

# CSE 482: Big Data Analysis (Spring 2022)

---

## General Information:

- Lecture Hours: Mon, Wed 10:20am - 11:40am
- Classroom: STEM 1202
- Zoom Link: <https://msu.zoom.us/j/96074129548> (passcode: 486076)
- Instructor and office hour:
  - [Pang-Ning Tan \(email\)](#) Mon, Wed 12:00-1:00pm
- Teaching assistants and office hours:
  - Ethan Fulcher ([email](#)) Thur 1:00-2:00pm, Fri 9:30am-10:30am
  - Li Liu ([email](#)) Tue 3:00-4:00pm, Fri 2:00-3:00 pm

## Course Description:

This course is intended for senior undergraduate students who are interested to gain hands-on experience applying computational techniques to solve big data problems. The course will cover a variety of topics from data collection and preprocessing to data analysis and large-scale database querying. Students are expected to already have some prior background in Python.

There is no main textbook for the class. Lecture notes will be posted on D2L. You may use the following reference books for some of the topics covered in this class.

- Introduction to Data Mining (Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar), 2nd Edition, Addison Wesley, 2018. Sample chapters are available [here](#)
- Mining of Massive Datasets (Jure Leskovec, Anand Rajaraman, Jeff Ullman), 3rd Edition, Cambridge University Press, 2020. Draft chapters from the book are available [here](#)
- Hadoop: The Definitive Guide (Tom White), O'Reilly Media, Fourth Edition, 2015.
- Python for Data Analysis: Data Wrangling with Pandas, Numpy, and iPython (Wes McKinney), 2nd Edition, O'Reilly, 2017.

## Course Prerequisites:

CSE 331 and (STT 351 or STT 380 or STT 430 or STT 441) and MTH 314 and (MTH 234 or MTH 254H or LB 220).

## Course Requirements:

- Internet connection (DSL, LAN, or cable connection desirable)
- Access to [Desire2Learn](#) (D2L) and [Piazza](#).
- Ability to scan and upload documents.
- Laptop or desktop computer.

## Course Delivery:

This course will be delivered in hybrid mode. The first 3 weeks of lectures are scheduled to be held online (per university policy). The instructor plans to have some in-person lectures during the semester (though the online zoom option will always be made available using the zoom link given in the general information section above). Exam is currently scheduled to be held in person (subject to change). More information will be given as the semester progresses per guidance from the university. You will need your MSU NetID to login to the course content on the D2L homepage (<http://d2l.msu.edu>). In D2L, you will access the course materials, exercises, assignments, and additional resources. [Piazza](#) will be used as the discussion forum for students to ask questions. Modules from [Amazon Web Service \(AWS\) Academy](#) will also be used to supplement the materials and exercises for this class.

## D2L Technical Assistance

If you need technical assistance at any time during the course or to report a problem you can:

- Visit the Distance Learning Services Support Site.
- Visit the Desire2Learn [Help Site](#).
- Or call Distance Learning Services: (800) 500-1554 or (517) 355-2345.

## Class Recordings

The lectures will not be recorded. However, the power-point slides for the lecture have pre-recorded audio, which can be replayed.

## Course Outline:

The topics to be covered in this course include:

- Data and preprocessing (2 weeks)
- Overview of data analysis techniques (~6 weeks)
- Big data programming techniques (~6 weeks)
- Applications (1 week)

## Course Assessment:

Students will be assessed based on the following:

- Homework (55%)
- Weekly exercises (25%)
- Exams (20%)

The final grade will be determined based on the following scale:

### GPA Grade

4.0 [90,100]

## GPA Grade

- 3.5 [85,90)
- 3.0 [80,85)
- 2.5 [75,80)
- 2.0 [70,75)
- 1.5 [65,70)
- 1.0 [60,65)

## Important Dates:

Note that the first lecture will begin on Jan 10 via zoom. There is no lecture on Jan 17 for MLK day. The last day to add a class is Friday, Jan 14 while the last date to drop a class with tuition refund is Friday, Feb 4. More details about the academic calendar can be found at <https://reg.msu.edu/ROInfo/Calendar/Academic.aspx>.

