COMP 5118F / COMP 4900H

Trends in Big Data Management

Fall 2026

We, the people of the Faculty of Science at Carleton University, acknowledge that our campus is located on the traditional, unceded territories of the Algonquin Anishinabeg people. Miigwetch for your hospitality and stewardship of this territory and the teachings that come from it. We are grateful for this land, the air that we breathe, and the water that sustains us all as well as for the animals, plants and other living beings: these enable us to research, teach, mentor, support, study, and learn. We recognize our responsibility to our natural environment and to reconciliation with Indigenous peoples.

Course Instructor: Ahmed El-Roby **Email:** ahmed.elroby@carleton.ca **Student Hours:** Via email **Best ways to contact:** in class or via email **Course website:** people.scs.carleton.ca/~ahmedelroby/COMP5118_F26/	**Office Location:** HP 5433 **Class Location:** Check Central **Class Time:** Monday, 11:35 AM to 2:25 PM **Prerequisites:** COMP 3005, for students registering in COMP 4900H **Preclusions:** none **Department/Unit:** School of Computer Science

Graduate courses only:

- Brightspace access for University of Ottawa students; please see information here: Faculty of Graduate and Postdoctoral Affairs, access to Brightspace.
- University of Ottawa students who need access to SCS IT resources, such as OpenStack and Nextcloud, must submit a request to SCS Tech Support at SCS.Tech.Support@cunet.carleton.ca. The request must be sent from their @cmail.carleton.ca email address and the email should say which resource is required and for which course, including section.

Important dates and deadlines can be found here: Dates, Deadlines, and Regulations, Registrar's Office, including class suspension for fall and winter breaks, and statutory holidays.

Course Calendar Description

In-depth study of recent research articles in the field of data management, with focus on data integration, Internet of Things, large scale data management, recommendation systems, text processing, and question answering. Students will work on a term-long project.

Prerequisite(s): COMP 3005 for students registering in COMP 4900H.

Learning Material(s) and Other Course/Lab-Related Resources

No textbooks are needed in this course. All readings are research papers that are publicly available or accessible through the Carleton University Library.

Learning Material	**Options for Purchasing**	**Approximate Cost**
Scientific papers	Publicly available	$0

Students are not required to purchase textbooks or other learning materials for this course.

Topics Covered and Learning Outcomes

This is a seminar course. Every student reads and reviews the assigned papers before class, a subset of students presents them, and the whole class discusses them. Participation is therefore essential, and the course is structured so that every voice in the room is heard. Students come to this course from different backgrounds, and no prior exposure to any particular subfield of data management is assumed. If any aspect of the course, including the readings, the discussion format, or the project, presents a barrier to your participation, please contact me early in the term so we can find an arrangement that works.

Topics to be covered:

- Graph data management
- Natural language to SQL
- Query processing and optimization
- Vector search
- LLM applications
- Unstructured and document data
- Entity resolution
- Time series
- Geospatial and scientific data
- Healthcare analytics
- Industrial and cloud data systems
- Sports analytics

The full reading list, with a link to each paper, is maintained on the course website. Papers are assigned to specific dates as the term progresses and as presentation slots are filled.

Date	Topic/content	Reading/Prep for Class
Sep 14	Course introduction, logistics, how to read and review a paper	None
Sep 21	Graph data management	The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing (2017)
Sep 28	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Oct 5	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Oct 12	No class, Thanksgiving	
Oct 19	No class, instructor traveling	
Oct 23	**Project proposal due, 11:59 PM**	
Oct 26	No class, Fall break	
Nov 2	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Nov 9	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Nov 16	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Nov 23	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Nov 30	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Dec 7	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Dec 11	To be assigned from the reading list	Two papers, posted on the course website at least one week in advance
Dec 18	**Project paper due, 11:59 PM**	

Note that December 11 is a Friday on which Monday classes are held. No work is assigned during the Fall break.

