CS1555/2055

Database Management Systems

Instructor: Panos K. Chrysanthis
Recitations: Constantinos Costa
TA: Xioazhong Zhang

Fall 2019 (2201, 20-1)
db.cs.pitt.edu/courses/cs1555/current.term

Why CS1555/2055?

❑ **Top reasons to take CS1555/2055**

- My friends are taking it too
- Works with my schedule

- I wanna know how database systems work
- I wanna do research in database systems
- I want a job in/with database systems / data science

Some Recent Salary Numbers

	Database Administrators	Database administrators (DBAs) use specialized software to store and organize data, such as financial information and customer shipping records. They make sure that data are available to users and secure from unauthorized access.	Bachelor's degree	$87,020
	Information Security Analysts	Information security analysts plan and carry out security measures to protect an organization's computer networks and systems. Their responsibilities are continually expanding as the number of cyberattacks increases.	Bachelor's degree	$95,510
	Network and Computer Systems Administrators	Computer networks are critical parts of almost every organization. Network and computer systems administrators are responsible for the day-to-day operation of these networks.	Bachelor's degree	$81,100
	Software Developers	Software developers are the creative minds behind computer programs. Some develop the applications that allow people to do specific tasks on a computer or another device. Others develop the underlying systems that run the devices or that control networks.	Bachelor's degree	$103,560
	Web Developers	Web developers design and create websites. They are responsible for the look of the site. They are also responsible for the site's technical aspects, such as its performance and capacity, which are measures of a website's speed and how much traffic the site can handle. In addition, web developers may create content for the site.	Associate's degree	$67,990

[US Department of Labor: Friday, April 13, 2018]

Databases everywhere

"Big" Data is an important problem
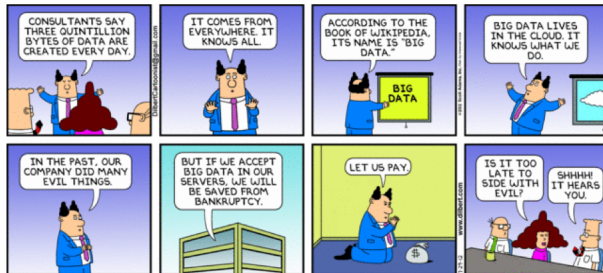
- You know Big Data is an *even more* important problem if... It has a Dilbert cartoon!

Focus of the CS1555/2055 course

- CS1555/2055: Application-oriented study of databases
  - Introduce fundamental concepts of data management
  - Design and use of a database system
  - Provide practical experience in applying these concepts using commercial DBMS

- CS1655: Advanced topics of databases
  - XML, XML Query Optimizations, applications & Web

- CS1656: Introduction to Data Science
  - Different data manipulation & data analysis techniques beyond traditional data management

- CS2550: System-oriented study of databases
 - Design of database management system

Quick Facts - Lectures

- When: Tue & Thu 2:00 pm – 3:15 pm

- Where: CL 208B

- What:
  - *Lecture Notes*
  - *Fundamentals of Database Systems*, R. Elmasri and S. B. Navathe, 7th edition, 2015
  - *Oracle10g Programming: A Primer*, R. Sunderraman, Addison-Wesley, 2007
  - *Learning SQL*, Alan Beaulieu, O'Reilly 2009 (Available online from campus computers through Safari Bookshelf)

Recitations

- When & Where
  - Friday, 11:00 - 11:50 am
  - Friday,   2:00 - 2:50 pm
  - @ 6110 Sennott Square Building

- What: **Complements** & **goes beyond the lectures**
  - solve problems & homework,
  - study more Examples,
  - practice with PostgreSQL, Oracle
  - it is required... attendance and participation in lecture and recitation may be used to decide borderline grades.

