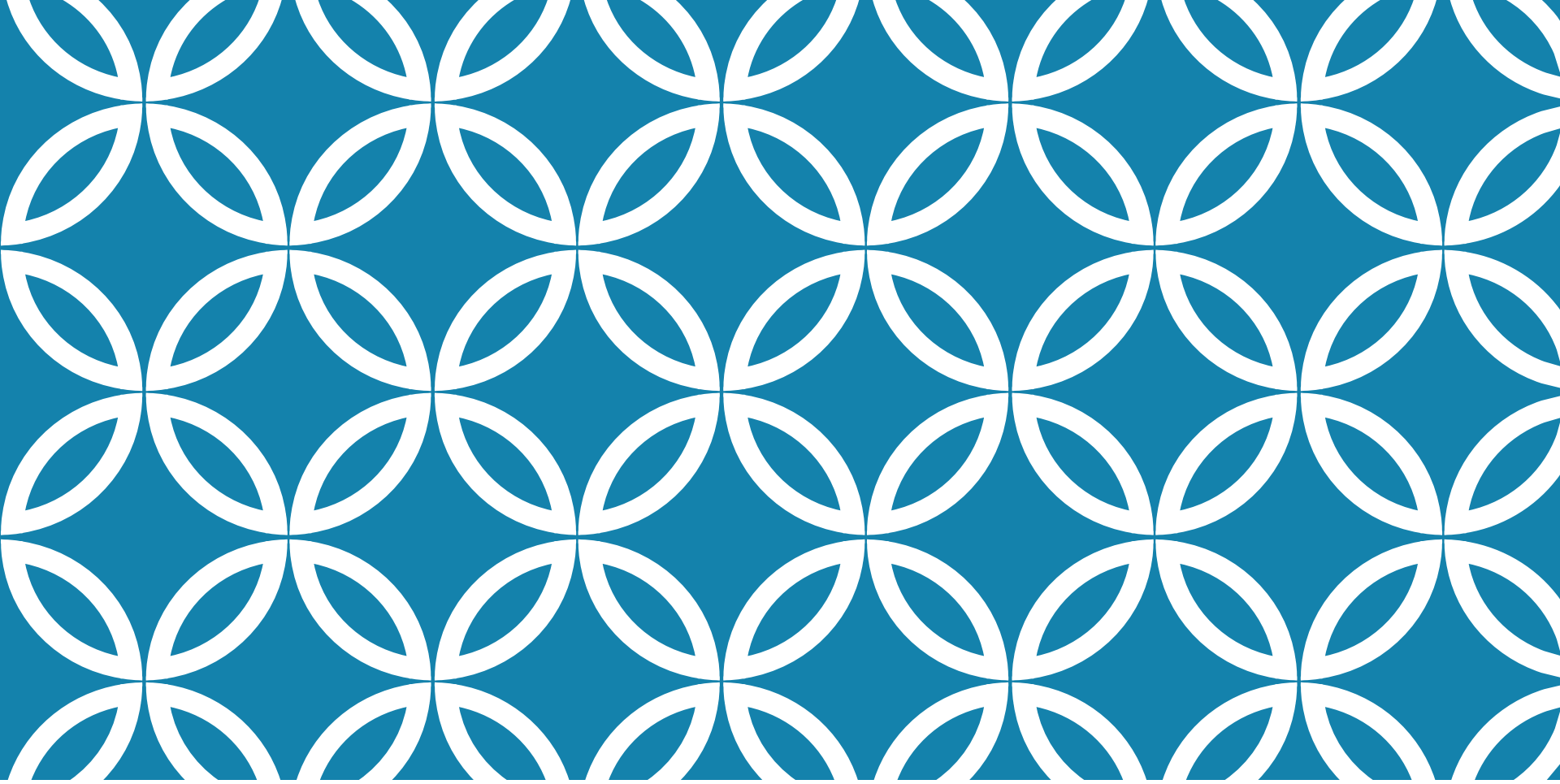


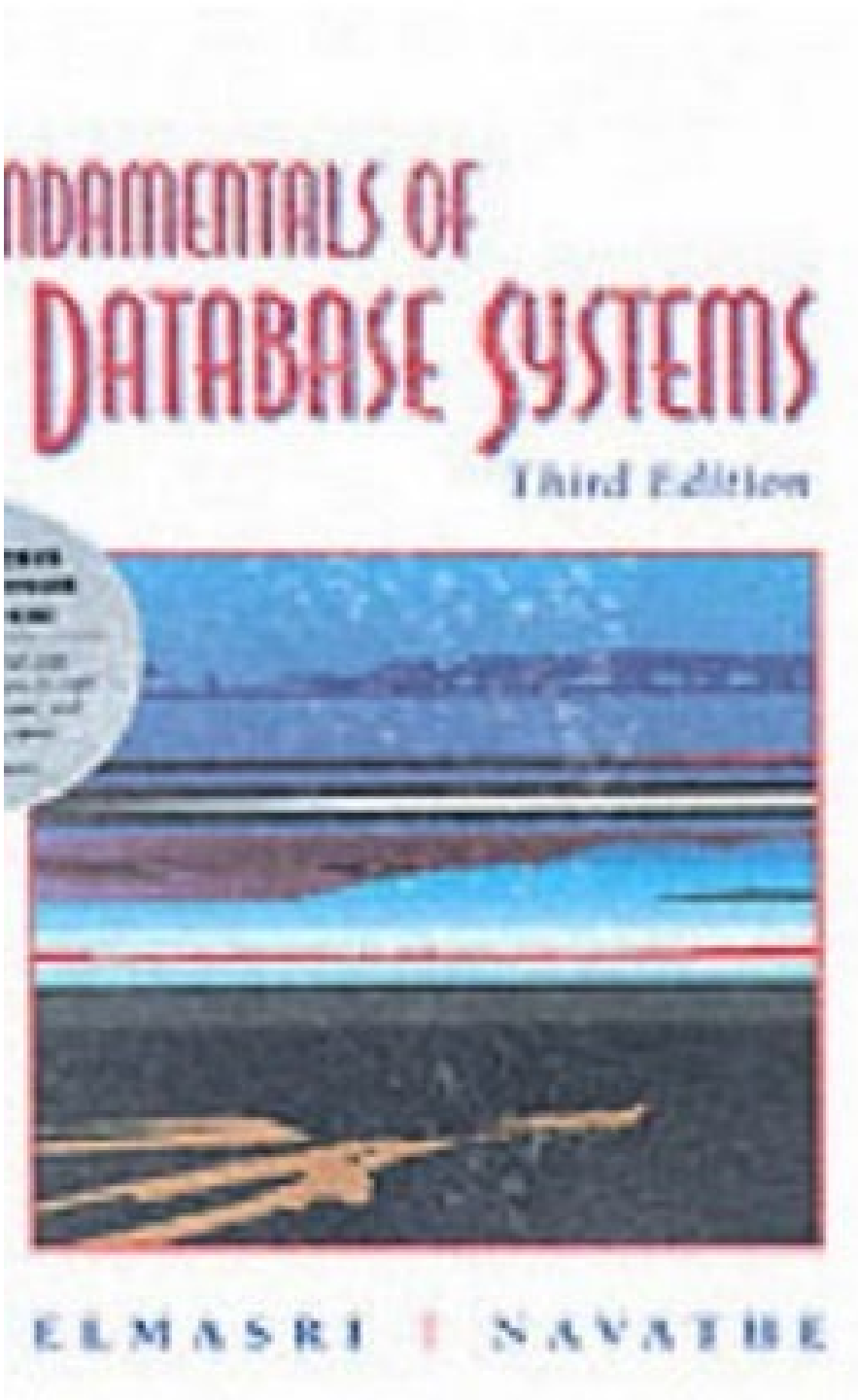
DATABASE SYSTEM INTRODUCTION

Topic: Fundamentals *(1 of 2)*

BLACKBOARD "QWICKLY" SIGN IN . . .

