Goals
- Define “big data” and discuss its basic characteristics
- Understand ways to store information
- Understand the value of a Database Management System
- Explain the basic structure and components of a relational database

Big Data
- 1GB storage, around 1985, cost $5,000
- 1GB storage, today, costs around 3 cents
- The Encyclopedia Britannica, 2010 edition, weighs 129 pounds and contains 1GB text, <5GB with pictures
- Today, 5,000 GB of storage costs $300
- Storing information is not the challenge
- Managing information, and condensing it into meaningful information is the challenge
Focus on Information

Different schemes for managing data
- MS Access projects
- Database
- WordPress blog
- mySQL database
- Your “music” database
- ASCII text file
- Other schemes
  - Word processor
  - Spreadsheet
  - Open ended systems (OneNote, Lotus Notes)

Spreadsheets

Strengths
- Very simple data storage tasks
- Relatively easy to use
- Require less planning
- Good for analyzing and displaying information visually

Weaknesses
- Data integrity problems include inaccurate, inconsistent and out of date data
- Formulas could be incorrect

Databases

Strengths
- Methods for keeping data up-to-date and consistent
- Data is of higher quality than data stored in spreadsheets
- Good for storing and organizing information

Weaknesses
- Requires more planning and designing
The Database

- One of the best methods
- One of the most difficult methods

Databases and Database Management Systems

A database is an organized collection of data.
- These databases can store different types of “information” including text, numbers, documents, images and videos.
- Databases are managed by database management systems (DBMS).
- A database management system provides the means for creating, maintaining and using databases.

Multi-tiered Architecture

- WordPress
Various Schemes

- Hierarchic model
- Network model
- Object model
- Relational model

Relational Model History

- Dr. Edgar F. Codd, 1969
- Most significant contribution to database systems
- Made database management a science
- Defined both a relational algebra and relational calculus for dealing with data (e.g., join)
- Introduced the concepts of “normalized” forms
- Introduced a language, SQL, for dealing with data (Structured Query Language, not “sequel”)

Relational Databases

- A relational database stores data in the form of connected tables. Tables are made up of records (rows) and fields (columns)
- A record is a set of fields that all pertain to the same thing, while the fields represent some characteristic of the thing
Other Database Characteristics

- **Primary Key**
  - A unique record identifier
  - Each table in a database has a primary key
  - Sometimes a primary key is made up of more than one field, called a composite primary key

- **Foreign Keys**
  - Foreign keys are fields that reference a primary key in a related table
  - This cross-referencing is called a relationship.
  - Cross-referencing foreign keys make it easy to combine data contained in multiple tables

Related Tables

- Cross-referencing foreign keys make it easy to combine data contained in multiple tables.

Normal Forms

There are some rules that govern relational database design called “normal forms”

- These rules are in place to ensure data consistency by eliminating unnecessary redundancy.
- A particular row in a table can be related to at most one row in a related table.
- For example, in most businesses a specific order can only be related to one customer.
Normal Forms

- First normal form
- Second normal form
- Third normal form
- Boyce/Codd normal form
- Fourth normal form
- Fifth normal form

Relationships

- One-to-one
- One-to-many
- Many-to-many
  - In many-to-many relationships a specific row can be related to multiple rows in a related table.
  - In contrast to the one-to-many relationship, this is true in both directions.
  - Many-to-many relationships require creating a new table that links the two related tables.
  - These are called linking or intersection tables.

Database Diagrams

As you might imagine, using the actual data to show the structure of a database (as in the previous figure) only works for very small databases. For larger databases, we illustrate structure using database schema diagrams.

- Entity-relationship diagrams (ERD)
- Database schemas
Data Warehouse and Data Marts

- Online databases that are quite useful for performing research
- Databases serve a different purpose than the databases we discussed earlier
- These online databases contain pointers to sources of information.
- Most of these resources are available through libraries.