

## Databases & Information Systems

### What is this?

**Task:** Store, manage, & retrieve large amounts of data.

Sounds boring, eh? Surprisingly not. (Well, maybe.)

Change in emphasis from rest of CS.

**CS:** Using computers to compute things.  
(*programming*)

**DB:** Using computers to manage data.  
(*querying, ...*)

*Data-driven.*

## Databases & Information Systems

### Database: storage

- **store data (permanently)**

- Handle lots of data.  
*Petabytes?*
- Have a logical structure to the data.  
*schema design / modeling*
- Don't allow logical corruption of the data set.  
*integrity & transaction management*
- Be able to recover with no loss from practically any crash / disaster.  
*crash recovery*
- Let 1000's of users use the same database at the same time.  
*concurrency control & transaction management*
- Hide certain information from certain users.  
*security*

## Databases & Information Systems

### Database: models & design

- **data models**

What information does a given database store?

- Need a general *model* for what a DB looks like.

- \* O-O: classes and pointer

- \* relational: tables and constraints

What should the model be?

- **schemas**

The “format” for a given database is called a *schema*.

- **design:** What should the schema be, given a particular *domain* we want to store data about?

## Databases & Information Systems

### Database: queries

- **retrieve data**

How to find what you need out of a petabyte?

*Paradigm:* Ask the database a question (*query*).

DB will *answer* the query by composing the relevant data from the database.

- query language (**FRONT-END**)

Our interface to the DB.

- \* *very expressive: a subset of first-order logic*

- \* *declarative: I express what I want, but not how it is to be done.*

- \* *examples:*

- relational algebra & calculus

- query-by-example (QBE)

- SQL

- query language (**BACK-END**)

- \* *How to answer (evaluate) the question (query) efficiently and correctly?*

## Databases & Information Systems

### Database: updates

- **update data**

Can easily update the information in the database.

- **integrity**: Updates must respect the database's integrity.
- **consistency**: Updates should be consistent with respect to one another.
- **security / privacy**: Only certain users should be allowed to update certain data; only certain users should be allowed to read certain data.

## Databases & Information Systems

### Database Systems (p.1)

- **Database Management Systems**

What is a database *system*?

- Supports the creation / alteration / deletion of databases.
- Should it help with schema design?
  - \* *Supports and enforces schemas.*
  - \* *Does not really help with design.* (Sigh)

## Databases & Information Systems

### Database Systems (p.2)

- **Database Management Systems**

- Supports all the properties we want for our databases.
  - \* *permanence*
    - crash recovery
    - back-ups
  - \* *updates & transactions*
    - transaction management (consistency) & concurrency control
    - integrity checking
    - security / privacy
  - \* *queries*
    - a powerful, declarative query language
    - interface to programming languages

## Databases & Information Systems

### Why interesting?

This is a microcosm of CS itself!

- **Logic**
  - modeling / schema design
  - query languages
- **AI**
  - query evaluation & optimization
  - data mining
- **Theory**
  - complexity of queries
  - ...of query answering
- **Data structures & Algorithms**
  - databases = advanced data structures
  - new challenges for algorithms
    - \* E.g., Sort a list larger than main memory.
- **Programming & Software Design**
  - query languages
  - application programming
- **Systems / OS**
  - building database systems
    - \* resource management
    - \* software / hardware interface
- **Networks**
  - distributed information systems
  - Web

## Databases & Information Systems Areas

- Database systems (primarily *relational*)
- Information Retrieval
  - search engines
- Data Mining (Knowledge Discovery in Databases / KDD)
  - How to derive automatically interesting patterns (“knowledge”) from large databases.
- Information Systems at large
  - E-commerce Systems
    - \* *multi-tier architectures*
    - \* *focus: business solutions*
  - Web Technologies
    - \* *back-ended by database systems*
    - \* *data-driven*
    - \* *making information ubiquitous*
  - . . .

## Databases & Information Systems Curriculum at York

- Database systems
    - CSE-3421: Databases “Programming”
      - \* *database design*
      - \* *SQL*
    - CSE-4411: Database Systems “OS”
      - \* *database system architecture*
      - \* *query optimization*
  - Data Mining
    - CSE-4412: Data Mining
      - \* *data warehousing & OLAP*
      - \* *algorithms for data mining*
      - \* *machine learning*
  - E-commerce
    - CSE-4413: E-commerce Systems
      - \* *standards*
      - \* *multi-tiered systems*
      - \* *advanced application programming*
- 
- Background?
    - standard CS
    - logic!!
    - statistics & probability

## Databases & Information Systems Research

- How to build a better system?
  - handle more data (scaling up)
  - handle more complex queries
- How to support more applications?
  - more expressive query languages
  - more & better tools
- How to integrate into more complex systems?
  - integrating multiple databases
  - integrating into larger software systems

E.g., How to ask *best-match* queries in databases?  
How to include *preferences* in queries?

### Skyline Queries

```
select * from RestaurantRating
      skyline of Service max, Food max,
              Decor max, Price min
```

## Databases & Information Systems Sources

- ACM (Groups & Conferences)
  - SIGMOD
    - <http://www.acm.org/sigmod/>
  - KDD
    - <http://www.acm.org/sigkdd/>
- VLDB (Organization & Conference)
  - <http://www.vldb.org/>
  - Conference: 30 August – 3 September 2004,  
*Toronto!!*
- On-line Resources
  - DBLP Archive:
    - <http://www.informatik.uni-trier.de/~ley/db/>
  - DBWORLD mailing list
- Journals
  - TODS
  - IEEE TKDE
  - VLDB Journal
  - JIIS (Journal on Intelligent Information Systems)
  - . . .