RELEVANT CHAPTER IN CORE TEXT FOR TODAY'S LECTURE:

Chapter 1 & Chapter 2



DATA AND INFORMATION AND KNOWLEDGE

"Drowning in Data,

Dying of Thirst for Knowledge"

STORAGE AND RETRIEVAL

"two sides of the same coin"

HOW DO WE DEFINE "GOOD" DATA AND INFORMATION?

HOME › REGISTRY › ACADEMIC TERM DATES

ACADEMIC TERM DATES

2022-2023

2021-2022

2020-2021

TIMETABLES

Overview

Welcome

Contact Us

Current Students

Academic Calendar
ComputerDISC
Contact Points
Deferral of Examinations
Handbooks
Laptops
Getting Your IT Account
Software Available For

i Timetables We are looking forward to welcoming you to campus in September. Timetables are currently being finalised - the lniks below are for the timetables that have been finalised thus far.

These timetables are based on currently-available information and current national and university guidelines on implementation of public health measures. The most recent timetables can always be found here, and have a date stamp and version number.

iLaptops: NUI Galway School of Computer Science requires all students to have exclusive use of a laptop for use in lectures and labs, for home use of online materials and for participation in online sessions.
Follow this link for a guide for minimum and recommended spec laptops.

B.Sc. in Computer Science and Information Technology

1BCT 2BCT 3BCT 4BCT

AP

National ENQUIRER
HOLLYWOOD'S HOTTEST WEEKLY

GENERAL CHARACTERISTICS OF GOOD DATA AND INFORMATION:

- accurate / error free
- complete
- economical
- flexible

- reliable
- relevant
- simple
- timely
- verifiable
- secure

QUESTION:

Which (top 3) are most important to you in your daily life?

- accurate / error free
- complete
- economical
- flexible
- reliable
- relevant
- simple
- timely
- verifiable
- secure

QUESTIONS:

Which ones (if any) are easy/simple things to guarantee?

- accurate / error free
- complete
- economical
- flexible

- reliable
- relevant
- simple
- timely
- verifiable
- secure

Where does information on the WWW and Cloud fit in?

There are many database systems online and on cloud storage

Database System approaches are not sufficient on their own to deal with more free text data and with all multimedia data, however they are widely used

DATABASE DEFINITION
(CHAPTER 1 E&N)

Collection of
related data
describing some
"mini-world"

DATABASE EXAMPLE

The "mini-world" of a University:

lecturers, courses, modules, schools, colleges, lectures, timetables, students, exams, staff, results, fees, buildings, contractors, research projects, publications, etc.

HOW TO STORE THIS INFORMATION?

Registration

StudentDetails.csv

CourseDetails.csv

ModuleDetails.csv

Buildings Office

StaffDetails.sql

RoomDetails.csv

LecturerDetails.xlsx

Science & Engineering

staffDetails.csv

programmeDetails.csv

schoolDetails.csv

deferralsDetails.xlsx

Exams Office:

StudentDetails.csv

ExamDetails.csv

FeesDetails.csv

LecturerDetails.csv

Salaries

Staffdetails.txt

Staffgrade.txt

Staffleave.txt

Staffholidays.txt

Fees

StudentDetails.csv

CourseDetails.csv

FeeDetails.csv

OTHER “MINI-WORLDS”?

