JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Computer Science EN 601.420, and EN 601.620, Fall 2022
Parallel Computing for Data Science
Semester, Year (3 credits, E)

Online Resources
Please refer to http://parallel.cs.jhu.edu/ for all materials related to this course. This page has up to date information on office hours, course announcements, readings, etc.

Instructor
Professor Randal Burns, randal@jhu.edu, www.cs.jhu.edu/~randal

Head Teaching Assistant
Ariel Lubonja, ariel@cs.jhu.edu

Teaching Assistants
Brian Wheatman, bwheatman@cs.jhu.edu
Meghana Madhyastha, mmadhya1@jhu.edu

Meetings
Monday, Wednesday 4:30–5:45 pm

Textbook
The course does not follow a textbook. Lectures will refer to specific material from the following books that are available through the JHU Libraries.

Required: N. Matloff, Parallel Computing for Data Science. CRC Press, 2015. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.9918&rep=rep1&type=pdf

Required: Mattson, T. G., B. A. Sanders, and D. L. Massingill. Patterns for Parallel Programming. Addison-Wesley, 2004. https://learning.oreilly.com/library/view/patterns-for-parallel/0321228111/

Required: Herlihy, M. and N. Shavit. The Art of Multiprocessor Programming. Morgan-Kaufmann, 2008. https://learning.oreilly.com/library/view/the-art-of/9780123973375/

Course Information

- This course studies parallelism in data science, drawing examples from data analytics, statistical programming, and machine learning. It focuses mostly on the Python programming ecosystem but will use C/C++ to accelerate Python and Java to explore shared-memory threading. It explores parallelism at all levels, including instruction level parallelism (pipelining and vectorization),

shared-memory multicore, and distributed computing. Concepts from computer architecture and operating systems will be developed in support of parallelism, including Moore's law, the memory hierarchy, caching, processes/threads, and concurrency control. The course will cover modern data-parallel programming frameworks, including Dask, Spark, Hadoop!, and Ray. The course will not cover GPU deep-learning frameworks nor CUDA. The course is suitable for second-year undergraduate CS majors and graduate students from other science and engineering disciplines that have prior programming experience. [Systems]

- **Prerequisites**
  Intermediate Programming (EN 601.120 or the equivalent)
  Data Structures (EN 601.226 or the equivalent)
  Computer Systems Fundamentals (EN 601.333 or the equivalent)
  Familiarity with Python.

- **Elective**

Course Goals

Specific Outcomes for this course are that

- Take a computational task and construct an implementation that maximizes parallelism.
- Analyze and instrument an implementation of a computer program for its speedup, scaleup, and parallel efficiency.
- Reason about the loss of parallel efficiency and attribute that loss to factors, including startup costs, interference, and skew.
- Work with a diverse set of programming tools for different parallel environments.
- Analyze how locality, latency, and coherency in the memory hierarchy influence parallel efficiency and improve program design based on the properties of memory.

This course will address the following CSAB ABET Criterion 3 Student Outcomes

Graduates of the program will have an ability to:

1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
3. Apply computer science theory and software development fundamentals to produce computing-based solutions.

Course Topics

- Amdahl's law and strong scaling

- Data science in Python: dataframes, numpy, scipy
- Machine learning in Python: scikit-learn
- Multicore architectures
- Instruction-level parallelism
- Shared-memory parallelism and programming with threads
- Memory hierarchy and caching
- Synchronization and concurrency control
- Gustavson's law and weak scaling
- Distributed computing
- Data-parallel distributed computing: dask, spark
- Distributed actor programming: ray

Course Expectations & Grading

The course includes of six programming activities that span one to two weeks of course time. Activities will be graded for completion of the assignment. Activities that are incomplete or do not fulfill the stated objectives may be resubmitted with permission from the instructor. The goal of the activities is for the student to gain skills with the algorithms, programming tools, and principles presented in the class. Credit for the assignment does not depend on providing correct answers to each question. Answers that are incorrect or programs that do not meet the assignment objectives will be either (1) be marked as incorrect to provide feedback to the student or (2) be returned to the student for resubmission. Every student will have the opportunity to receive all credit for all activities. Activities makes up 30% of the course grade.

There has three exams: two midterms and a final. Each exam counts for 20% of the course grade.

The remaining 10% of the course grade is based on course participation and completion of in-course exercises. Attendance is not required, but credit for this portion of the grade can only be accumulated during the course.

