# CS 4604: Introduction to Database Management Systems

Virginia Tech CS 4604 Spring 2021

Instructor: Yinlin Chen

VIRGINIA TECH

# Topics

- Course overview
- Overview of database systems

# Course Information

- Instructor: Yinlin Chen: ylchen@vt.edu
  - Include string **CS 4604** in subject in any email you send me
  - Zoom Address: https://virginiatech.zoom.us/my/yinlinchen
- Teaching Assistants:
  - Hongjie Chen: jeffchan@vt.edu
    - Zoom Address: https://virginiatech.zoom.us/my/jeffchan
  - Monjura Rumi: mrumi@vt.edu
    - Zoom Address: https://virginiatech.zoom.us/j/91963146619
- Office hours will be announced by end of this week
- Keeping in Touch
  - Course syllabus: http://courses.cs.vt.edu/~cs4604
  - Canvas: https://canvas.vt.edu/courses/125160
  - Both updated regularly through the semester

# Course Format

- Lectures (Tue and Thur)
  - 3:30 - 4:45pm
  - You should attend the class
  - We will take attendance
- Online meeting with Zoom
  - 5%: Class attendance and participation
  - Class videos with captions will be available on Canvas
- Turn on your Canvas notification (email)
- Check Canvas announcements

# Uncertainty

- Zoom was down
- Have been dropped by Zoom
- Was unable to join Zoom
- Be clam, I will show up!
- Check announcement in Canvas or email notification

# 2021 Spring Break Day

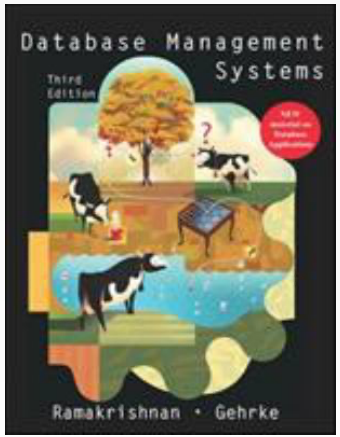
- February 25 (Thur) and April 6 (Tue)
- No class meetings
- No assignment due
- No exams
- No office hours
- Mark your calendar

# Require by Virginia Tech

- You MUST complete these:
  - **Understanding the Code:** https://canvas.vt.edu/enroll/YN3BXF
    The course opens at 8:00am on January 25 and closes at 11:59pm EST on January 29.
  - **Online Academic Integrity Module**: https://canvas.vt.edu/enroll/HKRXKT
    The course opens at 8am on January 18th, and will remain open all year.

# Textbook

- Required
  - Database Management Systems, by Raghu Ramakrishnan and Johannes Gehrke. 3rd Ed. McGraw Hill.
  - Web page for the book (with errata): http://pages.cs.wisc.edu/~dbbook/
- Optional:
  - Garcia-Molina, Ullman and Widom, 3rd Ed.
  - Silberschatz, Korth and Sudarshan, 6th Ed.



# Assignments

- Electronic Homework
  - Written problems
  - Due at 11:59pm
  - Exercises based on class material and textbook
- Team Projects
  - Programming problems
  - Due at 11:59pm
- Honor code. Do you best! Don’t cheat.
- Zero-tolerance philosophy regarding plagiarism or other forms of cheating.

# Course Project

- We will put project overview at next class (first project assignment)
  - 4 persons per project
- Project runs the entire semester with regular assignments
- Midterm project presentation
- Final project presentation and deliverables

# Labs

- Labs give you an opportunity to practice the applied aspects of this course.
- Short - both in time to complete and number of questions
- Asynchronous (somewhat) with class or off-class
  - Self-contained - everything you need to know to complete the lab is included in the lab instructions.
  - Off-class lab can be started any time
- Labs are due or will be expired on the date indicated in Canvas
- GitHub: https://github.com/VTCourses/CS4604_Labs

# Exam Dates

- Midterm (03/11): 3:30-5:30pm
  - Online with time limit, open book
- Final (05/08): 7:45-9:45am
  - Online with time limit, open book
- You must attend both midterm and final exam in order to pass the course
- No makeup exams
- Let us know (emails) if you cannot attend the exams due to personal reasons at least two weeks early, except emergency.
- Mark your calendar!