...	...	...	...	...	...	...	...	...	...	...	...	...
70	65	Alakazam	Psychic	NaN	55	50	45	135	95	120	1	False
71	65	AlakazamMega Alakazam	Psychic	NaN	55	50	65	175	95	150	1	False
72	66	Machop	Fighting	NaN	70	80	50	35	35	35	1	False
73	67	Machoke	Fighting	NaN	80	100	70	50	60	45	1	False
74	68	Machamp	Fighting	NaN	90	130	80	65	85	55	1	False
75	69	Bellsprout	Grass	Poison	50	75	35	70	30	40	1	False
76	70	Weepinbell	Grass	Poison	65	90	50	85	45	55	1	False
77	71											
78	72											
79	73											
80	74											
81	75											

```
" quest for camelot " is warner bros . ' first feature-length , fully-animated attempt to steal clout from disney's cartoo
the only other recent challenger to their throne was last fall's promising , if flawed , 20th century fox production " anas
this year , it's no contest , as " quest for camelot " is pretty much dead on arrival .
even the magic kingdom at its most mediocre -- that'd be " pocahontas " for those of you keeping score -- isn't nearly as d
the story revolves around the adventures of free-spirited kayley ( voiced by jessalyn gilsig ) , the early-teen daughter of
kayley's only dream is to follow in her father's footsteps , and she gets her chance when evil warlord ruber ( gary oldman
with the help of hunky , blind timberland-dweller garrett ( carey elwes ) and a two-headed dragon ( eric idle and don rickl
 " quest for camelot " is missing pure showmanship , an essential element if it's ever expected to climb to the high ranks
there's nothing here that differentiates " quest " from something you'd see on any given saturday morning cartoon -- subpar
 ( compare kayley and garrett's run-in with the angry ogre to herc's battle with the hydra .
i rest my case . )
even the characters stink -- none of them are remotely interesting , so much that the film becomes a race to see which one
in the end , it's a tie -- they all win .
that dragon's comedy shtick is awfully cloying , but at least it shows signs of a pulse .
at least fans of the early-'90s tgif television line-up will be thrilled to find jaleel " urkel " white and bronson " balki
a few scenes are nicely realized ( though i'm at a loss to recall enough to be specific ) , and the actors providing the vo
but one must strain through too much of this mess to find the good .
aside from the fact that children will probably be as bored watching this as adults , " quest for camelot " 's most grievou
and personality , we learn from this mess , goes a very long way .
```

ACCESSING THIS INFORMATION?

Data Access **methods**

- File system approach
- DataBase Management System (DBMS) approach
- Many others . . .

(COMPUTER) FILE SYSTEM APPROACH (NO DBMS)

Each user stores the information they require to do their job in their own files on their own machines/servers and writes own programs/applications to access and update data

e.g.,

```
pd.read_csv(io.BytesIO(uploaded['weather_data.csv']))
```

or

```
infp = fopen("weather.txt", "r");
```

For example, in the University mini-world scenario:

Registration

StudentDetails.csv

CourseDetails.csv

ModuleDetails.csv

Buildings Office

StaffDetails.sql

RoomDetails.csv

LecturerDetails.xlsx

Science & Engineering

staffDetails.csv

programmeDetails.csv

schoolDetails.csv

deferralsDetails.xlsx

Exams Office:

