COM106: Introduction to Databases

Introduction and Module Overview

So …. **Why** are we learning about Databases?

First - What is 'data'?

Dictionary Definition:-
- facts and statistics collected together for reference or analysis.

Computer data is information processed or stored by a computer.

May be in a *variety of forms* from bits and bytes to text documents, images, audio clips, software programs, or other types of data.
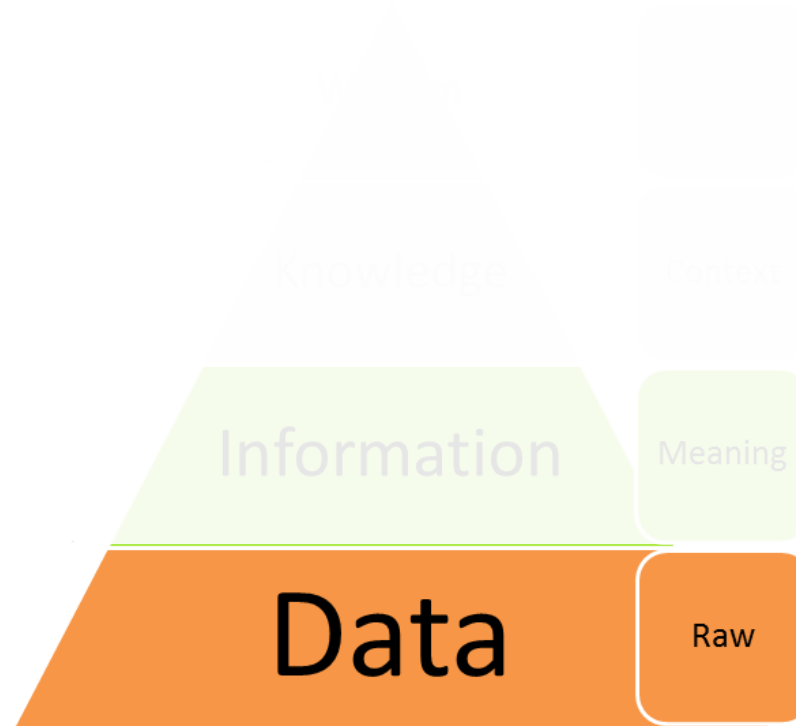
What is a 'database'?

A database is an *organised* collection of data.

Information is held such that it can easily be accessed, managed, and updated.

So…. **Why** are we learning about Databases?

Data – Information – Knowledge - Wisdom

So.... **Why** are we learning about Databases?

- Databases are of *fundamental importance* in a digital world.
- Necessary in any situation where:-
  - information (organised data) is needed ......
  - to build knowledge ......
  - and make decisions about future actions (wisdom)
- *Databases are also of fundamental importance in Computing and Business, and something all Computing/Business Professionals need to know about.*

So.... **Why** are we learning about Databases?

There's an unimaginably large amount of data in the digital universe :
- estimated at **16 Zettabytes** (10^{21} Bytes) in 2017, 90% of which was created in the last two years.
    - *For more info, and some truly staggering stats, google '**amount of data in the world 2023**'*
- Increasing need to analyse data in **near real-time** to derive **value** from it

