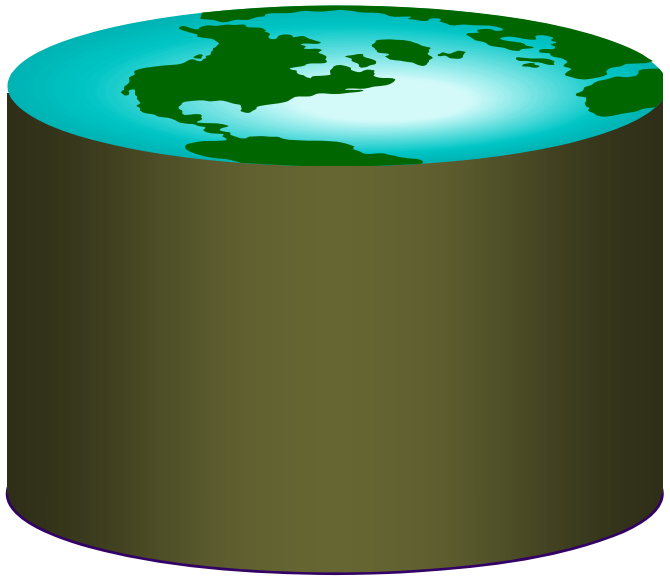
CS 405G: Introduction to Database Systems

Lecture 1: Introduction



Topics for Today

- Topics
  - Introduction
  - What is a database?
  - What is a database management system?
  - Why take a database course?
  - How to take the class?

Who am I?

- Instructor
  - Jinze Liu
  - Associate professor in the department of Computer Science at UK
  - Research area: Data mining and Bioinformatics
  - Office: 235 Hardymon building
  - Email: liuj@cs.uky.edu

About the course: Information

- Class web page is at
  - http://protocols.netlab.uky.edu/~liuj/teaching/CS405G_F13/
  - Syllabus, homework, grading policy, etc. available from class web page
- Textbook
  - Fundamentals of database systems (sixth edition)
  - Ramez Elmasri and Shamkant B. Navathe
  - Can get it from the bookstore
- My Office Hours:
  - 235 Hardymon building, Wednesday 10AM-12PM
  - Email: please include CS405G in the subject line for fast response
- Class mailing list
  - Will send emails to everyone once set up.
  - Will be used for announcement/clarification of assignments/answering questions

About the Course – Workload

- 6 homework assignments
  - Including programming assignment
  - Building blocks for your project
- 4 quizzes
- 1 Programming project
- Exams
  - 1 Midterm & 1 Final
- Cheating policy: zero tolerance

About the course: Grading

- Weights
  - 6 Homework assignments 25%
  - Project 20%
  - Midterm exam 25%
  - Final exam 25%
  - Quizzes 5%

- More information is in the syllabus
  - Final grade
  - Late homework
    - Will be penalized
  - Academic mis-conduct
    - You are expected to do the assignment independently
    - Discussions if allowed should be acknowledged

