Introduction to Big Data and High-Performance Computing

Lecture – 9:30 a.m., Tuesday
Labs – 10:00 a.m. – noon Thursday & 2-4 p.m. Friday

Provide the information below to one of us to sign up for 3 credits of one of these:

- **CSC 592** - Joan Peckham, Computer Science & Statistics, joan@cs.uri.edu
- **OCG593** – Yang Shen, Graduate School of Oceanography, yshen@gso.uri.edu
- **BCH523** – Ying Zhang, Cell and Molecular Biology, yingzhang@mail.uri.edu
- **CHE591 or CHE491**, Angelo Lucia, Chemical Engineering, lucia@egr.uri.edu

**Prerequisites: Permission of Instructor**
Because of the limited lab space and available HPC resources, the class capacity is set to 25 students. In order to best serve students with Big Data and HPC interests and needs, students must provide a written statement of interest demonstrating one or more of the following:

1) Adequate programming experience using MatLab/ R/C++ or some other scientific computer language (e.g., FORTRAN).
2) Have a credible basic research or application need that requires the use of HPC or generates Big Data. This need can be in any discipline (science, engineering, computer science, mathematics, economics, etc.).
3) Prior hands-on research experience with either HPC or Big Data. This experience can include, but is not limited to, algorithmic development, middleware or interface programming, applications, data mining, and/or visualization.

**Course Description:** In this interdisciplinary and project-based class, students will work on scientific, engineering, and social problems that involve a large amount of data and/or high performance computing (HPC). Scholars from various disciplines such as Oceanography, Cell and Molecular Biology, Computer Science and Statistics, Social Sciences, and Engineering will mentor students on projects. Students will be provided with materials and information that will help them to design and implement computation on HPC facilities to manage, process, and visualize large or complex data sets.
Introduction to Big Data and High-Performance Computing
CSC 592 / OCG593 / BCH523 / CHE591 / CHE491
Fall 2014; 3 credits

Instructors:
Joan Peckham. Office: Tyler 254; Phone: 874-4174; Email: joan@cs.uri.edu; Office Hours: By Appointment
Yang Shen. Office: 203 Horn Building, Bay Campus; Phone: 874-6848; Email: yshen@gso.uri.edu; Office Hours: TBA
Angelo Lucia. Office: Kirk 224; Phone: 874-2814; Email: lucia@egr.uri.edu; Office Hours: TBA
Ying Zhang. Office: CBLS 487; Phone: 874-4915; Email: yingzhang@mail.uri.edu; Office Hours: TBA

Class Meeting Times: Lecture - 9:30 am, Tuesday
Labs - 10:00 am – noon, Thursday, and 2:00-4:00 pm, Friday

Classroom: Lectures: TBA
Labs – Graphics Lab, Tyler 053

Prerequisites: Permission of instructor
Because of the limited lab space and available HPC resources, the class capacity is set to 25 students. In order to best serve students with Big Data and HPC interests and needs, students must provide a written statement of interest demonstrating one or more of the following:
1) Adequate programming experience using MatLab/ R/C++ or some other scientific computer language (e.g., FORTRAN).
2) Have a credible basic research or application need that requires the use of HPC or generates Big Data. This need can be in any discipline (science, engineering, computer science, mathematics, economics, etc.).
3) Prior hands-on research experience with either HPC or Big Data. This experience can include, but is not limited to, algorithmic development, middleware or interface programming, applications, data mining, and/or visualization.

Class Capacity: 25

Course Description: In this interdisciplinary and project-based class, students will work on scientific, engineering, and social problems that involve a large amount of data and/or high performance computing (HPC). Scholars from various disciplines such as Oceanography, Cell and Molecular Biology, Computer Science and Statistics, Social Sciences, and Engineering will mentor students on projects. Students will be provided with materials and information that will help them to design and implement computation on HPC facilities to manage, process, and visualize large or complex data sets.

Course Goals and Learning Outcomes:
This course will introduce basic computing and programming concepts, showcase Big Data and HPC applications in science, social sciences, and engineering, and provide training in managing, processing, and visualizing large and/or complex data sets, or modeling with HPC. Students will be able to understand basic Big Data and HPC computing concepts, gain hands-on experience using a high-performance computer, and develop programming, teamwork, and communication skills.

**Textbook and Recommended Readings:**

**Assignments:**
Labs – Weekly lab exercises working in small groups to explore various aspects of Big Data and HPC.
Final projects – Big Data and HPC Projects chosen by students

**Grading Policy:**
- Labs: 30%
- Final project, oral presentation and written report: 70%

**Topics and Schedule:**
- Topics are lectures/seminars presented by lecturers and invited speakers to provide application examples and inspiration.
- Assignments provide basic training in HPC and Big Data related skills and opportunities for students to work with Big Data and HPC.

