Why study databases?

- Interesting concepts and techniques.
- Spans computer science, including OS, languages, theory, AI, multimedia, logic.
- Databases have become increasingly important
  - shift from a focus on computation to information
  - data increases in volume and diversity.
- Jobs: In demand and well paid.
- Research: Many open problems.
Our first hour or so

- Some key concepts
- Examples to motivate the course
- Admin info
Databases and DBMSs

- Databases are everywhere, often behind the scenes.

- DBMS (Database Management System): “A powerful tool for creating and managing large amounts of data efficiently and allowing it to persist over long periods of time, safely.” [Ullman and Widom, FCDB]

- Database: a collection of data managed by a DBMS.
Data models

- Every DBMS is based on some data model: a notation for describing data, including
  - the structure of the data
  - constraints on the content of the data
  - operations on the data

- Some specific data models:
  - network & hierarchical data models — of historic interest
  - relational data model
  - semistructured data model
The relational data model

- Main concept is a “relation.”
  Based on the concept of relations in math.
- Can think of as tables of rows and columns.

<table>
<thead>
<tr>
<th>Teams</th>
<th>Name</th>
<th>Home Field</th>
<th>Coach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rangers</td>
<td>Runnymede CI</td>
<td>Tarvo Sinervo</td>
</tr>
<tr>
<td></td>
<td>Ducks</td>
<td>Humber Public</td>
<td>Maeve Mahar</td>
</tr>
<tr>
<td></td>
<td>Choppers</td>
<td>High Park</td>
<td>Tom Cole</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Games</th>
<th>Home team</th>
<th>Away team</th>
<th>Home goals</th>
<th>Away goals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rangers</td>
<td>Ducks</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Ducks</td>
<td>Choppers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Rangers</td>
<td>Choppers</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Choppers</td>
<td>Ducks</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
Example ...

- A dataset scraped from Twitter
- Defining a schema that expresses its structure
- Creating an instance that contains the data
- Writing some queries on the data
What a DBMS provides

- Ability to specify the logical structure of the data
  - explicitly
  - and have it enforced
- Ability to query or modify the data.
- Good performance under heavy loads (huge data, many queries).
- Durability of the data.
- Concurrent access by multiple users/processes.
Overall architecture of a DBMS

- The DBMS sits between the data and the users or between the data and an application program.
- Within the DBMS are layers of software for:
  - parsing “queries”
  - implementing the fundamental operations
  - optimizing queries
  - maintaining indices on the data
  - accessing the files that store the data and indices
  - management of buffers
  - management of disk space
A “semi-structured” example ...

- An xml dataset scraped from imdb.com
- No schema required, no instance made
- We can immediately write queries on the data
- A much looser approach
What this course is about

- csc443 is about implementation of the DBMS itself
- csc343 is about *using* DBMSs:
  - defining schemas and instances
  - writing queries
  - connecting to code written in a general-purpose language
  - rigorous underlying principles