CS 6104/5914: Algorithms for Big Data
(Fall 2025)

Prof. Ali Vakilian (vakilian@vt.edu)
Virginia Tech

1 Course Logistics

- **Schedule:** Tue/Thu, 12:30 – 1:45 pm (**MCB 240**).
- **Instructor:** Prof. Ali Vakilian (`vakilian@vt.edu`)
- **Office Hours:** To be announced on the course website.
- **Website & Canvas:** All announcements, slides, and assignments will be posted online.

2 Topics and Learning Objectives

This course surveys modern algorithmic techniques for massive data, including:

- Streaming & Sketching Algorithms
  - Vector statistics, such as frequency estimation and moment
  - Graph problems
  - Geometric problems
  - Linear sketching, such as CountMin and CountSketch
  - Lowerbounds
- Dimensionality Reduction
- Numerical Linear Algebra at Scale
  - Subspace embedding
  - Approximate matrix multiplication
  - Low-rank approximation and PCA
- Nearest Neighbor and Locality Sensitive Hashing
- Clustering and its Coreset Constructions
- Sublinear Algorithms for Graph Problems
- Learning–Augmented Algorithms
- Data-Driven Algorithms
- Distribution Testing
- Massively Parallel Computation Model (MPC)

Course Objectives

By semester's end you will be able to design and rigorously analyze space-efficient algorithms for high-volume streams and high-dimensional matrices.

3 Evaluation and Grading

Homework (45%) Three problem sets released, each 15%, roughly every five weeks.

Final Project (45%) Proposal (5%), checkpoint meeting (5%), presentation (10%), final report (25%).

Class Participation (10%) Active engagement is expected during lectures, and each student must contribute by scribing at least one lecture of their choice.

4 Assignments and Deadlines

Release and due dates appear on the course calendar. **No late submissions.** All work is due at the posted time; no slip days or late penalties apply.

5 Final-Project & Options

The final project is your opportunity to explore an area of modern "big-data" algorithms that excites you. Start thinking early and **discuss your ideas with the instructor**; we are happy to help you scope the work.

Project Styles

(i) **Survey:** Synthesize 3–5 recent research papers into a coherent mini-survey that highlights common themes, contrasting approaches, and open questions.
(ii) **Implementation:** Build and benchmark two (or more) competing algorithms on realistic data sets; evaluate trade-offs in accuracy, speed, and memory.
(iii) **Research:** Propose and develop a new theoretical or empirical result under close mentorship from the instructor.

Deliverables & Timeline

- **Proposal (5%) — due Week 5.** A ~1-page PDF describing your topic, motivation, and an initial plan of attack. Schedule a quick chat with the instructor to refine scope.
- **Checkpoint Meeting (5%) — Week 12.** A 15-minute meeting to review progress and adjust goals. Please bring preliminary results or a working demo.
- **Presentations (10%) — last three lecture slots.** Each team/individual will give a 15-minute talk (+2 min Q&A) to the class. These final sessions replace traditional lectures and are a chance for peer learning.
- **Final Report (25%) — due reading day (12/11/2025).** A 6–8 page write-up summarizing motivation, methods, results, and future work. Submit both PDF and any code/data via Canvas.

brainstorming with the instructor early and often is the best way to ensure your project is appropriately scoped. Except for **surveys**, which must be completed individually, all project types may be done in teams of up to two students.

6 Prerequisites & Background

Students are expected to be comfortable with undergraduate algorithms, discrete probability, and linear algebra; programming experience is helpful but not required.

Self-assessment. Please work on *Homework 0* during the first week as a readiness check. If any portion feels unfamiliar or tricky, bring your attempt to office hours to discuss it with the instructor.

7 Policies

Collaboration. Discuss concepts and collaborate in small groups (of size at most 3) but write solutions alone and list collaborators/resources you used.

Academic Integrity. All submitted work must be your own; violations are handled under the University Honor Code.

Well-being and Inclusion Your mental health matters; reach out early if challenges arise.

The course welcomes students of every background and identity; help us keep the classroom respectful and inclusive.

8 University Statements

We follow all institutional policies on accessibility, discrimination, harassment, and sexual misconduct. See the student handbook for full text.