Look up :
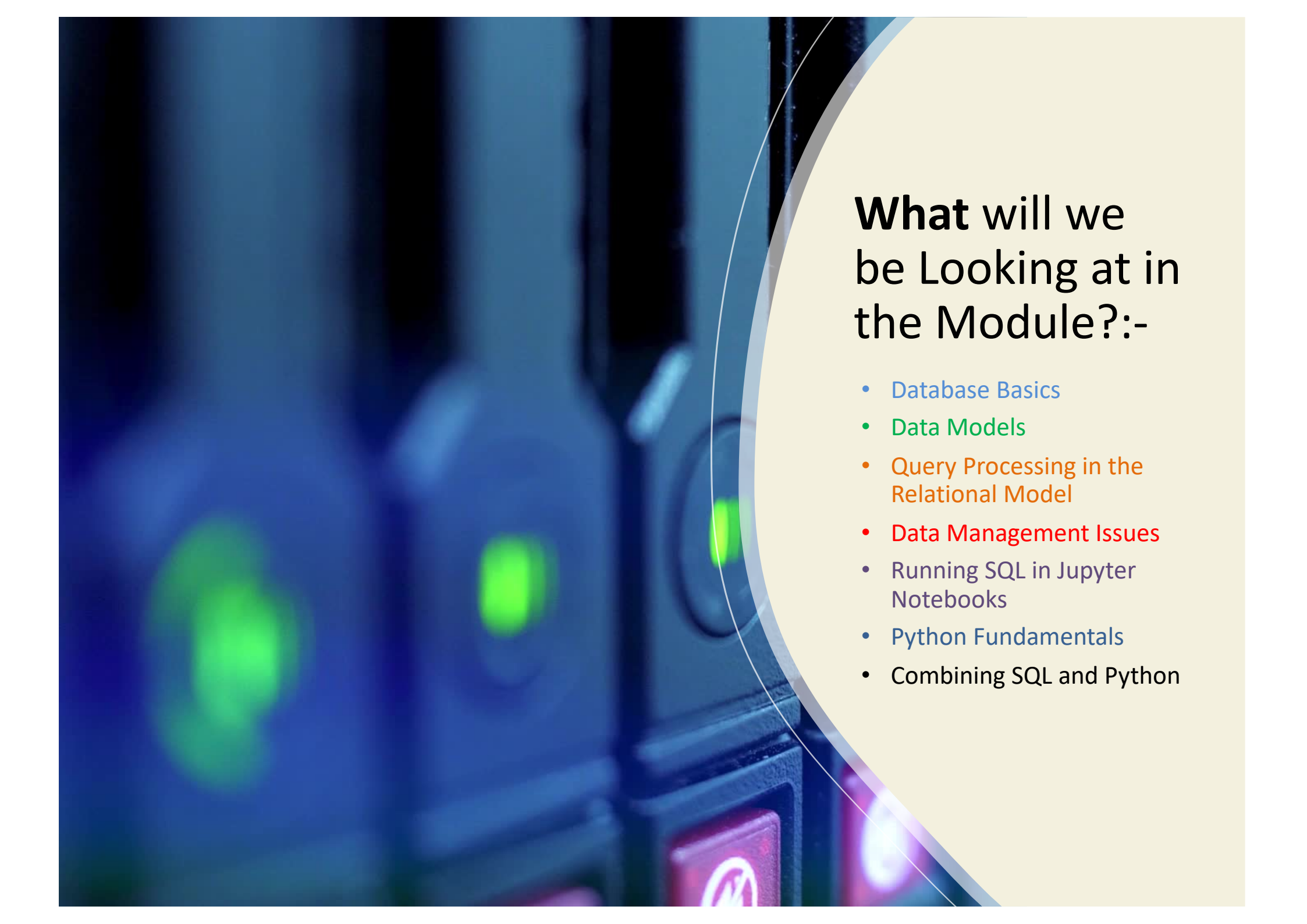- Data Analytics
- Big Data
- Data Mining
- Zettabyte

What will I be able to do at the end of this module?

A successful student will be able to:

- Explain the concepts underlying the design, structure, implementation, manipulation and management of relational databases in modern information systems.

- Apply appropriate models, processes and techniques in the design, implementation and use of database systems to meet an outline business scenario.

- Construct a database from a given relational schema using a commercial database management system.

