

CS 564: DATABASE MANAGEMENT SYSTEMS

Fall 2021

COURSE LOGISTICS

INSTRUCTOR

Paris Koutris

- paris@cs.wisc.edu
- Office hours
 - **WHERE:** hybrid (Zoom + office)
 - **WHEN:** Thursday 1:30-2:30 pm

ABOUT ME

- undergrad in Athens, Greece
- Ph.D. in University of Washington (the other UW)
- at UW-Madison since Fall 2015!

Research Interests

- massively parallel processing
- data pricing
- managing uncertain data
- data structures for query processing

MEET YOUR TAS

- Austen Fan
 - Office Hours: **Monday** 6:00-7:00 pm @ Zoom
 - Email: zfan64@wisc.edu
- Song Bian
 - Office Hours: **Wednesday** 6:00-7:00 pm @ CS 3252
 - Email: sbian8@wisc.edu
- Ruohui Wang
 - Office Hours: **Friday** 9:30-10:30 am @ CS 3269
 - Email: ruohui@cs.wisc.edu
- Yifei Yang
 - Office Hours: **Tuesday** 1:00-2:00 pm @ CS 3213
 - Email: yyang673@wisc.edu

COURSE FORMAT

- Lectures **M+W+F** 4:00-5:15 pm
- 3 programming projects (in groups of 3)
- 3 problem sets (individual)
- Midterm Exam
- Final Exam

CANVAS HAS EVERYTHING!

Fall 2021-2022

Recent Announcements

Home

Piazza

Lectures

Assignments

Grades

People

Course Summary

Course Syllabus
(AEFIS)

Pages

Kaltura My Media

Kaltura Gallery

Lectures: **MWF** 4:00-5:15 pm @ Social Sciences 6210

Instructor: [Paris Koutris](#)

- Office Hours: **Thursday** 1:30-2:30 pm or by appointment (Zoom or @ CS 4363)
- Email: paris@cs.wisc.edu

Teaching Assistants:

- Austen Fan
 - Office Hours: **Monday** 6:00-7:00 pm @ Zoom
 - Email: zfan64@wisc.edu
- Song Bian
 - Office Hours: **Wednesday** 6:00-7:00 pm @ CS 3252
 - Email: sbian8@wisc.edu
- Ruohui Wang
 - Office Hours: **Friday** 9:30-10:30 am @ CS 3269
 - Email: ruohui@cs.wisc.edu

View Course Stream

View Course Calendar

View Course Notifications

To Do

- | | | |
|--|--|--|
| | Problem Set 1: SQL | |
| | 75 points | |
| | Sep 26 at 11:59pm | |
| | Programming Project 1... | |
| | 75 points | |
| | Oct 10 at 11:59pm | |
| | Problem Set 2: Normal... | |
| | 50 points | |
| | Oct 22 at 11:59pm | |
| | Midterm Exam | |
| | 100 points | |
| | Oct 27 at 11:59pm | |
| | Programming Project 2... | |
| | 100 points | |
| | Nov 7 at 11:59pm | |
| | Programming Project 3... | |
| | 125 points | |

COMMUNICATION

Piazza: access through Canvas

- all course announcements
- questions (answer each other's questions!!)
- discussions

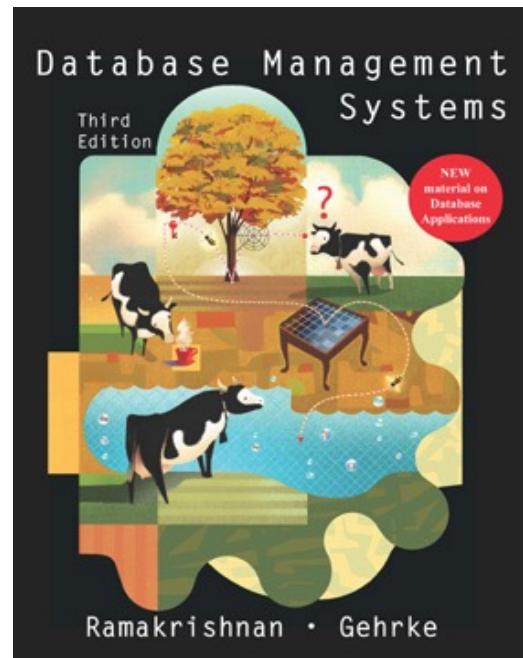
TODO: Enroll to Piazza!

LECTURES

- Lectures will be in person!
- All lectures will be recorded and uploaded to Kaltura

TEXTBOOK

- Database Management Systems (3d edition)
- **Come to the lectures!**
 - take notes
 - ask questions
 - participate



PREREQUISITES

- Data structures and algorithm background
- For the **programming projects**
 - C++ will be used for the database internals
 - Python is also required

C++

Brush up your C++ skills!

