Introduction to Database Systems: CS380
An Overview of Databases

Oliver Bonham-Carter

19 Jan 2018
Let’s discuss...

What is the function of a database?
Database Management System (DBMS)

Database Applications:
- Banking: transactions
- Airlines: reservations, schedules
- Universities: registration, grades
- Sales: customers, products, purchases
- Online retailers: order tracking, customized recommendations
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions
Application:
- Adding new students, instructors, and courses
- Registering students for courses, and generate class rosters
- Assigning grades to students,
- Computing grade point averages (GPA)
- Generating transcripts
To connect data in some meaningful way
Databases were not always used as they are typically applied now...

Disconnections in an organization

Such as, multiple silos of data - separate data-holding entities
Meanwhile at *Batman and Associates*...

One working databases containing all company information.
Batman and Associates takes on another DB

- Two teams and two working databases
- *Nearly identical* information.
- Not-consistently updated
Two new teams at *Batman and Associates*...
Batman and Associates confused about client’s address. What went wrong?! :-(

1st Address: 221b Baker Street
London, England

2nd Address: 1600 Pennsylvania Avenue,
Washington, DC

Dr. Vader’s actual address?
Database applications were once built directly on top of file systems (i.e., permanent records of information in various spread-out files)

- A firm may have managed several information sources!
- Anything wrong with this picture?
Messy Data Storage?
Risks of having several different DBs for same information

- **Drawbacks of using file systems to store data**
  - **Data redundancy and inconsistency**
  - Multiple file formats, duplication of information in different files - hard to share
  - Difficulty in accessing data
  - Need to write a new program to carry out each new task
  - **Data redundancy and inconsistency**
  - Data isolation - multiple files and formats
Drawbacks of using file systems to store data, continued

- Data redundancy and inconsistency
- Integrity problems - The data is not accurate and consistently stored
- Data redundancy and inconsistency
- Integrity constraints (i.e., account balance $\geq 0$) become “buried” in program code rather than being stated explicitly
- Hard to add new constraints or change existing ones

How many times did you see, “Data redundancy and inconsistency”, here??
Messy Data Storage?

- **Atomicity of updates**
  - Failures may leave database in an inconsistent state with partial updates carried out
  - Example: Transfer of funds from one account to another should either complete or not happen at all

- **Concurrency**: transactions at the same time
  - Access by multiple users
  - Access needed for performance - no slow-down
  - Uncontrolled concurrent accesses can lead to inconsistencies
    - Example: Two people reading a balance (say 100) and updating it by withdrawing money (say 50 each) at the same time
Messy Data Storage?

- Security problems
  - Hard to provide user access to some, but not all, data
  - Difficult to control how data is used
- Database systems offer solutions to all the above problems
Video: Finding Trends in Data

- Hans Rosling’s 200 Countries, 200 Years, 4 Minutes - The Joy of Stats
- https://www.youtube.com/watch?v=jbkSRLYSojo
1. What kind of data was likely being used for this study?
2. Where could a database serve in this project?
3. How was the data likely organized for the study?
4. What kinds of relationships were likely used to connect data points?
5. How could this study be done in absence of a database?