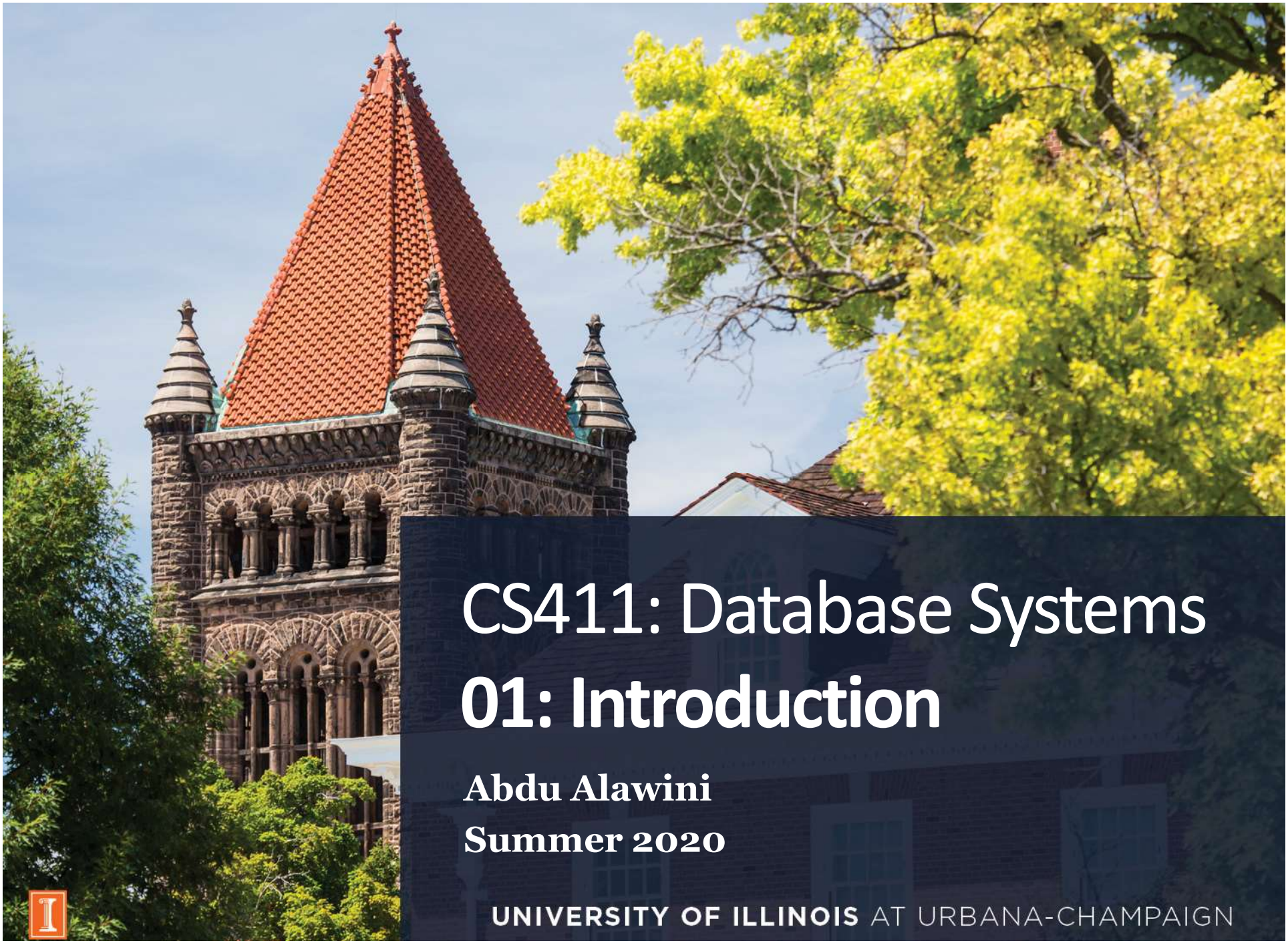


CS411: Database Systems

01: Introduction

Abdu Alawini

Summer 2020

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Teaching Staff

- Abdu Alawini
  - Teaching Assistant Professor

  twitter: @aalawini
  LinkedIn: aalawini
  web: http://www.alawini.com

- Research interests:
  - Database Systems
  - Scientific Data Management
  - Computer Science Education

