Handout 1
Introduction

Course Information

Title: *Data Management with SQL.*

Course information including the syllabus and course policies are posted on the course web page: [http://cis.bentley.edu/tbabaian/cs361/](http://cis.bentley.edu/tbabaian/cs361/)

SQL stands for **Structured Query Language** - most commonly used data definition and manipulation language.

Role of Data and Databases

For many applications data is a principal component - application consists mostly of data processing and presentation.

The data can be complex. An ad-hoc implementation of data storage and retrieval would be inefficient and simply would not work in many cases.

A **database** is a structured way of representing the data.

A **relational database** is the most widely used model in which data is organized in a set of tables according to certain rules.

For an example see pages 3,4 of the textbook.

A **Database Management System (DBMS)** is a software system that is used in creating, accessing and maintaining user defined databases.

Because of the complexity of the data database design and implementation cannot be approached in an ad-hoc manner.

Consider for example the following *CompProducts* database application.

Steps to take:

1. **Conceptual Data Modeling** - create a representation of data in terms of domain
entities, their attributes and relationships. Conceptual model is independent of any particular DBMS.

**Entity-Relationship (E-R)** diagrams are the most widely used tool for conceptual modeling. Richer than relational model.

2. **Logical Data Modeling** - from the conceptual model create a set of relations (tables) for implementation in a relational DMBS. Relations must be designed properly to enable efficient access and data integrity (i.e. validity).

3. **Implementation** using a DBMS, a. k. a. **Physical Data Modeling**.

**Review of Conceptual Modeling**

**Conceptual Data Model** must capture as many business rules as possible. *business rules are statements that define or constrain some aspect of the business.*

Some key terms:

**Entity instance** (entity) - a person, place, object, event, concept (often corresponds to a row in a table).

**Entity Type** - collection of entities (often corresponds to a table) with common characteristics (attributes). Denoted with a rectangle.

**Attribute** - property or characteristic of an entity type (often corresponds to a field in a table).

**Candidate key** - an attribute or a (minimal) set of attributes that uniquely identifies each instance of an entity type.

**Primary key** - a candidate key that has been used as the identifier for an entity.

**Relationship** - association between one or more instances of an entity types. Denoted with an annotated line or a diamond. Key characteristics of a relationship

1. **degree** - the number of *entity types* involved
   - unary - entity type related to itself, e.g. *course* is a prerequisite of another *course*,
   - binary - between two entity types; is most common, e.g. *professor* teaches *course*,
   - ternary - between three entity types, e.g. *person* maintains an *account* in a *bank*.

2. **cardinality** - describes number of *entity instances* involved in each relationship instance.
   - described with a minimum and maximum instances on each side:
specification of the maximum cardinalities: one-to-one (department-chair), one-to-many (manager-employee), many-to-many (student-course).

specification of the minimum cardinalities: Optional one (or zero or one), Mandatory one, Optional many (zero or many), Mandatory one.

Relationships can also have attributes. These describe features pertaining to the association between the entities in the relationship. Example: an employee works for a department. The works relationship may have following attributes that describe it: date hired, rank, etc.

Two entities can have more than one type of relationship between them (multiple relationships).

**Associative Entity** - combination of relationship and entity.

Review the Handout on conceptual modeling.

Why do conceptual modeling? A good Conceptual Model

1. captures fundamental business rules in a graphical format accessible to both the users and the analyst;
2. is an important communication and documentation tool

In addition to that entities and their relationship form a relatively stable (not prone to revision or change in requirements) base of an information system.

Review the conceptual model of the CompProducts database.

**Practice problem:** Categorize each of the following relationships and draw an E-R diagram depicting a complete specification of each of the following associations:

1. A country has a capital city.
2. A student takes a course from a professor.
3. A programmer uses a computer language on a project.

**Review of Logical Modeling**

**Logical Modeling** - create a set of relations to be implemented with a relational DBMS.

Goal: to create a properly structured (normalized) set of tables. Proper structure ensures minimal redundancy, no anomalies (insertion, deletion, update)

Consider the following example:

Some key terms:
Relation - a table with named columns (a.k.a. attributes or fields) and rows (a.k.a. records or tuples)

Important properties of relations

- values stored in columns are *atomic* (i.e. non-decomposable values of primitive types: number, string, char, etc.)
- values in the same column have the same type and are from the same domain,
- each row is unique (no duplicates allowed)

Records from different tables are associated with each other. **Foreign key** is an attribute of one relation that “connects” tuples from that relation to tuples from another relation by referring to the values of primary key of another relation.