Logistics

• Web page
  – http://www.cs.sunysb.edu/~hgupta/532

• Office Hours
  – TBD.
What is a Database?

- **Databases:**
  - Collection of large amounts of data

- **Examples:**
  - **Airline Reservation System**
    - Stores flights, airlines, reservations, price information, actual departure or arrival times, etc. Historical information.
  - **Banking System**
    - Stores account information, transactions, bank locations, etc.
  - **SOLAR System**
Managing Databases

• **Database Management System (DBMS)**
  – Software to manage databases.

• **DBMS’s job/purpose:**
  – Allow data definition (*schema*)
  – Manage data storage (access, backup)
  – Efficient Querying of data
  – Allow concurrent access
  – Support secure, atomic access
  – Crash Recovery
  – Etc.
DBMS Architecture

• **High-level Architecture:**

```
Query Processor

Storage Manager

Data/Schema Modifications

Queries

Transaction Manager

Data Metadata
```
Course Outline (Tentative)

- **Recap**: Relational Model, Relational Algebra, SQL (2 weeks)
- Datalog (1 week)
- Data Storage and Indexes (2 weeks)
- Query Processing/Optimization (3 weeks)
- Concurrency/Transactions (2 weeks)

- Semi-structured data; XML (1 week)
- Advanced Topics: Distributed DB, Mining, Search (1 week)
Data Model

• What is a data model?
  – Structure
  – Operations
  – Constraints

• DB Data Models
  – Relational
  – Semi-structured
  – Object Oriented
Relational Model

- Based on tables. E.g.,

<table>
<thead>
<tr>
<th>Course</th>
<th>Instructor</th>
<th>Semester</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 532</td>
<td>M. Kifer</td>
<td>Fall, 2014</td>
</tr>
<tr>
<td>CS 532</td>
<td>H. Gupta</td>
<td>Spring, 2015</td>
</tr>
</tbody>
</table>

- Used by *most* DBMS
- Why: Simple; Intuitive in terms of how we think of data.
- Terms:
  - Tables = Relations
  - Columns = Attributes
  - Row = Tuple
How to decide what tables to create?
Purpose of E/R Model

- An E/R diagram is a way to *formally* represent the meaning or ‘semantics’ of the data we wish to store.
- Converting E/R diagram to relational database – fairly mechanical techniques exist.
- In this class, we won’t discuss E/R modeling further.
Relational Terminology

• **Relation Schema:**
  – Name(attributes) + other structure info., e.g., keys, constraints.
  – Example: Movies(Name, Year)

• **Relation Instance:**
  – Set of rows/tuples for the relation

• **Database Schema:**
  – Collection of relation schemas

• **Key:**
  – A set of attributes such that no two (different) rows have the same values for these attributes.
Relational Algebra

• Next set of slides.