Teaching Staff: TAs

- Fareedah ALSaad (Ph.D., CS)
  - alsaad2@illinois.edu
- Iris Du (MS, CS)
  - yutingd3@illinois.edu
- Lifu Huang (Ph.D., CS)
  - lifuh2@illinois.edu
- Surya Kalidas (MS, CS)
  - kalidas2@illinois.edu
- Siyuan Zhang (MS, CS)
  - siyuan3@illinois.edu

Welcome to CS411

- **Course website**:
  - https://wiki.illinois.edu/wiki/display/CS411su20
  - Please read the class syllabus, policies, and schedule (tentative, will be updated constantly)
- **Pre-lecture videos**:
  - https://mediaspace.illinois.edu/channel/CS411+Summer+2020/166897321/subscribe
  - Please subscribe to the course media space channel to access the pre-lecture and discussion videos.
- **Course forum and live streaming**: campuswire
  - Signup link: https://campuswire.com/p/GA67004BF
  - Class join code: **6072**
- **Assignment and In-class Activities**:
  - PrairieLearn: SQL, MongoDB, Neo4j
  - Gradescope: Database Design, Indexing, Transactions, Query Processing and Optimization
- **Gradebook**: Compass 2g

CS411: All about “Databases”

DBMS (Data Base Management System) = Database Systems = Databases

System to manage, maintain, query, interact with, transact with data.

More loosely, database systems are used for “data management”

CS411 Goals: Two Perspectives of DBMS

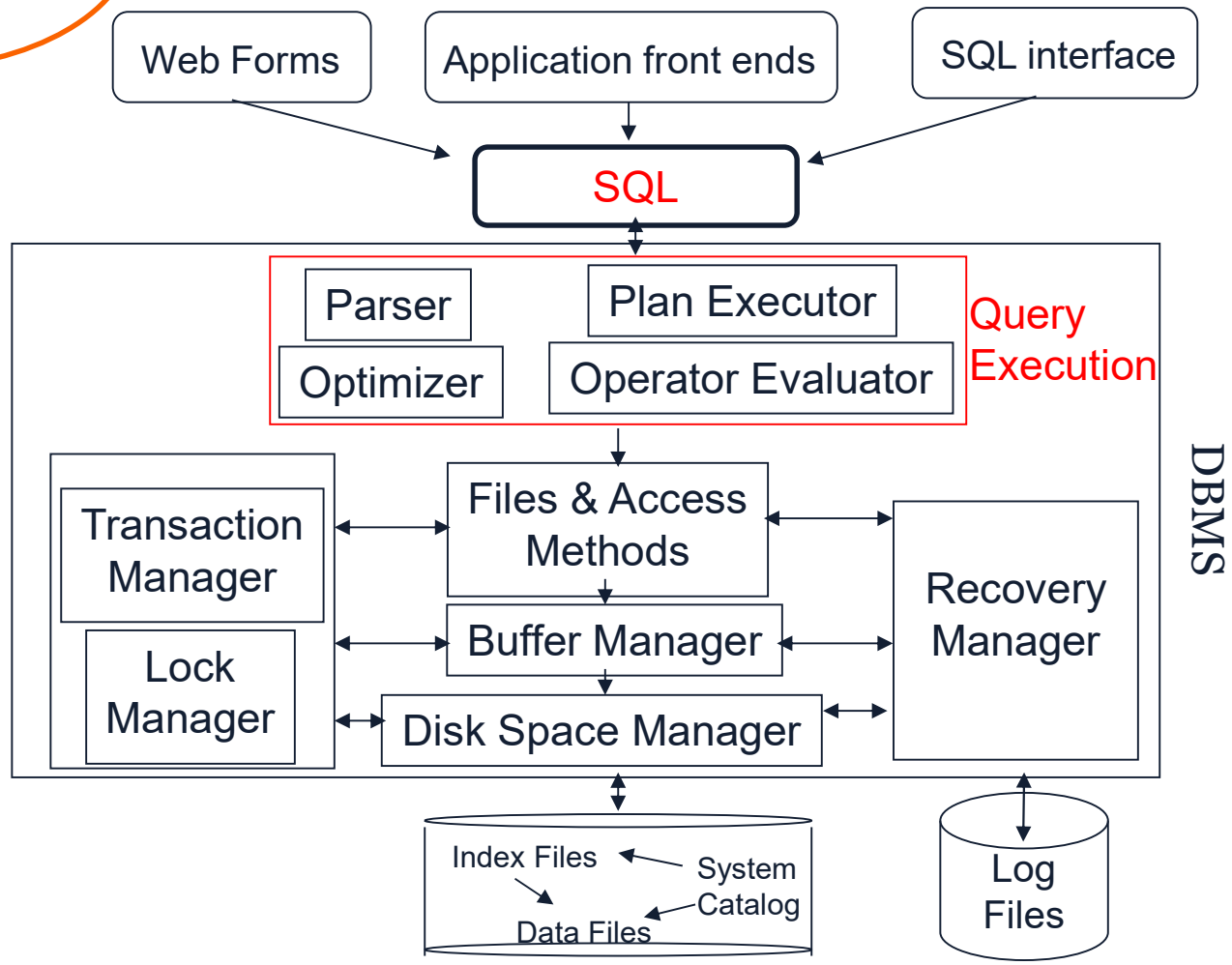
- USER PERSPECTIVE
  - **how to use a database system?**
  - conceptual data modeling, the relational and other data models, database schema design, relational algebra, SQL and No-SQL query languages.
- SYSTEMS PERSPECTIVE
  - **how to design and implement a database system?**
  - data representation, indexing, query optimization and processing, transaction processing, and concurrency control.
  - NOT COMPLETE: high-level view of implementation; CS511

