


# Introduction to Databases

Database Systems



# What is data?

- “a collection of facts from which conclusions may be drawn” ([www.cogsci.princeton.edu/cgi-bin/webwn](http://www.cogsci.princeton.edu/cgi-bin/webwn))
- “Programs, files, and other information stored in, communicated, or processed by a computer.” ([www.nrc.gov/site-help/eie/terms\\_id.html](http://www.nrc.gov/site-help/eie/terms_id.html))
- “A representation of facts, concepts, or instructions in a formal manner suitable for communication, interpretation, or processing by human beings or by computers.” ([cedar.web.cern.ch/CEDAR/glossary.html](http://cedar.web.cern.ch/CEDAR/glossary.html))
- However, we can say:
  - It is unprocessed information
  - Data is converted into information, and information is converted into knowledge
  - For the purposes of enterprise, data is a small unit of information, i.e. a learner's name or an exam mark

# Some Important Aspects of Data

- It is always an abstract representation
- Good measurement is the assignment of numbers to perceived attributes of objects or events according to rules that
  - are easily understood
  - are easily used
  - yield numbers that are as simply related as possible to as many other sets of measurements

# Some Important Aspects of Data

- Scalability
- Functionality
- Expandability
- Cost
- Usability
- Capability
- Interoperability
- Reliability

# Some Important Aspects of Data

- Scalability
- Functionality: store images, music, geographic data, drawings, etc.
- Expandability: capabilities in terms of application development, e.g., Java may be used in certain systems to code database logic
- Cost: Acquisition Cost and Maintenance Costs
- Usability: How easy or difficult the DBMS can be learned, modified, and used to accomplish key tasks that are performed frequently.

# Some Important Aspects of Data


- Capability
  - Breadth and depth of features and functions that a DBMS can perform.
- Interoperability
  - How well a DBMS supports other database access standards and defines the ability of software on multiple machines from multiple vendors to communicate.
- Reliability
  - Recover from unplanned outages, facilitate planned database maintenance to occur while the database is available to users, improve system serviceability, and disaster planning

# Database Management Systems

- A set of software programs that allows users to:
  - create database files
  - edit and update data in database files
  - store and retrieve data from database files
- DBMSs come in many shapes and sizes
- DBMSs vary in terms of their scalability
- DBMSs can be run on everything; Handhelds, Laptops, UNIX servers and cluster of mainframes.

# Basic Database Terms

- Character: A single symbol such as a digit, letter, or other special character (e.g., \$, a, 2, etc.)



| D          | E         | F       | G                   | H               |
|------------|-----------|---------|---------------------|-----------------|
| First Name | Last Name | D.O.B.  | Address             | Social Security |
| John       | Smith     | 6/12/82 | 321 Byberry Road    | 010-22-9432     |
| John       | Smith     | 5/9/40  | 268 Monroe Avenue   | 003-63-0037     |
| John       | Smith     | 12/4/57 | 8120 Venshire Drive | 020-45-9326     |
| Sally      | Smith     | 3/4/86  | 207 Congress Drive  | 289-56-4321     |
| Steve      | Smith     | 4/23/79 | 1519 Ashbury Lane   | 170-54-2334     |


- Field: Contains an item of data, that is, a character, or group of characters that are related.

| D          | E         | F       | G                   | H               |
|------------|-----------|---------|---------------------|-----------------|
| First Name | Last Name | D.O.B.  | Address             | Social Security |
| John       | Smith     | 6/12/82 | 321 Byberry Road    | 010-22-9432     |
| John       | Smith     | 5/9/40  | 268 Monroe Avenue   | 003-63-0037     |
| John       | Smith     | 12/4/57 | 8120 Venshire Drive | 020-45-9326     |
| Sally      | Smith     | 3/4/86  | 207 Congress Drive  | 289-56-4321     |
| Steve      | Smith     | 4/23/79 | 1519 Ashbury Lane   | 170-54-2334     |



# Basic Database Terms

- Record: A group of related fields.



| D          | E         | F       | G                 | H               |
|------------|-----------|---------|-------------------|-----------------|
| First Name | Last Name | D.O.B.  | Address           | Social Security |
| John       | Smith     | 6/12/82 | 321 Byberry Road  | 010-22-9432     |
| John       | Smith     | 5/9/40  | 268 Monroe Avenue | 003-63-0037     |

- Table (database file): A collection of related records.

| D          | E         | F       | G                   | H               |
|------------|-----------|---------|---------------------|-----------------|
| First Name | Last Name | D.O.B.  | Address             | Social Security |
| John       | Smith     | 6/12/82 | 321 Byberry Road    | 010-22-9432     |
| John       | Smith     | 5/9/40  | 268 Monroe Avenue   | 003-63-0037     |
| John       | Smith     | 12/4/57 | 8120 Venshire Drive | 020-45-9326     |
| Sally      | Smith     | 3/4/86  | 207 Congress Drive  | 289-56-4321     |
| Steve      | Smith     | 4/23/79 | 1519 Ashbury Lane   | 170-54-2334     |

This table is made up of multiple records containing multiple fields (First Name, Last Name, D.O.B., Address, and Social Security).

# Basic Database Terms

- Database: A collection of related tables

| First Name | Last Name | B.G.B.  | Address             | Social Security |
|------------|-----------|---------|---------------------|-----------------|
| John       | Smith     | 871,382 | 371 Sydney Road     | 010,21,3422     |
| John       | Smith     | 65940   | 280 Monroe Avenue   | 0204624022      |
| John       | Smith     | 1,74857 | 81,20 Vanhook Drive | 020465,0026     |
| Sally      | Smith     | 34400   | 207 Congress Drive  | 200460-4021     |
| Steve      | Smith     | 402078  | 1819 Ashbury Lane   | 17044,2004      |

# Data Modeling – An Introduction

- Data modeling is the act of defining data-oriented structures
  - Data models can be used for a variety of purposes, from high-level conceptual models to physical data models.
  - In data modeling you identify entity types
  - Attributes are assigned to entity types
  - There are associations between entities – relationships, inheritance, composition, and aggregation are all applicable concepts in data modeling.
- Data modeling can be one of the most challenging tasks that a DBA can be involved with on a development project

# Data Models Used in Practice

- You are likely to see three basic styles of data model:
  - Conceptual data models. These models explore domain concepts with project stakeholders. Conceptual data models are often created as the precursor to LDMs.
  - Logical data models (LDMs). LDMs are used to explore the domain concepts, and their relationships, of your problem domain. LDMs depict the entity types, the data attributes describing those entities, and the relationships between the entities.
  - Physical data models (PDMs). PDMs are used to design the internal schema of a database, depicting the data tables, the data columns of those tables, and the relationships between the tables.

# How to Model Data

1. Identify entity types
2. Identify attributes
3. Apply naming conventions
4. Identify relationships
5. Apply data model patterns
6. Assign keys
7. Normalize to reduce data redundancy
8. Denormalize to improve performance

Thank You

