Information Management

Dr Marilyn Rose McGee-Lennon
mcgeemr@dcs.gla.ac.uk

What is Information Management about

- **Aim**: to understand the ways in which databases contribute to the management of large amounts of data.

What is Information Management about

- **Objectives**:
  - understand the *nature of applications* built using programs clustered around databases and other large collections of data.
  - understand the overall *architecture* of a database management system.
  - be able to carry out all the *operational tasks* of setting up and using a relational database.

Why IM is Valuable to You

- Databases are a key technology
  - Used in a variety of applications to manage data
  - Growing in importance – amount of data increasing

Data vs Information

data	52
information	J Smith's score on the final exam is 52%
knowledge	I've passed!

structured representation (encoding)

data + meaning

true belief

Data Management

- We need an efficient and reliable means of storage & access to large amounts of data

Issues in managing data

- Consider the example of handling the billing & monitoring of all UK household telephone calls
  - Write down as many issues as you can think of associated with managing all this data and the tasks associated with it
  - Especially if you had to handle it all on paper in filing cabinets

Issues in managing data

- Managing the billing & monitoring of telephone calls
  - Billions of transactions per day
  - Need to hold terabytes of data
  - Thousands of users at the same time
  - Processing distributed around the world
  - Need access to data for monitoring and looking for significant patterns
  - Reliability and security are paramount

Data Management

- Data storage method must provide:
  - Data definition (data structuring)
  - Data entry (to add new data)
  - Data editing (to change existing data)
  - Querying (a means of extracting data by a description)
  - Persistence (data existing beyond a single operation or program invocation)

Strategies for Data Management (1)

- A **program** where all the data is held in the program's memory
  - Like in the programming course
  - No persistence between invocations of a program
  - Data is reconstructed (or re-entered) at each invocation of the program

Strategies for Data Management (2)

- **Files** of data can be accessed by different applications
  - Each application is responsible for its own representations of the data
  - Its difficult to coordinate between applications

A better way…

- Combine together all the functions of data storage and access for a related set of tasks, e.g.
  - handling student records
  - stock control in a warehouse
  - account management in a bank
- This is known as a
  Database Management System (DBMS)

What is a database?

What is a database?

- A database is "One or more structured sets of persistent data." (The Free On-line Dictionary of Computing, http://www.foldoc.org/, Editor Denis Howe. Available via Burks 6.)
- Usually
  - held on computer
  - associated with software to update and query the data, such as a database management system (DBMS).

What is a Database Management System (DBMS)?

- A DBMS is "A suite of programs which typically manage large structured sets of persistent data, offering ad hoc facilities to many users.

  A database management system (DBMS) can be an extremely complex set of software programs that controls the organisation, storage and retrieval of data (fields, records and files) in a database. It also controls the security and integrity of the database. The DBMS accepts requests for data from the application program and instructs the operating system to transfer the appropriate data. "

  From the Free On-line Dictionary of Computing, http://www.foldoc.org/, Editor Denis Howe. Available via Burks 6.

Key Functions of the DBMS

- Sharing and integration of data
- Multiple views of the same data
- Controlled concurrent access
- Management of security and integrity

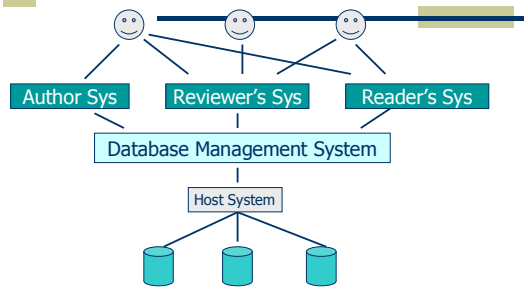
Summer School 2007

Example: An on-line magazine system

- Stores magazine articles & related data throughout the creation, review and publication process
- Author writes an article, saves it, edits it
- Reviewers can extract articles ("the latest version of John Smith's article") and annotate them
- Readers can extract articles on different criteria

