#### COS 597D:

Principles of Database and Information Systems

Professor Andrea LaPaugh

#### Traditional database system?

- · Large integrated collection of data
- Uniform access/modifcation mechanisms
- Model of data organization
- · Can relax in various ways

## **Relational Model**

- · Dominant DB model
  - Fully structured: informally, collection of tables
  - Formal underpinnings
  - SQL most widely used DB language
- Historical staying power
- Introduced 1970 by Edgar Codd What his motivations?- next time! How do they compare to modern concerns?
- Flat model
- vs older hierarchical and newer XML tree models

# Other Current Database Models Entity relationship model External "information" view conceptual "NoSQL" Key-value(s) model Unstructured value(s) various types value 1 bit string variable number columns XML model

# XML model Semi-structured versus fully structured

- Hierarchical

### Levels of Abstraction

- 1. Logical (e.g. relational) model
- 2. Data organization
  - indexing
- 3. Physical model
  - File organization
  - File storage
  - Determines access and manipulation methods

# Why study?

- importance of data in our lives
  - ubiquitous
  - huge volume
    - 2.5 exabytes of "high velocity" data created per day (source IBM "Big Data" Web page)
      - social media, sensors, video, transaction records ...
    - 2.8 zetabytes of "global data" 2012
    - (source BBC Future June 21, 2013)
- importance of techniques developed

**ISSUES?** 

#### Our syllubus: Models

- Structured Database model
  - relational model
    - Algebra, calculus and SQL
    - complex data relationships and constraints
- Unstructured data
  - "NoSQL" models
  - "Big Data": Information Retrieval
- Semi-structured data
  - XML and the tree model
    - bridging database systems and IR systems

#### Our syllubus: Storing & Retrieving

- File Organization
- Access Methods

   indexing
   Disk
- Distributed Storage
- Query Evaluation
- relational query evaluation methods
- Highly-parallel query evaluation
- MapReduce
- •

#### Our syllubus: Maintaining

- Correctness
  - concurrency & ACID properties
     consistency for distributed stores
- Fault tolerance
  - crash recovery
    - logging
  - distributed replication

#### Our syllabus: Specialized Databases (if time)

- Examples

   graph data
  - streaming data

#### Our syllubus: Current Research Student Presentations

- · advances in fundamentals and applications
- any aspect of data/information storage and use

### Graduate Focus

- Emphasize fundamental models and methods
  - expressiveness of languages
  - relationships through constraints
  - effectiveness and efficiency
- De-emphasize how use standard DB systems

   still opportunity to do so

#### **Graduate Focus**

- Explore interaction with "other" research areas
  - research techniques applied to database/info systems
    - · example: advanced data structures
    - example: caching in information systems
  - database/info system concepts applied to research
    - example: how integrate heterogeneous data sets in genomics
    - example: how structure data for network monitoring

#### Course logistics- overview

- Web page has all: READ!! http://www.cs.princeton.edu/courses/archive/fall13/cos597D/
- · Texts
  - Required: Database Management Systems by Ramakrishnan and Gehrke, 3rd Ed., McGraw-Hill, 2003
  - reserved books in library
- online readings

#### · Graded work

- Written exercises and questions on primary source reading about 10 of these (approx. 15%)
- Class Participation (approx. 15%)
- Oral presentations (approx. 20%)
- Design Project (approx. 50%)

3