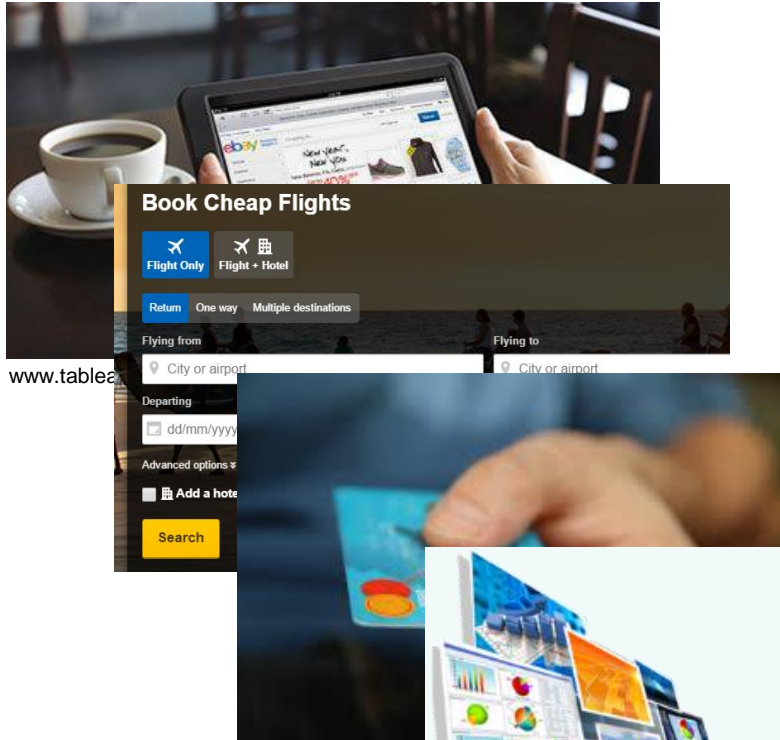# INFO20003 Database Systems

Dr Renata Borovica-Gajic
Dr Ida Asadi Someh

Lecture 01 – Subject Introduction

Subject Administration

What are Database Systems?

[The Economist]

www.tablea...

Book Cheap Flights

http://reportlogix.com/reporting.html

Data Amounts

Year

50-fold from 2010-2020*

* "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", 2012, IDC

And grows exponentially…

Equals to finding the needle in a haystack

Organisational Description and Problem Area

An investment bank wants to have a database to provide it with the ability to store information about its trading operations. The bank essentially works with customers by providing the capability for trading stocks, shares and other commodities. The bank has three branches in which exist a number of departments. Departments have a department manager who supervises a number of staff within the department. A set of accounts are used to store information about the currency of the organisations operations. Accounts can be customer accounts or internal "house" accounts, each of which allow trades to be made upon them. There are a number of account types. There are many customers and customers may have one or more contacts. Customers have a facility for lending money to pay for their purchases of stocks and commodities. Staff make deals on the behalf of their customers using a funding source and keeping track of settlements on the deals being made. There are many types of deal to be made. Settlements are full or partial payments of the deals and are recorded whenever a payment is made.

Please note that this section is purely made up and by all means is a very short description of a real investment bank (although many details have been left out and wide ranging assumptions have been made.
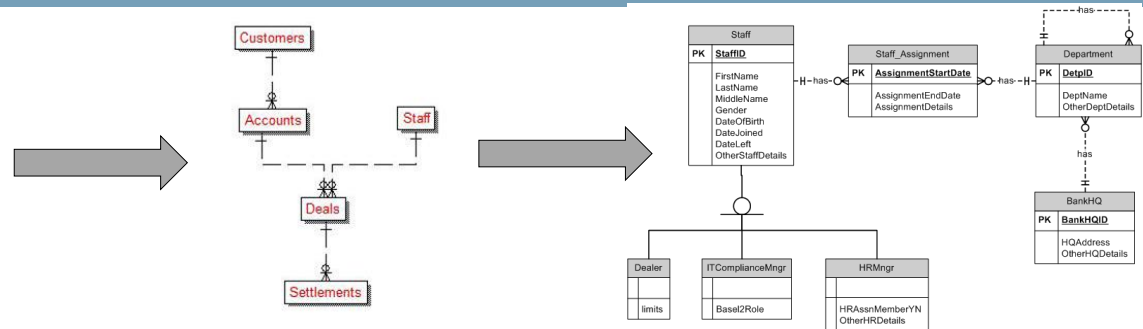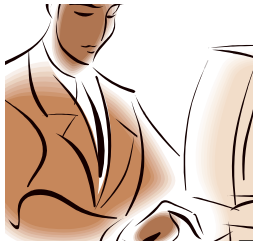
Results

Process

Access

Store

Database System

SQL Queries

select val from sales
where id = max;

- How to design & build a database application
  - Model and write SQL queries

+A brief look "under the hood" of a DBMS

Why?