CS411 Modules

- **Module 1: Data Models and Query Languages**
  - **Relational Data Model and Structured Query Language**
  - **Document-oriented Data Model and MongoDB Query Language**
  - **Label Property Graph Data Model and Cypher (Neo4j) Query Language**
- **Module 2: Relational Database Design**
  - **Conceptual Design: ER and UML**
  - **Schema Refinement: Functional Dependencies and Normal Forms**
- **Module 3: Relational Database System Internals**
  - **Storage and Indexing**
  - **Transaction Management: Concurrency Control**
  - **Query Processing and Optimization**

Pretty complex piece of software!!

DBMS Architecture



Why is This Course in the Curriculum?

- It integrates CS concepts

  Languages, data structures, concurrency
- Most CS courses concentrate on code – our interest is managing and representing data and data-centric computation
- It teaches valued job skills
  - DB design and modeling, SQL, NoSQL, Web technologies.
- Important to learn data management
- Example of the practical power (query optimizers) of an underlying theory (relational algebra)
- Databases are now part of a larger data management ecosystem underneath the Web

Prerequisites

- Must have data structure and algorithms background
  - CS 225 or equivalent assumed

- Good programming skills
  - Project will require lots of programming
  - Need C++, Java, Python, or PHP … to communicate w/ the DB
  - Your project group picks the language
  - We cannot help with debugging for your language
    - (i.e., pick wisely)

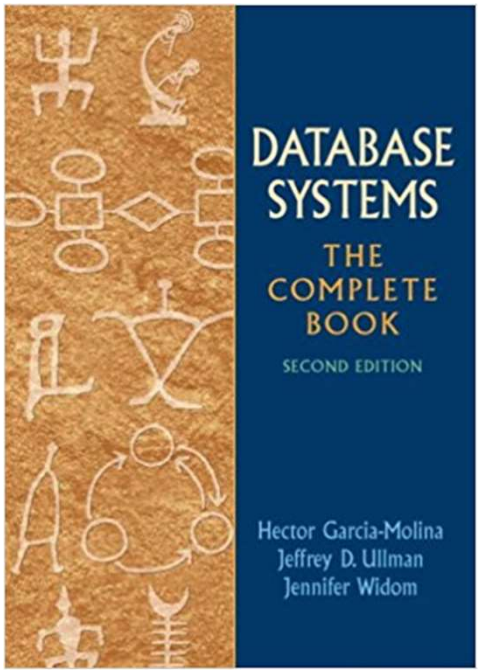
Textbook

- Textbook:

  Database Systems: The Complete Book, 2/e,
  by Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer D. Widom

  All readings will come from this book.

