Course Description & Objectives

This project-based seminar-like course will be covering a number of advanced topics in modern database systems. The course will be organized as a mix of lectures and paper presentations on the recent development of distributed database systems. The purpose of this course is to prepare graduate students the philosophy, methods, and thinking of modern database systems for their career in data science or data management research.

Required Materials

- Course notes and the reading list will be available on Canvas.

Prerequisites

Students are expected to understand the fundamentals of databases (e.g., COMP3115), algorithms (e.g., COMP4030), and operating systems (e.g., COMP4270) each at least at the level of an introductory course.

Workload & Mark Breakdown

Project (55%)

The students are encouraged to come up with their own projects. However, I am happy to suggest some projects as well. The projects are intended to be mini-research projects in a topic of the course. Example formats include (but not limited to):

- Build a new application on top of one of the systems.
• Build a new feature/extension to an existing system.
• Implement a new algorithm that replaces a core part of a system (e.g., scheduling algorithm of a system).

The most important thing is to pick a project on a topic you really like! You will meet with me several times during the term to discuss your progress. You have about 4 weeks to decide on the project. By September 20th you need to give a proposal of your project.

**Project Deliverables:** project proposal (10%), project presentation (15%), project report (20%), and source code (10%)

**Paper Presentation (30%)**

Each student will present two papers from the given reading list (which will be available on eCourseware). Please consult the instructor first to if you plan to present a paper out of the list.

• Please use slides for presentation. The content in your slides should be your own but you can use others’ materials, e.g., figures from the paper we are reading, when necessary and by crediting your source on your slide.

• Your presentation should answer the following questions:
  – What is the problem?
  – Why is it important?
  – Why is it hard? Why don’t naive/previous methods work?
  – What is the solution to the problem the authors propose?
  – What interesting research questions does the paper raise?
  – How does the paper relate to the existing work?
  – In your opinion, what is the major contribution and impact of the work?

• If you are presenting a system, it would be good to show a little demo of how to configure and run that system.

**Participation (15%)**

Your participation in the class is very critical to everyone’s learning. That is why it will be a significant part of your grade. Please ask questions and make comments throughout the discussions that we will have.
Course Schedule (Tentative)

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overview &amp; Introduction</td>
</tr>
<tr>
<td>2, 3</td>
<td>Distributed query &amp; transaction processing</td>
</tr>
<tr>
<td>4</td>
<td>P2P &amp; Multidatabases</td>
</tr>
<tr>
<td>5, 6</td>
<td>1st round paper presentation</td>
</tr>
<tr>
<td>7</td>
<td>HDFS &amp; MapReduce</td>
</tr>
<tr>
<td>8</td>
<td><em>Fall break week</em></td>
</tr>
<tr>
<td>9</td>
<td>Spark &amp; Streaming</td>
</tr>
<tr>
<td>10, 11</td>
<td>2nd round paper presentation</td>
</tr>
<tr>
<td>12</td>
<td>Graph processing and NoSQL systems</td>
</tr>
<tr>
<td>13</td>
<td>Project presentation</td>
</tr>
<tr>
<td>14</td>
<td><em>Thanksgiving week</em></td>
</tr>
<tr>
<td>15</td>
<td>Project presentation</td>
</tr>
</tbody>
</table>

Assessments & Grading

7000-level v.s. 8000-level

Students enrolled in 7000 and 8000 sessions will have the same amount of workload, but 8000-level students are expected to work on more challenging problems for the course project.

Grading Scale

We will calculate final letter grades in two different ways; then each student will receive the higher of the two letter grades. One way is a fixed grading scale, with the following cutoffs:

\[
A \geq 92\% \quad A- \geq 85\% \quad B+ \geq 80\% \quad B \geq 75\% \quad B- \geq 70\% \quad C+ \geq 65\% \quad C \geq 60\%
\]

The other way is a curve, with the following percentages of students receiving each grade:

\[
A : 18\% \quad A- : 18\% \quad B+ : 18\% \quad B : 18\% \quad B- : 18\% \quad C+ : 5\% \quad C : 5\%
\]

However, we will feel free to give an F to any student who clearly did not put effort into the course (or an A+ to any student with truly exceptional performance).

Course Policies

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student’s own work. The plagiarism is incurred when any part of anybody else’s work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.
If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor’s discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Office of Student Conduct for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to: http://www.memphis.edu/studentconduct/misconduct.htm