

Fall 2020

## CS 644-103: Introduction to Big Data

Yao Shen

Follow this and additional works at: <https://digitalcommons.njit.edu/cs-syllabi>

---

### Recommended Citation

Shen, Yao, "CS 644-103: Introduction to Big Data" (2020). *Computer Science Syllabi*. 130.  
<https://digitalcommons.njit.edu/cs-syllabi/130>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Computer Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact [digitalcommons@njit.edu](mailto:digitalcommons@njit.edu).

# Course Syllabus

## CS644 – Fall 2020 – Introduction to Big Data

**Class Schedule:** Thursday 6-8:50pm, KUPF210

**Instructor:** Yao Shen, email: [yshen@njit.edu](mailto:yshen@njit.edu)

**TA:** Haoran Liu, email: [hl425@njit.edu](mailto:hl425@njit.edu)

**WebEx link:**

Meeting link:

<https://njit.webex.com/njit/j.php?MTID=m21adb8610735ae91272c61d4bb78362e>

Meeting number:

120 724 5124

Password:

2hKJ6ws2SkZ

Host key:

220539

**Course description and format:**

This course provides an in-depth coverage of various topics in big data from data generation, storage, management, transfer, to analytics, with focus 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 and workflows in various domains (particularly in the sciences) are introduced as use cases to illustrate the development, deployment, and execution of a wide spectrum of emerging big-data solutions.

**Prerequisite:**

Some basic knowledge of python programming, SQL programming, Unix, algorithms, and machine learning. If in doubt about the prerequisites, please consult with the instructor for permission to take the class.

**Attendance:**

You are supposed to attend all the classes. Participation is highly encouraged to make the class more interactive. Class attendance and participation are taken into consideration by the instructor for the evaluation of the students. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class be sure to consult one of your classmates about the content of the lecture and visit the course web page and Canvas ([canvas.njit.edu](http://canvas.njit.edu)) to get notes, exercises, assignments, deadlines and announcements.

**Textbooks:**

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

Hadoop: The Definitive Guide. By Tom White

Spark in Action. By Petar Zecevic etc.

## Resources:

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

## Grading:

The requirements of this course will consist of participating in lectures, homework, project, midterm exam, and final exam. The grading breakdown is the following:

- Attendance (5%)
- Homework (15%)
- Projects (20%)
- Midterm (30%)
- Final (30%)

## Collaboration and Honor Code:

Students may discuss problems together but must write up their own solutions. When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note in particular that copying homework or programming assignments, in full or in part is forbidden. Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an “F” grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified.

## Tentative course schedule (Subject to changes according to progress):

Date	Topic
Week1	Introduction
Week2	Big Data Ecosystem, Hadoop HDFS
Week3	Yarn and MapReduce
Week4	MapReduce
Week5	Hadoop Streaming and MRjob
Week6	Computing Trends
Week7	Midterm
Week8	NoSQL Database
Week9	Hbase
Week10	Pig
Week11	Spark
Week12	Machine Learning Basics
Week13	SparkMlib, Sparksql and Big Data Visualization
Week14	Final Exam