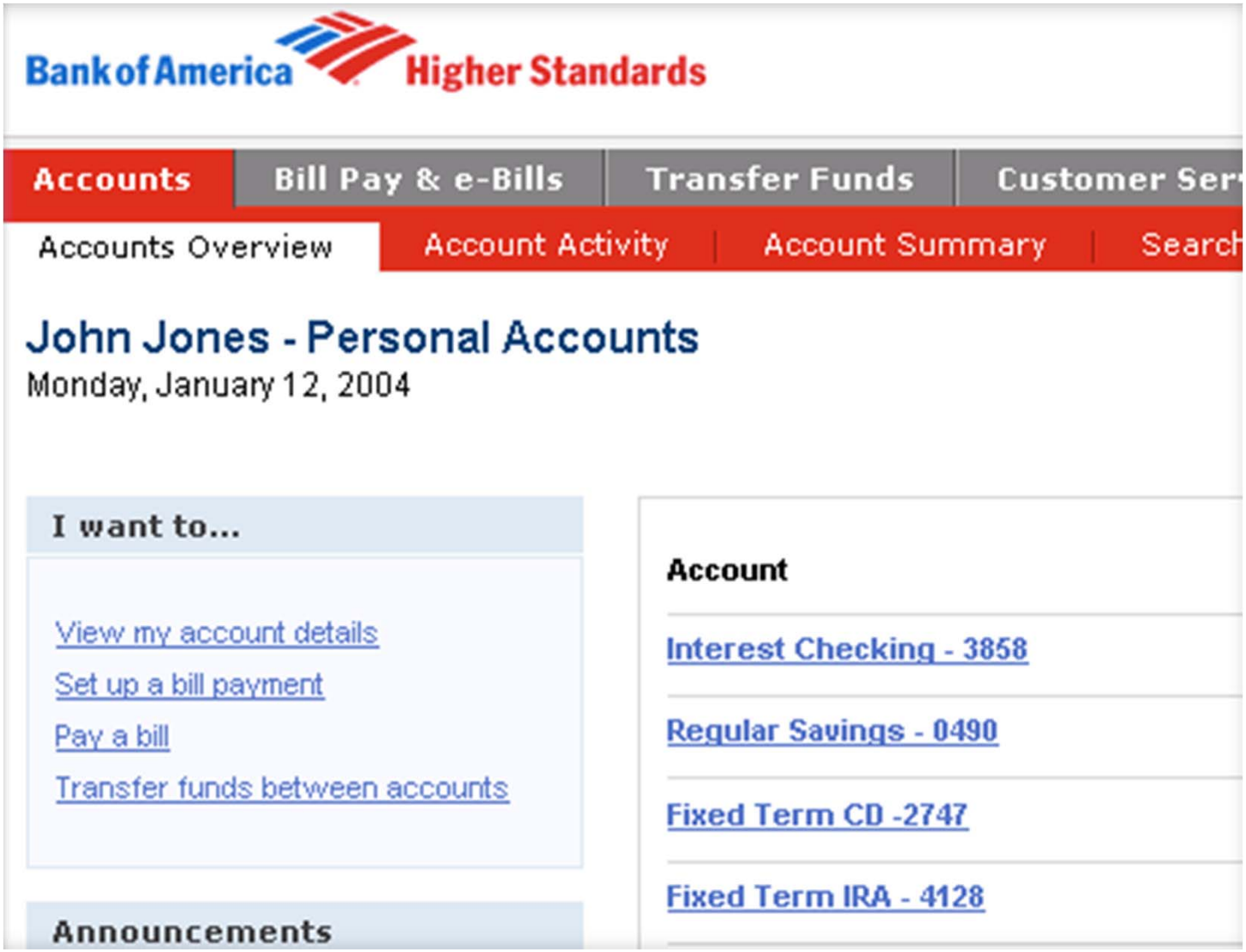
About the course: Project

- Programming projects have a practical, hands-on focus:
  - A relational DBMS for a particular application
  - Projects are to be done in teams of 2
  - Pick your partner ASAP!