StudentDetails.csv

ExamDetails.csv

FeesDetails.csv

LecturerDetails.csv

Salaries

Staffdetails.txt

Staffgrade.txt

Staffleave.txt

Staffholidays.txt

Fees

StudentDetails.csv

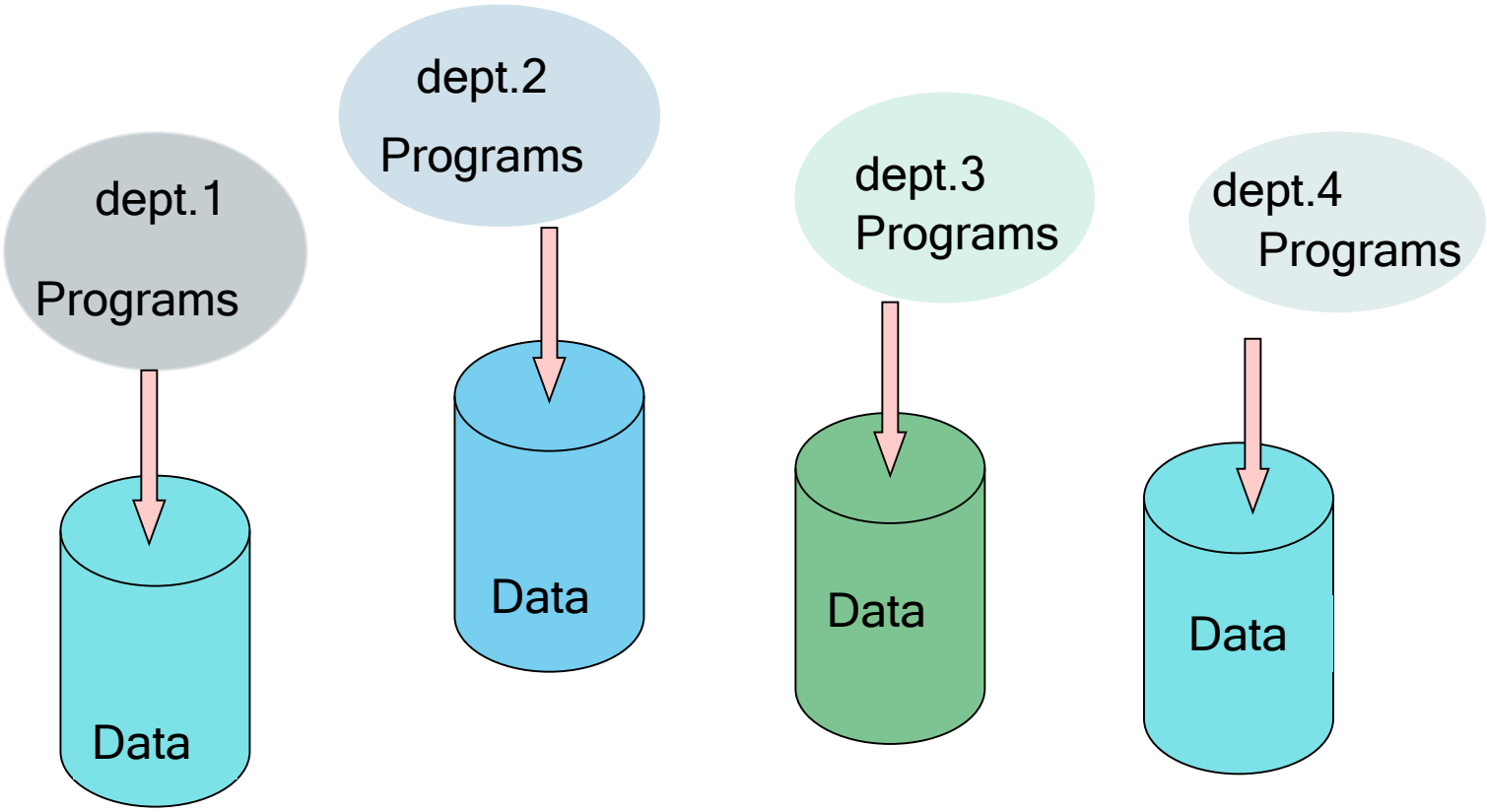
CourseDetails.csv

FeeDetails.csv

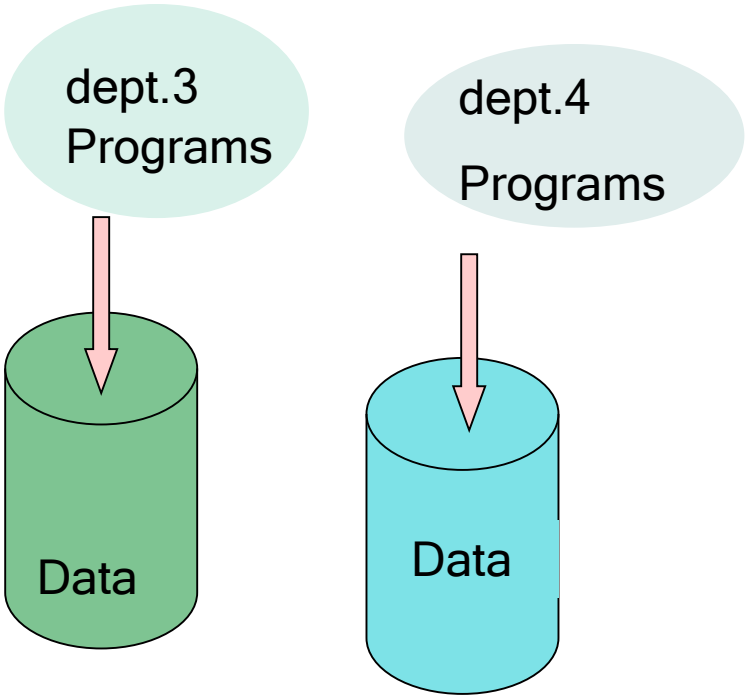
Characteristics of File System Approach:

Files separate: **not** cross-indexed (**"flat files"**)

Programs separate: each program "owns" and accesses its own files (**file processing**)



FILE PROCESSING



This is what we *often* do when we write programs in Java or C or Python to open a text file and work with the data in the file in some way or what we do with shared documents or sheets (e.g. Google docs)