Course level learning outcomes. By the end of this course, students will be able to:

1. Critically analyze recent research papers from top data management conferences and journals, such as SIGMOD, VLDB, CIDR, and ICDE, evaluating the novelty, methodology, and significance of each contribution.
2. Identify emerging trends and key challenges in big data systems, including cloud and industrial data platforms, vector search, query optimization, LLM-integrated data management, and processing of unstructured and document data.
3. Compare and contrast different system designs and techniques proposed in the literature, articulating their trade-offs and applicability in various contexts.

4. Present and defend technical content from recent papers in a clear and structured manner, both orally and in writing.
5. Review research papers, providing constructive feedback through structured peer-review exercises.
6. Design and conduct a research-oriented course project that contributes to an active research program, and report the results as a paper in ACM conference proceedings format together with a public code repository.
7. Apply rigorous scientific reasoning to assess the validity of experimental results and claims presented in contemporary data management literature, including the design of honest evaluation protocols.

Assessments

Grade Breakdown

Component	Grade Value	Date
Project	45 %	Proposal due October 23; project paper due December 18
Presentations	20 %	Weekly, September 21 to December 11
Paper Reviews	20 %	Weekly, due at 11:00 AM on the day of class
Class Participation	15 %	Weekly

Project (45 %). The course project is tied to the lab's research program in real-time soccer analytics, described in the document *From Pixels to Decisions: A Research Program in Real-Time, Large-Scale Soccer Analytics*, which is linked from the course website. Each project must be one of the Stage 1 sub-projects listed in that document. Stage 1 sub-projects have no dependencies on any other stage, run mostly on public data, and can start in the first week of term. Each sub-project section states the problem, the proposed approach, the concrete steps, and the exit criteria that define completion. Projects may be done individually or in groups, and assessment takes group size into account. The deliverables are:

- **Project proposal:** maximum 2 pages in the ACM conference proceedings format, specifically the `sigconf` option of the `acmart` class. LaTeX is mandatory. The proposal names the Stage 1 sub-project and states which of its exit criteria the project expects to meet. Due October 23 at 11:59 PM. Early submissions are strongly encouraged.
- **Project paper:** minimum 7 pages in the same format. Due December 18 at 11:59 PM.
- **Source code:** publicly available on GitHub with a clear README. The link must appear in the project paper.

Presentations (20 %). Each presentation is 30 to 45 minutes, followed by a 30 to 45 minute discussion that the presenter leads. Students indicate which papers they would like to present using the form linked from the course website. Presentation slots are assigned from those preferences.

Paper Reviews (20 %). A review is due at 11:00 AM on the day of the class in which the paper is discussed, through the submission link on the course website. Each review contains a summary, at least two strong points, at least two weak points, and any additional comments. Points must be enumerated and justified rather than written as a single paragraph, and must engage with the technical substance of the paper rather than its writing style. The two lowest review grades are dropped.

Class Participation (15 %). This is a seminar course, so participation is essential. Students are expected to ask questions, respond to questions from others, and comment on the papers under discussion.

No summative tests or final examinations are held in this course. The project paper is a term-long deliverable rather than an examination, and it is due after the last day of classes.

Late and Missed Work Policies

Late Work

No late submissions are accepted for paper reviews. A review submitted after 11:00 AM on the day of the class receives a grade of zero, because the review is the basis for that day's discussion.

The project proposal and the project paper are due at 11:59 PM on the dates given above. Late submissions of project work are accepted only in special cases and must be approved by the instructor in advance.

Missed Work

Short-term, five days or less: the two lowest paper review grades are dropped, which is approximately the lowest 20 % of the reviews. Short-term missed work due to personal circumstances is therefore absorbed by this policy and no documentation is required.

Long-term, more than five days: the project must be submitted in order to pass the course. Students facing circumstances that prevent submission should contact me as early as possible and follow the process for longer-term accommodation.

Undergraduate Academic Advisors

The Undergraduate Advisors for the School of Computer Science are available in Room 5302 HP, or by email at scs.ug.advisor@cunet.carleton.ca. The undergraduate advisors can assist with information about prerequisites and preclusions, course substitutions and equivalencies, understanding your academic audit, and the remaining requirements for graduation. The undergraduate advisors will also refer students to appropriate resources such as the Science Student Success Centre, Learning Support Services, and Writing Tutorial Services.

