7 quiz (1 hour)	7) HBase (~ 6 hours) 8) Spark SQL (~ 4 hours)	5) Spark MapReduce
8		*No New Content – **Midterm** Exam preparation and Exam.*		7 practice quizzes (30 mins. each) Midterm exam is 90 minutes (ProctorU)	----------	----------
9		Spring Break	----------	----------	----------	----------
10		Cloud Based Analytics: Data Cube, Columnar storage, Data Lake	2	Week 10 quiz (1 hour)	9) Cloud Analytics and Visualization (~ 6 hours)	6) AWS RDS & ElastiCache

11		Graph Processing, Graph Databases and Machine Learning in the Cloud	5	Week 11 quiz (1 hour)	10) Spark - GraphFrames and MLLib (~ 6 hours)	7) HBase
12		Fast Data Processing and Streaming in the Cloud	5	Week 12 quiz (1 hour)	11) Storm and Flux (~ 6 hours)	8) SparkSQL
13		Virtualization, Containers and Docker	4	Week 13 quiz (1 hour)	12) Containerization & Kubernetes (~6 hours)	9) Cloud Analytics and Visualization
14		Container Orchestration Part 1: Docker Swarm, ECS and ACI	3	Week 14 quiz (1 hour)	---------	10) Spark - GraphFrames and MLLib
15		Container Orchestration Part 2: Kubernetes	2	Week 15 quiz (1 hour)	---------	11) Storm and Flux
16		Future Developments in the Cloud, Course Wrap-up	2:30	Week 16 quiz (1 hour)	---------	12) Containerization & Kubernetes
17		*No New Content –**Final** Exam preparation and Exam*		7 practice quizzes (30 mins. each) Final Exam is 90 minutes (ProctorU) Course Final Feedback (30 mins.)	---------	---------

Course Description

Welcome to CS 498: Cloud Computing Applications! This 17-week course is designed to give you a comprehensive view of the world of cloud computing and Big Data. Each week has between 3 to 5 hours of learning material (video lectures + readings), a quiz (usually 1 hour), and for most weeks a programming assignment. **You should expect to spend 10-12 hours per week on this course in average**.

Please note that the times specified for the programming assignments are *averages*. This means that some students may finish the assignment in 2 hours, and a few others may need to spend 20 hours or more on an MP. Likewise, some weeks have more training material than others. Since all course material is available from day 1, we expect you to proactively plan your schedule in advance. The programming assignments are a major component of the learning process of this course, and to some extent they are designed as self-learning and exploration opportunities. They are all auto-graded, and you have unlimited number of tries before each assignment's deadline.

In this course we cover a multitude of technologies that comprise the modern stack of cloud computing. Cloud computing is an information technology revolution that has impacted many enterprise computing systems in major ways, and it will change the face of computing in the years to come.

We start by introducing some major concepts in cloud computing, the economical foundations of cloud computing, and the concept of Big Data. We also cover the concept of software defined architectures, and how cloud service providers organize their offerings. We also will compare Infrastructure as a Service offered by the big three: Amazon, Google, and Microsoft, including Infrastructure as a Service, Platform as a Service, and Software as a Service along a few others.

Serverless computing has gained massive popularity in recent years, as it is both economical and easy to use and deploy. We cover serverless computing, serverless storage, and middleware required to weave on-site or end-user applications to serverless resources. We then shift our focus slightly to the topic of big data programming, and how Big Data systems are now mainly deployed in cloud environments. We cover MapReduce programming in both Apache Hadoop and Apache Spark.

The next three weeks focus on cloud storage services. We introduce cloud object storage systems, virtual hard drives (block storage), and virtual archival storage options, including a discussion of Dropbox. This course also introduces large-scale data

storage, along with the difficulties and problems of consensus in enormous stores that use large quantities of processors, memories, and disks. We also present distributed key-value stores and in-memory databases used for caching layers (e.g. Memcached and Redis) used in data centers for performance. Next we present NewSQL and NoSQL Databases in the cloud. We visit HBase, the scalable, low latency database that supports database operations in applications that use Hadoop. Then, we will show how Spark SQL can program SQL queries on huge data and present Distributed Publish/Subscribe systems using Kafka, a distributed log messaging system that is finding wide use in connecting Big Data and streaming applications together to form complex systems.

Right after the midterm exam (which is proctored by the ProctorU service), we switch to higher end applications in the cloud. We start by exploring how the Cloud opens up data analytics to huge volumes of data that are static or streamed at high velocity and represent an enormous variety of information. Cloud applications and data analytics represent a disruptive change in the ways that society is informed by, and uses information. We also introduce some common enterprise-level analytics applications that are offered by major cloud providers. We then look at graph processing, graph databases and machine learning in the cloud. We introduce the ideas of graph processing and present Pregel, Giraph, and Spark GraphX, as well as machine learning. Spark ML and Mllib continue the theme of programmability and application construction. We also cover the Machine Learning lifecycle, and how different cloud services contribute to it.

