INFO20003 Database Systems

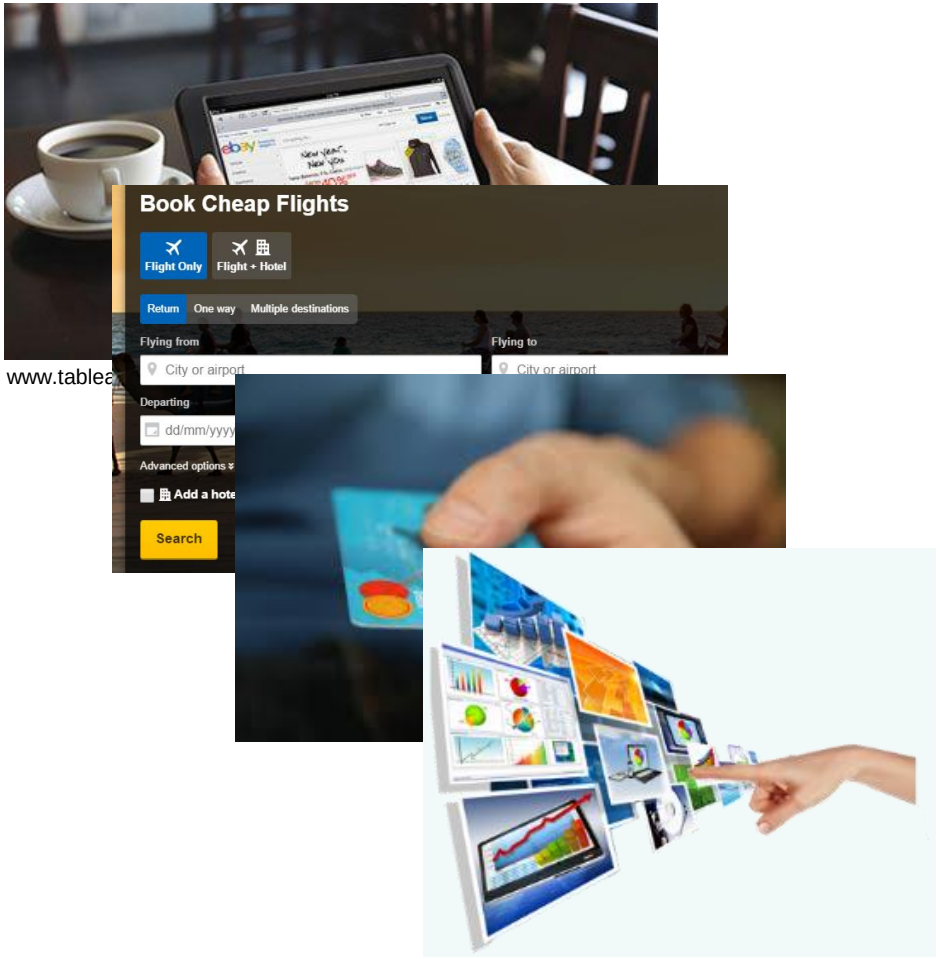
Dr Renata Borovica-Gajic

Lecture 01 – Subject Introduction
What are Database Systems?

Semester 1 2018, Week 1

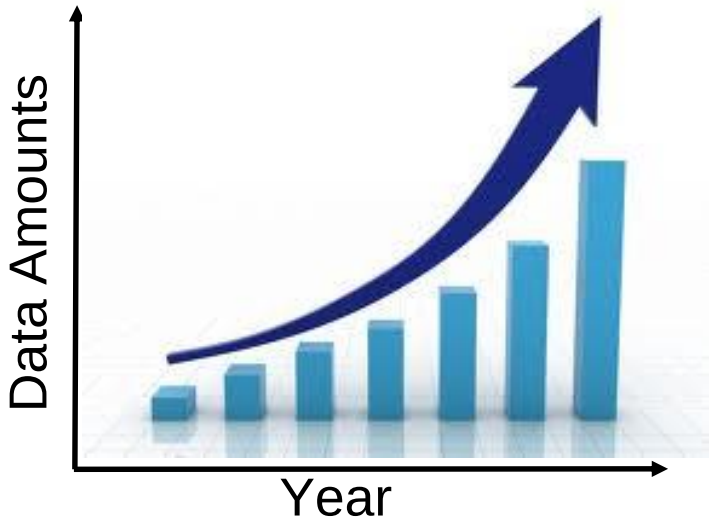
Data, data everywhere…

[The Economist]