  - The best application writers & database administrators **understand** DBMS internals
  - DBMS technology is still very much in evolution in industry
    - e.g. MapReduce-databases hybrid, NoSQL movement

- Introduction to INFO20003 Database Systems
  - Expectations
  - Week by week plan
  - Assessment

- Introduction to Databases
  - From files to relational database
 - Query Language (SQL)
  - DBMS (MySQL)

- Lecturers: Renata Borovica-Gajic (me) and Ida Asadi Someh
- Head tutor: Farah Khan
  - Email: headtutor.info20003@gmail.com
- Senior tutor: David Eccles
  - Email: eccles.d@unimelb.edu.au
- Tutors/Lab demonstrators

- Interaction
  - Attend classes & tutorials
  - Use LMS discussion forum
  - Email – only if personal
    - For questions about subject material, we want to answer where everyone can participate, not in a private email chain.

- Textbook: Ramakrishnan and Gehrke, 3rd Edition
- Lectures and lecture notes

- **Grading policy**
  1. Assessments x3   30% total (10% each)
  2. Midterm   10%
  3. Final Exam   60%
        Quizzes   (exam preparation)
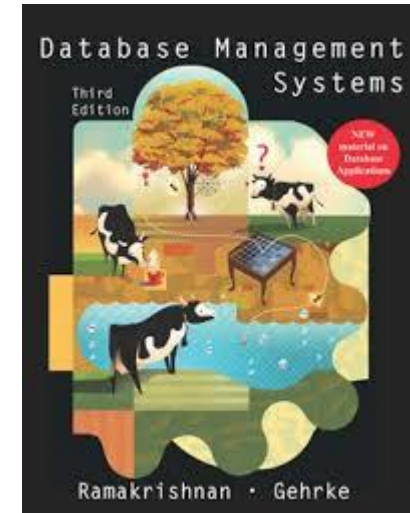
- On LMS: Subject information

- Assignments:
1. A1: post W4, due W6 (modelling)
2. A2: post W6, due W10 (SQL)
3. A3: post W8, due W12 (Query processing/optimisation)

- In the lectures
  - Turn up
  - Be prepared to answer the questions we put to you
    - GUESS– most of what we do is logical and guesses are often correct
    - THERE ARE NO DUMB QUESTIONS
    - If you thought of the question it is likely someone else might have the same thought
- In the workshops
  - Turn up
  - Follow / try it for yourself
  - One-on-one feedback
- Student representative
  - Nominations here or via email

What do you expect from the subject?

Introduction to Database Systems

- Data
  - known facts stored and recorded
  - can include: text, numbers, dates, plus images, sound, video, and other complex objects
- Information
  - Data presented in context (can be summarised data)
  - Data that has been processed increasing the users knowledge
- Data vs Information
  - *Data* is known and available; *Information* is processed and more useful

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145
Sisneros, Michael	392416582

Database Systems Assignment 4 Marks Semester 3 2014		
Student Name	**Student ID**	**Grade**
Baker, Kenneth D.	324917628	H1
Doyle, Joan E.	476193248	H2B
Finkle, Clive R.	548429344	H3
Lewis, John C.	551742186	H2A
McFerran, Debra R.	409723145	P
Sisneros, Michael	392416582	H3

Mark Distribution
Database Systems (Semester 4, 2012)

