Introduction to Data Management

Lecture #1
(The Course “Trailer”)

Instructor: Mike Carey
mjcarey@ics.uci.edu
Today’s Topics

- Welcome to my biggest class ever!
- Read (and live by!) the course wiki page:
- Also follow (and live by) the Piazza page:
  - [https://piazza.com/uci/spring2018/cs122a/home](https://piazza.com/uci/spring2018/cs122a/home)
- Let’s peek at the wiki page, and then let’s also preview what lies ahead...
- **Note:** There will be a quiz in this week’s initial discussion sessions...!
  - **Note:** You must attend the one you registered for.

What is a Database System?

- What’s a *database*?
  - A very large, integrated collection of data
- Usually a model of a *real-world enterprise*
  - *Entities* (e.g., students, courses, Facebook users, ...)
    - with attributes (e.g., name, birthdate, GPA, ...)
  - *Relationships* (e.g., Susan is taking CS 234, Susan is a friend of Lynn, ...)
- What’s a *database management system* (DBMS)?
  - A software system designed to store, manage, and provide access to one or more databases
**File Systems vs. DBMS**

- Application programs must sometimes *stage large datasets* between main memory and secondary storage (for buffering huge data sets, getting page-oriented access, etc.)
- *Special code needed* for different queries, and that code must be (stay) correct and efficient
- Must *protect data from inconsistency* due to multiple concurrent users
- *Crash recovery* is important since data is now the currency of the day (corporate jewels)
- *Security and access control* are also important(!)

**Evolution of DBMS**

<table>
<thead>
<tr>
<th>Files</th>
<th>CODASYL/IMS</th>
<th>Relational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Coding</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
  - Byte streams  |
  - Majority of application development effort goes towards building and then maintaining data access logic  |
| Early DBMS Technologies  |
  - Records and pointers  |
  - Large, carefully tuned data access programs that have dependencies on physical access paths, indexes, etc.  |
| Declarative approach  |
  - Tables and views bring "data independence"  |
  - Details left to system  |
  - Designed to simplify data-centric application development  |

Database Management Systems 3ed, R. Ramakrishnan and J. Gehrke
Why Use a DBMS?

- Data independence.
- Efficient data access.
- Reduced application development time.
- Data integrity and security.
- Uniform data administration.
- Concurrent access, recovery from crashes.

Why Study Databases?

- Shift from computation to information
  - At the “low end”: explosion of the web (a mess!)
  - At the “high end”: scientific applications, social data analytics, ...
- Datasets increasing in diversity and volume
  - Digital libraries, interactive video, Human Genome project, EOS project, the Web itself, ...
  - Mobile devices, Internet of Things, ...
  - ... need for DBMS exploding!
- DBMS field encompasses most of CS!!
  - OS, languages, theory, AI, multimedia, logic, ...
Why Study Databases (Really)?

Big Data! 😊

To Be Continued...