Information Management (IM)

CS1Q/CF2

IM Lecture 1

Dr. Craig Macdonald

Who am I?

Dr. Craig Macdonald

Lord Kelvin Adam Smith Research Fellow

Room M101, Lilybank Gardens

Email: Craig.Macdonald@glasgow.ac.uk

Lecturer for Information Management part of CS1Q

CS1Q Overview

Week 1 - week 5: **Information Management**
Week 6: **Professional Issues (PI)**
Week 7 - week 12: **HCI**
Week 13: Class Test
Week 14, 15, 16: Christmas Holiday

Week 17-week 27: **Systems**

Week 28, 29, 30: Easter Holiday

Week 32-36: Exams

What's in a name

4

- This course is called CS1Q and CF2
  - Lectures & labs for both are at the same time and place!
- You take CS1Q if:
  - If you are studying 1st year Computing Science and taking CS1P
- You take CF2 if:
  - If you are in 2nd year and took CS1CT in first year (and not CS1Q)
  - If you are in the first year of a "fast track degree"

If you are neither, then see me!

CS1Q Assessment

	CS1Q		CF2	
Class Test	December	10%	n/a	
Exam	May	70%	December	70%
Practical Work	Ongoing	20%	Ongoing	30%
Resit Exam	*August*	*70%*		

For CS1Q, this equates to

- **Information Management Practical** is worth **6%**
- **Human Computer Interaction Practical** is worth **6%**
- **Professional Issues Practical** is worth **2%**
- **Systems Practical** is worth **6%**

Overview of Information Management

Week	Lecture (Tue)	Lecture (thur)	Tutorial	Lab	Hand in
1	Introduction	DB modelling		Familiarisation	
2	ER Diagrams	Relational Model	ER diagram		**ER Diagram**
3	ER to Tables	Sets and Relations		Tables	
4	Relational Algebra	Basic SQL	Sets		
5	Querying	More SQL		Populate and Query	
6	*HCI*	*HCI*		**Assess DB**	**DB report**
7	*HCI*	*HCI*	Feedback		

Lectures will be posted on **Moodle** - **http://moodle2.gla.ac.uk/course/view.php?id=2890**

Lab and tutorial sheets
will be given out by tutors at your tutorial/lab and also posted weekly on Moodle

There is also an overview document (available on Moodle)
Note - You should bring this document to all 1Q IM sessions.

Familiarisation Lab

Turn up to ONE of the following familiarisation labs **this week** (week 1)

They will be held in the Level 1 Computer Labs (Room 715 Boyd Orr Building):
Tuesday 23rd (2-4); Wednesday 24th (10-12); Wednesday 24th (2-4)

You should bring your Student ID card

You will be issued with: (1) a username and password, (2) important lab documentation and (3) a lab worksheet for this week

What you will do in the lab:
(1) Log in and find your way around Moodle
(2) A short exercise (CS1CT or CS1P) - upto 1 hour

It is your job to turn up to a suitable lab group

The Information Management Module

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

The Information Management Module

- Another aim: to understand the mathematics necessary behind databases and how to manage and query them
- Objectives
  - understand sets & relations
  - understand relational algebra, SQL

Why are Databases Important?

- Databases are a key technology
  - Used in a wide variety of applications to manage data
  - Growing in importance – amount of data we store is increasing
- Consider the management of all of Facebook's user accounts and data……..
    - This is a lot of data and a lot of data management involved!

Lectures

- IM Lectures from week 1 through to week 5
- Topics
  - Issues in Data & Information Management
  - Data Modelling and ER Diagrams
  - Design & Implementation of a (Relational) Database Application
  - Querying a Database
  - Sets, Relations, and Relational Algebra

Tutorials & Labs

- Tutorials & Labs will be closely linked to the practical assessment and the exam
- You have been allocated to a lab group – you can find your allocated group on MyCampus.
  - It is your job to find your lab group and turn up!
- Practical Assessment
  - worth 6% of 1Q module mark
  - involves designing, building, testing and reporting on a simple database system (using a MySQL database server)

How this module is assessed

CS1Q

- Practical = 20% (this database practical = 6%)
- Class Test = 10% (*HCI* + IM)
- Degree Exam = 70% (*HCI* + IM + *Systems*)

