**Databases..**

* Now, central component
* Manipulate data
* Use in many fields: Telco, IT etc etc

**Data is structured**

* Often structured
* Efficiency is becoming really important
* DBMS software is highly successful as a commercial technology
  + Oracle, DB2 MS SQL
* Made a lot contribution to CS
  + Banks, human genome, e-commerce

**Address book**

* Solution
  + Create text file
* Advantages
  + Easy to add and modify
  + Easily copied
  + Shareable
  + Substring searchable
  + Powerful, programmable tools
* But there can be complications

**Complication 1: File gets very large**

* Problem
  + Searching gets slow and imprecise
  + Search for “elm street” yields “Wilhelm streeter”
* Solution
  + Add indexes over fields commonly searched upon
  + Structure data into fields
    - Search for street=”elm street”

**Complication 2: Data Redundancy**

* Why?
  + Large families, drequent moves
  + Want space economy, single point of update
* Solution
  + Separate residences from names: 2 files, one for persons, one for residence
  + But how do we associate?
    - Consistency
    - Keys

**Complication3: multiple associations of persons and residences**

* Examples
  + Many-one, one-one etc

**Complication 4: Need to add information for new purposes**

* Requirements
  + Adding fields and/or new tables
    - **Schema evolution**

**Complication5: Doing ad hoc analysis and retrieval**

* Example
  + Implementation that performs analysis and retrieval correctly and effciency
    - **Query language**
    - **Query optimization and execution**

**Complication 6: Want to organize the data differently for some users**

* Concepts
  + **Joins**
  + **Views**
  + **Security**

**Complication 7: Required existence of associated data**

* Examples:
  + Can’t send in

**Complication 9: Multiple updates on all or none basis**

* Concepts
  + Transactions
  + Atomicity

**Database VS File System**

* Drawbacks on cb on DS
  + Transactions

http://vgc.poly.edu/~juliana/courses/BigData2014/Lectures/intro-to-db.pdf

**Relational data model**

* Page 33
  + Patient(id,name,age

**Problem 2:**

* Sort dictionary using key
* “how to sort dictionaries by key”
* If AA, AA=1
* if not next row

**Problem 3:**

* Sort by number of complaints (different key?)

**Problem 4:**

**Problem 5:**

**Problem 6:**

**Problem 7:**

* Just assume you always the zip
  + DOB
    - 13131
      * 1
    - 14131
      * 2

**Problem 8:**

* If ‘cesar’ in tel

**Problem 7-8:**

* Data cleaning

**Mean date**

* Huge value