N,5%
P,15%
H1,13%
H2A,23%
H3,20%
H2B,25%
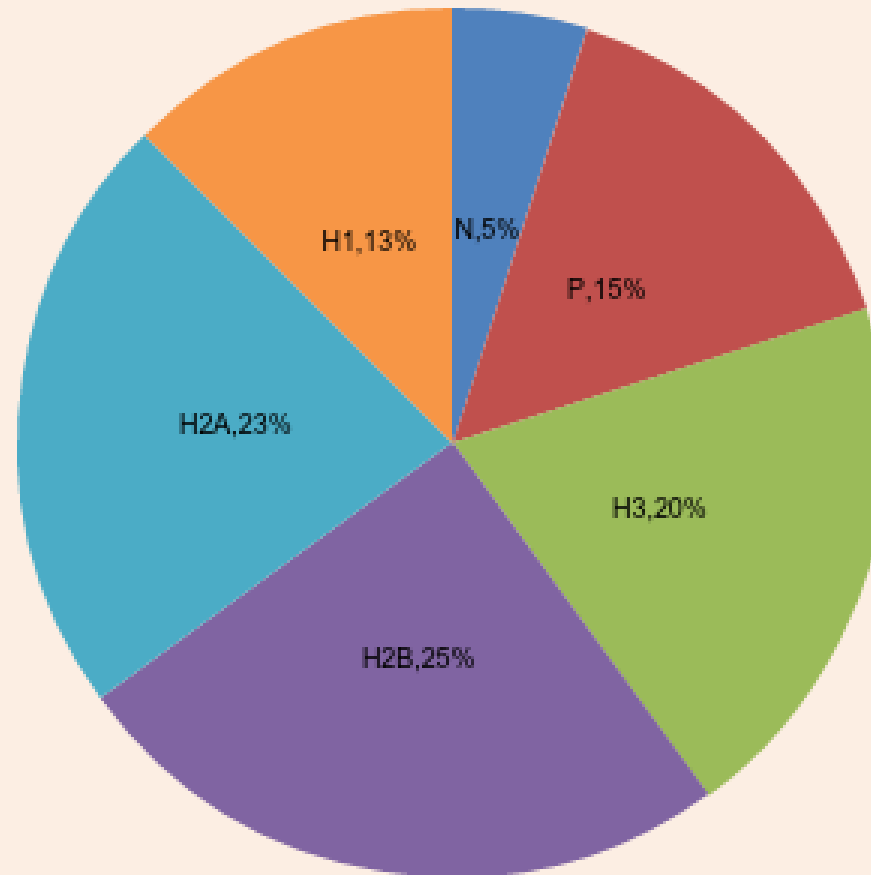
- eg. Definitions:

Name	Type	Length	Description
Course	Alphanum	30	Course ID
Tutorial	Integer	2	Tute no.
Name	Alphanum	30	Student name

- Can include:
  - structure, rules, constraints
- Why do we need Metadata
  - Consistency
  - Meaning
- We generate a **data dictionary** as part of the analysis of system requirements

a large, integrated, structured collection of data

- **Usually intended to model some real-world enterprise**

- **<u>Example</u>: a university**
  - *Entities …* such as courses, students, professors
  - *Relationships …* such as enrollment, teaching

A *Database Management System (DBMS)* is a software system designed to store, manage, and facilitate access to databases.

- A Query Language (e.g. Structured Query Language – SQL or Sequel)

  – Data **Definition** Language (DDL)

    » To define and set up the database

  – Data **Manipulation** Language (DML)

    » To maintain and use the database

  – Data **Control** Language (DCL)