## Course Policies:

- **Attendance and Participation:** Students are expected to attend all the lectures (virtually or in person, if the latter option becomes available) and must inform the instructor if they are unable to do so due to emergency or other legitimate reasons. Students are expected to participate in all activities for the course as listed on the course syllabus and calendar. The instructor will monitor student participation in the class using D2L and zoom tracking tools to assess their level of participation.
- **Weekly Exercises:** Weekly exercises are due on **Fridays by 11:59pm** unless stated otherwise. These are short programming exercises to help you familiarize with the concepts taught in the lectures and/or the tools/programming environments used for the class. In some cases, additional supplementary materials will be provided to help you complete the exercises. Only the top-10 highest scores from the exercises will be used to compute your final grade. You must use [D2L](#) to submit your solution. **No late submission will be accepted** for weekly exercises.
- **Homework Assignment:** Homework assignments are due before midnight (11:59pm) on the due date unless specified otherwise. You must use [D2L](#) to submit a soft copy of your solution. Late homework assignments will be penalized 25% of the total assignment grade (if submitted the next day after the deadline) or 50% (if submitted the second day after the deadline). Assignments submitted later than that will not be accepted.
- **Exam Information:** There will be a midterm and a final exam. Exams will be open book and notes. Exams are currently scheduled to be held in person, unless there is a change in the university policy. Make-up examinations can be arranged if your absence is due to documented illness or personal emergency. A written explanation (including supporting documentation) must be submitted to the instructor; if the explanation is acceptable, an alternative to the examination will be arranged. Except for emergency cases, make-up arrangements must be made in advance, at least a week before the scheduled exam.
- **Discussion Forum:** We will be using [Piazza](#) for class discussion. Rather than

emailing class-related questions to the instructor and TA, you are strongly encouraged to post them on Piazza. However, do not post your solution or attempted solution (e.g., for programming exercises, assignments or project) on Piazza.

- **Academic Honesty:** You are forbidden from sharing the solution of your programming exercises and homework assignments with anyone else, except with the instructor and TAs. All submitted assignments must be your own work. You are also prohibited from posting questions and/or answers related to the homework assignments on websites such as Chegg, Reddit, and other similar websites. You are allowed to use only the CSE482 Piazza website to post any questions you have about the exercises and assignments. Students who were found copying or sharing their solutions with others (including posting the information online without approval from the instructor) will automatically receive a zero for the entire assignment, and possibly, a fail grade for the class depending on severity of the offense. In addition, an Academic Dishonesty Report will be filed and submitted to the university.
- **Resources for Persons with Disabilities (RCPD):** Michigan State University is committed to providing equal opportunity for participation in all programs, services and activities. Requests for accommodations by persons with disabilities may be made by contacting the Resource Center for Persons with Disabilities at 517-884-RCPD or on the web at [rcpd.msu.edu](http://rcpd.msu.edu). Once your eligibility for an accommodation has been determined, you will be issued a verified individual services accommodation ("VISA") form. Please present this form to me at the start of the term and/or two weeks prior to the accommodation date (exam, homework, project, etc). To make an appointment with a specialist, contact: (517) 353-9642 or TTY: (517) 355-1293. The website for RCPD is located at <http://MYProfile.rcpd.msu.edu>.
- **Integrity:** All students are expected to be responsible users of the computer system provided by the Department of Computer Science and Engineering and College of Engineering. The Department of Computer Science adheres to the policies on academic honesty as specified in General Student Regulations 1.0, Protection of Scholarship and Grades, and in the all-University Policy on Integrity of Scholarship and Grades, which are included in Spartan Life; Student Handbook and Resource Guide.
- **Religious Observance:** Michigan State University has long had a policy to permit students, faculty/academic staff, and support staff to observe those holidays set aside by their chosen religious faith. If you wish to be absent from class to observe a religious holiday, make arrangements in advance with the instructor.
- **Tolerance and Civility** MSU strives to build an academic community with living and learning environments that expects tolerance of viewpoints and civility toward others, whether at public forums, athletic events, in residential communities, classrooms or laboratories. We call upon all who participate in university events to promote tolerance and civil behavior and to hold themselves to high standards that reflect the university's commitment to respect viewpoints that may be different from their own. Only by respecting individuals with diverse perspectives and ideas can we build an environment of civility that is conducive to advancing knowledge and transforming lives.