The final letter grades do not depend solely on the achievement of a target score over all assignments and exams. Grades will be determined based on the achievement of learning goals. The course staff will determine a map of total scores to grades at the end of the semester. This policy lets instructors account for variance in exam scores, specifically when the exam scores are lower than intended or expected by the instructors.

Key Dates
Dates for assignments and examination are list on http://parallel.cs.jhu.edu/schedule.html.

Assignments & Readings
All assignments and readings are posted at http://parallel.cs.jhu.edu/schedule.html. This page is definitive and students are responsible for any material on this site. Cancas and gradescope **WILL NOT** have complete or up to date information.

Ethics

The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful, abiding by the *Computer Science Academic Integrity Policy*:

> Cheating is wrong. Cheating hurts our community by undermining academic integrity, creating mistrust, and fostering unfair competition. The university will punish cheaters with failure on an assignment, failure in a course, permanent transcript notation, suspension, and/or expulsion. Offenses may be reported to medical, law or other professional or graduate schools when a cheater applies.
>
> Violations can include cheating on exams, plagiarism, reuse of assignments without permission, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Ignorance of these rules is not an excuse.
>
> Academic honesty is required in all work you submit to be graded. Except where the instructor specifies group work, you must solve all homework and programming assignments without the help of others. For example, you must not look at anyone else's solutions (including program code) to your homework problems. However, you may discuss assignment specifications (not solutions) with others to be sure you understand what is required by the assignment.
>
> *If* your instructor permits using fragments of source code from outside sources, such as your textbook or on-line resources, you must properly cite the source. Not citing it constitutes plagiarism. Similarly, your group projects must list everyone who participated.
>
> Falsifying program output or results is prohibited.
>
> Your instructor is free to override parts of this policy for particular assignments. To protect yourself: (1) Ask the instructor if you are not sure what is permissible. (2) Seek help from the instructor, TA or CAs, as you are always encouraged to do, rather than from other students. (3) Cite any questionable sources of help you may have received.
>
> On every exam, you will sign the following pledge: "I agree to complete this exam without unauthorized assistance from any person, materials or device. [Signed and dated]". Your course instructors will let you know where to find copies of old exams, if they are available.

In addition, the specific ethics guidelines for this course are: Students are encouraged to consult with each other and even collaborate on all programming activities. This means that students may look at each other's code, pair program, and even help each other debug. Any code the was written together must have a citation (in code comments) that indicates who developed the code. Any code excerpted from outside sources must have a citation to the source (in code comments). Each assignment involves questions that analyze the assignment and connect the program to course concepts. The answers to these

questions must be prepared **independently** by each student and must be work that is **solely their own**.

For any activities that are done in pairs or teams, **all team members must be listed as collaborators in comments in the source code AND in any submitted documents (PDFs and notebooks)**. You should also specify the nature and scope of the collaboration, shared programming, discussion, debugging, consulting. Failure to state a collaboration and its scope is an ethics violation and will be treated as such.

Report any violations you witness to the instructor.

You can find more information about university misconduct policies on the web at these sites:

- For undergraduates: http://e-catalog.jhu.edu/undergrad-students/student-life-policies/
- For graduate students: http://e-catalog.jhu.edu/grad-students/graduate-specific-policies/

Personal Wellbeing

- If you are sick please notify me by email so that we can make appropriate accommodations should this affect your ability to attend class, complete assignments, or participate in assessments. The Student Health and Wellness Center is open and operational for primary care needs. If you would like to speak with a medical provider, please call 410-516-8270, and staff will determine an appropriate course of action based on your geographic location, presenting symptoms, and insurance needs. Telemedicine visits are available only to people currently in Maryland. See also https://studentaffairs.jhu.edu/student-life/student-outreach-support/absences-from-class/illness-note-policy/
- The Johns Hopkins COVID-19 Call Center (JHCCC), which can be reached at 833-546-7546 seven days a week from 7 a.m. to 7 p.m., supports all JHU students, faculty, and staff experiencing COVID-19 symptoms. Primarily intended for those currently within driving distance of Baltimore, the JHCCC will evaluate your symptoms, order testing if needed, and conduct contact investigation for those affiliates who test positive. More information on the JHCCC and testing is on the coronavirus information website.
- All students with disabilities who require accommodations for this course should contact me at their earliest convenience to discuss their specific needs. If you have a documented disability, you must be registered with the JHU Office for Student Disability Services (385 Garland Hall; 410-516-4720; http://web.jhu.edu/disabilities/) to receive accommodations.
- Students who are struggling with anxiety, stress, depression or other mental health related concerns, please consider connecting with resources through the JHU Counseling Center. The Counseling Center will be providing services remotely to protect the health of students, staff, and communities. Please reach out to get

connected and learn about service options based on where you are living this fall at 410-516-8278 and online at http://studentaffairs.jhu.edu/counselingcenter/.