    » To control access to the database
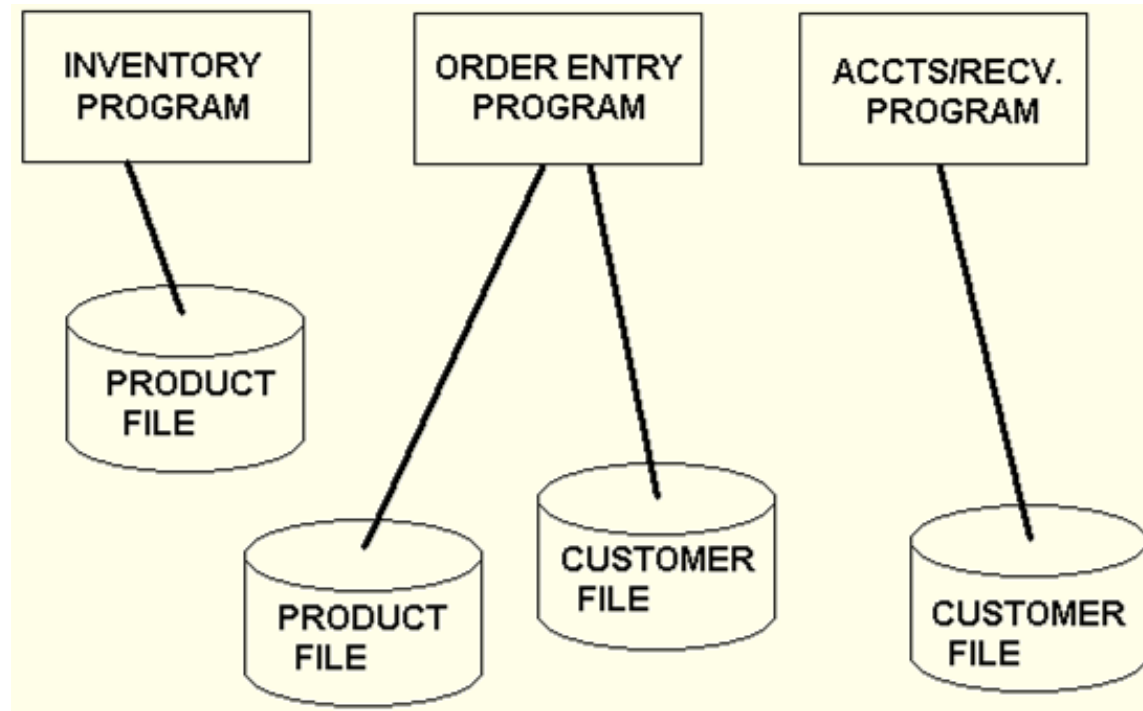
- ## What are the problems you can see with this?
- (Diagram adapted from Hoffer p. 42)

- **Program-data dependence**
  - If the file structure changes, so does the program
  - Program "knows" too much about low-level data structure
  - What if you change data structure for one program
    - Do you change it for all?
- **Duplication of data**
  - wasteful, inefficient, loss of data integrity
  - loss of metadata integrity
    - same name different data
    - same data different name
- **Limited data sharing**
  - data tied to application, hard/slow to create adhoc reports

What's Wrong Here?

- Lengthy development times
  - application has to do low level data management, figure out file format each time

- Excessive program maintenance
  - up to 80% of development time in traditional file based organisations is for maintenance

Database Systems as Solution

- Manage data in a **structured** way
- Many models (hierarchical, network, etc), but **relational** dominant since ~1980
  - Relational Model
    - Rows & Columns forming Relations
    - Keys & Foreign Keys to link Relations

Enrolled

sid	cid	grade
53666	Carnatic101	5
53666	Reggae203	5.5
53650	Topology112	6
53666	History105	5

Students

sid	name	login	age	gpa
53666	Jones	jones@cs	18	5.4
53688	Smith	smith@eecs	18	4.2
53650	Smith	smith@math	19	4.8
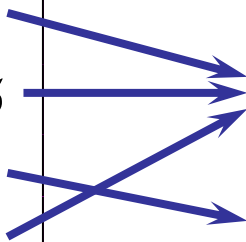
- A **data model** is a collection of concepts for describing data

- A **schema** is a description of a particular collection of data, using a given data model

- The **relational model of data** is the most widely used model today
  - Main concept: **relation**, a table with rows and columns
  - Every relation has a **schema**, which describes the columns, or fields (metadata)