  (Looks intimidating, but it's not very dense…)



- Good references:
  - *Database Management Systems*, by Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill
  - *Database System Concepts*, by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan, McGraw Hill

Course Format

- For all students
  - Pre-lecture recordings and Quizzes / week
    - Quizzes are pass/no pass.
  - ~4 Discussion sessions / week
  - 7 graded assignments planned (MPs/Paper-based)
  - project track 1 (significant): For all students
    - Several stages
  - 1 midterm and 1 final exam
    - Cheat sheet allowed, possibly in both
- Graduate students: 4 credits option
  - Project track 2: write a survey paper

Pre-lecture Videos/Quizzes

- We will post 1-3 short pre-lecture videos (and quizzes) before each discussion session.
  - You should watch these videos before the discussion session (even if you are not attending the session)
- Pre-lecture recordings will cover the basic concepts that we will discuss in class.
- Pre-lectures are important for guiding your reading of textbook (and will be covered in exams and homework assignments)
- As you watch these lectures, please keep notes of any questions you may have. You will have a chance to ask questions during the discussion sessions.

Discussion Sessions

- We will hold discussion sessions during class meetings (MWTH 11am–12:15pm).
- Discussion sessions will be structured as follows
  - First 15 minutes:
    - review of the content covered in the recorded lectures
    - answer any questions about the content covered in the pre-lecture videos.
  - The next 45 minutes: active learning activities.
  - The last 15 minutes: answer questions related to active learning activities.
- The discussions will help you learn the material by
  - working on activities,
  - interacting with your classmates,
  - and asking questions to instructors/TAs.
- Discussion sessions will be recorded.

Participation

- Each discussion will have several activities to encourage collaboration and engagement with the class material.
- Attendance is **NOT required** to receive participation credits.
- You can do the activities during or after the class meeting time.
- You **MUST** complete each day's activities **before** the end of the day in which these activities were posted.

Participation

- Participation is worth 10% of the class grade and is distributed as follows:
  - 2% for watching the pre-lecture videos and taking the video quizzes before the deadline.
  - 8% for participating in "live" or "offline" activities.
- You must complete all questions in each activity to receive participation credits.
- Answering students' questions on Campuswire will be rewarded with **1% extra credit** of the entire course grade.

Campuswire

- Campuswire replaces Piazza, Slack, and (possibly) Zoom.
- You should've received my invitation to campuswire. If not, use the following sign up link:
  - Signup link: https://campuswire.com/p/GA67004BF
  - Class join code: **6072**
- Campuswire Features:
  - Q&A and Chatrooms
  - Live sessions and lectures with classroom responses system **without** iClickers
  - Live sessions with advanced queuing and group collaboration tools

Homework Assignments

- 7 weekly programming (prairielearn) and paper-based (gradescope) assignments
- Assignments will be posted on the wiki page
- Submit through Prairie Learn or gradescope
- **No** late homework will be accepted
  - Late = missing
  - OK to submit partial work
- For more details, see policy page

Course Projects

- There will be a semester-long project, which involves significant database application programming.
- The project will be structured with several milestones due in the course of the semester, leading to a demo and write-up near the end of the semester.
- We will use CATME to form teams (HW 0)
- I'll talk about course projects on Thursday.

