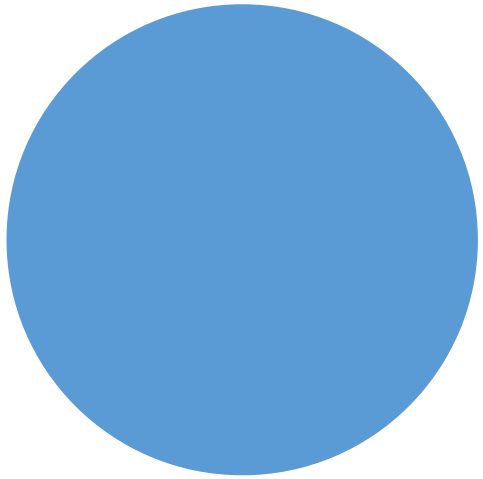
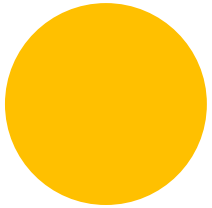
Databases

Leanne Wu

lewu@ucalgary.ca

Department of Computer Science

What databases do you use now?

Database

- A **Database Management System** (DBMS) is a general purpose system to manage data
- Can be accessed using a variety of methods
  - Programmatically using different languages (Java, C#, Python,…)
  - Directly using natively-provided languages and APIs
  - Indirectly using pre-built applications
- Used to isolate logical units of data from their physical representation

Physical/Conceptual/Logical Representations

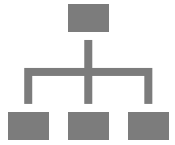
Physical

- Where is the data, on what computer, in what file?

Logical

~~Conceptual~~

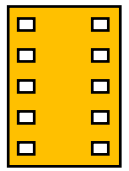
- How is this data organized in the system?



Conceptual

~~Logical~~

- How does this data look to your end users?



Why do we need databases?

- Operations
  - Writing
  - Reading
  - Updating
  - Deleting
- Transactional support
- Availability
- Performance
- Security

Physical Representations: Storage formats

Format	Description	Pros	Cons
Paper	Printed paper, punch cards	Long lifetime	Difficult to manage in volume
Tape	Reels of media (usually magnetic)	Long lifetime, durable	Writing speeds
Magnetic disk	Traditional HDDs	Cheap, read speeds	Lifetime, write speeds
Modern flash-based technologies	Solid-state drives	Fast	Expensive, lifetime

Kinds of databases

- Relational
  - SQL (Structured Query Language)
  - Relies on *relations*
    - Relations are essentially tables which have *attributes* (columns) and *records* (rows)
  - Certain formal guarantees can be provided about the state of the data:
    - how it will be retrieved
    - how to keep data consistent
    - how data is backed up and recovered
    - ACID (Atomic, Consistent, Isolated, Durable)
- Others
  - NoSQL
  - Graph
  - Stream
  - Temporal
  - …

NoSQL Databases

- Literally “No SQL”
  - Sometimes “Not Only SQL”
- Different types
  - key-value
  - document
  - column-oriented
- BASE (Basically Available, Soft state, Eventually Consistent)
- Operations

Other ways to classify databases

- Typical uses (workload)
  - Transactional (OLTP), Batch-oriented, Decision Support, Analytic
- Number of users
  - one, a few, a lot
  - users with specific access, public, open
- Where the computing is located
  - one location, distributed (homogeneous, heterogeneous), cloud
  - virtualized

What are we storing?

- Most relational databases depend upon some kind of metadata
- What kind of data might you need to store about your data?
    - does this work for a relational database?
    - does this work for a non-relational database?

Data management: Metadata