www.tablea

http://reportlogix.com/reporting.html



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…

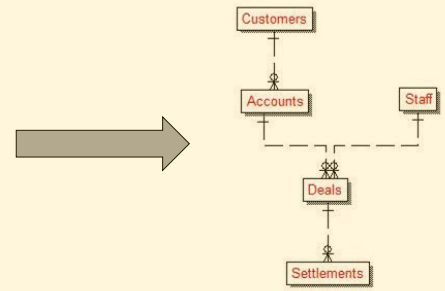
Finding useful information

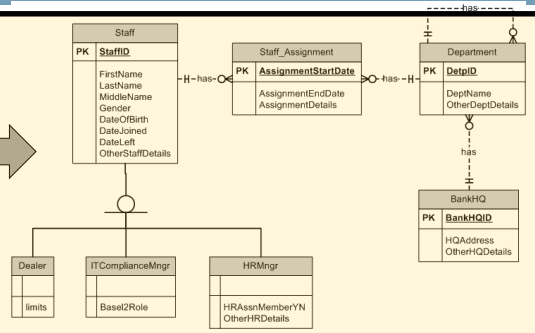
Equals to finding the needle in a haystack

What this subject is all about

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.

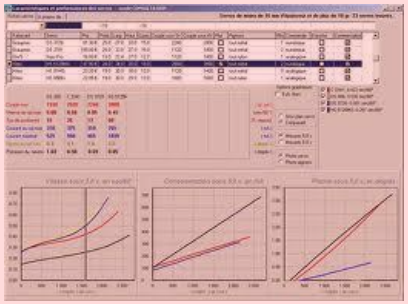




MODELLING

SQL

ARCHITECTURE / INTERNAL WORKINGS



Results

SQL Queries

```
select val from sales
where id = max;
```

Process

Access

Store

Database System

Subject scope

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

2. 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 (plenty of job opportunities)

In this lecture

- Introduction to INFO20003 Database Systems
  - Organization
  - Week by week plan
  - Assessment
- Introduction to Databases
  - From files to relational database
  - Query Language (SQL)
  - DBMS (MySQL)

Organization

1. LECTURES — Teach concepts

2. TUTORIALS — Apply

3. LABS — Practice

COMPLEMENTARY

Assessments:

1. Assignment 1: 10% (ER modelling)
2. Assignment 2: 10% (RA & SQL)
3. Assignment 3: 10% (Query Processing/ Optimisation)

} Hurdle 1 (15%)

4. Mid-Semester Test: 10% (ER modelling, RA, SQL)
5. Final Exam: 60%

} Hurdle 2 (35%)

BOTH REQUIRED

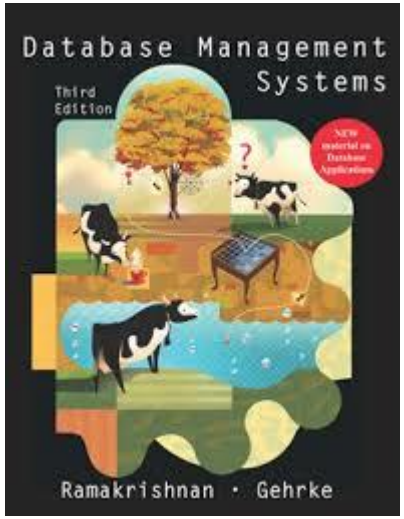
Week by week schedule

- **LMS -> Subject information** (up to date with links to material)
- Here: allow minor changes