Recitation Attendance Vs. Scores
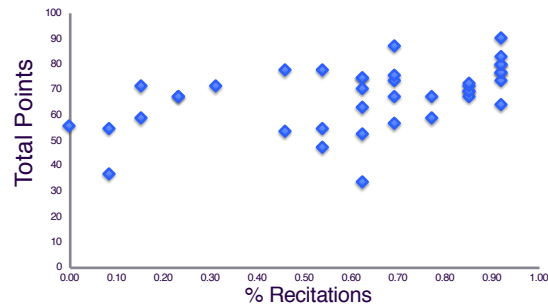
Assessments - Grading

Assessment	Percentage	Dates
Homeworks	20%	Multiple
Term Project	15%	3-phases Demos: Dec 11-13, 2019
Midterm Exam	30%	Oct. 22, 2019
Final Exam	30%	Dec. 10, 2019
Participation (Class & Recitations)	5%	

Class Notes

❑ web page: check often!

  http://db.cs.pitt.edu/courses/cs1555/current.term

  http://db.cs.pitt.edu/courses/cs2055/current.term

❑ Credentials:
  ▪ db1555.20-1
  ▪ dbest&Fun

Meta-notes

● These notes are intended for use by students in CS1555 at the University of Pittsburgh. They are provided free of charge and may not be sold in any shape or form.
● These notes are incomplete and NOT a substitute for material covered during course lectures. If you miss a lecture, you should definitely obtain both these notes and notes written by a student who attended the lecture.
● Material from these notes is obtained from various sources, including, but not limited to, the following:
  ○ Fundamentals of Database Systems by Elmasri and Navathe
  ○ Various online resources (see notes for specifics)
  ○ Improvements of instructor's notes by colleagues (M. Sharaf, N. Farnan, C. Costa)

Communication

- class email: **`cs1555-staff@cs.pitt.edu`**
  - For *confidential* matters only
  - use keyword **`cs1555`** in all emails to instructor/TA (as part of the subject line)
  - it works *only* within pitt.edu
- Piazza:
  - for all clarifications to lectures, recitations and assignments
- assignments:
  - To be submitted electronically
  - **No** Piazza or email clarifications 4 hours prior a deadline

---

Piazza Guidelines

- Remember that **everything you post is public**
- Please do not post any assignment code on Piazza, even as a private message. Visit office hours for any code related questions.
- Keep posts on Piazza course-related.
- Please read all questions and responses that are on Piazza before asking a question. Utilize the excellent Piazza search facilities.
- Use a meaningful subject heading.
- Make sure your questions/posts are in the appropriate folder (e.g., hw0, hw1, hw2, ...).
- Tag your post with all the applicable tags.
- Please don't post things to the group that give no useful information.
- Please keep complaints about the course out of the newsgroup. If you have a concern about anything to do with the course, please talk to the instructor.
- **Please be respectful of your peers and others in your posts.**

Overview of Database Management Systems

What is a Database ?

4

What is a database?

- A very large, integrated collection of **related** data
  - data is raw facts on some aspect of the world
- Models a real-world enterprise (e.g., university)
  - **Entities** (e.g., students, courses)
  - **Relationships** (e.g., Bob took CS 1550)

Students

SID	Name	Age	GPA
546007	Susan	18	3.8
546100	Bob	19	3.65
546500	Bill	20	3.7

Courses

CID	CName
CS 1555	DB
CS 1530	SW
CS 1550	OS

Enrollment

SID	CID	Grade
546007	CS 1550	A
546007	CS 1530	B+
546100	CS 1550	B

Integrated Data

- All data are stored and manipulated in a *uniform way* on a secondary storage

  - Databases store large amounts of data that cannot fit in main memory.
  - Data are stored for long and indefinite period
  - Data are shared across multiple applications

What is a Database Management System?

- ***Database Management System (DBMS):***
  - A general purpose software package designed to store and manage databases *conveniently & efficiently*
- **DBMSs:**
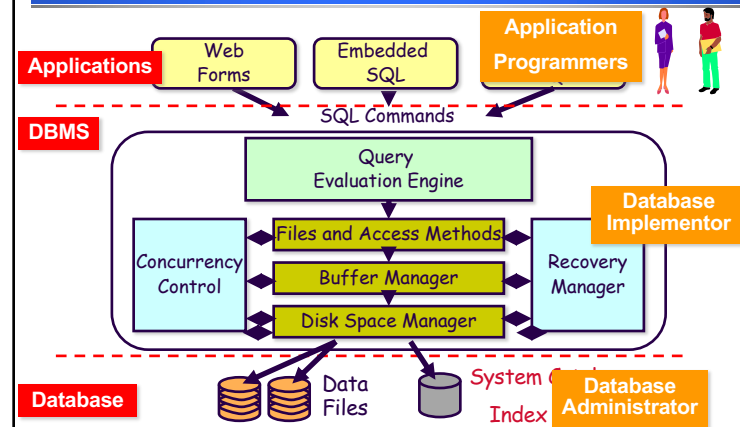 - Oracle, IBM DB2, SQLServer, MySQL, PostgreSQL …
- **Usage:**