- Perform basic data manipulation and information retrieval operations using SQL.
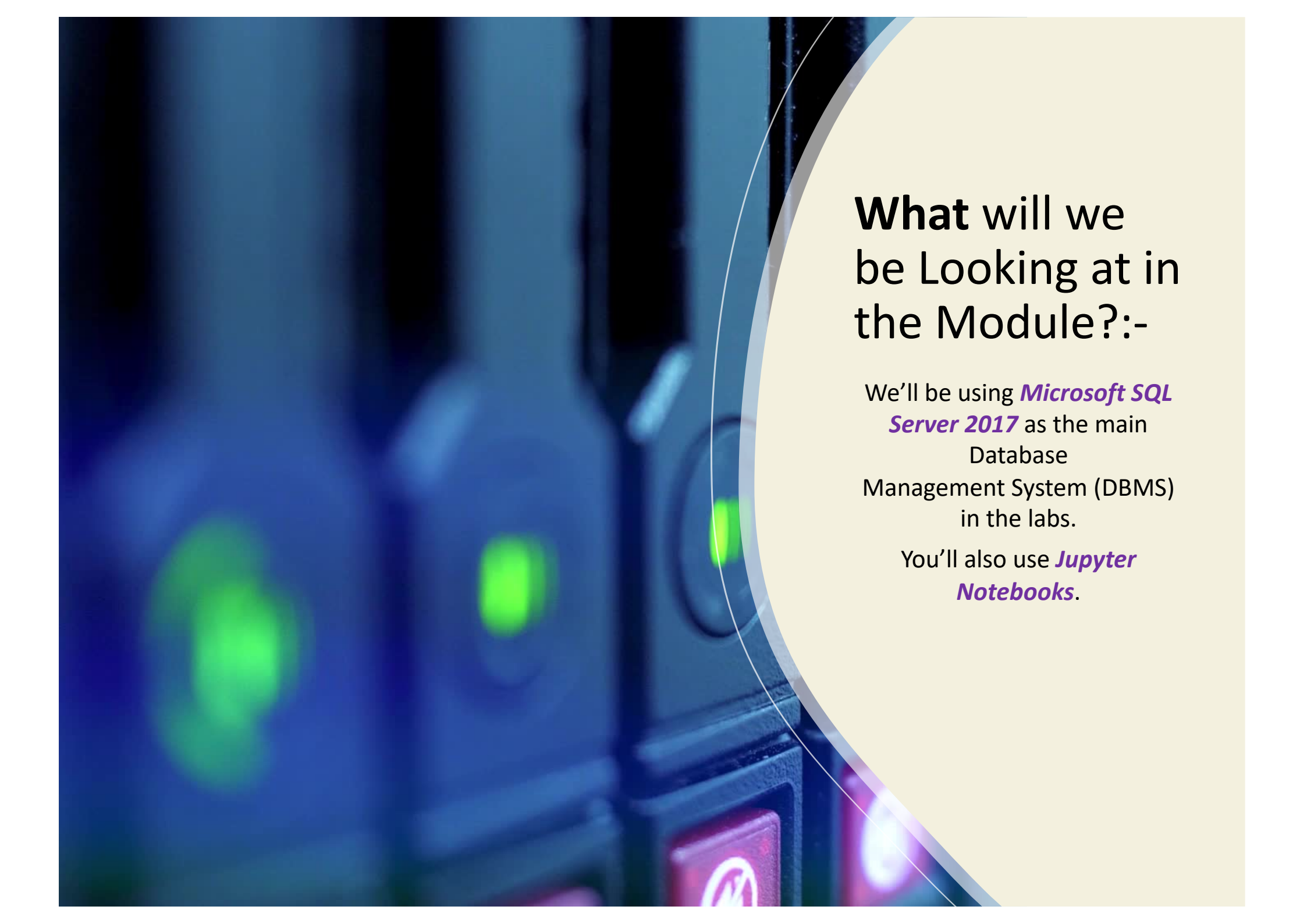
You will have:

- an understanding of the theoretical basis, structure, management and workings of database systems and provide some insight into recent developments and future trends

- a knowledge and understanding of the techniques of database design and the ability to apply these techniques to implement such designs on a relational database management system

- an understanding and practical experience of data manipulation and query in SQL

- experience in using a commercial database management system.

What will we be Looking at in the Module?:-

- Database Basics
- Data Models
- Query Processing in the Relational Model
- Data Management Issues
- Running SQL in Jupyter Notebooks
- Python Fundamentals
- Combining SQL and Python

What will we be Looking at in the Module?:-

We'll be using *Microsoft SQL Server 2017* as the main Database Management System (DBMS) in the labs.

You'll also use *Jupyter Notebooks*.

How Do _**YOU**_ Learn the Material

Lots of material and support available:-

- Lectures & Practicals – attendance is compulsory and <mark>engagement</mark> is monitored
- 'Homeworks' – every week, in preparation for the next week.
- Blackboard Learn: https://learning.ulster.ac.uk
- Library – books suitable for an Introductory Database Course
- Lots & Lots of external on-line resources available:-
  - books, blogs, videos, entire courses, simulators, etc.
  - we'll be using some and recommending others

This module requires <mark>200 hours total effort</mark> (at least)

- Lectures - 24 hrs
- Lab - 36 hrs
- Independent Study - _**140 hrs.**_ Plan to allocate between <mark>ten to twelve hours for private study per week</mark>.