Week	Lecture 1	Lecture 2	Tutorial	Lab	Assessments
1. 26 Feb	1. Introduction to the Subject and Database Systems	2.The Database Development Process	My SQL Overview/ Installation		
2. 5 Mar	3. Introduction to ER Modelling	4. Relational Model & Introduction to SQL	Tutorial: Introduction to Database Development	Lab: ER modelling with MySQL workbench	
3. 12 Mar	5. ER Example with MySQL Workbench	6. Extended Entity Relationship Models	Tutorial: Conceptual and Logical Modelling (ER)	Lab: ER modelling examples	A01 ER post
4. 19 Mar	7. Relational Algebra	8. SQL	Tutorial: ER and EER	Lab: EER & Introduction to the SQL case study	
5. 26 Mar	9.SQL Summary	10.Storage and Indexing	Tutorial: Relational Algebra and translation to SQL	Lab: Referential Integrity (scott.sql), SQL	**A01 ER DUE/** A02 SQL & RA post
Break 2 Apr	Mid Semester Break No Class	Mid Semester Break No Class	Mid Semester Break No Tutorial	Mid Semester Break No Lab	
6. 9 Apr	11.Query Processing-Part 1 (Selection & Projection)	12.Query Processing-Part 2 (Joins)	Tutorial: Indexing and Storage	Lab: More SQL Skills	
7. 16 Apr	13. Query Optimization-Part 1	14. Query Optimization-Part 2	Tutorial: Query Processing	Lab: Even More SQL Skills	**A02 SQL & RA DUE**
8. 23 Apr	ANZAC Day Holiday NO LECTURE	16. Normalization	Tutorial: Query Optimization skills	Lab: Query Optimization using Execution Plan	A03 QP/QO post
9. 30 Apr	**MID SEMESTER TEST**	18. Database Administration	Tutorial: Normalization	Tutorial: Normalization	**MST**
10. 7 May	19. Transactions	20. Data Warehousing	Tutorial: Database Admin: backup and Recovery	Lab: Database Admin: backup and Recovery	
11. 14 May	21. Introduction to NoSQL	22. Database applications	Tutorial: Data Warehouse Tutorial	Lab: Transaction exercise using MySQL workbench	**A03 QP/QO DUE**
12. 21 May	23. Cutting edge research in databases	24. **Wrap up and Review**	Tutorial: NoSQL and exam FAQs	Lab: MongoDB	

Teaching material

- Lectures and lecture notes
- Tutorials
- Labs
- Textbook: Ramakrishnan and Gehrke, 3rd Edition (if you wish)



Staff & Interaction

- **Lecturer**: Renata Borovica-Gajic
  - Email: renata.borovica@unimelb.edu.au
- **Senior tutor**: David Eccles (assessments: contact for extensions)
  - Email: eccles.d@unimelb.edu.au
- **Tutors**: Farah, Alan, Mina, Ibrahim
  - Consultation hours will be posted shortly
- **Interaction**:
  - Attend classes & workshops
  - Use LMS discussion forum
  - Email to us (only if personal)
  - Our emails to you (for assignments clarifications – to guide you)

Your Engagement…

- In the lectures:
  - Turn up
  - Be prepared to answer the questions we put to you
    - GUESS – it's a great preparation for the exam
    - 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

Now your turn…

What do you expect from the subject?

Introduction to Database Systems

Readings: Chapter 1, Ramakrishnan & Gehrke, Database Systems

Data, Information, Knowledge

- 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

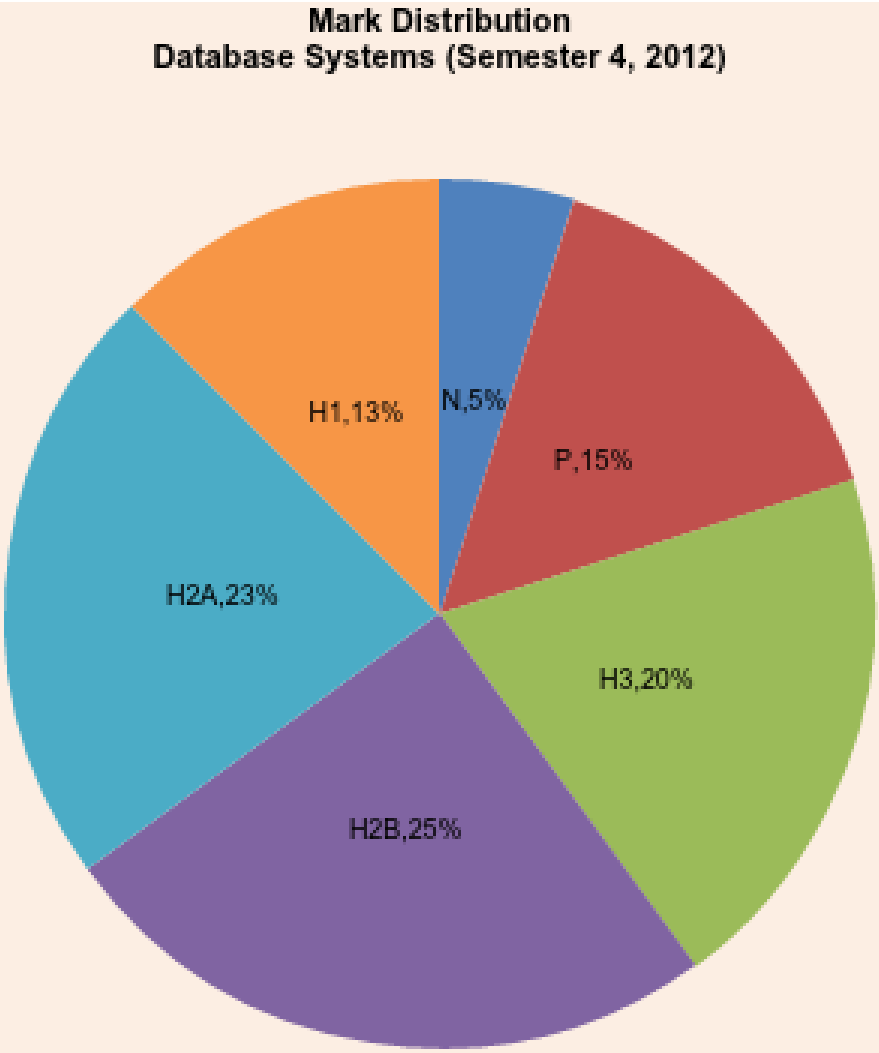
Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145
Sisneros, Michael	392416582

