CSCI 620 Introduction to Big Data

Section 02 Fall 2241

Course Syllabus

Instructor

Michael Mior
mmior@mail.rit.edu

Prerequisites

CSCI-603 or CSCI-602 or CSCI-605 or CSCI-604 or CSCI-243 or equivalent course. Students may not take CSCI-620 for credit if they have already taken CSCI-320 and CSCI-420.

Course Description

This course provides a broad introduction to the exploration and management of large datasets being generated and used in the modern world. The course is divided into two components: data management and data exploration.

The first component consists of analyzing relational and document-oriented databases to store, retrieve, and manage data in the real world: relational and document-oriented models, SQL and aggregation pipeline, normalization and document embedding, data integrity, transactions, and access control. Additional types of NoSQL databases may be covered. It also analyzes the steps to combine several datasets into a single consolidated dataset: modeling issues, data integration, and data cleaning. Finally, the course will also cover some modern data processing systems.

The second component introduces practical techniques used in exploratory data analysis and mining. Topics include descriptive statistics and data visualization over consolidated data, clustering, and frequent itemset mining. Additional descriptive or predictive techniques may be covered, such as decision trees or neural networks. Case studies will examine issues in data capture, organization, storage, retrieval, visualization, and analysis in diverse settings such as urban crime, drug research, census data, social networking, and space exploration. Big data exploration and management projects, a term paper and a presentation are required.

Office Hours

Office hours will be Monday and Wednesday from 11am-12pm in my office, 70-3517. Any deviations from this schedule will be announced on myCourses.

Course Outcomes

After taking this course, a student will be able to demonstrate the following outcomes:
- Describe the basic concepts, policies, and mechanisms for big data exploration and data mining
- Design, implement, and explain a practical solution for making sense of big data
- Explain how large amounts of data are stored, retrieved and analyzed in modern data management systems
- Explain basic concepts, policies, and mechanisms used in handling large amounts of data in modern real-world settings
- Describe legal, privacy, and ethical issues in handling big data
- Explain basic concepts underlying contemporary data management topics

Wellbeing

Success in this course depends heavily on your personal health and wellbeing. Recognize that stress is an expected part of the college experience, and it often can be compounded by unexpected setbacks or life changes outside the classroom. Your other instructors and I strongly encourage you to reframe challenges as an unavoidable pathway to success. Reflect on your role in taking care of yourself throughout the term, before the demands of exams and projects reach their peak. Please feel free to reach out to me about any difficulty you may be having that may impact your performance in this course as soon as it occurs and before it becomes unmanageable. In addition to your academic advisor, I strongly encourage you to contact the many other support services on campus that stand ready to assist you.

Textbooks and other resources

- Database System Concepts, 7th Ed., by Silberschatz, et al., McGraw-Hill, 2019. (optional)
- Additional material will be provided as handouts or online/offline references
- All class material (slides, lectures, discussions, handouts, student presentations, and web/book references) may be used for exam questions. Some topics are not from the textbook, and not all chapters in the textbook will be covered.
- Feel free to use the tremendous amount of web resources on database systems, but cite any resources used in all papers, presentations, projects, exams, or other assignments. This includes your own previous work in other courses. Failing to do so will be considered plagiarism.

Assessment and Grading

Component	Weight	Comments
Assignments (9)	45%	**Individual work:** There will usually be one assignment every ten days. You need to start working on your assignment from day one; to avoid problems completing it on time. Each student must provide an individual solution. To be submitted to myCourses. Late submissions will not be allowed. Check myCourses for specific dates. Questions about the current assignment will not be allowed when there are less than 24 hours before the submission deadline.
Quizzes (10)	10%	**Individual work:** There will be a 20 minute quiz for each unit.
Final exam	20%	**Individual work:** There will be a 2.5 hour final exam.
Team project	15%	**Group work**: The teams will be usually composed of four students. It consists of three phases. The grade for team members does not necessarily need to be the same for all the team members. A project submission will consist of a report and code implementation.
Feedback	5%	**Individual work:** You will need to provide feedback regarding your teammates and team presentations.
Final presentation	5%	**Teamwork:** Each team will present their project work (see Project Description for more details).

Note: For group work, each student's individual score will be computed from the initial group score by reducing or increasing it based on personal contributions using my observations, peer feedback, and other mechanisms. Little contribution, little credit!

Grading Policy

Letter grade conversions from total scores will done as follows:

90–100 is A	73-75.99 is C+
86-89.99 is A-	70-72.99 is C
83-85.99 is B+	67-69.99 is C-
80-82.99 is B	60-66.99 is D
76-79.99 is B-	less than 60 is F

Your course grade will not be more than two +/- letter grades higher than the weighted average of your individual work. See the Group Work section below to understand this grading policy better.

