COMP 6730 Advanced Database Systems
Big Data Analytics

Fall 2016

Meeting Time and Place: Wednesday 5:30PM-8:00PM, in Olsen Hall 402
Instructor: Tingjian Ge
Instructor Email: ge@cs.uml.edu
Course Web Page: http://www.cs.uml.edu/~ge/#teaching/673
Instructor Office Hours: Wed, Thu, 2:30 - 4:00 PM, Olsen Hall 204

Textbook:

Mining of Massive Datasets (Second Edition)
By Jure Leskovec, Anand Rajaraman, Jeff Ullman.
Cambridge University Press. http://www.mmds.org/.

Optional:
Probability and Computing:
Randomized Algorithms and Probabilistic Analysis
By Michael Mitzenmacher and Eli Upfal.
Cambridge University Press. ISBN: 9780521835404.

1. Course Description

We are surrounded by data. Managing big data, as appeared in social networks, smart devices, sensors, web, and cloud computing environments, has drawn much attention lately in the data management research. Big data analytics – data combined with algorithmic analysis – in a broad sense captures much of the essence of computer science, and is transforming the world. In this course, we study a few recently established and trending topics in this area, including data streams, link analysis, finding similar items, web advertising, recommendation systems, social-network graphs, and large-scale machine learning. The course will also teach students techniques in applying probability in computing. There will be homework assignments and a semester-long course project.

2. Learning Outcomes

- A successful student, after learning the course, should understand in depth a number of data analysis and mining problems on massive datasets in the real world.
- The student should be able to grasp some analytical skills on such datasets, including algorithm design and analysis.
- The student should be able to write programs for a parallel and distributed big data cloud environment.
- The course should get the student prepared for research work in this area.

3. Prerequisites

COMP 5730 Database I or an equivalent course. Undergraduate algorithm and probability courses are helpful.

4. Participation in Classes

The best learning experience is arguably through active participation in classes. To encourage proper participation, we factor it as a part of the final grade. Attendance at lectures is mandatory and you are expected to show up prepared to answer questions and participate in discussion. Excellent participation in class discussions will be recorded and reflected in your final grade. Likewise, poor participation or absence from classes will have a negative impact. As shown below, 15% of your final grade is determined by class participation.

5. Grading

Grades are assigned based on class participation, written homework assignments, and a semester-long project. The grading breakdown is as follows:

- Class Participation: 15%
- Written Homework Assignments: 40%
- Project Assignment: 45%

Note that missing classes or being late will hurt your participation grade. If there is a special occasion that you have to miss a class, you need to let me know. Your final grade will be approximately distributed as follows (subject to change), where the numbers are weighted averages of all grade components. A+ or A: 100 – 87, A-: 86 – 84, B+: 83 – 81, B: 80 – 77, B-: 76 – 74, C+: 73 – 72, C: 71 – 60, E: below 60.

6. Academic Conduct

Students are expected to complete all assignments independently. Honest and ethical behaviors are always expected. There will be no tolerance for plagiarism or other academic misconduct. The minimum punishment is an E grade that cannot be removed by the repeat option.

7. Schedule

The following schedule is tentative and is subject to change.

Week	Date	Topic
1	Sep 7	Introduction, Background
2	Sep 14	MapReduce
3	Sep 21	Background
4	Sep 28	Finding Similar Items
5	Oct 5	Background
6	Oct 12	Data Streams
7	Oct 19	Background
8	Oct 26	Link Analysis
9	Nov 2	Background
10	Nov 9	Advertising on the Web
11	Nov 16	Background
12	Nov 23	Recommendation Systems
13	Nov 30	Background
14	Dec 7	Social-Network Graphs