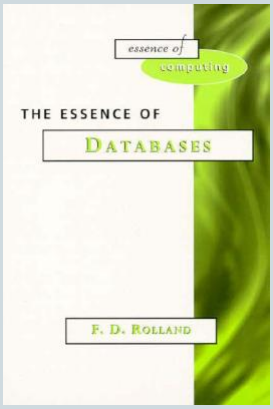
CF2

- Practical = 30% (this database practical = 13%)
- Degree Exam = 70% (*HCI* + IM + *Systems*)

Don't be fooled! The practicals are important for your understanding of the material, and 75% submission is required for progression

The Course Textbook

- Rolland, F. D. *The Essence of Databases*.
  Prentice-Hall, 1998. ISBN 0-13-727827-6
- This is an *essential* book for this part of CS1Q
- Lectures will be closely tied to this textbook



Understanding Databases

- On this course you will design, build, populate and query your own database using **MySQL**
- It is important that you understand all of the **theory,** as well as how to use **MySQL** so that you will easily be able to design, build and use other database applications in the future

Issues in Data Management

What is data?

17

What is data?

18

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

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

Issues in managing data

- Consider the example of handling the billing & monitoring of all UK household telephone calls

Issues in managing data

- Consider the example of handling the billing & monitoring of all UK household telephone calls
  - What kinds of data would you have to store?

Issues in managing data

- Consider the example of handling the billing & monitoring of all UK household telephone calls
  - What kinds of data would you have to store?
    - People's names
    - Addresses
    - Phone numbers
    - Post codes
    - Account Codes
    - Bank details
    - Money owed
    - ……

Issues in managing data

- Consider the example of handling the billing & monitoring of all UK household telephone calls
  - Can you think of any issues associated with managing all of this data and the tasks associated with it?
    - On paper?
    - In filing cabinets?
    - In spreadsheets or text documents?

Issues in managing data

- What about managing all of the data stored on Facebook?
  - Billions of uploads/updates per day
  - Need to hold terabytes of data
  - *1.28 billion* users, with thousands of users active at the same time
  - Processing distributed around the world
  - Need access to data for monitoring and looking for significant patterns
  - Reliability and security important

Managing Data

- If we have lots of data to store, we need a really good way to store that data
- We also need good ways to be able to access that data quickly and easily
- And we don't want lots of different versions of the data all over the place - we want to avoid REDUNDANCY

Managing Data

- Our data storage tool must provide these features:
  - 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
- But no data persistence between invocations of a program
- The data is reconstructed (or re-entered) at each invocation of the program

Strategies for Data Management 2

- **Files** of data that can be accessed by different applications
  - Each application is responsible for its own representations of the data
  - Difficult to coordinate between applications
  - Might be different versions of the data

Strategies for Data Management 3

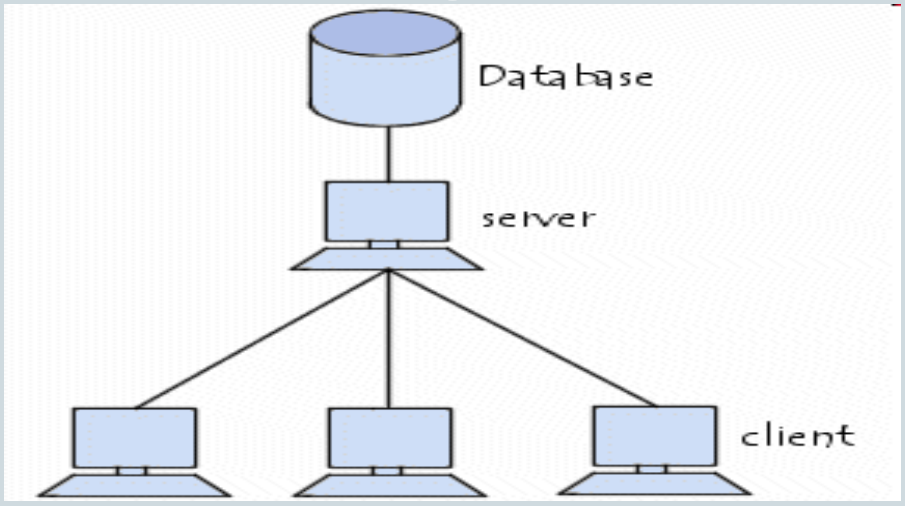
- Combine together all the functions of data storage and access for a related set of tasks, e.g. separate "databases" for:
  - handling student records
  - stock control in a warehouse
  - account management in a bank

What is a database?