W and I Grade Policy

RIT policy allows you to withdraw from a course with a grade of W on or before the date shown in the current RIT semester calendar. After this date, I cannot give you a W, but I must assign you a grade based on your work completed in the course.

Incomplete grades, therefore, will be given only in truly exceptional circumstances, and only by prior arrangement with me, as I have the final say in this matter.

myCourses Policy **(check it frequently!)**

The myCourses page for the course will be used to share information and must be checked frequently, as many announcements may not be made in class or by email. This may include corrections to course content or assigned work. Students are responsible for being aware of any updates.

Email Policy

Sending messages to me:
- Please include the course code [CSCI-620] in the subject of your message.
- You must use an RIT email account for all email, which must include the course number in the subject. (If you don't have an RIT email account, get one immediately.)
- Adhere to this policy as I receive many emails daily: using an RIT email account and with an appropriate subject title will prevent your message from getting removed as spam.

Responses from me:
- Replies will be sent to your official RIT email account.
- While I will respond to messages as quickly as possible, delays are possible based on workload, schedule, and the number of messages—which increases as deadlines approach!
- Responses outside of regular working hours, especially weekends, are not guaranteed.
- If you don't get a reply to a message you sent from an RIT account, please raise the question in class. Some messages may be responded to via announcements in class or on myCourses.
- Messages about program errors or problems must include all relevant information.

Assignments and email:
- Assignments must be submitted on myCourses and follow the provided guidelines.
  Note: email submissions are never accepted.

Submissions Policy

All submissions will be on-line and should be submitted by the due date in myCourses. Most submissions will have a due date of Friday at 5pm, but submissions will be accepted until Sunday at midnight without penalty. Any submissions after this date will not be accepted. Some submissions may specify an alternative time which will take priority. Again, submissions are never accepted by email, and such email will be ignored.

Late Policy

No late submissions! Start early! An assignment due in the last teaching week or finals week of the semester will not be accepted late.

Make-ups Policy

In general, there will be no make-ups for graded events such as quizzes, tests, exams, or assignments, and a zero score will be assigned for the missed event unless the I accept the reason for missing the graded event as valid (e.g., a doctor's note, a religious reason, or similar), a grade might be assigned based on future performance on a similar graded events. Make-ups *may* be offered only if my schedule permits it.

Alternative Final

If you have a final exam conflict, follow the Registrar's rules to schedule an alternative exam.

Extra Credit

The policy is simple: there will be no extra credit assignment; please do not ask for any.

Re-scoring policy

For score adjustments, you must request regrading within one week of when the graded material was returned. As re-grading is done individually, your score may go up or down when re-graded.

Assignment Submission Format

Each assignment should include a written report in PDF format. For assignments where code is to be submitted, the code should be in text format with any necessary instructions for execution. Please do not include compiled programs or large datasets in your submission unless explicitly requested. Further guidelines will be available on myCourses. Marks may be deducted for violating these guidelines.

Rehabilitation Act (Section 504), and Americans with Disabilities Act (ADA) Statement

RIT is committed to providing reasonable accommodations to students with disabilities. If you would like to request accommodations such as special seating or testing modifications due to a disability, please contact the [Disability Services Office](). It is located in the Student Alumni Union, Room 1150. After you receive accommodation approval, it is imperative that you see me during office hours so that we can work out whatever arrangement is necessary." If approved for appropriate accommodations, please set up a meeting with me to discuss the needed accommodations as soon as possible. Note that if you don't discuss your plans for the use of their test related accommodations at least two weeks in advance of the test or assignment, it may not be possible to make arrangements for the needed accommodations.

In-classroom Electronic Devices Policy

Unless there is a verified documented reason such as an NOA letter from the DSO or prior permission has been obtained, the use of any device (e.g., tablets, laptops, smartphones, or MP3 players) for media (audio/image/video) recording in the classroom is prohibited. Laptop use will be permitted for note-taking purposes and participation in any in-classroom activities.

Group Work Policy

The intent of this policy is to ensure you understand that group work is beneficial, and a student's score on group work may differ from others in the same group.

Each group member is expected to understand the group work in its totality, e.g., all parts of a software project design and implementation, even if parts were assigned to other group members. While some grading for group work is collective, the final score is individual based on the collective work, exam questions, peer evaluation, and my observation of individual contributions to group work. If needed, I may schedule individual meetings to discuss each member's contribution to the group work.

While group work is typically beneficial, I recognize that groups sometimes become dysfunctional. If you are in such a group, inform me privately ASAP so that corrective action can be taken. If you wait too long, it then becomes your own problem!

A student's individual score on group work may differ from others in the same group.

General Conduct and Academic Integrity

The intent of this policy is to ensure you understand that any violation, however "small," may result in a course grade of F.