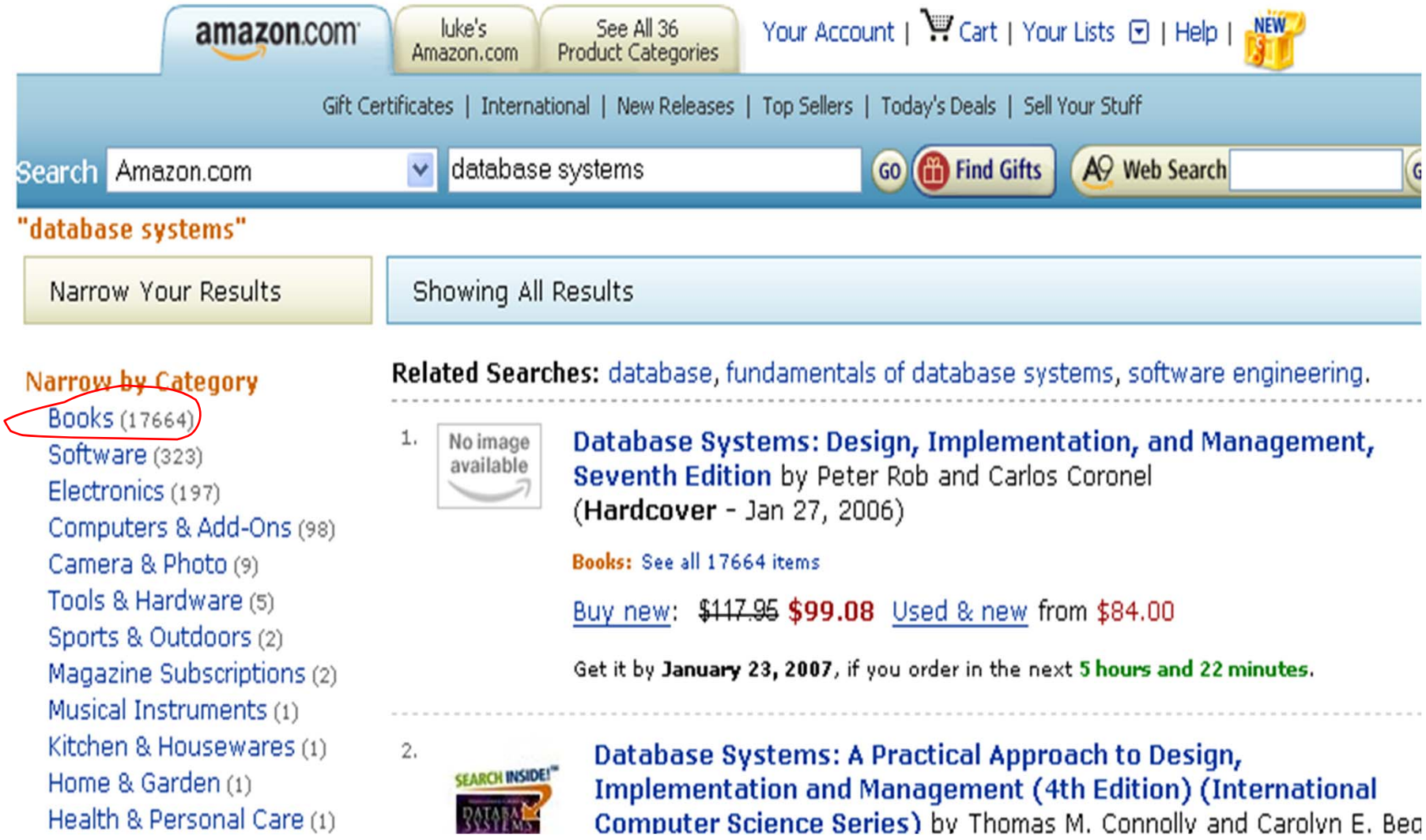
Database Systems?

- Name a few!