What is a database?

- A database (abbreviated *DB*) is an entity in which data can be stored in a **structured manner**, with as **little redundancy** as possible
- A database centralises users access to data, which they can view, enter, or update
  - within the limits of the access rights granted to them
- It is viewable by many users at the same time (**controlled concurrent access**)

What is a database?



Types of database

- **Hierarchic databases** (older)
- **Network databases** (older)
- **Relational databases** (very common)
- **Object Oriented Databases** (newer)
- **NoSQL databases** (very new)

Databases

- Local database
  - can be used on one machine by one user only
- Distributed database
  - the information is stored on remote machines and can be accessed over a network
- Databases become all the more useful as the amount of data stored continues to grow
- It is associated with software to add, update, and query the set of data (DataBase Management System: **DBMS**)

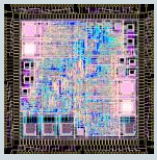
What is a Database Management System (DBMS)?

- The DBMS is a suite of services (software applications) for managing databases, which involves:
  - enabling simple access to data
  - allowing multiple users access to the information
  - manipulating the data found in the database (inserting, deleting, editing)
- 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

Database Management Systems (DBMSs)

Your Applications Go Here

DBMS



Raw Resources (bare metal)

Database abstractions allow this interface to be cleanly defined and this allows applications and data management systems to be implemented separately.

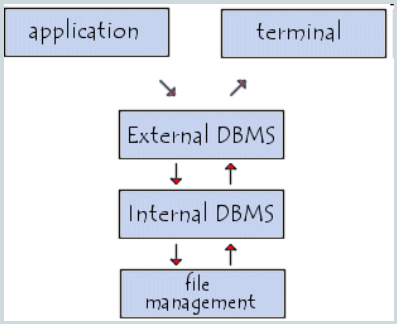
36

What is a Database Management System (DBMS)?

37

The DBMS can be broken down into three subsystems:

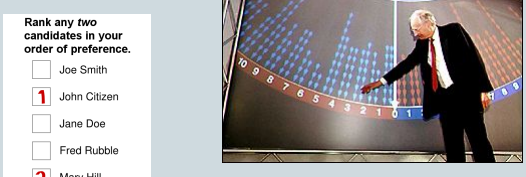
- The file management system:
  for storing information in a physical medium
- the internal DBMS:
  for placing information in order
- the external DBMS:
  represents the interface of the database to the outside world



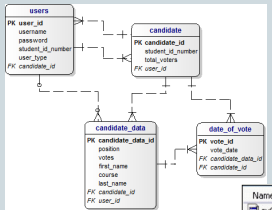
Three-level Architecture: Election Example

38

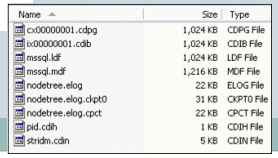
External Schema



Conceptual Schema



Internal Schema



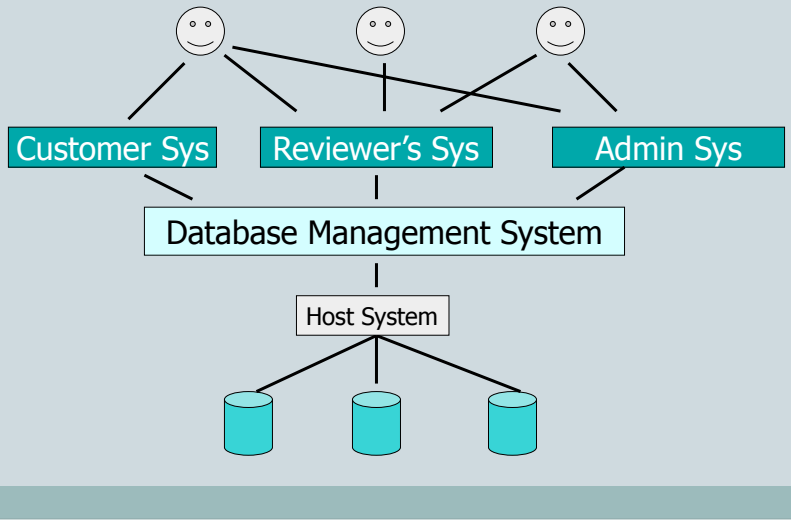
Different Types of DBMS

- Filemaker
- IBM DB2
- Ingres
- Interbase
- Microsoft Access
- *Postgres*
- Microsoft FoxPro
- Oracle
- Sybase
- **MySQL**
- Microsoft SQL Server 11