# Course Grading

- 5%: Class attendance and participation
- 25%: Homework
- 25%: Project
- 20%: Midterm exam
- 25%: Final exam

# Class Policies

- Make sure you go through the detailed policies on website:
  - http://courses.cs.vt.edu/~cs4604/Spring21/
- Lectures: Inform me and TA in advance if you can't attend the class or have to leave a class early or come late for any reason
- Assignment late policy
  (Penalized score) =
  (Your raw score) * (1 - 0.1 * (# of days past deadline))
- Exams: You MUST take midterm and final exam

# Course Objective

- Intended to be a first course in database systems for advanced undergraduates in computer science.
- Introduction to the design and programming of database systems.
- We will cover:
  - The ER (entity-relationship) approach to data modeling
  - The relational model of database management systems (DBMSs)
  - The use of query language SQL
  - Relational algebra & calculus
  - Database normalization
- Will touch upon query processing and the role of transaction management
- Will also devote some time to current topics of research such as semi-structured databases, database security, cloud databases, etc

# Course Topics

- The Relational Model
- Relational Algebra and SQL
- E/R Models
- Storing Data and Indexes
- Hashing/Sorting and Query Processing
- Query Optimization
- Function Dependencies and Normalization
- Transactions and Concurrency Control
- Logging and Recovery
- NoSQL, Data Mining, Data-warehousing, Cloud databases, etc.

# What Will You Learn in CS 4604?

- Implementation
  - What is under-the-hood of a DB like Oracle/MySQL/PostgreSQL?
- Design
  - How do you model your data and structure your information in a database?
- Programming
  - How do you use the capabilities of a DBMS?
- CS 4604 achieves a balance between
  - a firm theoretical foundation to designing moderate-sized databases
  - creating, querying, and implementing realistic databases and connecting them to applications

# Software Stacks

- Python: https://github.com/VTCourses/Python_tutorial
- GitHub: https://github.com
- SQLite, MySQL, Postgres
- Cloud databases: AWS
- Jupyter Notebook
- Google G-suite
- And more…

# Why Study Database?

- Why not study databases?
- The global Database Management System (DBMS) market will be expected to reach USD 50940 million by 2025, from USD 42980 million in 2019
- Why you study databases?

