New Jersey Institute of Technology

Digital Commons @ NJIT

Summer 2020

CS 644-J31: Introduction to Big Data

David Bader

Recommended Citation

CS644: Big Data

[✎ **Edit**]

Course Description

Data science ↗ **(https://datascience.njit.edu/)** is a rapidly emerging discipline at the intersection of machine learning, algorithms, and statistics. Solving real-world problems in domains such as **financial services** ↗ **(https://www.wired.com/brandlab/2016/11/financial-services-2020-bank-branches-wither-ai-comes/)**, **cybersecurity** ↗ **(https://www.technologyreview.com/topic/computing/cybersecurity/)**, and **health informatics** ↗ **(https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care)** requires the generation, storage, management, and analytics, on massive data sets. This course will take you on a journey of developing high performance data analytics for massive data sets using accelerated computer architectures, state-of-the-art software frameworks, and real-world case studies. In-depth topics will focus on mathematical modeling, scalable algorithms, and high-performance computing, for fast and practical solutions. Together we will explore real-world data modeling and prediction from the **Data Science Design Manual** ↗ **(http://www.data-manual.com/)** and learn **How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results** ↗ **(https://www.wiley.com/en-us/Big+Data+in+Practice:+How+45+Successful+Companies+Used+Big+Data+Analytics+to+Deliver+Extraordinary+Results-p-9781119231387)** .

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) The Data Science Design Manual** ↗ **(http://www.data-manual.com/)** , Steven S. Skiena, ISBN 978-3-319-55444-0, Springer, 2017.
- ↗ **(http://www.data-manual.com/) Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results** ↗ **(https://www.amazon.com/Big-Data-Practice-Successful-Extraordinary/dp/1119231388)** , Bernard Marr, ISBN-13: 978-1119231387, Wiley, 2016.

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)**

Additional Reading Materials:

- TBD

Course Summary:

Date	Details	
Mon May 18, 2020	📅 **CS 644: Big Data lecture 1** (https://njit.instructure.com/calendar?event_id=5853&include_contexts=course_11885)	5:45pm to 9:30pm
Mon May 25, 2020	📅 **CS 644: Big Data lecture 2** (https://njit.instructure.com/calendar?event_id=5854&include_contexts=course_11885)	5:45pm to 9:30pm
Mon Jun 1, 2020	📅 **CS 644: Big Data lecture 3** (https://njit.instructure.com/calendar?event_id=5855&include_contexts=course_11885)	5:45pm to 9:30pm
Mon Jun 8, 2020	📅 **CS 644: Big Data lecture 4** (https://njit.instructure.com/calendar?event_id=5856&include_contexts=course_11885)	5:45pm to 9:30pm
Mon Jun 15, 2020	📅 **CS 644: Big Data lecture 5** (https://njit.instructure.com/calendar?event_id=5857&include_contexts=course_11885)	5:45pm to 9:30pm
Mon Jun 22, 2020	📅 **CS 644: Big Data lecture 6** (https://njit.instructure.com/calendar?event_id=5858&include_contexts=course_11885)	5:45pm to 9:30pm
Mon Jun 29, 2020	📅 **CS 644: Big Data lecture 7** (https://njit.instructure.com/calendar?event_id=5859&include_contexts=course_11885)	5:45pm to 9:30pm

Date	Details	
Mon Jul 6, 2020	📅 **CS 644: Big Data lecture 8** **(https://njit.instructure.com/calendar?event_id=5860&include_contexts=course_11885)**	5:45pm to 9:30pm
Mon Jul 13, 2020	📅 **CS 644: Big Data lecture 9** **(https://njit.instructure.com/calendar?event_id=5861&include_contexts=course_11885)**	5:45pm to 9:30pm
Mon Jul 20, 2020	📅 **CS 644: Big Data lecture 10** **(https://njit.instructure.com/calendar?event_id=5862&include_contexts=course_11885)**	5:45pm to 9:30pm
Mon Jul 27, 2020	📅 **CS 644: Big Data lecture 11** **(https://njit.instructure.com/calendar?event_id=5863&include_contexts=course_11885)**	5:45pm to 9:30pm
Mon Aug 3, 2020	📅 **CS 644: Big Data lecture 12** **(https://njit.instructure.com/calendar?event_id=5864&include_contexts=course_11885)**	5:45pm to 9:30pm