Fall 2019

CS 644-1J1: Big Data

David Bader

CS644: Big Data

Jump to Today ✎ Edit

Required Background

- Programming Skills
  - Java or C/C++ in Linux
- Prerequisite Courses
  - CS 610: Data Structures and Algorithms
  - Or permission of instructor

Textbook

(https://www.amazon.com/Big-Data-Analytics-Enterprise-Integration-ebook/dp/B00EVSOWVA)

- Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph (https://www.amazon.com/Big-Data-Analytics-Enterprise-Integration-ebook/dp/B00EVSOWVA). 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.

Week	Topic
1	• Introduction
2	• In-class Presentation on 4 V's of Big Data Applications
3	• Trends of Computing for Big Data ◦ High-performance Computing (Supercomputers and Clusters) ◦ Grid Computing ◦ Cloud Computing ◦ Mobile Computing
4, 5	• Big Data Overview ◦ Drivers of Big Data ◦ Big Data Attributes ◦ Data Structures ◦ Big Data Ecosystem ◦ Examples of Data Analytics
6, 7	• Big Data Tools, Techniques, and Systems ◦ Exascale Computing

	◦ HDFS, HBase, and NoSQL (Document Store, Graph DB, etc.) ◦ MapReduce, Spark, Oozie, Tez, Hive, Pig, etc. ◦ Hadoop 1 and Hadoop 2 (YARN)
8, 9	• Review and Midterm Exam
10, 11	• Advanced Analytical Theory and Methods ◦ Hadoop/Mahout ◦ Recommendation ◦ Clustering ◦ Classification ◦ Regression
12, 13, 14	• Advanced Topics ◦ Big Data Volume and Information Visualization ◦ High-performance Networking for Big Data Movement ◦ Big Data Scientific Workflow Management and Optimization ◦ Massive-Scale Graph Analytics
15	• Review

Evaluation

Grading components:

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

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:

“Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at:

http://www.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf (http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf.)

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu" (mailto:dos@njit.edu%E2%80%9D)

Reading Materials:

- How To Kill A Supercomputer: Dirty Power, Cosmic Rays, and Bad Solder? (http://spectrum.ieee.org/computing/hardware/how-to-kill-a-supercomputer-dirty-power-cosmic-rays-and-bad-solder) By Al Geist, Feb. 2016
- Global Big Data Market from the Ricerca Alfa (https://www.ricercaalfa.com/), a leading market research firm estimated that the overall Big Data market hold a market size of approximately US$ 65 billion for the year 2018 globally.
- Hadoop installation
- Oozie installation instructions: slides

Course Summary:

Date	Details	
Mon Sep 9, 2019	Lecture (https://njit.instructure.com/calendar?event_id=571&include_contexts=course_8497)	6pm to 9pm
Mon Sep 16, 2019	Lecture (https://njit.instructure.com/calendar?event_id=572&include_contexts=course_8497)	6pm to 9pm
Mon Sep 23, 2019	Lecture (https://njit.instructure.com/calendar?event_id=573&include_contexts=course_8497)	6pm to 9pm
Mon Sep 30, 2019	Lecture (https://njit.instructure.com/calendar?event_id=574&include_contexts=course_8497)	6pm to 9pm
Mon Oct 7, 2019	Lecture (https://njit.instructure.com/calendar?event_id=575&include_contexts=course_8497)	6pm to 9pm
Mon Oct 14, 2019	Lecture (https://njit.instructure.com/calendar?event_id=576&include_contexts=course_8497)	6pm to 9pm

Date	Details	
Mon Oct 21, 2019	Lecture (https://njit.instructure.com/calendar?event_id=577&include_contexts=course_8497)	6pm to 9pm
Mon Oct 28, 2019	Lecture (https://njit.instructure.com/calendar?event_id=578&include_contexts=course_8497)	6pm to 9pm
Mon Nov 4, 2019	Lecture (https://njit.instructure.com/calendar?event_id=579&include_contexts=course_8497)	6pm to 9pm
Mon Nov 11, 2019	Lecture (https://njit.instructure.com/calendar?event_id=580&include_contexts=course_8497)	6pm to 9pm
Mon Nov 18, 2019	Lecture (https://njit.instructure.com/calendar?event_id=581&include_contexts=course_8497)	6pm to 9pm
Mon Nov 25, 2019	Lecture (https://njit.instructure.com/calendar?event_id=582&include_contexts=course_8497)	6pm to 9pm
Mon Dec 2, 2019	Lecture (https://njit.instructure.com/calendar?event_id=583&include_contexts=course_8497)	6pm to 9pm
Mon Dec 9, 2019	Lecture (https://njit.instructure.com/calendar?event_id=584&include_contexts=course_8497)	6pm to 9pm