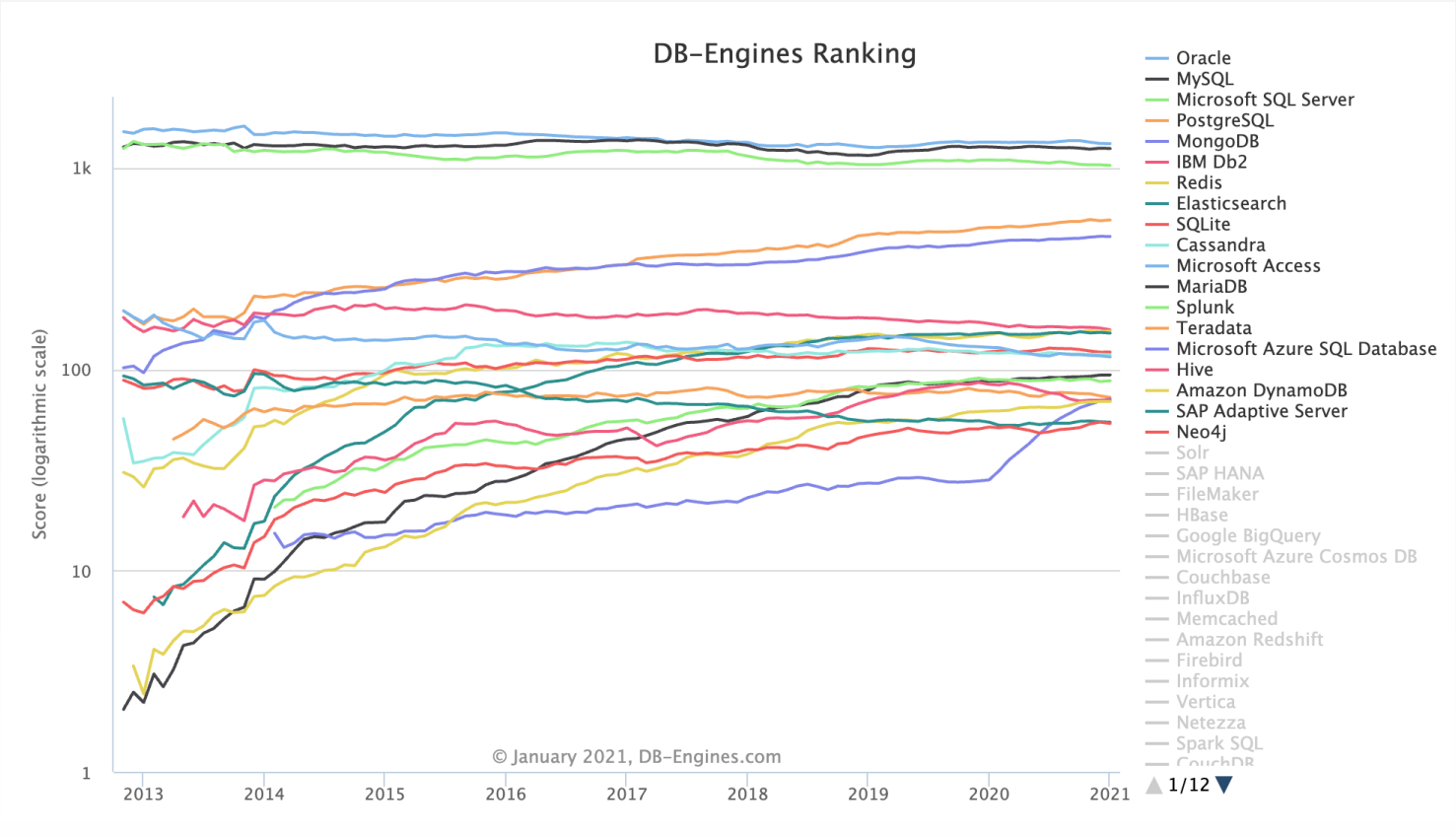
SQLite
ORACLE
PostgreSQL
db isam
TURBODB
Embedded
Microsoft
SQL Server
Microsoft
SQL Azure
elevatedb
MySQL

**RDBMS**

Redis
Neo4j
the graph database
Cassandra
HBASE
CouchDB
relax
mongoDB
RAVENDB
membase
riak
elasticsearch.
ORACLE
Coherence
Apache
Solr

**NoSQL**

# DB-Engines Ranking - Trend Popularity




VIRGINIA TECH

# RDBMS

| Commercial | Open source |
| --- | --- |
| | |
| | |
| | |

# SQLite

- Most popular embedded DB in the world
  - iPhone (iOS), Android, Chrome….
- (Very) Easy to use: no need to set it up
- Self-contained: data + schema
- DB on your laptop: useful for testing, understanding….

# NoSQL Databases

- Wide column: Cassandra, Hbase, etc
- Document: Apache CouchDB, MongoDB, etc
- Key–value: Apache Ignite, Berkeley DB, Amazon DynamoDB, MemcacheDB, Redis, etc
- Graph: Apache Giraph, Neo4J, Virtuoso, etc
- In-memory: Memcached, Redis, etc
- Search: Elasticsearch, Apache Solr, etc

# Cloud Databases

- Amazon Web Service (AWS)
- Microsoft Azure
- Google Cloud Platform
- Oracle Database
- IBM DB2
- MongoDB Atlas
- …and more

# Things to do This Week

- Check Syllabus, Canvas, and GitHub
- Mark exam dates in your calendars
- Mark Spring break dates in your calendars
- Do University online courses
- Create a GitHub account if you don’t have one
- Finish student survey if you didn’t. https://tinyurl.com/yxwm5qye
- Ask questions if you have
- Office hours start next week

# Reading and Next Class

- Next class:
  - Introduction to database systems
  - Architecture & Classification
  - Reading: Ch1, Ch2