Geography 970 Distributed computing for GIScience

Instructor: Prof. Qunying Huang
Office: 426 Science Hall
Tel: 608-890-4946
E-mail: qhuang46@wisc.edu
Office Hours: T, R, and F 1:00-2:00 PM or by appointment

TA: TBD

Lectures

9:30AM - 10:45AM, Tuesday/Thursday, 360 Science Hall

Labs

Tuesday 11:00 AM – 1:00 PM, Wensday 6:00 – 8:00 PM, 380 Science Hall

Course Overview:

The scientific and engineering advancements in the 21st century pose computing intensive challenges in managing Big Data, using complex algorithms to extract information and knowledge from Big Data, and simulating physical and social phenomena. Distributed computing is required to address these computing challenges. This course covers general introductory concepts about distributed systems, covering all the major computing paradigms such as Cloud Computing, Grid Computing, Cluster Computing, Graphic Processing Unit (GPU) Computing, and Citizen Computing. The strategies and considerations how to choose different computing paradigms for various applications will also be covered. The main content of this course will include:

- Distributed programming models, libraries and tools (Hadoop, MPICH, Cuda, etc).
- Cloud platform/cloud services, such as Amazon Elastic Cloud Computing (EC2), Microsoft Azure, and Google App Engine.
- Development, deployment and migration GIScience applications onto Cloud platforms
- Open-source cloud solutions, e.g., Eucalyptus, CloudStack, and OpenStack etc
- Geospatial processing and application intra- and inter-clouds
- Big data management, backup and synchronization on clouds
- Scalability, discovery of services and data in Cloud computing infrastructures

Course Goals:

Upon the completion of the course, students are able to:

- Build a cluster (MPICH2, or Hadoop framework) for different applications;
- Use real Cloud Systems such as Amazon EC2/S3, Eucalyptus;
- Leverage distributed computing to support various applications;

Course Requirements:

It is required that students have taken at least one introductory GIS course (Geog170, Geog 370 or Geog377), and programming course (Geog 378, Geog 575, or programming course provided by the computer science).

Recommended Textbook (Optional)

- Yang C., Huang Q., Li Z., Xu C., Liu K., 2013, Spatial Cloud Computing: A Practical Approach, CRC Press/Taylor & Francis, 304p. ISBN: 978-1466593169.
- Distributed and Cloud Computing: Clusters, Grids, Clouds, and the Future Internet (DCC) by Kai Hwang, Jack Dongarra & Geoffrey C. Fox

Additional reading materials will be distributed through the Learn@UW.

Evaluation:

Your grade in this course is based on two exams, ten labs, several ad-hoc quizzes, and final project. The points assigned to each component are as follows:

Items	Percentage	Date
Exam#1	15%	03/11
Quizzes (6, dropping the lowest)	5%	ad-hoc
Lab (10)	35%	throughout
Exam#2	15%	04/29
Final project	30%	5/17 (noon)

Quizzes: Quizzes will be administered during any class period - immediately after a lecture, at the beginning or end of a class, etc. **Make-up quizzes will not be given at any circumstance, no exceptions.** However, the lowest quiz scores will be dropped.

Lab: You will complete ten lab assignments throughout the semester. Most labs will consist of exercises and small projects using open sources to practice and reinforce your understanding about database, and programming concepts. Plagiarismis not tolerated. As with other evaluated items, any offense results in a zero for the lab assignment and disclosure of the impropriety to the Department and University. Assignment must be submitted to ***Learn@UW dropbox*** prior to the start of class on the day it is due. **Late labs will be marked down 10% a day;** submission of an assignment the day it is due, but after the deadline (e.g., following your lab that day), counts as one day late**.** Technical complications (e.g., disk errors, printing problems) are not reason for extension; be sure to back up copies of all of your work and version meticulously, as forgetting to save and back up your interactive map is the easiest way to lose your work. Requests for grade changes must be submitted in writing (via email) within ***24 hours*** of receiving your feedback.