We then turn our attention to Fast Data systems, such as Apache Storm and Flux. We discuss real-time data streaming and introduce Storm technology that is used widely in the industry. We continue with Spark Streaming, Lambda and Kappa architectures, and a lesson on a complete streaming ecosystem. After that, we move on to the topics of virtualization and containers, which is a fundamental technology behind many cloud-based services. We cover virtualization and containers with a deeper focus, including lectures on Docker, Docker Compose, ECS, Kubernates and Infrastructure as Code. Finally, we wrap up the course by talking about the future trends, and wraps up with an interview with an industry-expert cloud architect.

Course Goals and Objectives

Upon successful completion of this course, you will be able to:

- Understand what cloud computing is and why it is important.
- Get a picture of the economics of cloud computing.
- Describe Big Data and the challenges of working with it.
- Learn about many fundamental technologies that enable cloud computing, such as software defined architectures, virtualization, and containers.
- Learn about many "glue" technologies that enable access to clouds, such as web middleware, JSON, REST API, RPC, etc.
- Learn about the different levels of clouds services, which include IaaS (Infrastructure as a Service), PaaS (Platform as a Service), SaaS (Software as a Service), MaaS (Metal as a Service), FaaS (Function as a Service (server-less architecture)), MBaaS (Mobile Backend as a Service (server-less architecture)), and Amazon Lambda.
- Learn about many types of cloud-based storage services, including object storage, block-level storage, archival storage, and Big Data file systems.
- Become familiar with the key concepts underlying Big Data and data streaming applications on the Cloud.
- Describe the concerns of storage, processing, parallelism, distribution, consensus, and scalability as they relate to the Cloud.
- Understand key benefits and limitations of the various technologies available in the Cloud.
- Utilize the course content to select technologies you wish to use in your work or company.

Textbook

There are no required textbooks for this course.

Course Components

Lecture videos

- Each week, your instructors will teach you the concepts you need to know through a collection of short video lectures. You may either stream these videos for playback within the browser by clicking on their titles, or you can download each video for later offline playback by clicking the download icon.

- The videos usually total 3 to 5 hours each week. The actual amount of time needed to digest the content will naturally vary according to your background.

Quizzes

- Each week will include one for-credit quiz. You will have unlimited attempts for each quiz and your highest score will be used toward your final grade.
- **Note: Each late day after the quiz deadline results in 20% grade deduction.**

Machine Problem Assignments

- This course consists of 12 Machine Problems, which are an opportunity for you to practice your programming skills and apply what you've learned in the course. Set aside 8-16 hours to work on each of the MPs. In past semesters, most MPs averaged about 6 to 8 hours per student. However, a typical student also needs to dedicate longer times to at least a few of them. You may need to budget more time if you are not familiar with the language or framework.
- We allow both Python and Java as allowable languages in most of the MPs, so that you can have your choice, **BUT some MPs are only in one language (Python)**. For the best learning experience, we suggest trying both. Even though Python has gained a lot of traction lately, Java is still the language of choice for the backend in enterprise and Big Data software platforms (although Go and Rust are making good traction lately, but we stick to Java and Python here). Gaining working knowledge with both languages will definitely improve your future job prospects.
- Amazon typically offers our students a $100 code for this course, with the details announced in the beginning of the semester. But this is not guaranteed. Please note: **A) The course staff (TAs and instructors) cannot help you in redeeming the credits, and B) you should consider the cloud fees that you will spend in the course of solving the programming assignments just like textbook fees or ProctorU exam fees.** We think $100 should be about enough to cover all AWS assignments, but we have had students in the past that had to pay significantly more. Unfortunately, there is nothing the course staff can help you with in this regard. Also, you may need to use your credit card to register an account on AWS.
- Speaking of which, please be very careful with how you experiment with Amazon services. If you are not careful, you will burn through your credit in no time! **MAKE SURE TO TURN OFF ANY RESOURCE YOU ARE NOT USING AFTER EACH WORK SESSION, SCALE YOUR EXPERIMENTS SLOWLY (don't run 20 instances all at once, build your way up), AND DEFINITELY AFTER YOU HAVE SUBMITTED THE FINAL SOLUTION AND RECEIVED THE GRADE FROM THE AUTOGRADER SYSTEM.** If you forget to do so, amazon will happily keep charging your credit card!
- Programming assignments (otherwise known as Machine Problems, or MPs) are a **MAJOR** component of the learning experience of this course. Note that there are a total of 12 MPs in this course. **Late homework assignments will have 10% grade deductions per late day submitted (for a total of 9 allowed days, on the 10th day the penalty reaches 100%)**. This is equal to 0.5 point off your final grade for every day you are late. With hundreds of students enrolled, there is **NO EXCEPTION** for this policy.
- **NOTE: To help you with the course workload, our system will automatically drop ONE lowest grade from your assignments.** This means that the top 11 programming assignments will be counted towards your final grade.
- If you need help in solving the programming assignments, please check the course forum as your first line of defense! The course TAs constantly monitor the forum, and many times your fellow students might nudge you in the right direction.

Exams

- This course will have two (2) 90-minute exams – a midterm exam and a final exam. The exams will be taken using ProctorU.
- Note that Additional ProctorU fees may apply.