Graduate Academic Advisors

The Graduate Advisors for the School of Computer Science are available in Room 5302 HP, or by email at grad.scs@carleton.ca. The graduate advisors can assist with understanding your academic

audit and the remaining courses required to meet graduation requirements.

SCS Computer Laboratory

Students taking a COMP course can access the SCS computer labs. The lab schedule and location can be found at https://carleton.ca/scs/tech-support/computer-laboratories/. All SCS computer lab and technical support information can be found at https://carleton.ca/scs/tech-support/. Technical support staff may be contacted in person or virtually; see this page for details.

Mental Health and Wellness

Graduate seminar courses can be demanding, and a term-long research project can be stressful in ways that are easy to normalize. If you are struggling, please reach out, either to me or to the university's support services. The Carleton Wellness website is a good starting point and lists the counselling, health, and peer support options available to students.

Academic Accommodations and Regulations

Academic Accommodation

Carleton is committed to providing academic accessibility for all individuals. You may need special arrangements to meet your academic obligations during the term. The accommodation request processes are outlined on the Academic Accommodations website (https://students.carleton.ca/course-outline/).

Chat GPT / Generative AI Usage

AI use in this course: students may use AI tools for basic word processing and formatting functions, including:

- Grammar and spell checking, for example Grammarly or Microsoft Word Editor.
- Basic formatting and design suggestions, for example Microsoft Word's formatting tools or the PowerPoint Design editor.

You can also use generative AI in writing code (e.g., Codex, Claude, etc.).

You **cannot** use generative AI in paper reviews nor in the project deliverables.

Documenting AI use: it is not necessary to document the use of AI for the permitted purposes listed above. If you have questions about a specific use of AI that is not listed above, please consult your instructor.

Why have I adopted this policy? This policy ensures that student voices and ideas are prioritized and authentically represented, maintaining the integrity of the work produced by students while allowing basic support to enhance clarity, correctness, layout, and flow of ideas. The goal of adopting a limited use of AI is to help students develop foundational skills in writing and critical thinking by practicing substantive content creation without relying on AI support.

This matters in particular for the paper reviews, which are the one place in this course where you practice forming and defending your own judgment about a piece of research. A review produced by an AI tool is worth nothing to you and nothing to the discussion.

As our understanding of the uses of AI and its relationship to student work and academic integrity continues to evolve, students are required to discuss their use of AI in any circumstance not described here with the course instructor, to ensure it supports the learning goals for the course.

Academic Integrity

Misconduct in scholarly activity will not be tolerated and will result in consequences as outlined in Carleton University's Academic Integrity Policy. A list of standard sanctions in the Faculty of Science can be found here. Additional details about this process can be found on the Faculty of Science Academic Integrity website. Students are expected to familiarize themselves with and abide by Carleton University's Academic Integrity Policy.

If you are unsure of the expectations regarding academic integrity, such as how to use and cite references, or whether unauthorized collaboration with classmates is permitted and if so to what degree, then you must ask your instructor. Sharing assignment or quiz specifications or posting them online, to sites such as Chegg, CourseHero, or OneClass, is always considered academic misconduct. You are never permitted to post, share, or upload course materials without explicit permission from your instructor. Academic integrity offences are reported to the Office of the Dean of Science. Information, process, and penalties for such offences can be found on the ODS webpage: https://science.carleton.ca/students/academic-integrity/.

In this course specifically, paper reviews must be written in your own words. Copying text from the paper under review, from another student's review, or from any other source without attribution is an academic integrity offence. Project papers must cite all prior work they build on, and any code reused from another project must be attributed in the README.

Student Rights and Responsibilities

Students are expected to act responsibly and engage respectfully with other students and members of the Carleton and the broader community. See the Student Rights and Responsibilities Policy for details regarding the expectations of non-academic behaviour of students. Those who participate with another student in the commission of an infraction of this Policy will also be held liable for their actions.