How – Socrative

- Socrative is a student response system that allows students to actively engage during lectures:- http://www.socrative.com/

- We'll be using it throughout the course.

- You'll need to bring your phone, tablet, laptop, etc with you to class.

- Ensure you can connect to the wi-fi network (eduroam):

- *For Help See -* **https://www.ulster.ac.uk/ds/services/wifi-services/eduroam-wifi**

- Go to your app store and search for 'socrative' *(or use the web version)*.

- Download and install – make sure you install the student version.

Students enter their teacher's unique **Room Code**

Students will see this waiting screen until the teacher starts an activity

Okay ….. Let's Explore ….

Go to Socrative.com and enter the Room Code
COM106

Spot Questions:

True or False:

Data are discrete objective facts that have no meaning in isolation

Fill in the blank:

A database is a(n) _____ collection of data.

Example of an in Class Quiz

1. What is the world's longest river?

    A - The Mississippi River

    B - The Nile River   Use Room Code **COM106**

    C - The Danube River

    D - The Amazon River

    E - The Yangtze River

2. The diameter of Earth is 8,000 miles (12,875 km)?

    A - True   B - False

3. This is an outline of which country?

    A - France

    B - Italy

    C - Canada

    D - Vietnam

    E - Sweden

4. Iceland is the world's biggest island.

    A – True   B - False

5. The world's largest ocean is the _____ ocean.

How – Books - Library

Any introductory text on Databases will cover the material in this module.
  • There are lots and lots available – both hardcopy and electronic

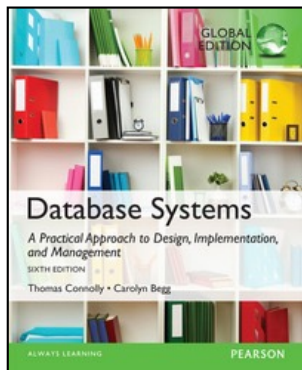Check out the library catalogue available through the portal
  • or directly at http://catalogue.library.ulster.ac.uk/

And the Subject guide for Computing (access through the portal)
  • or at http://guides.library.ulster.ac.uk/computing

There are a number of *recommended text books* for the module. It's not essential that you buy one, but it **IS ESSENTIAL** that you regularly use one!

Be aware - the slides *won't* give all the information
you'll need for the module.

Database systems: a practical approach to design, implementation, and management
Connolly, Thomas M; Begg, Carolyn E
Pearson Education, 6th ed., Global ed. 2015
ISBN: 1292061189, 9781292061184
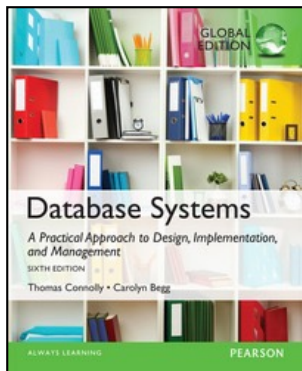Shelfmark: 005.74/CON

Hard copy available in the library
Also available as an ebook
Earlier editions are also available

How – Books - Library

There are a number of ***recommended text books*** for the module. It's not essential that you buy one, but it **IS ESSENTIAL** that you regularly use one!

Be aware - the slides *won't* give all the information you'll need for the module.

Database systems: a practical approach to design, implementation, and management
Connolly, Thomas M; Begg, Carolyn E
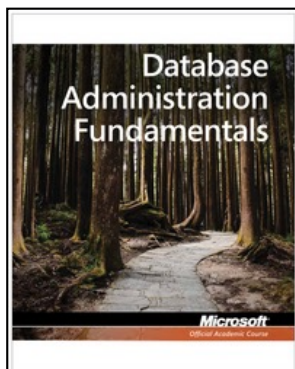Pearson Education, 6th ed., Global ed. 2015
ISBN: 1292061189, 9781292061184
Shelfmark: 005.74/CON

Hard copy available in the library
Also available as an ebook
Earlier editions are also available

More details are available
in the **Additional Resources Folder**

These are the main text books that will be used and referenced throughout the module