Information: data in context

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

Information: Summarisation!



Metadata - Data about data

Name	Type	Length	Description
Course	Alphanum	30	Course ID
Tutorial	Integer	2	Tutorial number
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

What is a “database” ?

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

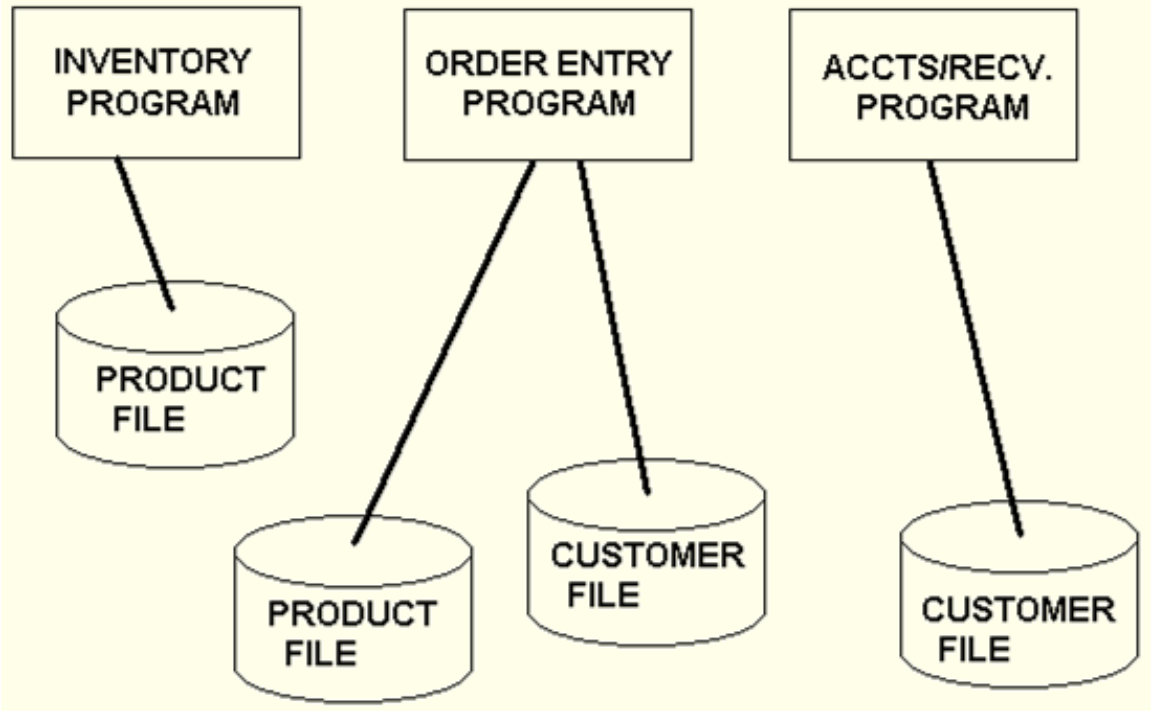
What Is a Database System?

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

How do we manipulate with data?

- 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

File Processing Systems



- What are the problems you can see with this?
- (Diagram adapted from Hoffer p. 42)

What's Wrong Here?