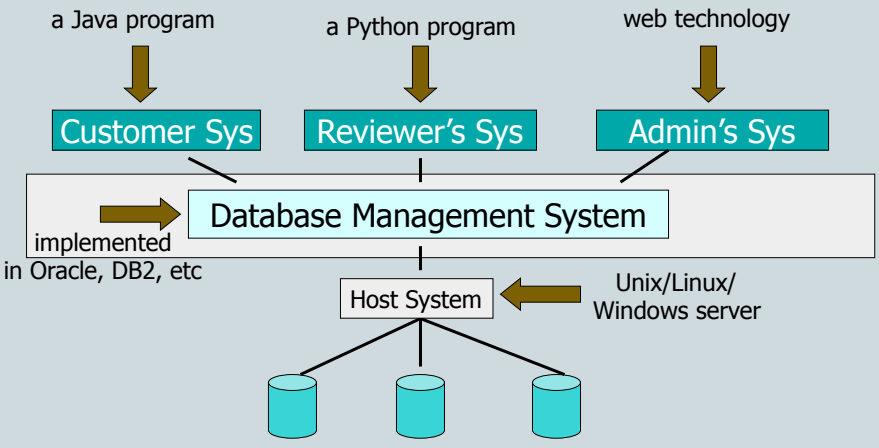
Example: Amazon

- Stores data about products and their related details (name, price, colour, product code…….)
- A customer can view products, search products, buy a product and rate reviews
- An admin person can upload products & edit product info
- A reviewer can write reviews
- …….

An example Database system - AMAZON



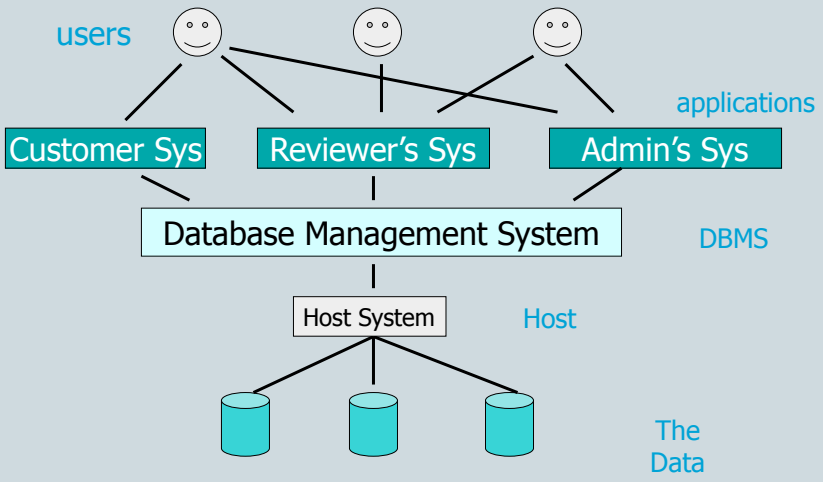
An example Database system - AMAZON



Main components of a database

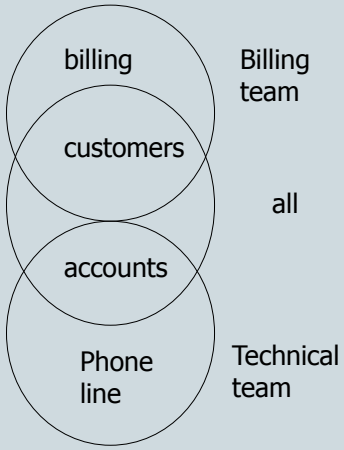
- Users
- Applications
- DBMS
- The Data
- The Host system

An example Database system - AMAZON



Databases avoid Redundancy

- Ambiguity
  - Same thing with different name in different files
- Inconsistency
  - If data changes in one place it should also change in the other files is exists in
- Wasted effort
  - Should share data where possible to save time and effort



Key Functions of the DBMS

46

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

Essential Reading…

47

- After this lecture
  - Rolland, Chapter One
- Before next lecture
  - Rolland, Chapter Two
    - 2.1
    - 2.3.1