CS 306 Database Systems

Introduction
GENERAL INFORMATION

- Office: FENS 2081
- Email: ysaygin@sabanciuniv.edu
- Tel: 9576
- Web: http://people.sabanciuniv.edu/~ysaygin/
- Instructor Office Hours:
  - Monday 10:40-13:30
  - Thursday 11:40-14:30
  - Or by appointment
- Use sucourse for discussions regarding the course.
  (Your TAs will check the messages regularly and inform me for any urgent matter)
GENERAL INFORMATION

- TAs: Barış Altop, Berkay Dinçer
- UG TA: Yağız Nizipli
- TA OFFICE HOURS:
  - Tue, Wed, Thu, Fri:....
GENERAL INFORMATION : Grading

- Midterms: MT1 - 30% + MT2 - 40%
- Quizes and HWs: 15% (5% ALS, 10% HW)
- C++/DB Implementation: 10%
- DB Application Project: 5%
  - Groups of max 4 students.
  - Different project for each group.
  - Will consist of several steps. First step is to decide what to do!
  - You may use the Oracle Database Management System to develop your database application.
Form your group

Step 1: Write a one page report describing what you want to do for this project. For example, a library database, bank database, school database, etc. Write your report using any editor and submit it through sucourse

If you submit you will get 0

Otherwise you will get -1

Deadline: TBA.
An example database application

(Taken from the book “Fundamentals of Database Systems” by Elmasri and Navathe)

**Company** database keeps track of a company’s employees, departments and projects.

1. The company is organized into departments. Each department has a unique name, a unique number and a particular employee who manages the department. We keep track of the start date when that employee began managing the department. A department may have several locations.

2. A department controls a number of projects, each of which has a unique name, a unique number and a single location.

3. We store each employee’s name, Social Security number, address, salary, sex and birth date. An employee is assigned to one department but may work on several projects, which are not necessarily controlled by the same department. We keep track of the number of hours per week that an employee works on each project. We also keep track of the direct supervisor of each employee.

4. We want to keep track of the dependents of each employee for insurance purposes. We keep each dependent’s first name, sex, birth date and relationship to the employee.
Learning objectives

- Learning how to **design and implement** a database application
  - Conceptual design (ER model)
  - Relational Model
  - Mapping ER to Relational Model
  - Schema refinement and normal forms
  - Querying (SQL)

- Learning **database system** concepts
  - Concurrency Control and Transaction Management
  - Recovery
  - Storage and Indexing
What you need to do:

- Attend the lectures! (ALSs)
- Ask any question you like during the lectures (or during my office hours)
- Don’t be scared to ask questions or contribute with your comments
- Work harmoniously for the group project
- Pass this course with a good grade!
  - Min 45 out of 100
  - Please take advantage of the first midterm (it will be easier than the second midterm, and there are less topics)
INTRODUCTION TO DATABASES
Why do we need DBMSs?

- Computers were originally developed for number crunching.
- By time, data storage and processing became as important as scientific computing.
- Assume that you need to store information about students, courses, and enrollment of students to courses.
<table>
<thead>
<tr>
<th>Students.txt</th>
<th>Courses.txt</th>
<th>Enrollment.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td>00009374 Süha Orhun Mutluergil</td>
<td>CS306 Database Systems</td>
<td>00009374 CS306</td>
</tr>
<tr>
<td>00011749 Francesco Verdoja</td>
<td>CS201 Intro to Comp</td>
<td>00011749 CS306</td>
</tr>
<tr>
<td>00009054 Umut Öztok</td>
<td>CS204 Advanced Prog.</td>
<td>00009054 CS306</td>
</tr>
<tr>
<td>00010562 Selen Başol</td>
<td>MS304 Manuf Sys</td>
<td>00009054 CS201</td>
</tr>
<tr>
<td>00010499 Ozan Erdem</td>
<td></td>
<td>0010562 CS306</td>
</tr>
<tr>
<td>00009068 Sinan Eğilmez</td>
<td></td>
<td>00009374 CS201</td>
</tr>
<tr>
<td>00008918 Yiğit Emin Köksal</td>
<td></td>
<td>0011749 CS201</td>
</tr>
<tr>
<td>00008333 Kamer Ali Yüksel</td>
<td></td>
<td>0009374 CS201</td>
</tr>
<tr>
<td>00009125 Yaşar Andaç Efe</td>
<td></td>
<td>0010562 CS201</td>
</tr>
</tbody>
</table>

Query: What are names of the courses that Umut Öztok enrolled?

What are the names of students enrolled in CS201
Another Example
<table>
<thead>
<tr>
<th>ID</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ahmet</td>
</tr>
<tr>
<td>2</td>
<td>Ali</td>
</tr>
<tr>
<td>3</td>
<td>Burak</td>
</tr>
<tr>
<td>4</td>
<td>Baha</td>
</tr>
<tr>
<td>5</td>
<td>Cengiz</td>
</tr>
<tr>
<td>6</td>
<td>Ceyda</td>
</tr>
<tr>
<td>7</td>
<td>John</td>
</tr>
<tr>
<td>8</td>
<td>Mary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID1</th>
<th>ID2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
Why do we need DBMSs?

- To abstract the data model and storage from querying
- Amount and types of data increased:
  - Image/audio/video data
  - Genome data
  - Customer transactions
- Database Management Systems were developed to manage this data.
A new popular term

- BIG DATA
- Research project for you
Example for Big Data

- Google Flu-Trends
  - Keyword searches
  - Number of hospital checkins with flue
Why do we need people who know about databases?

- Most of software projects have a database component
- Someone should design the database
- Someone should maintain/tune the database
- Someone should design queries to retrieve data from the database
- Someone should deal with the recovery process when the system crashes (Ex: Customs control problems last week)
What do DBMSs do?

- They are system programs like the operating system.
- Database Management Systems (DBMSs) enable us to:
  - Create a database
  - Populate the database
  - Query the database
  - Let multiple users use the database at the same time
  - Recover the data when something goes wrong
What do DBMSs do?

- Lets say you would like to create a database for the library:
- You need to keep information about the books and who borrowed them
- What do we need to know about a book?
  - Name
  - Author
  - Publisher
  - Year
- For each book we need to keep that information
  - Book (Name, Author, Publisher, Year)
“A Relational Model of Data for Large Shared Data Banks”

E. F. Codd

CREATE TABLE book (name CHAR(20),
author CHAR(20),
publisher CHAR(20),
year INTEGER)

INSERT INTO book ('Kyle’s Mom', 'Eric McCarthy', 'Chef', 2001)

<table>
<thead>
<tr>
<th>Name</th>
<th>Author</th>
<th>Publisher</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kyle’s Mom</td>
<td>Eric McCarthy</td>
<td>Chef</td>
<td>2001</td>
</tr>
<tr>
<td>Death becomes him</td>
<td>Kenny Grave</td>
<td>Mr. Garrison</td>
<td>2000</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>...</td>
</tr>
</tbody>
</table>
What do DBMSs do?

```sql
SELECT name
FROM book
WHERE publisher = 'Chef'
```
What do DBMSs do?
Syllabus

1. Introduction to Database Systems
2. Entity Relationship Model
3. Relational Model
4. Relational Algebra
5. SQL Queries, Embedded SQL, Triggers, and Stored Procedures
Syllabus

- 6. Schema Refinement and Normal Forms
- 7. Data Storage and Indexing
- 8. Transaction Management
- 9. Concurrency Control
- 10. Crash Recovery
- 11. Introduction to Internet Databases
- 12. Data Warehousing and Data Mining