Today ...

- Survey
- Course Overview
- Intro to database systems

Homework

- HW1 (out)
Course Overview

Course webpage

- [http://www.cs.gonzaga.edu/bowers/courses/cpsc321](http://www.cs.gonzaga.edu/bowers/courses/cpsc321)
- Be sure to check frequently (notes, readings, assignments, etc.)

Course discussion forum

- Piazza ...

Please be sure to **carefully** read the syllabus

- **Office hours:** Mon 11-12, Wed 1-3, Thur 12:30-1:30, or by appointment
- **Grading:** 30% homework, 10% group project, 15% quizzes, 45% three midsemester exams
- **Reading:** You are expected to do the reading assignments
- **Homework late policy:** Hand in during class period due for credit (no exceptions)
- **Missing class:** more than 4 absences results in a V grade

Textbook:

Academic Integrity

- Quizzes and exams must be your own work
- You are responsible for learning and understanding course content
- You will be tested on material from assignments (reading & hw)
- Okay to work with others on homework (but see above)
- Must turn in homework individually

Quizzes

We’ll have frequent quizzes

- Cannot “make up” a quiz
- Given at the beginning of class (so don’t be late!)

Why quizzes?

- Give you a feel for the types of questions I ask on exams
- Give you feedback on your understanding of material
- Give me feedback on your understanding

Be sure to study for quizzes

Homework

Assignments ...

- Mainly SQL and DB-design assignments, a few written assignments
• Reading is part of most homework assignments
• Larger group project (implement end-to-end database application)
• Assignments due in class on the due date!
Course Topics

This course covers various aspects of database management systems

- Relational database systems and SQL
- Logical database design (ER, normalization)
- Database internals (storage, indexes, optimization)
- Physical database design
- Transactions and recovery (as time allows)

Goals of the course

1. Know when and how to use database systems
2. Read and write SQL
3. Design databases, including basic database optimization
4. Use database systems within applications (basics)
5. Gain hands-on experience with a DBMS (MySQL) and embedded SQL
In-class exercise

Your first database: Storing registrar data

Instructions

1. Organize into groups of 3–4. Select one person to present your answers.

2. Pick 2 types of information (besides “First Name” and “Last Name”) that a registrar’s office would store for every student.

3. Pick 2 types of information that would (most likely) not be stored in a registrar’s database (e.g., “Favorite Movie”).

4. Create a “table” where each column contains a type of information and each row contains the corresponding values of the column for a particular student. Your columns should include “First Name”, “Last Name” as well as “Hometown”, and your answers above.

5. Fill in your table with one row for each person in your group.
Why study database management?

Obvious answer: Data is important!

• E.g., to business, science, government, etc.
  – Critical to existence of many companies

• Important skills for computer scientists / developers today
  – Many tech companies built on managing data
  – E.g., as a business asset: Google, Amazon, Facebook, Twitter, ...
  – Or as a business product: Oracle, Microsoft, IBM, SAP, ...

Database systems span many areas in CS

• File, memory, process management (Operating Systems)
• Languages, algorithms, complexity (Theory)
• Information modeling, formal logic (Artificial Intelligence)
• Application development (Software Engineering)
• Optimization, concurrency, distribution (Systems Programming)
Database management systems

We’ll focus on **Relational Database Management Systems**

- Abbreviated RDBMS or just DBMS
- Data stored in structured **relations** (i.e., tables)
- **Structured Query Language** – SQL – to manage and retrieve (query) data
  - Pronounced “es que el”
  - Sometimes as “sequel” (SQL’s “predecessor” …)

Some characteristics of a typical (R)DBMS:

- Data is **persistent**
  - DBMS can be stopped and restarted without losing data
  - Data stored across applications
  - Requires reliability
- Data accessed by **many users**
  - Data accessed concurrently by many users
  - E.g., flight reservations, product catalogs, facebook pages
  - Requires safe “transactions”
- Data impractical or inefficient to manage **in-memory**
  - Alternative is to manually manage data using files
  - But done in an *ad hoc* way (each application different)
  - Requires efficiency
- Data sets may be really **large**
  - Some examples (dated and approximate)
- Walmart’s data warehouse stores > 30 petabytes of data (2015)
- Bank of America’s warehouse stores > 170 petabytes of data (2016)
- Facebook’s warehouse stores > 300 petabytes of data (2014)
- Twitter stores > 500 petabytes of data (2017)
- Google stores 10–15 exabytes of data

1 petabyte ≈ 1,000,000,000,000,000 bytes

1 exabyte ≈ 1,000 petabytes
Some Basic Terminology

“Database”

- A database (DB) is a (structured) collection of persistent data
- In a relational DBMS, a database is a set of tables

“Database Management System”

- A database management system (DBMS) is software that supports the definition, population, and query of databases
- MySQL, PostgreSQL, MS SQLServer, Oracle (Database), and IBM DB2 are some examples of relational DBMSs
Basic DBMS Architecture

We’ll cover each these throughout the course