

## CS 698: Special Topics in Big Data

### General Information

Instructor: Chase Wu  
Office/Lab: GITC 4107  
E-mail: [chase.wu@njit.edu](mailto:chase.wu@njit.edu)  
Phone: 973-642-4579

Department office: GITC 4400  
Department phone: 973-596-5778

Course website:

[https://web.njit.edu/~chasewu/Courses/Spring2016/CS698BigData/CS698\\_BigData\\_Spring16.html](https://web.njit.edu/~chasewu/Courses/Spring2016/CS698BigData/CS698_BigData_Spring16.html)

### Course Description

This course provides an in-depth coverage of special topics in big data from data generation, storage, management, transfer, to analytics, with focuses on the state-of-the-art technologies, tools, architectures, and systems that constitute big-data computing solutions in high-performance networks. Real-life big-data applications in various domains (particularly in sciences) are introduced as use cases to illustrate the development, deployment, and testing of a wide spectrum of emerging big-data solutions.

### Required Background

Programming Skills

- C/C++ or Java in Linux

Prerequisite Courses

- CS 610: Data Structures and Algorithms
- CS 630: Operating System Design
- Or permission of instructor

### Textbook

- Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. By David Loshin, Elsevier, August 23, 2013.

### Resources

Additional reading materials including reference books and online resources will be assigned for some advanced topics as the course proceeds.

### Evaluation

Grading components:

Attendance	10%
Homework	10%
Project	20%
Midterm	30%
Final	30%

Grading scale\*:

Grade	CS 698
A	90 – 100
B	80 – 89
C	70 – 79
D	60 – 69
F	59 and below

\*Final grades will not be curved unless necessary.

### Late Policy

Students are expected to complete work on schedule. Late work is not accepted unless prior arrangements are made with the instructor.

### **Academic Integrity and Student Conduct:**

***Plagiarism or cheating*** behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature (including the internet) on their assignments, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Dean of Students for further disciplinary action. For further information on the Code of Student Conduct and academic discipline procedures, please refer to: <http://www.njit.edu/doss/policies/conductcode>.

### **Course Syllabus**

Week	Topic
1	<ul style="list-style-type: none"><li>• Introduction</li></ul>
2	<ul style="list-style-type: none"><li>• In-class Presentation on 4 V's of Big Data Applications</li></ul>
3	<ul style="list-style-type: none"><li>• Trends of Computing for Big Data<ul style="list-style-type: none"><li>○ High-performance Computing (Supercomputers and Clusters)</li><li>○ Grid Computing</li><li>○ Cloud Computing</li><li>○ Mobile Computing</li></ul></li></ul>
4, 5	<ul style="list-style-type: none"><li>• Big Data Overview<ul style="list-style-type: none"><li>○ Drivers of Big Data</li><li>○ Big Data Attributes</li><li>○ Data Structures</li><li>○ Big Data Ecosystem</li><li>○ Examples of Data Analytics</li></ul></li></ul>
6, 7	<ul style="list-style-type: none"><li>• Big Data Tools, Techniques, and Systems<ul style="list-style-type: none"><li>○ Exascale Computing</li><li>○ HDFS, HBase, and NoSQL (Document Store, Graph DB, etc.)</li><li>○ MapReduce, Spark, Oozie, Tez, Hive, Pig, etc.</li><li>○ Hadoop 1 and Hadoop 2 (YARN)</li></ul></li></ul>
8, 9	<ul style="list-style-type: none"><li>• Advanced Analytical Theory and Methods<ul style="list-style-type: none"><li>○ Hadoop/Mahout</li><li>○ Recommendation</li><li>○ Clustering</li><li>○ Classification</li><li>○ Regression</li></ul></li></ul>
10, 11	<ul style="list-style-type: none"><li>• Review and Midterm Exam</li></ul>
12, 13, 14	<ul style="list-style-type: none"><li>• Advanced Topics<ul style="list-style-type: none"><li>○ Big Data Visualization</li><li>○ High-performance Networking for Big Data Movement</li><li>○ Big Data Scientific Workflow Management and Optimization</li></ul></li></ul>
15	<ul style="list-style-type: none"><li>• Review</li></ul>