e.g., In Python:

```
pd.read_csv(io.BytesIO(uplo
aded['weather_data.csv']))
```

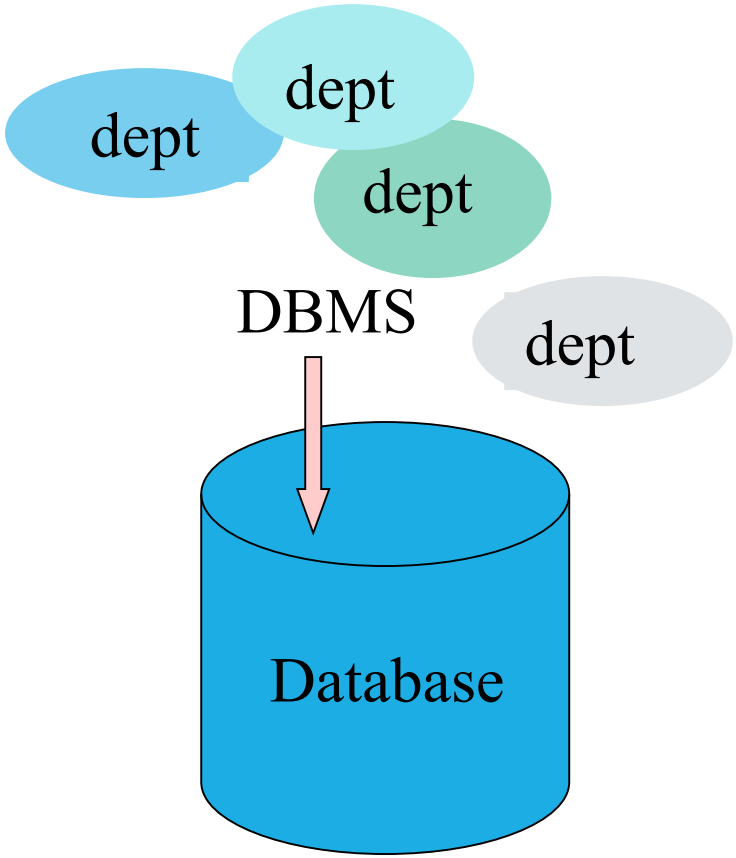
POTENTIAL PROBLEMS

Within a medium to large organisation with many users many departments and distributed data:

1. Duplicated data in different files and locations.
2. Difficulty in accessing data.
3. Programmer effort for any new tasks.
4. Data isolation - Due to multiple files and formats.
5. Integrity problems - Part of code.
6. Update Inconsistencies - Failures may leave data in an inconsistent state.
7. Concurrent Access - May not be supported.

