CS 498: Cloud Computing Applications Syllabus

Instructor: Reza Farivar

Course Outline

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	1/18 - 1/23	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	1/24 – 1/30	Networking, IP, HTTP, REST, VPC	4	Week 2 quiz (1 hour)	2) AWS Load Balancer (~ 6 hours)	1) Java/Python Warmup
3	1/31 - 2/6	Serverless Computing	2	Week 3 quiz (1 hour)	3) AWS Lex & Lambda (~ 6 hours)	2) AWS Load Balancer
4	2/7 - 2/13	Big Data Programming: MapReduce model, Hadoop, YARN, Spark and HDFS	5	Week 4 quiz (1 hour)	4) Hadoop MapReduce (~ 8 hours)	----------
5	2/14 - 2/20	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	2/21 - 2/27	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	2/28 - 3/6	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	3/7 - 3/13	*No New Content – **Midterm** Exam preparation and Exam.*		7 practice quizzes (30 mins. each) Midterm exam is 90 minutes (ProctorU)	----------	----------
9	3/14- 3/20	Spring Break	----------	----------	----------	----------
10	3/21 - 3/27	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	3/28 - 4/3	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	4/4 - 4/10	Fast Data Processing and Streaming in the Cloud	5	Week 12 quiz (1 hour)	11) Storm and Flux (~ 6 hours)	8) SparkSQL
13	4/11 - 4/17	Virtualization, Containers and Docker	4	Week 13 quiz (1 hour)	12) Containerization & Kubernetes (~6 hours)	9) Cloud Analytics and Visualization
14	4/18 - 4/24	Container Orchestration Part 1: Docker Swarm, ECS and ACI	3	Week 14 quiz (1 hour)	----------	10) Spark - GraphFrames and MLLib
15	4/25 - 5/1	Container Orchestration Part 2: Kubernetes	2	Week 15 quiz (1 hour)	----------	11) Storm and Flux
16	5/2 - 5/4	Future Developments in the Cloud, Course Wrap-up	2:30	Week 16 quiz (1 hour)	----------	12) Containerization & Kubernetes
17	5/6 - 5/13	*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 provide you with a comprehensive understanding of cloud computing and Big Data. Each week includes 3 to 5 hours of learning material (video lectures and readings), a quiz (typically lasting 1 hour), and, for most weeks, a programming assignment. **You can expect to spend an average of 10-12 hours per week on this course**.

Please note that the times specified for the programming assignments are averages. This means some students may complete the assignment in 2 hours, while a few others may need to spend 20 hours or more on an MP. Additionally, some weeks have more training material than others. Since all course material is available from day one, we expect you to actively plan your schedule in advance. The programming assignments are a major component of this course's learning process and are designed to provide self-learning and exploration opportunities. They are all auto-graded, and you have an unlimited number of attempts before each assignment's deadline. In this course, we cover the multitude of technologies that comprise the modern cloud computing stack. Cloud computing is an information technology revolution that has significantly impacted many enterprise computing systems and will change the face of computing in the years to come.

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

Serverless computing has gained massive popularity in recent years due to its economical benefits and ease of use and deployment. We discuss serverless computing, serverless storage, and the middleware needed to connect on-site or end-user applications to serverless resources. We then slightly shift our focus to the topic of big data programming and how Big Data systems are now primarily 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.

Immediately following the midterm exam (which is monitored by the ProctorU service), we transition to advanced cloud applications. We begin by examining how the cloud enables data analytics for vast amounts of static or high-velocity streamed data, which encompasses a tremendous variety of information. Cloud applications and data analytics represent a revolutionary shift in how society accesses and utilizes information. Additionally, we introduce several common enterprise-level analytics applications provided by major cloud vendors. Next, we explore graph processing, graph databases, and machine learning in the cloud. We present the concepts behind graph processing and introduce Pregel, Giraph, and Spark GraphX, along with machine learning. Spark ML and MLlib continue to emphasize programmability and application development. We also discuss the machine learning lifecycle and the various ways different cloud services contribute to it it.

We then turn our attention to Fast Data systems, such as Apache Storm and Flux. We discuss real-time data streaming and introduce the Storm technology used widely in the industry. We continue with Spark Streaming, as well as Lambda and Kappa architectures, and we provide a lesson on a complete streaming ecosystem. After that, we move on to the topics of virtualization and containers, which are fundamental technologies behind many cloud-based services. We cover virtualization and containers in greater depth, including lectures on Docker, Docker Compose, ECS, Kubernetes, and Infrastructure as Code. Finally, we wrap up the course by discussing future trends and conclude with an interview featuring an industry expert in cloud technology 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 includes 12 Machine Problems, giving you the opportunity to practice your programming skills and apply your knowledge. Set aside 8 to 16 hours for each MP. In previous semesters, most MPs averaged about 6 to 8 hours per student. However, typical students may need to allocate extra time for some of them. If you are not familiar with the language or framework, you should budget additional time.
- The main difference between the three-credit and four-credit versions of the course is the number of programming assignments you must complete. The first programming assignment is optional for all students. In the four-credit version, you are required to complete all 11 remaining programming assignments. In the three-credit version, the system automatically drops the lowest 4 out of 11, considering only your top 7 assignments in the grading formula
- We allow both Python and Java as allowable languages in most of the MPs, giving you the freedom to choose. However, some MPs are only available in one language, specifically Python. For the best learning experience, we recommend trying both. Although Python has gained significant traction lately, Java remains the language of choice for backend development in enterprise and Big Data software platforms. While Go and Rust are also gaining popularity, we will focus on Java and Python here. Acquiring working knowledge of both languages will definitely enhance your future job prospects
- **You should take into account the cloud fees you will incur while completing the programming assignments, just as you would for textbook fees or ProctorU exam fees**
    - Make sure to turn off any resources you are not using after each work session. Scale your experiments gradually (don't run 20 instances all at once; build up your usage), especially after you have submitted the final solution and received the grade from the autograder system. If you forget to do this, Amazon will happily continue charging your credit card!
- Late homework assignments will incur a 10% grade deduction for each day they are submitted late, up to a maximum of 7 days; on the 8th day, your maximum grade for that MP would be capped at 25%. With hundreds of students enrolled, there are no exceptions to this policy policy.
- If you need assistance with the programming assignments, please check the course forum as your first line of support! The course TAs regularly monitor the forum, and often your fellow students may guide 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.