  Database system = *DB* + *DBMS* + *Application Logic*
  - Resource Planning Applications:
    - PeopleSoft, SAP, …
  - Web-based Applications:
    - amazon, ebay, orbitz, trip-advisor, yelp, …

Database Management System (DBMS)

Applications
- Web Forms
- Embedded SQL
- **Application Programmers**

SQL Commands

DBMS
- Query Evaluation Engine
- Concurrency Control
- Files and Access Methods
- Buffer Manager
- Disk Space Manager
- Recovery Manager
- **Database Implementor**

Database
- Data Files
- System Catalog
- Index
- **Database Administrator**

5

Approaches to management of data

❑ File system approach
- Traditional (flat) files +
  C (Java, ...) programs to access them
- E.g., use one (or more) UNIX/DOS files, with student records and their courses
- Decide on a layout for the student records, etc.

❑ Database approach

File Approach to Data Management

- Write everything to text files!
  - Have your applications directly read/write to these files
- Problems?
  - Slow!
  - Have to constantly enforce layout of data
  - What if multiple instances of your application need to read/write to the same file at the same time?
  - What if one of them crashes?

Database Vs. File Systems Approaches

❑ Abstraction
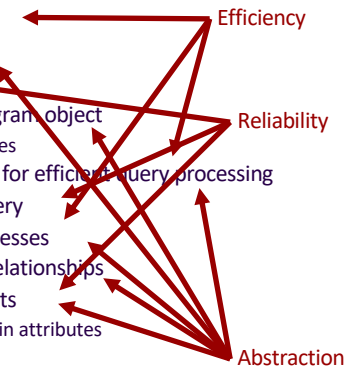- Data
- Execution

❑ Reliability

❑ Efficiency/Performance


How Are these Achieved?

- Control over redundancy
  - Via *data normalization*
- Access controls
- Persistent storage for program object
  - Object-oriented databases
- Structures and techniques for efficient query processing
- Provide backup and recovery
- Provide for concurrent accesses
- Represent complex data relationships
- Enforce integrity constraints
  - e.g., uniqueness of certain attributes

Efficiency

Reliability

Abstraction

6

Topics for the Term

1	Relational Model
2	DDL
3	Normalization
4	ER Modeling
5	Relational Algebra
6	DML & VDL
7	Access Control
8	Transactions
9	Integrity Constraints
10	Database Programming
11	File Storage & Indexing
12	Query Processing
13	Concurrency Control
14	Recovery

Theory

Systems

SQL/Programming

Database Management System (DBMS)
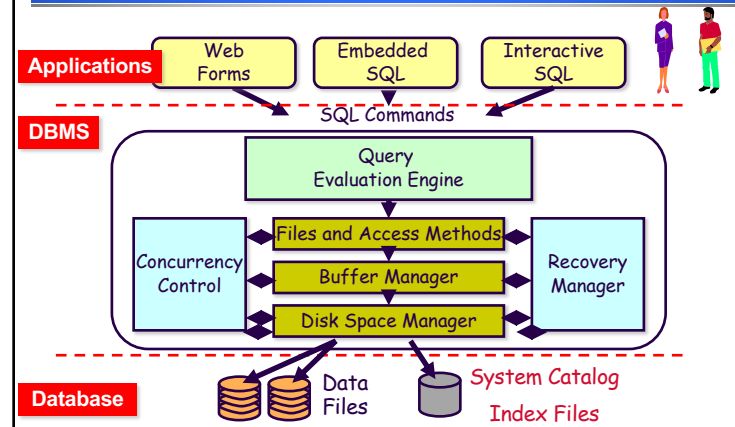
Agenda for 2nd Class

- ❑ Introduction to Relational Model
- ❑ Introduction to the used DB Technologies
  - ▪ PostgreSQL
  - ▪ DataGrip IDE

- ❑ Please bring your laptops to do the installation of these software together during class.

7