- **Views** describe how users see the data

- **Conceptual** schema defines logical structure

- **Physical** schema describes the files and indexes used

Users

| View 1 | View 2 | View 3 |

Conceptual Schema

Physical Schema

DB

Database Advantages
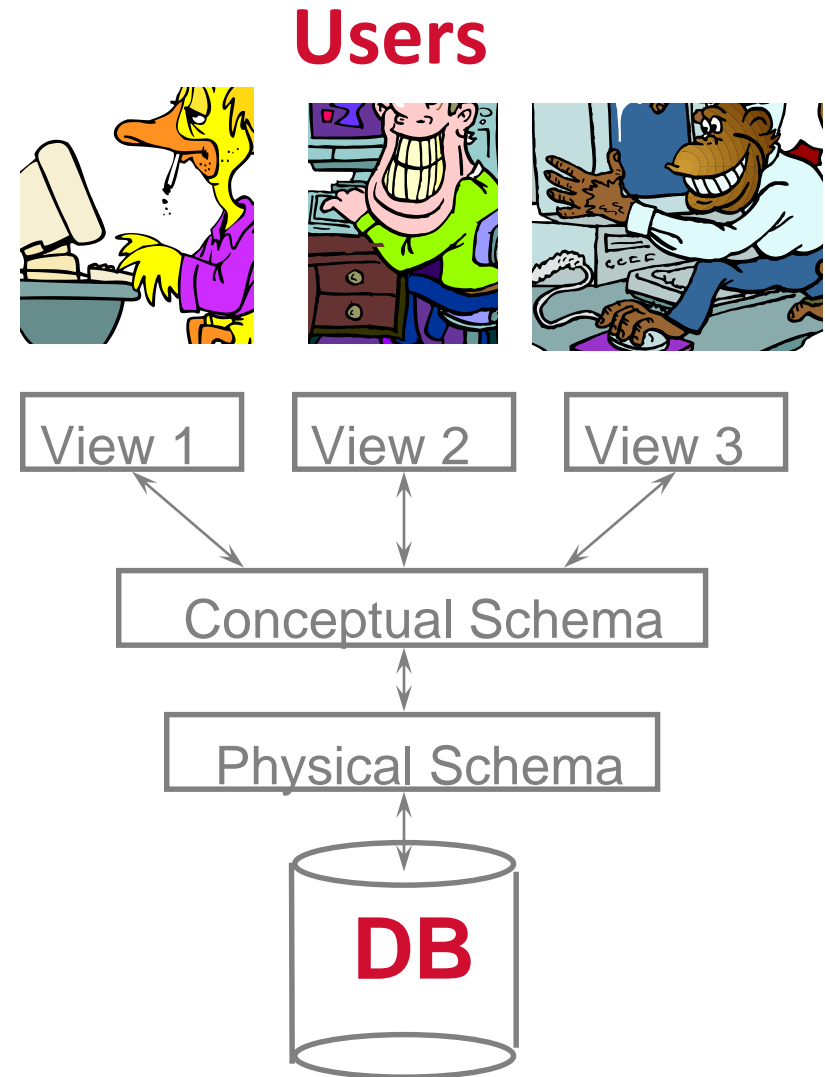
- Data Independence
  - separation of data and program, application logic
  - central data repository, central management
- Minimal Data redundancy
  - redundancy can be controlled (normalization)
- Improved data consistency
  - single store: no disagreements, update problems, less storage space
- Improved data sharing
  - data is shared, a corporate resource, not a necessity for an application
  - external users can be allowed access
  - multiple views of data, arbitrary views of data
- Reduced program maintenance
  - data structure can change without application data changing
    - external *Views* can be maintained, despite changes to the underlying physical structure.

