EECS 433: Database Systems
Spring 2020   Tu/Thu 8:30 - 9:45 am,   Olin Building 313

Instructor: Yinghui Wu

Assistant Professor, Department of Computer and Data Sciences
Case Western Reserve University, 515 Olin Building
(216) 368-8829, E-mail: yxw1650 at case dot edu
Office Hours: Tu, Thu 2- 3 pm (or by appointment) Olin 515

Teaching Assistant: Sheng Guan
Office Hours: Thu 1:30 – 2:30 pm, Olin 406
Contact info: sxg967 at case dot edu

Web Page: from canvas.case.edu/

Prerequisites:
- Basic knowledge of databases in general, and relational databases in particular (database design, E/R model, relational algebra(RA), tuple relational calculus (TRC), SQL (EECS 341 or equivalent)
- Basic knowledge on data structures (stacks, lists, queues, trees) and algorithms (basic searching and sorting, iteration, recursion) (EECS 233) and discrete structures (graphs, trees, sets) (EECS 302).

 Objectives:
- Knowledge on the challenges with querying and managing large, heterogeneous and diverse data; and known data management techniques and the current state of the art to deal with these challenges.
- Knowledge of XML, XQuery, information integration, data warehouses, RDF, SparQL, Column stores, Parallel Databases, Distributed Databases, Cloud-Based Databases.
- Paper presentations and critiquing.
- Experience on project design (survey of a research area, identifying a research problem, developing it as much as possible), report writing (project proposal, progress report, and final project report), demonstration and presentation, as well as experience in working in groups.

Grading:

Midterm Exam 20 %

Project (project reports, project presentations) 35 %

Assignment/Reviews (~6) 35 %

Class Participation 10%

Important Dates:
Midterm Exam: Thu March 5, 2020
Project Proposals Due: March 17, 2020
Final Project Reports Due: May 7, 2020.
Project Presentations: During the last week of classes (May 5)

Grading Policy: No Collaboration is allowed for Exams and Survey Projects.
R&D Projects can be done in groups.

Topics to be covered (tentative list—detailed Weekly Coverage will be posted)

- Review of basic RA, SQL, relational query optimization (Chapters 3-6, 12-13 of SKS, chapters 2-6, 15-16 of GMUW all very briefly)
- Semistructured Data, XML, XPath, XQuery (Chapters 11-12 of GMUW, Chapter 23 of SKS) RDF and SPARQL ;
- Approximate Query Processing
- Parallel and Distributed Databases, cloud computing (Chapter 20 GMUW, 18-19 of SKS)
- Information Integration, data warehouses, OLAP (Chapter 21-22 of GMUW, chapters 20-21 of SKS)
- Database Systems and Information Retrieval (Chapter 23 of GMUW, Chapter 21 of SKS)
- No-SQL/New-SQL databases; Graph Databases
- Data & Information Quality

Books:

Either one of the following books can be used as the textbook for this course.
- [GMUW] Database Systems: The Complete Book, Second Edition, H. Garcia‑Molina, J.D. Ullman and J. Widom, Prentice Hall, 2009
- [SKS] Database System Concepts, 6th Edition, Silberschatz, Korth and Sudarshan, McGraw Hill, 2011.

We will also use papers mostly from the ACM and IEEE Digital Library (free access from Case). Papers will also be posted on Canvas.