UCLA

CS143
Data Management Systems

Professor Junghoo “John” Cho

Course Staff

- Instructor: Junghoo “John” Cho
  - Email: cho@cs.ucla.edu
  - Office: Boelter Hall 3531H
  - Research area: natural language processing, machine learning, search engine
- TA
  - Vivian Cheng (viviancheng1993@gmail.com)
  - Tyler Davis (tylerdavis@g.ucla.edu)
  - Song Jiang (songjiang@cs.ucla.edu)
  - Jason Teoh (jteoh@cs.ucla.edu)
  - Ariel Wu (wuyue0715@ucla.edu)

Lecture Q&A

- Ask questions via sli.do during class
  - #808652
  - https://app.sli.do/event/otspn2lm
- Questions with 3 or more “likes” are answered in class
  - Other questions are answered after the class

PTE

- All CS students who submitted ECR have been added
- No more PTEs will be issued
  - We are way over capacity

Introduction

- Course objective
- Course administration
- What is a database system?

Course Objective

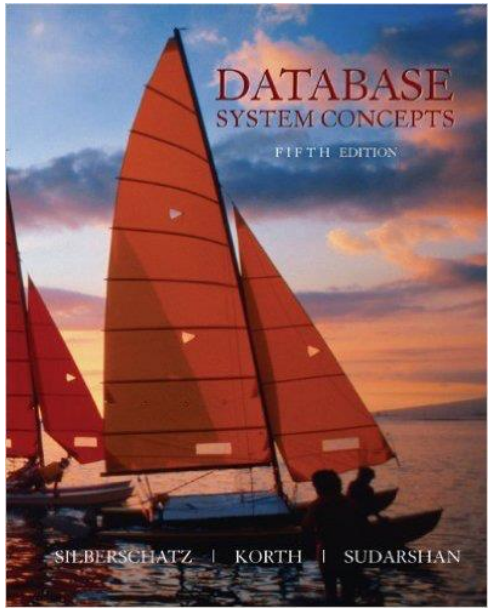
- Learn how to use data management systems
  - Learn data model
  - Learn SQL (Structured Query Language)
  - Learn Relational Database Management Systems (RDBMS)
  - Learn basics of other non-relational DBMS (NoSQL)
- By the end of the quarter you will know
  - How to store data in DBMS
  - How to retrieve and interact with data in RDBMS using SQL
    - In particular, you learn how to use MySQL
  - How DBMS works

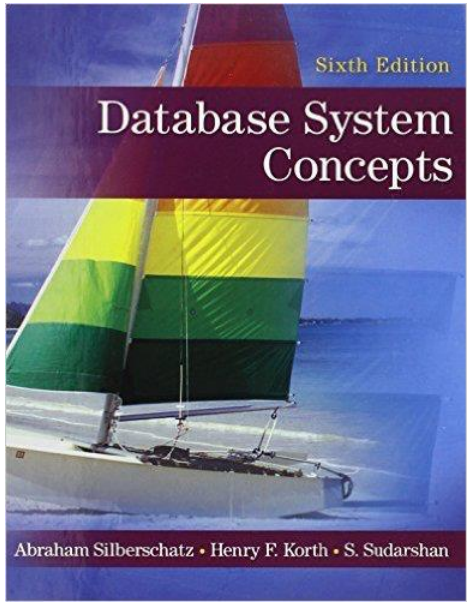
Tentative Course Topics

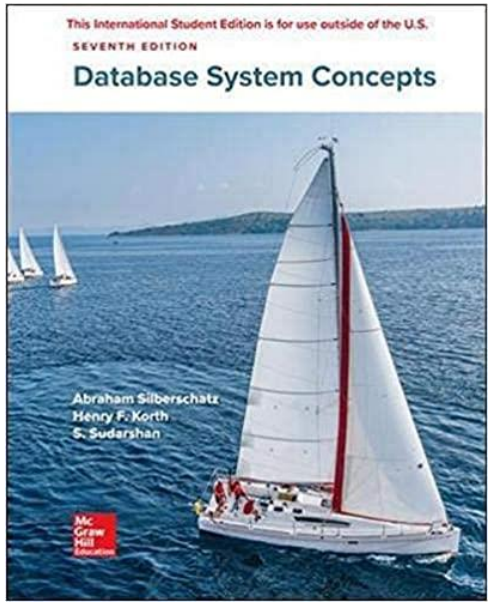
- Relational model
- Relational algebra
- SQL
- Entity-relationship model
- Relational design theory
- Key data structure and algorithms for DBMS
- Query execution
- Transactions
- NoSQL Database
- Non-relational data processing

Textbook

- Database System Concepts by Silberschatz, Korth, and Sudarshan
- 5th or later editions are OK







Grading

- Homework 10%
- Project 40%
- Quiz 50%

Quiz

- At the following hours
  - 10/30 (5th week Saturday) 10-11:59 AM
  - 12/04 (10th week Saturday) 10-11:59 AM
- Closed-book and closed notes
  - Cheat-sheet allowed
  - Calculator allowed
- If you cannot be physically present
  - Submit your reason and timezone info via Google Form
  - https://forms.gle/DskLuqYf6eEYKgn49
  - “I am attending the class remotely from South Korea”
  - “Saturday 10-11:50AM is Sunday 2-3:50AM in Seoul, Korea (GMT+9)”

Homework

- Review of what is learned in the lectures
- Checked by completion, not by correctness
  - You will get full credit as long as you submit your work and demonstrate reasonable efforts
- Tentative schedule (due on Wednesdays)
  - Week 2: Relational algebra
  - Week 3: SQL
  - Week 5: ER
  - Week 6: Normalization and integrity
  - Week 8: Disk
  - Week 9: Index
  - More on Join, transaction

