Overview
This course covers the fundamental concepts of database systems. Topics include data
models (ER, relational, and others); query languages (relational algebra, SQL, and
others); implementation techniques of database management systems (index
structures, concurrency control, recovery, and query processing); management of
semistructured and complex data; distributed and noSQL databases.

Prerequisites
Computer Science E-22 or the equivalent, and strong programming skills in Java.
To get a sense of the level of difficulty involved in some of the programming
assignments, we encourage you to review the following sample assignment:
http://sites.harvard.edu/~cscie66/problem_sets/ps2_preview.shtml

Instructor (see the course website for office hours)
David G. Sullivan, Ph.D. (sullivan@post.harvard.edu)
Master Lecturer on Computer Science, Boston University

Teaching Assistant (see the course website for office hours)
Eli Saracino (esaracin@bu.edu)

Key Dates
midterm exam: Wednesday, March 23, 2022
final exam: Wednesday, May 11, 2022
See the detailed schedule at the end of this document for more information.

Meeting Times and Places
lectures: Wednesdays, 7:40-9:40 pm Eastern time, or on demand. Students can
participate in live web conferences, or they can watch recorded lectures on demand.

sections: optional weekly one-hour meetings; time TBA, or on demand; also held via
web conferences. We encourage you to attend or watch sections because they will
reinforce the concepts covered in lecture and prepare you for the assignments.

Course website: https://sites.fas.harvard.edu/~cscie66/
Requirements

1. Problem sets: five assignments, including a combination of written exercises and programming problems. Some of the programming problems must be completed in Java.
2. Midterm exam (March 23; see below)
3. Final exam (May 11; see below)

Important note: The problem sets – especially the programming-intensive ones – tend to be fairly time-consuming. You should plan on devoting approximately 10-20 hours of work per week. If you have other major time commitments, you should reconsider whether to take this course.

Graduate-credit students: Students taking the course for graduate credit must complete additional homework. On most problem sets, the problems required of all students will be worth a total of 100 points; grad-credit students will complete one or two additional problems worth a total of 10 points. These grad-credit problems are typically more challenging than the other problems, and thus grad-credit students should plan to spend approximately 20% more time on the homework.

Grading Policies

Late penalties: Homework is due prior to the start of lecture. If it is submitted more than 10 minutes after the start of lecture, it will be considered a full day late. There will be a 10% deduction for homework that is up to four days late, and a 20% deduction for homework that is 5-7 days late. We will not accept any homework that is more than 7 days late. Plan your time carefully, and don't wait until the last minute to start an assignment so you have time to ask questions and get help.

Determining the final grade: problem sets 35%, midterm 25%, final exam 40%

Your final-exam grade will replace your lowest problem-set grade if doing so improves your final grade. A letter grade will be given in accordance with the Extension School's grading policy. The final grades are not curved. The performance of the class as a whole is taken into account when assigning letter grades, but this can only improve your grade.

Extensions and makeups will only be given in documented cases of serious illness or other emergencies. You cannot redo or complete extra work to improve your grade.

An EXT (extension) grade will be granted only in extreme circumstances (e.g., serious illness), and only when appropriate documentation has been provided. Please bring any such circumstances to Dr. Sullivan's attention as soon as possible.

Exam Policy

The exams will be administered online using Canvas and the Proctorio online proctoring platform. They must be completed within the 24-hour period that begins with the start of lecture on the date specified in the schedule below. Students are expected to have a web cam, microphone, and reliable internet access for the exams.
Academic Conduct

Unless otherwise stated, all work submitted as part of this course is expected to be your own. You may discuss the main ideas of a given problem with other students (provided that you acknowledge doing so in your solution), but you must write the actual solution by yourself. This includes both programming assignments and other types of problems that we may assign.

Prohibited behaviors include:

- copying all or part of another person's work, even if you subsequently modify it
- viewing all or part of another student's work
- showing all or part of your work to another student
- consulting solutions from past semesters, or those found in books or online.

You are also responsible for understanding Harvard Extension School policies on academic integrity. Not knowing the rules, misunderstanding the rules, running out of time, submitting the wrong draft, or being overwhelmed with multiple demands are not acceptable excuses. There are no excuses for failure to uphold academic integrity.

If we believe that a student is guilty of academic dishonesty, we will refer the matter to the Administrative Board of the Extension School, who could require withdrawal from the course and suspension from all future work at the School.

Other Extension School Policies

We also expect you to know and adhere to the general policies and procedures of the Extension School. You can find more information here: https://extension.harvard.edu/for-students/student-policies-conduct/

Accessibility Services

The Extension School is committed to providing an accessible academic community. The Accessibility Services Office offers a variety of accommodations and services to students with documented accessibility issues. For more information, please visit: https://extension.harvard.edu/for-students/support-and-services/accessibility-services/

Course Materials

- Computer Science E-66 coursepack. This will be available for download from the course website. More information will be given during the first lecture.

### Schedule (tentative)

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January 26</td>
<td>Introduction; ER diagrams and the relational model</td>
</tr>
<tr>
<td>2</td>
<td>February 2</td>
<td>Relational algebra and SQL</td>
</tr>
<tr>
<td>3</td>
<td>February 9</td>
<td>SQL (cont.) Storage and indexing</td>
</tr>
<tr>
<td>4</td>
<td>February 16</td>
<td>Storage and indexing (cont.) Implementing a logical-to-physical mapping</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Problem Set 1 due</em></td>
</tr>
<tr>
<td>5</td>
<td>February 23</td>
<td>Transactions and schedules; concurrency control</td>
</tr>
<tr>
<td>6</td>
<td>March 2</td>
<td>Concurrency control (cont.)</td>
</tr>
<tr>
<td>7</td>
<td>March 9</td>
<td>Semi-structured data and XML</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Problem Set 2 due</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Spring break. No lectures or sections.</em></td>
</tr>
<tr>
<td>8</td>
<td>March 23</td>
<td><strong>Midterm exam</strong> (first hour) Semi-structured data and XML (cont.) (second hour)</td>
</tr>
<tr>
<td>9</td>
<td>March 30</td>
<td>Distributed databases and replication; processing distributed data using MapReduce</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Problem Set 3 due</em></td>
</tr>
<tr>
<td>10</td>
<td>April 6</td>
<td>MapReduce (cont.) NoSQL</td>
</tr>
<tr>
<td>11</td>
<td>April 13</td>
<td>NoSQL (cont.)</td>
</tr>
<tr>
<td>12</td>
<td>April 20</td>
<td>Recovery and logging</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Problem Set 4 due</em></td>
</tr>
<tr>
<td>13</td>
<td>April 27</td>
<td>Recovery and logging (cont.) Performance tuning; wrap-up and conclusions</td>
</tr>
<tr>
<td>14</td>
<td>May 4</td>
<td>Review session</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Problem Set 5 due; no late submissions after Sunday, May 9.</em></td>
</tr>
<tr>
<td>15</td>
<td>May 11</td>
<td><strong>Final exam</strong></td>
</tr>
</tbody>
</table>

### Other important dates:
- January 20: registration ends
- January 30: course change period ends
- April 22: last day to withdraw for a grade of WD (no refund)