- [CS 368: C++ for Java Programmers](#)
- [C++ tutorial](#)
- [Another C++ tutorial](#)

GRADE DISTRIBUTION

- Programming Projects (3)
- Problem Sets (3)
- Midterm: 20%
- Final: 30%

PROBLEM SETS

Individual assignments: Python + Jupyter notebooks

- **Problem Set #1**
 - *SQL*
- **Problem Set #2**
 - *Normalization & Relational Algebra*
- **Problem Set #3**
 - *I/O cost & Query Optimization*

PROGRAMMING PROJECTS

In groups of 3: Python and C++

- **Project #1**
 - *ER Modeling & Schema Design*
- **Project #2**
 - *Buffer Manager*
- **Project #3**
 - *B+ tree*

EXAMS

- **Midterm Exam**
 - *when*: October 27 (during class)
 - *where*: Social Sciences 6210
- **Final Exam**
 - *when*: December 20, 7:45 – 9:45 am
 - *where*: TBD

WHAT IS EXPECTED FROM YOU

- Attend the lectures
- Participate and ask questions
- Do the assignments (start early!)
- Study for the exams

PLAGIARISM

- You may not copy source code or your answers from other students or the web
- Plagiarism will not be tolerated!

COURSE OVERVIEW

Part A: Databases from the **user's** perspective

- Module **A1**: SQL
- Module **A2**: ER Model + DB Design
- Module **A3**: Relational Algebra

COURSE OVERVIEW

Part B: Database **internals**

- Module **B1**: Basics of DB Internals
- Module **B2**: Indexes
- Module **B3**: Query Processing
- Module **B4**: Transactions

BEFORE WE START

JUPYTER NOTEBOOK

- Jupyter notebooks are interactive shells which save output in a nice notebook format
- You'll use these for
 - in-class activities
 - interactive lecture supplements/recaps
 - problem sets, projects, ...



JUPYTER NOTEBOOK SETUP

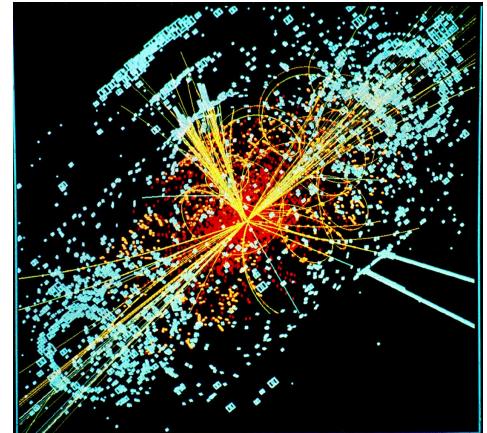
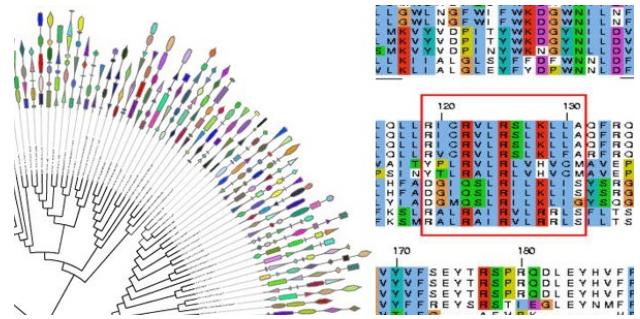
- Install on your laptop!
- We will be using **Python 3**
- Alternative methods are to use:
 - Ubuntu VM
 - CS Machines
- Come to office hours if you need help with the installation!

TODO: try Activity 1 to familiarize yourself with Jupyter!