Exams: Exams include a combination of multiple choices, True/False, and short answer questions, with an emphasis on the latter. A review is provided one or two days prior to the exam. The exams are not cumulative. While group studying is encouraged, cheating during the exam is not tolerated and results in a zero for the exam and disclosure of the impropriety to the Department and University. Exam must be taken at the scheduled time and date. **Make-up exams will not be given unless prior arrangements have been made with the instructor.** Make-up exams require a doctor's note or, in the event of planned travel, must be rescheduled ***4 weeks*** in advance. Make-up exams are in an essay format, rather than primarily short answer.

Term project: A project that utilizes programming technologies to solve problems is required. **A two page (double line) project proposal and a final project report (~ 3000 words) are required by the due day.** Each project will be carried out in a group of **three** or **four** students; graduate students are allowed to work alone only if the project is a component of their thesis research. Students will be required to present your project to the class at the end of the semester. Guidelines of the term project presentation and report will be released during the semester.

Grading criteria:

90-100% A
87-89.9% AB
83-86.99% B
80-82.99% BC
75-79.99% C
70-74.99% DC
60-69.99% D
< 60% F

Additional Course Information:

- Course information and copies of lecture material are disseminated in class or via the course website on Learn@UW. It is your responsibility to obtain this information and check the course website regularly. I intend to keep changes to the syllabus to a minimum, but do keep in mind that changes to the syllabus may occur. If changes are made, I will inform you in class and via the course website.
- A note on scholastic dishonesty: Academic honesty and integrity is expected at all times. All work, including assignments, quizzes and exams, must be completed individually by each student. It is expected that work submitted by a student reflects his or her original ideas and responses. Submissions that reflect substantially similar work by more than one student will be dealt with as an act of scholarly dishonesty and credit will be deducted from each assignment in question. Scholarly dishonesty includes: “cheating on an examination; collaborating with others in work to be presented, contrary to the stated rules of the course; submitting a paper or assignment as one’s own work when a part or all of the paper or assignment is the work of another; submitting a paper or assignment that contains ideas or research of others without appropriately identifying the sources of those ideas...” Please refer to the “Student Academic Misconduct Policy & Procedures” document produced by Student Advocacy & Judicial Affairs division of the Offices of the Dean of Students for further information.
- A note on class etiquette: Cell phones must be silenced when you are in lecture. In addition, please refrain from reading the paper, participating in social conversations and other disruptive behavior (including frequenting Facebook or other social networks on your laptop) during lecture.

Tentative Schedule:

<table>
<tr><th>Week</th><th>Module</th><th>Topic</th><th>Labs</th></tr>
<tr><td>1</td><td rowspan="9">HPC/HTC computing and programming models</td><td>CyberGIS, distributed/parallel computing</td><td></td></tr>
<tr><td>2</td><td>Cluster computing and programming model (MPICH2, and OpenMPI)</td><td>HPC for geospatial applications</td></tr>
<tr><td>3</td><td>Grid computing, middleware and public grids</td><td></td></tr>
<tr><td>4</td><td>Cloud computing I : Architecture, service types, service models, providers, enabling technologies</td><td rowspan="2">Deployment of geospatial applications onto the Amazon EC2</td></tr>
<tr><td>5</td><td>Cloud computing II: Public clouds (Amazon EC2 or Microsoft Azure)</td></tr>
<tr><td>6</td><td>Cloud computing III: Open-source clouds (Eucalyptus, Cloudstack etc.)</td><td></td></tr>
<tr><td>7</td><td>Cloud computing IV: Hapdoop</td><td>MapReduce tutorial</td></tr>
<tr><td>8</td><td>Others: Citizen, GPU, MIC computing</td><td></td></tr>
<tr><td>9</td><td>Spatial middleware and tools</td><td></td></tr>
<tr><td>10</td><td>CyberGIS frontiers</td><td>Contemporary and future research, challenges, CyberGIS in action</td><td></td></tr>
</table>