- Student Outreach & Support will be fully operational (virtually) to help support students. Students can self-refer or refer a friend who may need extra support or help getting connected to resources. To connect with SOS, please email deanofstudents@jhu.edu, call 410-516-7857, or students can schedule to meet with a Case Manager by visiting the Student Outreach & Support website and follow "Schedule an Appointment".

Classroom Climate

As your instructor, I am committed to creating a classroom environment that values the diversity of experiences and perspectives that all students bring. Everyone here has the right to be treated with dignity and respect. I believe fostering an inclusive climate is important because research and my experience show that students who interact with peers who are different from themselves learn new things and experience tangible educational outcomes. Please join me in creating a welcoming and vibrant classroom climate. Note that you should expect to be challenged intellectually by me, the TAs, and your peers, and at times this may feel uncomfortable. Indeed, it can be helpful to be pushed sometimes in order to learn and grow. But at no time in this learning process should someone be singled out or treated unequally on the basis of any seen or unseen part of their identity.

If you ever have concerns in this course about harassment, discrimination, or any unequal treatment, or if you seek accommodations or resources, I invite you to share directly with me or the TAs. I promise that we will take your communication seriously and to seek mutually acceptable resolutions and accommodations. Reporting will never impact your course grade. You may also share concerns with the Department Head (Randal Burns, randal@cs.jhu.edu), the Director of Undergraduate Studies (Joanne Selinski, joanne@cs.jhu.edu), the Assistant Dean for Diversity and Inclusion (Darlene Saporu, dsaporu@jhu.edu), or the Office of Institutional Equity (oie@jhu.edu). In handling reports, people will protect your privacy as much as possible, but faculty and staff are required to officially report information for some cases (e.g. sexual harassment).

Family Accommodations Policy

You are welcome to bring a family member to class on occasional days when your responsibilities require it (for example, if emergency childcare is unavailable, or for health needs of a relative). In fact, you may see my children in class on days when their school is closed. Please be sensitive to the classroom environment, and if your family member becomes uncomfortably disruptive, you may leave the classroom and return as needed.

The Office of Academic Support at JHU

All programs are free to students, please see below for specifics:

- **PILOT Learning** – Peer-Led Team Learning

    - Students are organized into small study teams who meet weekly to collaborate on faculty-developed problems-sets. Students work together as a team to solve problems.
    - A trained student leader acts as captain and facilitates the weekly meetings using various strategies to foster a collaborative learning environment.
    - Registration opens on August 31st at 9pm EST; registration will remain open throughout the semester if space allows.
    - Contact: Ariane Kelly - ariane.kelly@jhu.edu
    - Instagram: @jhupilot
- **Learning Den Tutoring Program** - Small Group Tutoring
    - Small group, tailored tutoring of 4 students or less which is headed by one tutor. Visit the website (above) to access zoom links for drop-in sessions
    - Tutors can assist with but are not limited to:
        - Review and strengthening of subject-specific material knowledge
        - Assist with homework-like problems
        - Course-specific study skills and exam preparation
        - Contact: Kaitlin Quigley – quigley@jhu.edu
        - Instagram: @jhulearningden
- **The Study Consulting Program**
    - Students work one-on-one with a study consultant to set academic goals and develop customized strategies for success. Areas addressed include but are not limited to:
        - Time management
        - Note taking and test preparation
        - Mastering large amounts of information
    - Contact: Dr. Sharleen Argamaso – sharleen.argamaso@jhu.edu
    - Instagram: @jhustudyconsulting
- **The Writing Center**
    - Undergraduate and graduate students in KSAS/Whiting School of Engineering can schedule 50-min sessions with a Writing Center tutor to look over a draft of written work (up to 10 pages) or a personal statement for graduate study
    - Contact: Robert Tinkle – rtinkle1@jhu.edu
    - Web Address: https://krieger.jhu.edu/writingcenter/