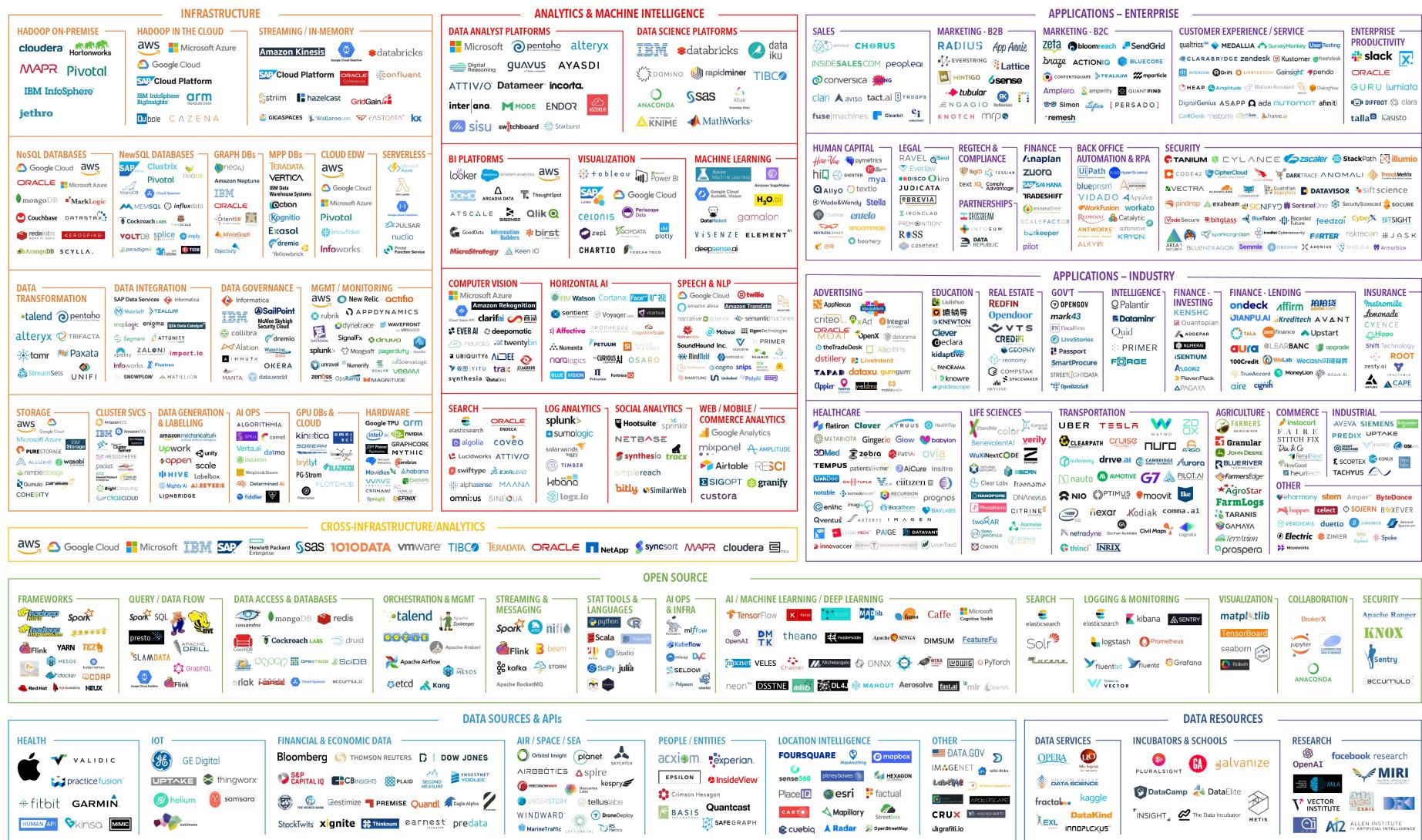
DATABASES: A SHORT INTRO

DATA IS EVERYWHERE!

- Our world is increasingly **data driven**
 - scientific discoveries
 - online services (social networks, online retailers)
 - decision making
- **Databases** are the core technology



DATA & AI LANDSCAPE 2019



July 16, 2019 - FINAL 2019 VERSION

© Matt Turck (@mattturck), Lisa Xu (@lisaxu92), & FirstMark (@firstmarkcap)

mattturck.com/data2019

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

CS 564 [Fall 2021] - Paris Koutris

WHAT IS THIS CLASS ABOUT?

The **fundamentals** of data management

- how we design and query a database?
- how do database management systems work?
- how do we build a DBMS?

DATABASES

*What is a database (**DB**)?*

an **organized** collection of **structured data**

What are examples of databases?

- payroll information
- product information (e.g. Amazon)
- airline reservations
- ...

DBMS

*What is a Database Management System (**DBMS**)?*

a **program** that allows us to manage **efficiently** a large DB and
allows data to **persist** over long periods of time

What are examples of DBMSs?

- SQL Server, Microsoft Access (Microsoft)
- DB2 (IBM)
- Oracle
- MySQL, PostgreSQL, SQLite

EXAMPLE: ONLINE BOOKSTORE

What data do we need to store?

- information on books: title, author, ...
- customer information
- orders: date, items, payment method
- user preferences and behavior
- ...



EXAMPLE: ONLINE BOOKSTORE

How will we use the data stored?

- search for a specific book
- make an order
- recommendation engine
- generate order history
- produce sales figures
- ...



WHY NOT USE A FILE?

We could store the information about books in a CSV file that the application manages

Problems:

- How can we find a particular book fast?
- What if two threads try to write at the same time?
- What happens if there is a crash during an update?
- How do we ensure that an attribute has a valid value?

WHAT CAN A DBMS DO?

- Automate a lot of boring operations on data
 - don't have to program over and over
 - write complex data manipulations in a few lines
- Make data retrieval very fast
- Scale up to very large data sets (100s of GB)
- Safely allow concurrent access to the data
- Protect from system crashes

KEY CONCEPTS

Data model: abstraction that describes the data

Schema: describes a specific database using the “language” of the data model

RELATIONAL MODEL

- The data is stored in **tables** (**relations** in the mathematical sense)
- A database is a set of tables

isbn	name	author	hardcover
007456	The Da Vinci Code	Dan Brown	yes
909405	Ender's Game	Orson Scott Card	no
...

schema

record/tuple

OTHER DATA MODELS

- Key-value (NoSQL)
- Graphs (Graph Databases)
- Matrix/Array (Machine Learning)
- Documents
- ...

KEY CONCEPTS

Query Language: high-level language to allow a user to store and retrieve data from the DB

- Declarative languages (SQL, Relational Calculus)
- Procedural languages (Relational Algebra)

Data Independence: the application does not change when the underlying data structure or storage changes