Database administration fundamentals, exam 98-364
Gambrel, Bryan; Microsoft Corporation
Wiley, 2011
ISBN: 0470889160, 9780470889169
Shelfmark: 005.74/MIC

An electronic copy is available on Blackboard (Additional Resources Folder)
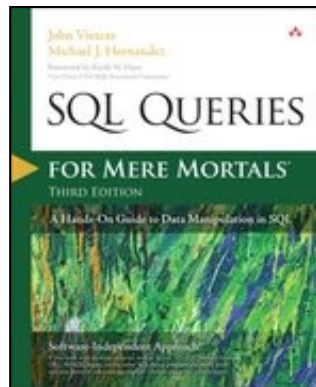
How – Books - Library
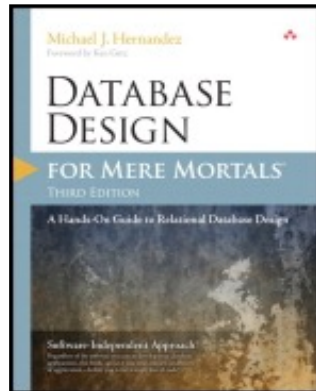
Database Design for Mere Mortals®: A Hands-on Guide to Relational Database Design
Michael J. Hernandez
Addison-Wesley Professional, 3rd ed. 2013
ISBN: 0-321-88449-3, 0-13-312228-X

Available in the library as an ebook

SQL Queries for Mere Mortals®: A Hands-On Guide to Data Manipulation in SQL
John Viescas; Michael J. Hernandez
Addison-Wesley Professional, 3rd ed. 2014
ISBN: 0-321-99247-4, 0-13-382488-8

Available in the library as an ebook

How – Books - Library

Any introductory text on Databases will cover the material in this module.
- There are lots and lots available – both hardcopy and electronic

Check out the library catalogue available through the portal
- or directly at http://catalogue.library.ulster.ac.uk/

And the Subject guide for Computing (access through the portal)
- or at http://guides.library.ulster.ac.uk/computing

How – On-line Resources

Lots and Lots of external on-line resources available:-
- books, blogs, videos, entire courses, simulators, etc.
- we'll be using some and recommending others

Easily found – search Google, YouTube, etc

As we go through the course, I'll be posting the resources we recommend in the Additional Resources Folder in Blackboard.

If you come across something new you think would be of benefit to other students - let me know and I'll put it up.

How – Get Your Own Copy of SQL Server 2017

Microsoft SQL Server 2017 Express Edition - the main relational database management system used throughout this module

Download and Install instructions are available in the Lab Resources Folder on Blackboard

DANGER
EXPECTATIONS

What You can Expect of Us

- Lots of learning material – easily accessible

- **Interactive classes**

- Answers to your questions - **ask**

- Monitoring your **engagement**

- Help when you need it – **ask**
    open door – but preferably email for an appointment

- Prompt feedback

What We Expect of You

- **Engage** – as an **Independent Learner**

    - Attend classes – on time

    - Contribute

    - Ask Questions

    - Complete all coursework – EC1

    - Use the available resources as directed and in your own independent study

---

We are here to facilitate

YOUR learning

How is the Module Assessed?

100% Coursework - Two pieces; Set Exercise and Assignment

Assignment (70%) - Given an outline business scenario and an unstructured file of sample data, design and implement a database capable of meeting the business need.

Submission, by **12:00 noon Friday 15th December**, of:

i) a written report, including the main design artefacts and code developed,

ii) a short video screencast showing a walkthrough demonstration of the code and its execution against a set of given test cases.

Set Exercise (30%) - A 50 min class test, held during allocated labs scheduled in Week 13 (**Week Commencing Monday 8th January 2024**).

Individual Assessments. No Group Work

How is the Module Assessed?

To Pass –

> **40% or more** aggregated over both coursework components.

It is not necessary to pass each component separately.