- Program-data dependence
  - If the file structure changes, so does the program
  - What if you change data structure for one program
- Duplication of data
  - wasteful, inefficient, loss of data integrity
- Limited data sharing
  - data tied to application, hard/slow to create adhoc reports
- 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

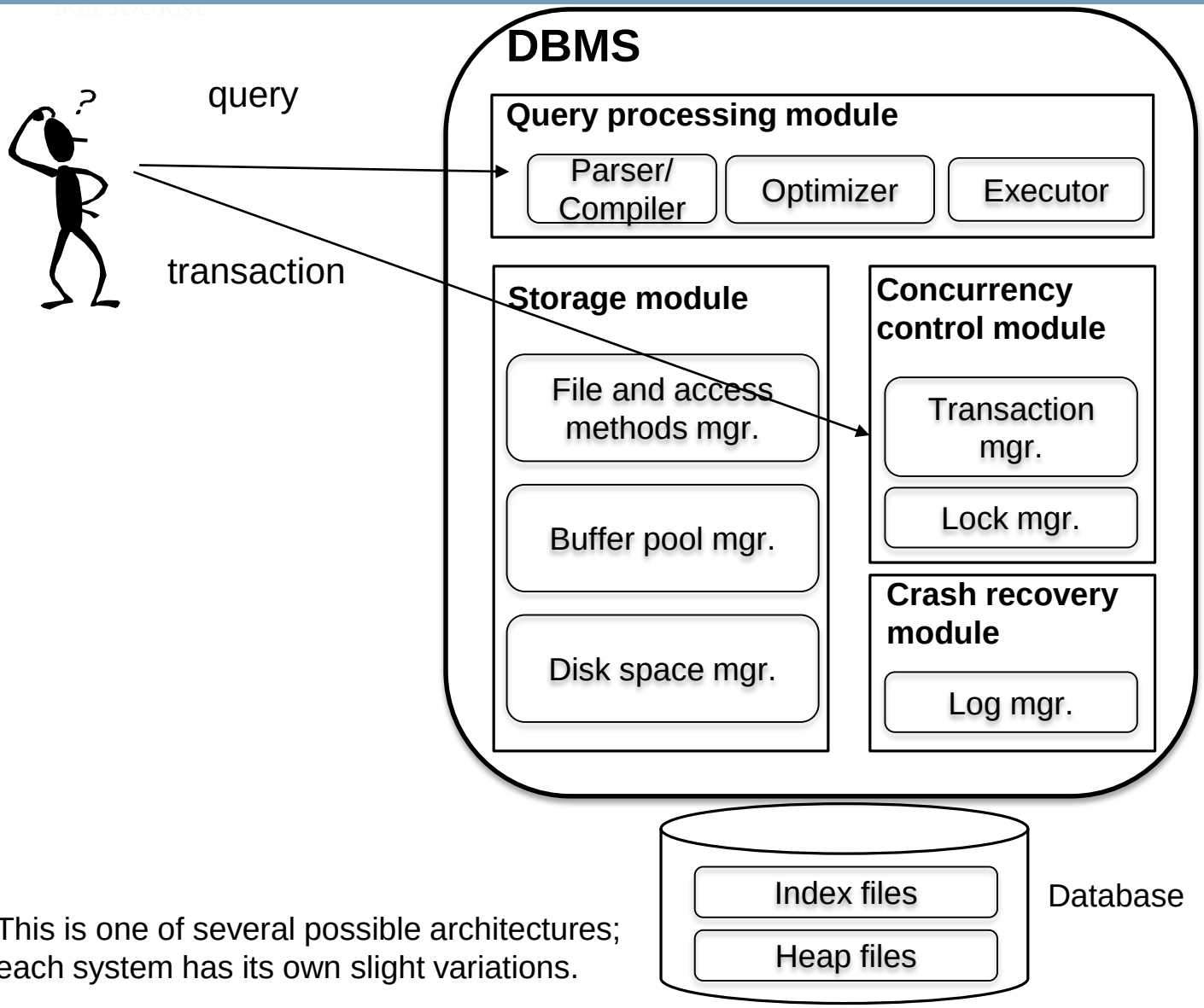
Students

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

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
- Novel ad hoc data access ‘without programming’
  - SQL

Components of a DBMS



This is one of several possible architectures; each system has its own slight variations.

DBMS Summary

- 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

- In this subject we will explore:
  1) **How to be a sophisticated user of DBMS technology**
  2) **What goes on inside the DBMS**

MySQL

- 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
- *Turn up to the first hour of this weeks tutorial: we will help you install MySQL*

What's Examinable?

- Difference between Data, Information and Knowledge
- Being able to discuss the advantages of Databases vs File Systems

Next Lecture

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