<table>
<thead>
<tr>
<th>Week</th>
<th>Topics</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Sept. 9</strong>&lt;br&gt;<strong>Introduction</strong>&lt;br&gt; a) What is high-performance computing?&lt;br&gt; b) What is Big Data?&lt;br&gt; c) How are a) and b) connected?&lt;br&gt; d) Algorithms – definition and importance for big data&lt;br&gt; e) Overview of applications</td>
<td><strong>Sept. 11 &amp; 12</strong>&lt;br&gt;<strong>Lab 1: Basic Linux - 1</strong>&lt;br&gt;(experienced linux users may opt out the lab)&lt;br&gt;TA</td>
</tr>
<tr>
<td>Week</td>
<td>Topics</td>
<td>Assignments</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>2</td>
<td><strong>Sept. 16</strong>&lt;br&gt;Data Mining&lt;br&gt;Dr. Gavino Puggioni</td>
<td><strong>Sept. 18 &amp; 19</strong>&lt;br&gt;Lab 2: Basic Linux - 2&lt;br&gt;(experienced linux users may opt out the lab)&lt;br&gt;TA</td>
</tr>
<tr>
<td>3</td>
<td><strong>Sept. 23</strong>&lt;br&gt;Data Visualization&lt;br&gt;Dr. Jean-yves Herve</td>
<td><strong>Sept. 25 &amp; 26</strong>&lt;br&gt;Lab 3: Matlab 1: The Basics&lt;br&gt;(experienced matlab users may opt out the lab)&lt;br&gt;TA</td>
</tr>
<tr>
<td>4</td>
<td><strong>Sept. 30</strong>&lt;br&gt;The Need for Big Data Analysis: A Case Study from the Research of Dengue Virus Infection&lt;br&gt;Dr. Carey Medin,</td>
<td><strong>Oct. 2 &amp; 3</strong>&lt;br&gt;Lab 4: Matlab 2: The parallel toolbox&lt;br&gt;(experienced matlab users may opt out the lab)&lt;br&gt;TA</td>
</tr>
<tr>
<td>5</td>
<td><strong>Oct. 7</strong>&lt;br&gt;Big Data: The Science of Patterns&lt;br&gt;Dr. Lutz Hamel</td>
<td><strong>Oct. 9 &amp; 10</strong>&lt;br&gt;Lab 5: Advanced visualization&lt;br&gt;(Paraview)&lt;br&gt;TA</td>
</tr>
<tr>
<td>6</td>
<td><strong>Oct. 14</strong>&lt;br&gt;Engineering Applications&lt;br&gt;a) Visualization of Enhanced Oil Recovery Processes</td>
<td><strong>Oct. 16 &amp; 17</strong>&lt;br&gt;Lab 6: Advanced visualization&lt;br&gt;(Paraview)&lt;br&gt;TA</td>
</tr>
<tr>
<td>7</td>
<td><strong>Oct. 21</strong>&lt;br&gt;Engineering Applications&lt;br&gt;b) Carbon Sequestration &amp; HPC</td>
<td><strong>Oct. 23 &amp; 24</strong>&lt;br&gt;Lab 7: Basic HPC&lt;br&gt;Ying Zhang</td>
</tr>
<tr>
<td>8</td>
<td><strong>Oct. 28</strong>&lt;br&gt;Applications in Earth and Environmental Sciences</td>
<td><strong>Oct. 30 &amp; 31</strong>&lt;br&gt;Lab 8: Toy modeling problem&lt;br&gt;Angelo Lucia</td>
</tr>
<tr>
<td>9</td>
<td><strong>Nov. 4</strong>&lt;br&gt;Applications in Earth and Environmental Sciences&lt;br&gt;Stephan Grilli, tsunami modeling</td>
<td><strong>Nov. 6 &amp; 7</strong>&lt;br&gt;Lab 9: Toy Big Data problem&lt;br&gt;Yang Shen</td>
</tr>
<tr>
<td>Week</td>
<td>Topics</td>
<td>Assignments</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 11 Veterans Day (classes do not meet)</td>
<td>Nov. 13 &amp; 14 Student presentations of final project abstracts/plans</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 18 Applications in Social Sciences</td>
<td>Nov. 20 &amp; 21 Lab 11: Final project workshop</td>
</tr>
<tr>
<td>12</td>
<td>Nov. 25 Big Data and HPC facilities  Mark Howison, Brown CCV</td>
<td>Nov. 27 &amp; 28 Thanksgiving Recess, classes do not meet</td>
</tr>
<tr>
<td>13</td>
<td>Dec. 2 Meeting with lecturers on issues related to final projects and presentations</td>
<td>Dec. 4 &amp; 5 Lab 13: Final project workshop</td>
</tr>
<tr>
<td>14</td>
<td>Dec. 9, reading period Meeting with lecturers by appointments</td>
<td>Dec. 11 &amp; 12, final exam period Project presentations</td>
</tr>
</tbody>
</table>