- Increased productivity of application development
  - data already collected & structures already known
  - DBMS provides many tools (particularly SQL language) to help access and manipulate the data…advantages
- Enforcement of standards
  - centralised data management
  - documented policy for data management
  - data definition and dictionary (metadata)
- Improved *data quality*
  - constraints built into the database
- Novel ad hoc data access 'without programming'
  - SQL
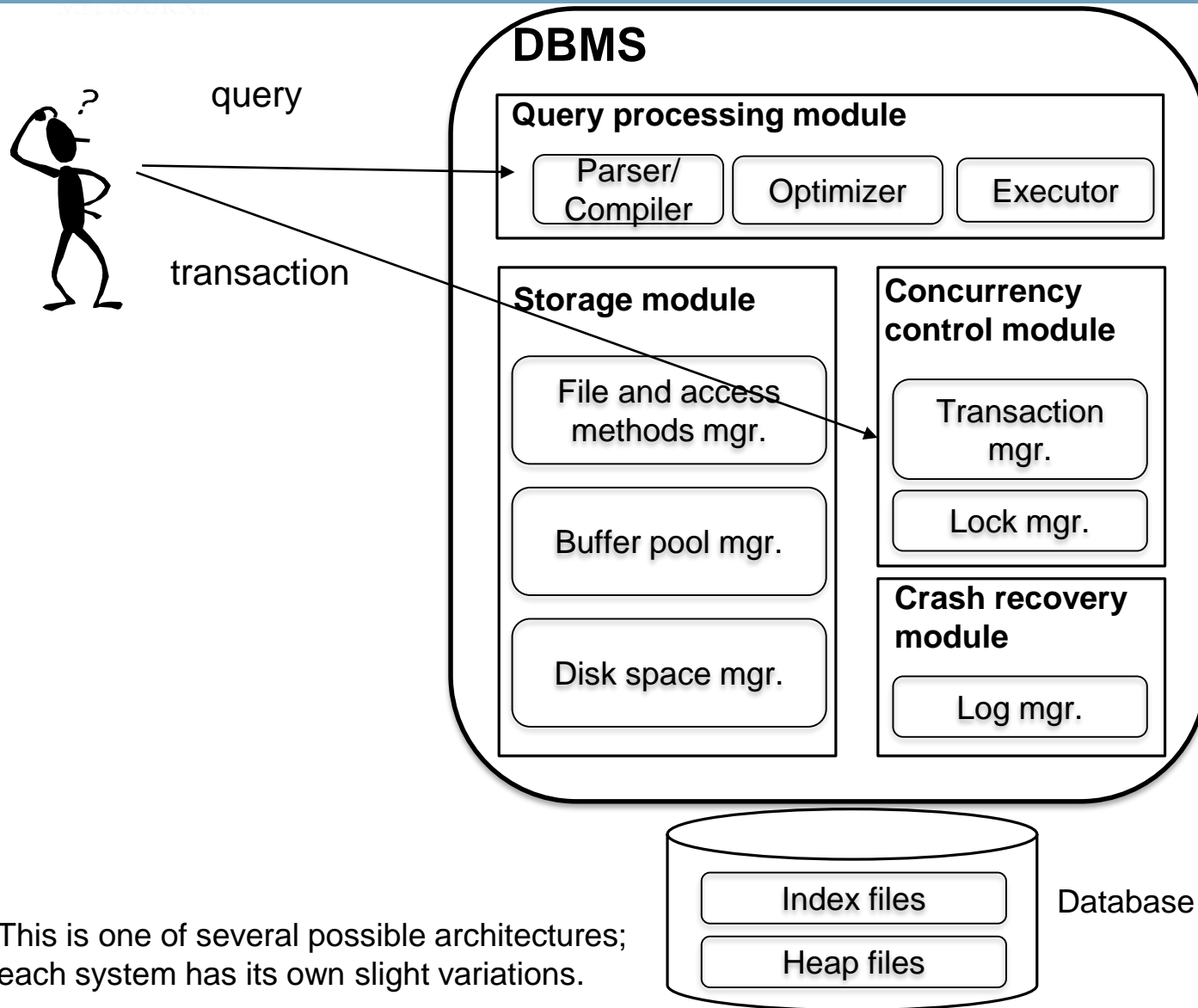
This is one of several possible architectures; each system has its own slight variations.

- DBMS used to maintain, and query large datasets
  - can manipulate data and exploit *semantics*
- Other benefits include:
  - recovery from system crashes,
  - concurrent access,
  - quick application development,
  - data integrity and security
- Levels of abstraction provide data independence
- In this course we will explore:
  1) **How to be a sophisticated user of DBMS technology**
  2) **What goes on inside the DBMS**

- We are using MySQL as the DBMS in this subject
- You can download and install MySQL at the following location
  - http://dev.mysql.com/downloads/
- You will be given a server name and address, along with a username and password for the database server we will be using in the subject in one of the labs in the next few weeks
- Whilst we encourage you to set up and use your own MySQL database on your own computer, you must complete all assessable work on our server!
  - Bear in mind, that saying that you couldn't hand in your assignment because you were trying to get it to work on our server is not an excuse for an extension.
  - Don't leave things until the last minute and there won't be any problems. (also remember Murphy's Law – the server will crash or be slow on the weekend before your assignment is due…)

- **Introduced you to the subject**
  - You now know when everything is due in and how much it is worth

- **Introduced you to the concept of data to information**

- **Introduced you to databases**
  - Why they are better than files
  - What a DBMS is
  - What is MySQL and what is SQL

- Difference between Data, Information and Knowledge
- Being able to discuss the drawbacks of file based systems
- Being able to discuss the advantages of Databases
- Database models, components, abstraction levels

- The database system lifecycle
  - With a focus on the design stage
    - Conceptual design
    - Logical design
    - Physical design

Enhancing career prospects for women in Engineering

Are you a female interested in undertaking an internship and not really sure where to start?

Come along to an information session to find out why gaining experience in the workplace during your Bachelor of Science degree can increase your graduate employability.

Hear from your academic champion for women in engineering, **Professor Sandra Kentish**, about what an internship can mean for you and how to access additional support to get you started.

Register through Careers Online:
https://careersonline.unimelb.edu.au/students/events/detail/997085

Time: 12-1pm
Date: Friday 28 July
Venue: C2 Theatre, Engineering C