Project

- Manage real-world data using data management/processing tools
- 4 grace days to accommodate emergency
  - Maximum 2 days late per project
  - 5^N% penalty once grace days are used up
- Tentative schedule (due on Sundays)
  - Week 2: System setup and MySQL bulk load
  - Week 4: Simple movie Web site
  - Week 6: Load Nobel Prize JSON Data to MySQL
  - Week 7: Load Nobel Prize JSON Data to MongoDB
  - Week 8: Frequent Book Pair on Spark
  - Week 9: Google n-gram processing using Unix text tools

Lectures and Discussion Sections

- Lectures
  - Monday, Wednesday: 10 – 11:50AM
  - https://ucla.zoom.us/j/96835293526
  - Meeting ID: 968 3529 3526, Passcode: 154806
- Three discussion sections (one in-person and two online)
  - Online: https://ucla.zoom.us/j/95871805961 (Passcode 557333)
    - Friday 10AM: Lecture and homework discussion
    - Friday 12PM: Project discussion
  - In-Person: Bunche 2209A
    - Friday 12-2PM: General Q&A
- Lectures and online discussion sections will be recorded

Office Hours

- Zoom link
  - https://ucla.zoom.us/j/92515789392
  - Meeting ID: 925 1578 9392, Passcode: 101707
- Hours
  - Monday 12-2PM (Vivian)
  - Tuesday 12-2PM (Ariel)
  - Wednesday 9-10AM (Jason), 2-3PM (John), 3-4PM (Jason)
  - Thursday 9-10AM, 3-4PM (Tyler)
  - Friday 4-6PM (Song)
- Office hours will not be recorded

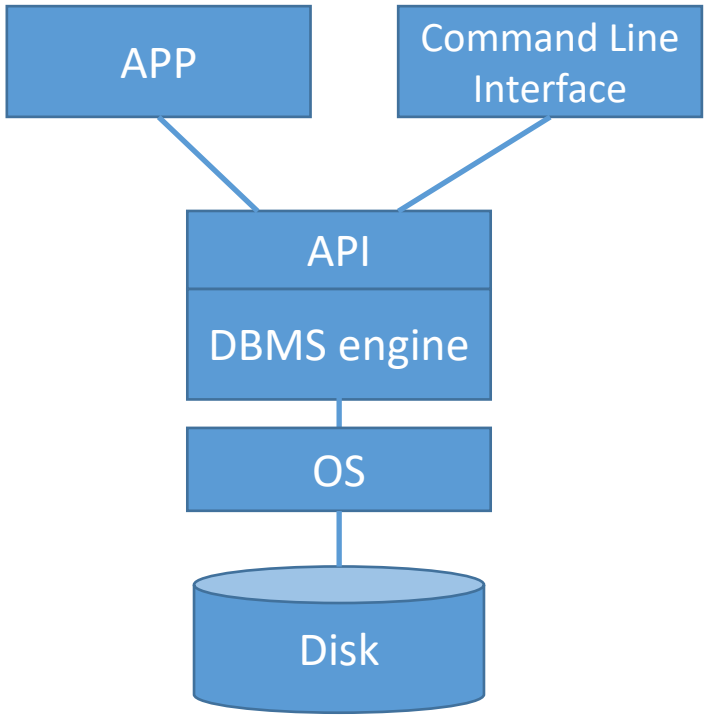
Other Communication

- Class videos will be available on CCLE
  - https://ccle.ucla.edu/course/view/21F-COMSCI143-1
- Other course materials are available on our Web site
  - http://oak.cs.ucla.edu/classes/cs143/
- Campuswire will be used for online discussion and general question
  - https://campuswire.com/c/G17AAD1D2
- Assignment submission via Gradescope
  - https://www.gradescope.com/courses/316671

Database Management System (DBMS)

- Q: What is a DBMS?
- A: A system that manages massive amounts of data and provides convenient, efficient, persistent, and safe access to such data to many users simultaneously
  - “Massive”
  - “Convenient”
  - “Efficient”
  - “Persistent”
  - “Safe”
  - “Many users”

Database Architecture



Popular DBMS Software

- Relational
  - Open source: MySQL, PostgreSQL, …
  - Closed source: Oracle, Microsoft SQL Server, IBM DB2, …
- Non-relational (NoSQL)
  - MongoDB
  - Spark
  - …
- Cloud
  - Amazon RDS, Azure SQL Database, Google Cloud Spanner, Oracle Database Cloud, …

Database is Everywhere

- All businesses manage various types of data
  - Online banking
  - Amazon, Walmart, …
  - Phone backend system
- SQL is *the* language to interact with business-related data
  - Primary focus of this class

MySQL, Docker, Unix

- In our class, students are required to learn and use
  - Docker
    - x86 Windows or Mac (M1 Mac may be used, but not tested or supported)
  - Unix command line
  - MariaDB (≅ MySQL)
- Why?
  - MySQL is most popular open-source RDBMS
  - Docker is easiest way to install, manage, and use software packages
  - Unix command line provides expressive power to deal with files and data
- They are difficult to learn initially, but it will pay off enormously

Announcements

- Make sure to be on Campuswire and Gradescope
  - Campuswire: https://campuswire.com/p/G17AAD1D2/ (code: 1767)
  - Gradescope: Send email to TA
- Start Project 1
  - http://oak.cs.ucla.edu/classes/cs143/project1/
  - Get familiar with Docker and Unix command line
  - MySQL Bulk-load command will be learned in next lecture
- Learn PHP (for Project 2) and Python (for Projects 3-5)