Practical Exam (30%)	Assignment (70%)	Module Mark	Pass/Fail
60%	50%	53%	Pass
75%	25%	40%	Pass
10%	55%	42%	Pass
39%	40%	40%	Pass
40%	39%	39%	Fail
80%	20%	38%	Fail

If you miss an assessment (**for valid reasons**) – submit an **EC1**

If you **FAIL** - Resit failed components in the next year (capped at 40%)

Introduction to Databases and the Relational Model

Data – Information – Knowledge - Wisdom

Decisions – based on Knowledge -
--- based on Data

Knowledge enables understanding or wisdom necessary for effective decision making.
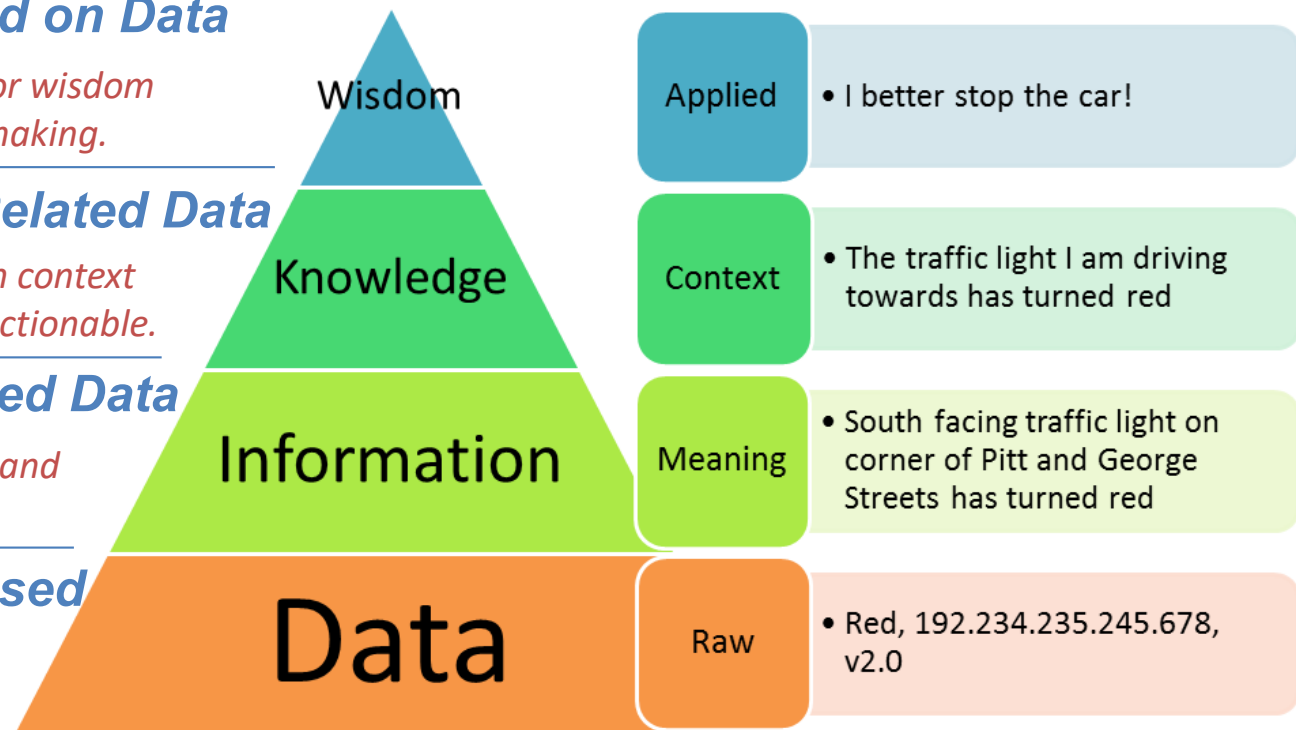
Extracted from Related Data

Knowledge is information placed in context based on facts and meaning – it is actionable.

Related Data

Information is data with relevance and purpose. It 'informs'

Disorganised

Data are discrete objective facts that have no meaning in isolation

Wisdom	Applied	• I better stop the car!
Knowledge	Context	• The traffic light I am driving towards has turned red
Information	Meaning	• South facing traffic light on corner of Pitt and George Streets has turned red
Data	Raw	• Red, 192.234.235.245.678, v2.0

Introduction to Databases & the Relational Model

So, whilst **DATA** is a **disorganized** collection of facts, numbers, audio clips, images, etc ……

a **DATABASE** is:

    a shared **collection** of **logically related data**

    stored in an **organised structure**

    which provides the ability to **interact** with data

    and to extract salient **information and knowledge**.