Exams

- One Midterms exam (CBTF remote)
- One Final exam (CBTF remote)
- The schedule is up. Check exam dates and mark them on your calendar
  - you should not have conflicts if you are able to take the class
  - generally no makeup exams unless exceptional cases (see policy page)

Tentative Grading Breakdown

- Homework: 25%
- Project: 30%
- Participation 10%
- 1st Midterm: 15%
- Final Exam: 20%

For 4 Credit students: 90% = the total of everything above, 10% = Survey Project

IMPORTANT: Plagiarism

- **We have a zero-tolerance policy on plagiarism.**
- We will be running automated plagiarism detection software on all your course submissions.
  - This has led to dozens of students being docked grades.
  - Do not do this!
- Similarly, do not attempt to copy in class
  - We will find you

IMPORTANT: Communications and Contacting the Staff

Communications: From us to you

Campuswire:

- vitally important!
- **announcements will be posted on Campuswire**
- make sure to check it regularly for questions/clarifications
- **Enable “notifications”**

Communications: From You to Us

If you have a question/problem

1. talk to people in your group first
2. post your question on compass
3. if it is of a sensitive nature, post your question *privately* to compass
4. if it is pertinent only to one TA, and of a sensitive nature, then send direct message (DM) to the TA
5. if it is sensitive & not easily conveyed electronically, then go to remote office hours to talk to TA or instructor

Campuswire Q&A

- Designed for you and your peers
  - to communicate and help one another
  - please do not post solutions
- TAs will monitor relatively regularly and try their best to help with your questions
  - **But don't expect responses in <24 hours.**
    - There will be many questions
    - May not be able to answer all of them in timely manner
    - Don't wait until the last minute to ask!
  - Not good for more complex questions – join remote office hours or contact TA

Campuswire: Incentivizing Participation

- Since it's hard for us to be present 24x7 on Campuswire, we want to incentivize the students who diligently answer other's questions
- Up to 1% extra credit of the grade for Campuswire participation
  - Grading criteria: informative, succinct, useful, clear answers are rewarded.
  - Campuswire grade will be based on the number of "instructor/TA" endorsed answers

Office Hours

- We will use Campuswire “Live Sessions” for hosting office hours (OH).
- OH are for any complex in-person questions and clarifications
- Teaching staff has OHs 5 times a week
- Mine: Friday 10:30-12:00
- TA OHs will start after we post HW1
- We will post TAs office hours on Campuswire

IMPORTANT: Registration Questions

- I get 2-3 emails a day about registration/wait-list questions.
  - Unfortunately, I won't be able to respond to all emails
- We will not maintain a waitlist
- Registration will be first-come-first-serve.

IMPORTANT: Grading Policy

- In the past we have tried two forms of grading:
  - Absolute, score-based grading

Total	Grade
90-100	A (A-, A, A+)
80-89	B (B-, B, B+)
70-79	C (C-, C, C+)
60-69	D (D-, D, D+)

  - And curving
- And then we take the best of both grades.
- Typically, curving leads to a higher grade.
- Separate curves for 4/3 credit, and ug/grad, but hasn't ended up mattering in the past.

How to Get the Most out of CS411?

- Read and think before/after class
  - readings are there for a reason
  - discuss assignments w/ others but write your own solution!
- Use lecture notes (slides) as a guide
  - a roadmap for what's important
  - lectures are **starting** points– they do not cover everything you should read
- Attend and participate in discussion sessions

Course Agreement!

Student Expectations

a. PLEASE BE ACTIVE AND PARTICIPATE.
b. Listen and respect others.
c. Be comfortable taking risks.
d. Complete all assignments.
e. Be punctual for all classes.
f. Discuss class concerns either after class or during designated office hours.
g. Be prepared for class by reading the assigned reading prior to lesson.

Instructor Expectations

a. BE ACTIVE AND ENTHUSIASTIC TO FACILITATE YOUR LEARNING.
b. Listen and respect your views.
c. Be available before and after each class.
d. Respond swiftly and effectively to your concerns.
e. Grade objectively, consistently, and in a timely manner.
f. Be prepared for class.
g. Accommodate differences in your learning

Source: Constructing a Learner-Centered Syllabus: One Professor's Journey

Questions?