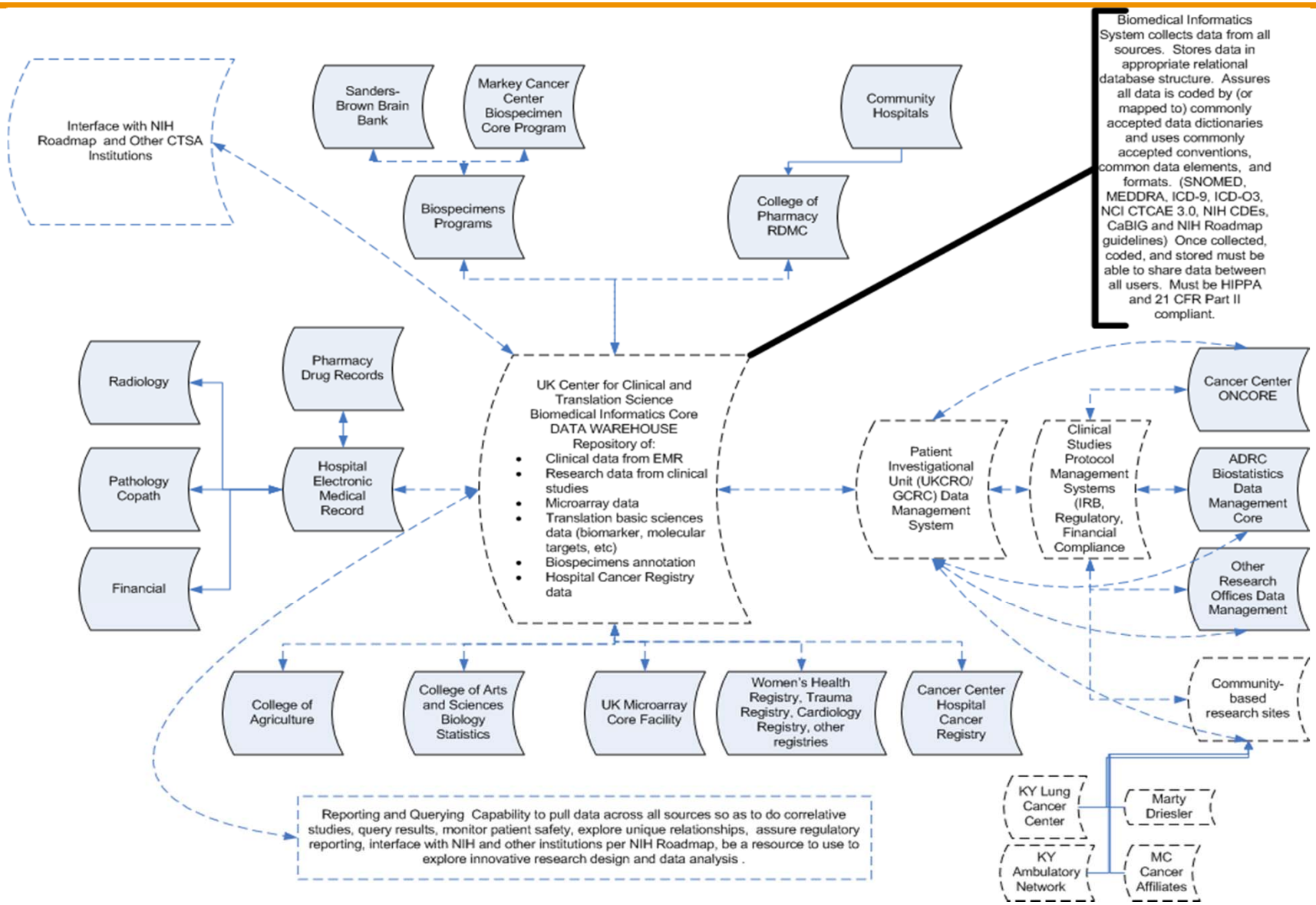
Database Systems: Bank Systems



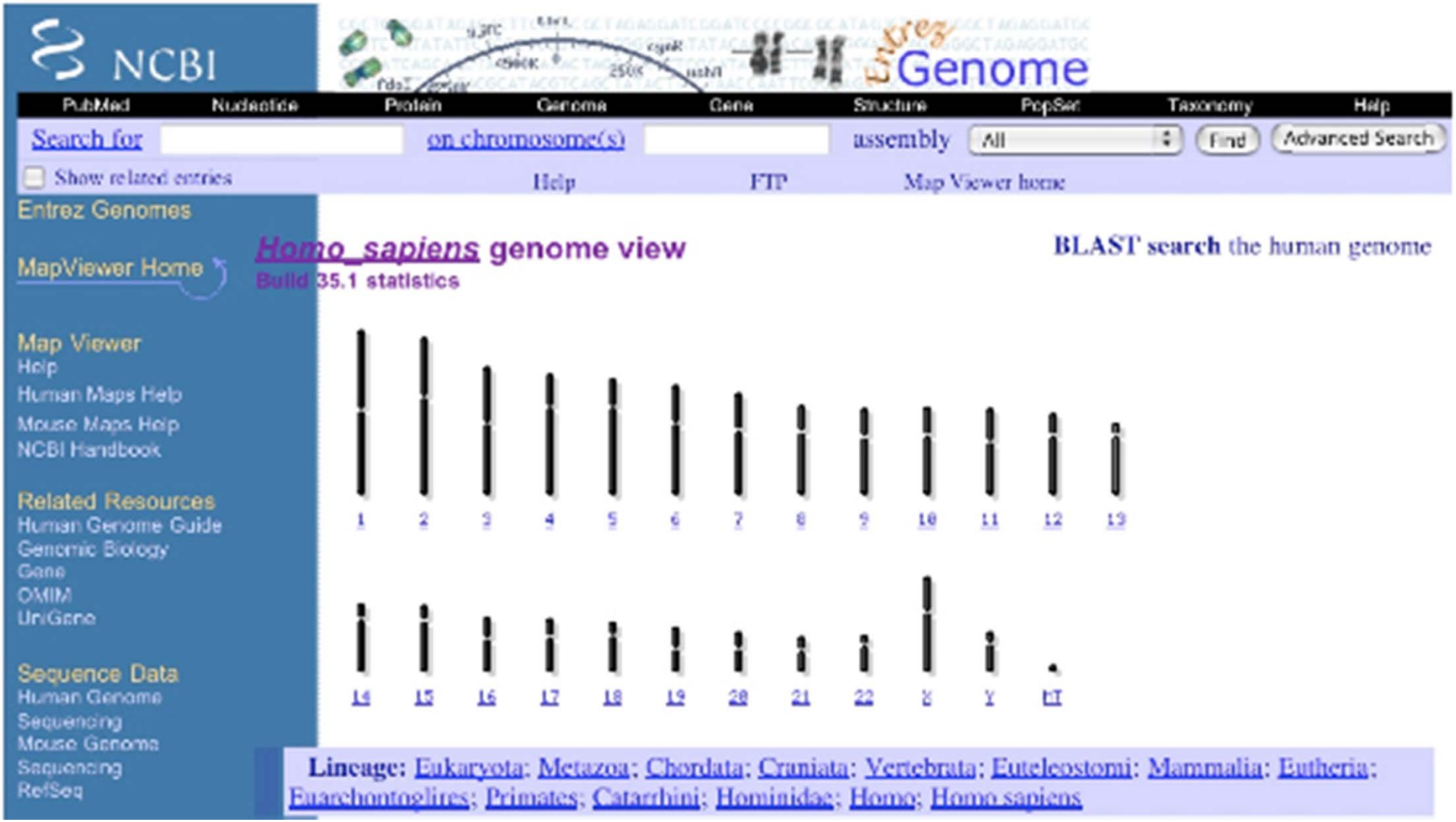
Database Systems - Ecommerce



Database Systems: Clinical Databases



Database Systems: Genome Bank



What is a Database?

- A *database* is an integrated collection of data.
  - Data is a group of facts that can be recorded.

- Typically a database is used to model a real-world "enterprise" (or a *miniworld*)
  - Entities (e.g., *basketball teams*, *games*)
  - Relationships (e.g. *UK's basketball team* beat ? yesterday)

What is a Database Management System?

- A Database Management System (DBMS) is a collection of programs that enable users to create and maintain databases
  - store, manage, and access data in a databases.

- Typically this term is used narrowly
  - Relational databases with transactions
    - E.g. Oracle, DB2, SQL Server
  - Mostly because they predate other large repositories
    - Also because of technical richness
  - When we say DBMS in this class we will usually follow this convention
    - But keep an open mind about applying the ideas!

Why take this class?

A. Database systems are at the core of CS
B. They are incredibly important to society
C. The topic is intellectually rich
D. ~~It isn't that much work~~
E. Looks good on your resume

Let's spend a little time on each of these

Why take this class?

A. Database systems are the core of CS

- Shift from computation to information
  - True in corporate computing for years
  - Web made this clear for personal computing
  - Increasingly true of scientific computing
- Need for DB technology has exploded in the last few decades
  - Corporate: retail swipe/clickstreams, “customer relationship mgmt”, “supply chain mgmt”, “data warehouses”, etc.
  - Web: not just “documents”. Search engines, e-commerce, blogs, wikis, other “web services”.
