INSTRUCTOR: Dr. Thomas S. Woodson  
E-MAIL: thomas.woodson@stonybrook.edu  
OFFICE PHONE: (631) 632-9974  
OFFICE HOURS: Tuesdays 2:30pm-5:30pm or by appointment  

COURSE DESCRIPTION:  
The infrastructure requirements and challenges of big data systems to support large-scale technology management applications will be discussed. Advanced topics in big data infrastructure such as data center operations, network and system security, data management and integration will be covered. Cloud computing platforms such as Iaas, Saas, and Pass will also be included. Other topics including advanced data mining and visualization techniques as they relate to customer relationship management (CRM), supply chain management (SCM), and global operations management applications will also be discussed.  

Prerequisites: Must be a Master’s student in Technology and Society or receive permission from the instructor.  

TEXTBOOK:  
Data and Goliath: The Hidden Battles to Collect Your Data and Control your World  
Bruce Schneier  
Big Data Analytics with R  
Simon Walkowiak  

HELPFUL BOOKS:  
• An Introduction to Applied Multivariate Analysis with R by Brian Everitt and Torsten Hothorn  

ASSIGNMENTS:  
Class participation and attendance: 10%  
Projects: 60%  
Book quizzes: 30%
GRADES
A: 94-100
A-: 90-93
B+: 87-89
B: 83-86
B-: 80-82
C+: 77-79
C: 73-76
C-: 70-72
D: 60-69
F: Less than 60

OTHER INFORMATION

Late assignments:
Unless otherwise noted, assignments are due BEFORE class on the day that they are due. If the assignment is turned in late, you automatically lose 10 points on the assignment and you continue to lose another 10 points each day. After 1 week, you will receive a 0 on that assignment.

Attendance/Late Policy:
Attendance to this class is mandatory. Having more than 1 unexcused absence will impact your grade. If you miss a class, you are still expected to do all the readings and assignments for that week. Be on time to class. If you are often late to class, you will lose class participation points.

Electronics Policy:
Silence/turn off your cell phones during the class. If you have an emergency where you need to keep your cell phone on, tell the professor before class. Please NO TEXTING during class. If you use a computer to take notes, please do not surf the web. It is distracting to the other students and the professor.

Academic Integrity:
Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Any suspected instance of academic dishonesty will be reported to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at www.stonybrook.edu/academicintegrity.

Disability Support Services:
If you have a physical, psychological, medical or learning disability that may impact your course work, please contact Disability Support Services, 128 ECC Building (631) 632-6748. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Students who require assistance during emergency evacuation are encouraged to discuss their needs with their professors and Disability Support Services. For procedures and information go
to the following web site: [http://www.ehs.sunysb.edu](http://www.ehs.sunysb.edu) and search Fire Safety and Evacuation and Disabilities.

**Critical Incident Management:**
Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Judicial Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures.

**Computers:** Please bring your laptop to class in order to do the class examples during class. If you do not have a laptop, you may share with another classmate.

**Software:** We will use the software program R for this course. R is an open source data analysis software that is growing in popularity. You can download it from [www.r-project.org](http://www.r-project.org).

**Extra Statistics resources:**
- DASL (The Data and Story Library): [http://lib.stat.cmu.edu/DASL](http://lib.stat.cmu.edu/DASL)
- Statlib-Datasets Archive [http://lib.stat.cmu.edu/datasets/](http://lib.stat.cmu.edu/datasets/)
- University of California, Los Angeles Case Studies [http://www.stat.ucla.edu/cases/](http://www.stat.ucla.edu/cases/)
- U.S. Census Bureau [http://www.census.gov](http://www.census.gov)
- Stats in the news, from George Mason University: [http://www.stats.org/](http://www.stats.org/)

**Online statistics textbooks and software:**
- Computing for Data Analysis [https://www.coursera.org/course/compdata](https://www.coursera.org/course/compdata)
- Data Analysis [https://www.coursera.org/course/dataanalysis](https://www.coursera.org/course/dataanalysis)
- SISA simple interactive statistical analysis
  http://www.quantitativeskills.com/sisa/

**Online resources for R:**
- The main R project site:
  www.r-project.org
- An R online textbook
  Kickstarting R: http://cran.r-project.org/doc/contrib/Lemon-kickstart/
- Website for the Sarkar book:
  http://lmdvr.rforge.r-project.org/figures/figures.html
- Quick R website (many helpful “how to” pages)
  http://www.statmethods.net/
- A U. of Wisconsin Stats professor’s site (B. Yandell)
  http://www.stat.wisc.edu/~yandell/software/
- Book on Regression and Anova on the R site:
  http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf

### Course calendar subject to change

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture 1</td>
<td>23-Jan</td>
<td>Intro/Ch.1</td>
<td></td>
</tr>
<tr>
<td>Lecture 2</td>
<td>30-Jan</td>
<td>Ch2/GGPlot</td>
<td></td>
</tr>
<tr>
<td>Lecture 3</td>
<td>6-Feb</td>
<td>Ch. 3</td>
<td>Book Quiz 1</td>
</tr>
<tr>
<td>Lecture 4</td>
<td>13-Feb</td>
<td>Clustering</td>
<td>GG Plot Project Due</td>
</tr>
<tr>
<td>Lecture 5</td>
<td>20-Feb</td>
<td>Regression</td>
<td></td>
</tr>
<tr>
<td>Lecture 6</td>
<td>27-Feb</td>
<td>Logistic Regression</td>
<td></td>
</tr>
<tr>
<td>Lecture 7</td>
<td>6-Mar</td>
<td>Review/Project Time</td>
<td>Project 2 Due</td>
</tr>
<tr>
<td></td>
<td>13-Mar</td>
<td>No class, Spring Break</td>
<td></td>
</tr>
<tr>
<td>Lecture 8</td>
<td>20-Mar</td>
<td>Text Mining</td>
<td></td>
</tr>
<tr>
<td>Lecture 9</td>
<td>27-Mar</td>
<td>Text Mining</td>
<td>Book Quiz 2</td>
</tr>
<tr>
<td>Lecture 10</td>
<td>3-Apr</td>
<td>Big Data Lecture with R</td>
<td>Text Mining Project Due</td>
</tr>
<tr>
<td>Lecture 11</td>
<td>10-Apr</td>
<td>Map Reduce/Hadoop</td>
<td></td>
</tr>
<tr>
<td>Lecture 12</td>
<td>17-Apr</td>
<td>R and relational</td>
<td></td>
</tr>
<tr>
<td>Lecture 13</td>
<td>24-Apr</td>
<td>Big Data Lecture</td>
<td>Book Quiz 3</td>
</tr>
<tr>
<td>Lecture 14</td>
<td>1-May</td>
<td>Review/Project Time</td>
<td>Big Data II Project Due</td>
</tr>
</tbody>
</table>