As an institution of higher learning, RIT expects students to behave honestly and ethically at all times. Familiarize yourselves with the following policies:
- [RIT Honor Code](#)
- [Academic Integrity Process](#)

In particular, note these definitions:
- **Cheating**: "any form of fraudulent or deceptive academic act, including falsifying of data, possessing, providing, or using unapproved materials, sources, or tools for a work submitted for faculty evaluation."
- **Duplicate submission**: "submitting of the same or similar work for credit in more than one course without prior approval of the instructors for those same courses."
- **Plagiarism**: "representation of others' ideas as one's own without giving proper credit to the original author or authors. Plagiarism occurs when a student copies direct phrases from a text (e.g., books, journals, and internet) and does not provide quotation marks, or paraphrases or summarizes those ideas without giving credit to the author or authors. In all cases, if such information is not properly and accurately documented with appropriate credit given, then the student is guilty of plagiarism." (In this course, this includes the submission of AI-generated content from systems such as ChatGPT.)

A grade of F will be assigned for any submission of work which is deemed to violate the academic integrity policies (see above). Depending on the nature of the violation, this may also result in a course grade of F.

Some examples of what is acceptable and what is not:
- Homework: you may freely discuss homework with one another as homework is meant to increase your understanding of the course material. You must, however, write the solutions individually, and list out all persons with whom you collaborated on the homework.
- All exams, tests, quizzes, and individual assignments, projects, and term papers, must be your unaided work as mentioned earlier. Specifically, you are not allowed to use any electronic device in any in-class quiz, test or exam; simple calculators may be used only when permitted.
- Abetting academic dishonesty, i.e., doing someone else's work for them, or making your solutions available for their use, will also be considered as academic dishonesty.
- If you are not sure or are confused about how a given activity will be viewed, you must contact me for clarification *before* you engage in that activity.

Plagiarism Checking Policy

All written assignments such as term papers, project reports, or online exams must be unaided work written in your own words and without the use of generative AI such as ChatGPT. Submissions may be checked using plagiarism checking software or services (such as turnitin.com), and attempts to circumvent such checking will be considered as flagrant violation of ethical standards, with penalties imposed as stated above.

Instructional Continuity

The intent of this policy is to ensure that you understand that any interruptions to normal functioning may require an alternative course delivery mechanism.

- Interruptions to instructional services may result from events such as illness (faculty or student), severe weather, loss of power, and other catastrophic events that may affect all or parts of the campus for varying durations. I am required to plan for instructional continuity if and when instructional services are interrupted due to such events.
- If the event is due to a student issue, the student should arrange to inform me or the institute as soon as possible so that appropriate alternatives can be developed.
- If the event is due to me, the institute or other issues, I, the department or the institute will contact you at your institute email address or other known contact information.
- If such an event occurs, classes, deadlines and grading schemes may change as needed to include alternative delivery methods, alternative student-instructor interaction, class materials, and/or teammates, a revised attendance policy, and a revised semester calendar and/or grading scheme.

Reasonableness Rule

A student needs to learn the material and demonstrate this learning to receive the deserved course grade. Please inform me when a genuine foreseen or unforeseen factor is likely to impact or has impacted either of these. Foreseen factors (such as religious events, work-related travel, or a wedding in the immediate family) must be notified before the event, and unforeseen factors (such as illness or other personal/work emergencies) must be notified as soon as possible during or after the event. Delaying notification will be regarded as unreasonable.
Reasonable requests made at reasonable times in reasonable ways will be handled reasonably!

Syllabus Changes

While every effort has been made to ensure accuracy in this syllabus document, I reserve the right to make reasonable changes to any aspect of the course when conditions require it to ensure that the course learning outcomes are achieved. Advance notice of any such change will be made through myCourses, or in class, or via email.

Policy C 6.0: Policy Prohibiting Discrimination and Harassment

RIT is committed to providing a safe learning environment, free of harassment and discrimination as articulated in our university policies located on our governance website. RIT's policies require faculty to share information about incidents of gender based discrimination and harassment with RIT's Title IX coordinator or deputy coordinators, regardless whether the incidents are stated to them in person or shared by students as part of their coursework.

If you have a concern related to gender-based discrimination and/or harassment and prefer to have a confidential discussion, assistance is available from one of RIT's confidential resources on campus (listed below).

- [The Women, Gender, and Sexuality Resource Center](): Campus Center Room 2010
- [CARES]() (available 24 hours/7 days a week) Call or text 585-295-3533
- [RIT Student Health Center]() – August Health Center/1st floor; 585-475-2255
- [Counseling and Psychological Services]() - August Health Center /2nd floor - 2100; 585-475-2261
- [The Ombuds Office]() – SAU Room 1114; 585-475-7200 or 585-475-7357 or 585-475-6424
- [Spirituality and Religious Life]() – Schmitt Interfaith Center/Rm1400; 585-475-5932
- [NTID Counseling & Academic Advising Services]() – 2nd Floor Lynden B. Johnson