Entrance Survey

bit.ly/398entrance
What this course is about

This is **Applied** Cloud Computing

- Very little theory
- Heavy focus on hands-on practice
What is this course *really* about?
Course Staff

Led by Professor Robert J. Brunner

Instructors: Ben Congdon, Tyler Kim

Course Assistants: Osmar Coronel, Sahil Bhatt
<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MapReduce and Hadoop</strong></td>
<td><strong>Spark and its Frameworks</strong></td>
<td><strong>Distributed Databases</strong></td>
<td><strong>Containerization and Infrastructure</strong></td>
</tr>
<tr>
<td>MapReduce Programming Paradigm, Distributed File Systems, Basic Cloud Parallelism</td>
<td>Modern Data-Intensive Cloud Applications, Stream Processing, Graph Processing</td>
<td>NoSQL Databases, Key-Value Stores</td>
<td>Basic Cloud Security, Commercial Clouds, Docker, Terraform, Kubernetes</td>
</tr>
</tbody>
</table>
Prerequisites

- Proficiency in programming
- Comfortable with programming in Python
- CS225 or equivalent
- Comfortable using CLI and git

Contact the instructors if you aren’t sure
Format

- Lectures
- Online Quizzes
- MPs
- Final Projects
# Grading/Cutoffs

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPs</td>
<td>55%</td>
</tr>
<tr>
<td>Online Quizzes</td>
<td>15%</td>
</tr>
<tr>
<td>Final Project</td>
<td>30%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-</td>
<td>90%</td>
</tr>
<tr>
<td>B-</td>
<td>80%</td>
</tr>
<tr>
<td>C-</td>
<td>70%</td>
</tr>
<tr>
<td>D-</td>
<td>60%</td>
</tr>
</tbody>
</table>
MP

- MP makes up 55% of the total grade.
- **10** MPs throughout the semester.
- Every MP is weighted equally (about \(\sim 6.11\%\) per MP)
  - Most MPs will run on the course cluster.
  - MPs will distributed through course gitlab repository
  - Must be done individually
AWS S3 Buckets

Master Node

Spark  Hadoop  YARN

AWS EMR Cluster

Worker

Worker

Worker

Worker
MP

- MPs will be distributed over the course Gitlab
  - Individual student repos will be emailed this weekend

- We run anti-plagiarism scripts; don’t cheat.
  - Standard academic code applies, see syllabus for details

- Don’t post your solutions in “plain view”
  - We reserve the right to retroactively lower your grade
Online Quizzes

- Online Weekly Quizzes makes up 15% of your grade.
- Content will cover current week’s lectures
- There will be 10-12 quizzes throughout the semester.
- All the quizzes are weighted equal.
- Must be done individually
Final “Team” Project

- 30% of total grade.
- Groups of 4 students
- More details/rubric will be discussed in class
- 10-15 minutes in-class presentation + peer grading.
Fastest-Growing Experiment Class at UIUC

276% growth from the past two semesters of CS199
Available Resources

- **Lectures**
  - Lecture Slides will be posted online at course website (after the lecture)

- **Asking Questions**
  - **Piazza**: Your first stop for all questions (Search First)
  - **Office Hours**: Hosted after each lecture.
  - **Email Staff**: mailing list (don’t email individual instructor/TA)
Communication

**Gradebook:** Moodle

**Weekly Quizzes:** Moodle

**Forum:** Piazza

**Course Website:** bit.ly/cs398cloud

**MPs:** GitLab + AWS/GCP

**Email:** cs398acc@lists.illinois.edu
Weekly Schedule

- **Monday**
  - Full-Lecture + Short Office Hours
  - MP Released after the lecture

- **Tuesday**
  - Previous Week’s MP due at 11:59pm
  - Online Quiz Released

- **Wednesday**
  - Short lecture + Office Hours

- **Friday**
  - Online Quiz Due at 11:55pm
Office Hours

Hosted after each lecture.

Email us at any time, but see the Email Policy on the website.
How to get most out of CS398?

- Do the MPs on your own!
- Read the Relevant Reading Materials posted
- Do the quizzes on your own
- Go to the office hours
- Ask questions?
bit.ly/cs398cloud
Motivations:

- Massive amounts of data
- Limitations of local compute resources
  - 1GB Dataset can be processed by your laptop, but what about 1 petabyte?
- Scalability
  - Every month your data scales by 2x. What now?
- Availability
  - Your customers / clients want 99.99% uptime
Clouds

Motivations:

● Distribution
  ○ Availability in multiple geographic regions
What is Cloud Computing?

Using someone else’s computers.
What is Cloud Computing?

- Popularized by the introduction of Amazon’s EC2 Service in 2006
- Connections to Distributed Systems and High-Performance Computing
What is Cloud Computing?

Two Important Perspectives:

- **Access**
  - Gives developers the ability to use pooled computer resources
What is Cloud Computing?

Two Important Perspectives:

- **Services**
  - Cloud utilities / frameworks / paradigms provide platforms for writing more complex applications
  - *Examples*: MapReduce, Spark, Hadoop, Distributed Databases
What can we use clouds for?

- Data storage
- Data processing
- Arbitrary computation or data intensive tasks
- Machine learning
- Serving web traffic
- Hosting software/services