The Evolution of Databases – Manual Filing Systems

    An organised collection of documents in separate files – logically related

    Held in locked filing cabinets, possibly over different premises

    Data can be manipulated – accessed, deleted, updated, etc

    And can be searched - first entry / indexing system
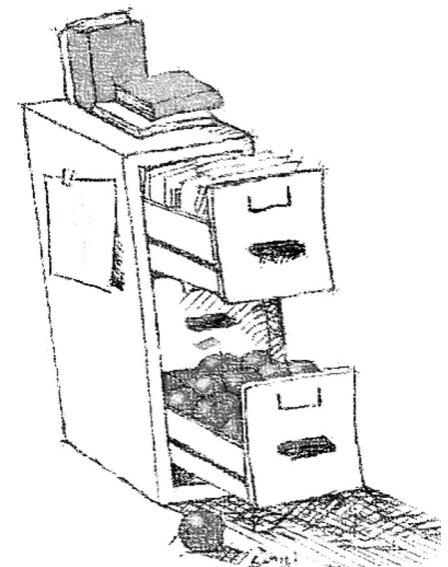
Can work well …. if there are a small number of entries

    or…. a large number of entries, but used to store and retrieve only.

BUT …. can be difficulties for anything more complex

    Data is distributed (**decentralised**) and may contain **duplicates**

    May be difficult to extract **information and knowledge**.
        e.g., Total Annual Salary Bill across all Departments

Introduction to Databases & the Relational Model

The Evolution of Databases – File-Based Approaches

First attempt at computerisation – modelled on the manual filing system

Initially, an application-centred approach, where an application program worked on associated data files – e.g., a payroll system.

Later, files could be shared among applications to reduce data duplication.

PROBLEMS:

No data centralisation (leading to data isolation and separation)

Poor file design can lead to data duplication (or data redundancy).

Updating multiple record copies is inefficient and may lead to loss of data integrity

Data dependence – some data in the system may be dependent on others and should be updated together

Application programs need a knowledge of file storage structures (queries are not independent of the data files or storage structures)

Incompatible file formats

Security (file level vs. data item level)

Data organisation and structure is limited …. so limited queries are possible

Introduction to Databases & the Relational Model

Explanation of Terms – Data Integrity

Accurate data is vital in all information systems

Data integrity is concerned with:

Consistency (is duplicate data identical?)

Correctness (is the stored data accurate?)

Maintaining data integrity involves:

ensuring data is designed correctly (data model)

ensuring data is entered correctly (validation)

ensuring data is free from corruption, modification or unauthorised disclosure (programming errors or hacks)

ensuring data correctness is maintained (backup, recovery, transaction integrity maintenance)

Using
Integrity
Constraints

Explanation of Terms – Data Dependency

A dependency occurs when information stored in a database table (or in a file) uniquely determines other information stored in the same table (or file).

i.e., knowing the value of one piece of data is enough to tell you the value of another in the same table (or file).

Introduction to Databases & the Relational Model

Some Examples

Log into Socrative **(COM106)** at socrative.com (or via the app)

Given the data table below, identify any **data dependencies** by selecting all of the statements that are **TRUE**

    A – Knowing **EName**, uniquely identifies **Manager**

    B – Knowing **Manager**, uniquely identifies **Dept#**

    C – Knowing **Salary**, uniquely identifies **DName**

✔ **D** – Knowing **DName**, uniquely identifies **Manager**

    E – Knowing **EName**, uniquely identifies **Dept#**

    F – None of the above

Is the reverse True?

  Does knowing **Manager** uniquely identify **DName** ?

    No

Does the table contain any more data dependencies?

Yes:-

    Dept# gives **DName** gives **Manager**

Notice the **data duplication** in the table

    Poor database design

Emp#	EName	Salary	Dept#	DName	Manager
1	Smith	15000	A	Sales	Kelly
2	Jones	21000	B	Admin	Whyte
3	Kane	18000	D	Finance	Coyle
4	Dwyer	20000	B	Admin	Whyte
5	Cook	16000	C	Personnel	Woods
6	Healy	15000	A	Sales	Kelly
7	Conroy	19000	B	Admin	Whyte
8	Price	22000	D	Finance	Coyle
9	Murphy	23000	C	Personnel	Woods
10	Jones	16000	A	Sales	Kelly