**Course Policies:**

*Group Work:*
Throughout the semester various problems will be worked out in class in groups. In the lab each week, students will work in small groups on the assigned lab exercise. These groups will be assigned by the instructor and will be different each week to give students an opportunity to work with as many classmates as possible.

*Individual Work:*
While the group work described above is meant to provide opportunities for students to collaborate on projects, there will also be times when individual work is required. Some assignments will be specified as individual work, and students will be expected to complete this work on their own. Questions regarding individual work should go to the instructor or the TA for the course.

*Assigned Readings and Videos:*
Students are expected to read any assigned readings and view any assigned videos before the class in which it will be discussed.

*Attendance:*
Students are expected to attend every class. Experienced linux and matlab users may opt out the introductory labs by written request. The grade for the labs is prorated according to the labs students do. Missed assignments may be made up only with prior permission.

**Final Projects:**
Students will be expected to spend significant amount of time working outside of formal lectures and labs in order to complete the projects for the course.

**University Course Policies:**

**Accommodations for Special Needs**
Any student with a documented disability is welcome to contact the instructors as early in the semester as possible so that we may arrange reasonable accommodations. As part of this process, please be in touch with Disability Services for Students Office at 330 Memorial Union, 401-874-2098.

**Academic Honesty**
All submitted work must be your own. If you consult other sources (class readings, articles or books from the library, articles available through internet databases, or websites) these MUST be properly documented, or you will be charged with plagiarism and will receive an F for the paper. In some cases, this may result in a failure of the course as well. In addition, the charge of academic dishonesty will go on your record in the Office of Student Life. If you have any doubt about what constitutes plagiarism, visit the following website:

    http://gervaseprograms.georgetown.edu/hc/plagiarism.html

the URI Student Handbook, and UNIVERSITY MANUAL sections on Plagiarism and Cheating at


Any good writer’s handbook as well as reputable online resources will offer help on matters of plagiarism and instruct you on how to acknowledge source material. If you need more help understanding when to cite something or how to indicate your references, PLEASE ASK.

**Academic Enhancement Center**
The work in this course is complex and intensive. To do the best you can, it’s a good idea to visit the Academic Enhancement Center (AEC) in Roosevelt Hall. The AEC offers a comfortable environment in which to study alone or together, with or without a tutor. AEC tutors can answer questions, clarify concepts, check understanding, and help you to study. You can make an appointment or walk during office hours -- Monday through Thursday from 9 am. to 9 pm, Friday from 9 am to 1 pm, and Sunday from 4 pm. to 8 pm. For a complete schedule For a complete schedule - including when tutors are available specifically for this class - go to www.uri.edu/aec, call (401) 874-2367, or stop by the fourth floor in Roosevelt Hall.
The Writing Center
The Writing Center is for “all writers, all disciplines, at all levels, and all stages of writing.” If an instructor suggests that you go to the Writing Center, it is not a punishment, and does mean that you are a terrible writer. It means the instructor wants you to receive more individualized attention to your writing than s/he is able to provide, given the constraints of the class. It will only improve your grade. If possible, call ahead for an appointment (874-4690). Drop-in tutorials are often available. You may make repeat appointments, requesting the same tutor each time if you wish. See their Web Page:

http://www.uri.edu/artsci/writing/center/index.shtml

for tips on how to make the best of your Writing Center visit.

Standards of Behavior
Students are responsible for being familiar with and adhering to the published “Community Standards of Behavior: University Policies and Regulations” which can be accessed in the University Student Handbook. If you must come in late, please do not disrupt the class. Please turn off all cell phones, pagers, or any electronic devices.

Religious Holidays
It is the policy of the University of Rhode Island to accord students, on an individual basis, the opportunity to observe their traditional religious holidays. Students desiring to observe a holiday of special importance must provide written notification to each instructor.

General Note: This syllabus is an outline of proposed events. It is subject to change; however, never without advanced notification.