  - Scientific: digital libraries, genomics, satellite imagery, physical sensors, simulation data
  - Personal: Music, photo, & video libraries. Email archives. File contents (“desktop search”).

Why take this class?

B. DBs are incredibly important to society

- "Knowledge is power." -- Sir Francis Bacon

- "With great power comes great responsibility." -- SpiderMan's Uncle Ben

Policy-makers should understand technological possibilities.
Informed Technologists needed in public discourse on usage.

Why take this class?

C. The topic is intellectually rich.

- representing information
  - data modeling
- languages and systems for querying data
  - complex queries & query semantics*
  - over massive data sets
- concurrency control for data manipulation
  - controlling concurrent access
  - ensuring transactional semantics
- reliable data storage
  - maintain data semantics even if you pull the plug
- data mining
  - Let your data speak

* semantics: the meaning or relationship of meanings of a sign or set of signs

Why take this class?

~~**D. It isn't that much work.**~~

- Bad news: It is a lot of work.

- Good news: the course is front loaded
  - Most of the hard work is in the first half of the semester
  - Load balanced with most other classes

Why take this class?

E. Looks good on my resume.

- Yes, but why? This is not a course for:
  - Oracle administrators
  - IBM DB2 engine developers
    - Though it's useful for both!
- It is a course for well-educated computer scientists
  - Database system concepts and techniques increasingly used "outside the box"
    - Ask your friends at Microsoft, Yahoo!, Google, Apple, etc.
    - Actually, they may or may not realize it!
  - A rich understanding of these issues is a basic and (un?)fortunately unusual skill.