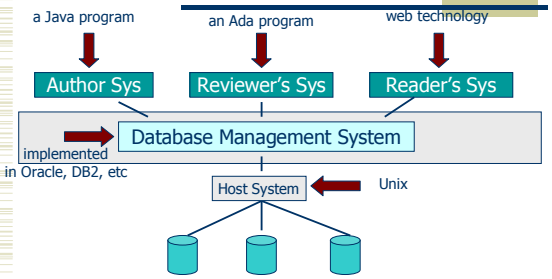
Summer School 2007

An on-line magazine system



Summer School 2007

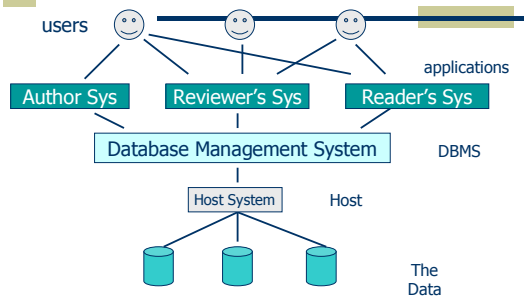
An on-line magazine system: Implementation 1



Main components of a database

- Users
- Applications
- DBMS
- The Data
- The Host system

An on-line magazine system



Database Design

Database Design

- How do we go about designing and building a database from scratch?

Database Design

- Creating a database involves:
  - **(1) Capture user requirements**
  - **(2) Represent them in a model**
  - **(3) Convert model into a schema**
  - *(4) Implementation on DBMS*
- Many different types of database BUT…
  - All require the DESIGN stages 1,2,3 above

Data Modelling

- A **data model** is an abstract representation of the data we wish to store
  - typically a set of concepts for capturing a **schema** - i.e. the structure of the database.
  - the **data types**, the **relationships** and the **constraints** in the DB

Schema: The glue that holds it all together

- A database schema is a description of a database
  - in particular, a description of its entire structure
- There are many different ways of representing a schema, including diagrams

A simple schema for the on-line magazine system example

Author

Id	Name	Address

Article

Title	Author	Keywords	Date	Contents

Reviewer

Id	Name	Address	Expertise

Review

Article	Reviewer	Date	Rating	Comments

Semantic Modelling

- Deals with the relationships between the words and the real world objects that those words refer to.
- Different types of database modelling (Functional, semantic objects, **Entity/Relationship (ER)**).

The Entity-Relationship Model

The Entity-Relationship Model

- Most common method for semantic modelling
- Simple and highly applicable
- Very popular, especially when linked to the relational model as its implementation
- Usually described using ***Entity-Relationship Diagrams***

The Entity-Relationship Model

- Data in the ER Model is described in terms of three key concepts:
  - Entities
  - Relationships
  - Attributes

Entities

- An **entity** is a uniquely identifiable object in the real world about which we wish to store data
  - For example: The Bank of Scotland, The University of Aberdeen, Tony Blair, Celtic Football Club, BBC, a car…….

Entities

- An entity is likely to have **properties** that describe it
  - BT: name, address, annual profit.
  - A hospital patient: name, address, hospital number, national insurance number, data of admission..
  - The CS1Q module: the lecturer, the meeting room, the times of lecture, etc.

Attributes

- Attributes are one kind of property that describe an entity
  - consist of one or more *atomic values*
- E.g. The entity "person" might have the attributes "name", "address", "age", "hobbies",……

Keys

- The Entity type will usually have a **key**:
  - one (or possibly more) of the attributes capable of identifying an entity
  - **candidate keys** - attributes whose values are unique and can thus be used to identify an entity
  - **primary key** - an attribute that is actually used for identification

Relationships

- Relationships represent the *interaction* between the entities
- For example the entities "person" and "Bank of Scotland" might interact through the relationship "works for".

From the Scenario to a Model

- Identify **Entities**, their **Attributes**, and all **Relationships** involved in the scenario
- Represent this in Entity-Relationship Diagram format
- This diagram (and model) can the be used to implement the actual relationship tables in the database

IM Notes