DATABASE APPROACH

A single repository of data (which may be distributed) is maintained that is defined once and then accessed by various users via a DBMS

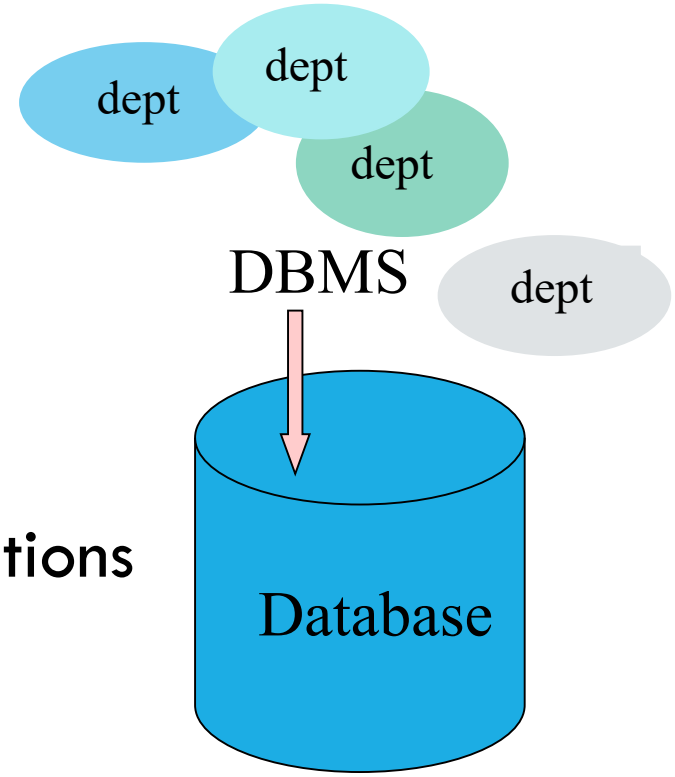


DATABASE MANAGEMENT SYSTEM (DBMS)

The DBMS is a collection of programs that facilitates the process of:

- defining
- constructing
- manipulating

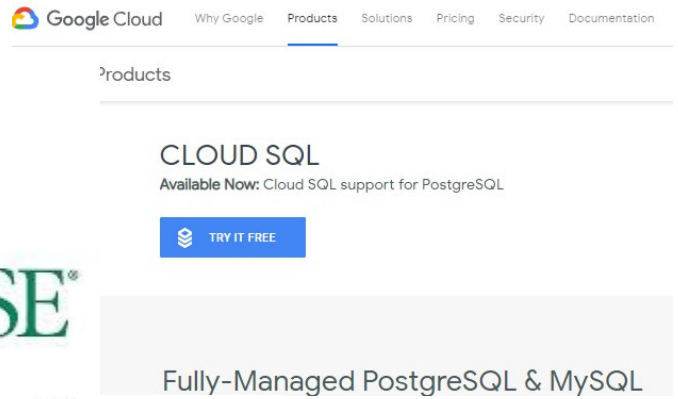
databases for various applications



DBMS CAPABILITIES:

1. Define database (DDL).
2. Manipulate database (SQL).
3. Control redundancy.
4. Restrict unauthorised access.
5. Enforce integrity constraints.
6. Provide multiple user interfaces/views.
7. Provide concurrent access.
8. Provide mechanism for recovery.
9. Provide back-up.
10. Allows representation of complex relationships between data (for efficiency and optimisation reasons).

EXAMPLES OF GENERAL PURPOSE DBMS



SYBASE

Informix SOFTWARE The database company

Microsoft Office Access

IBM DB2

Microsoft SQL Server

ORACLE

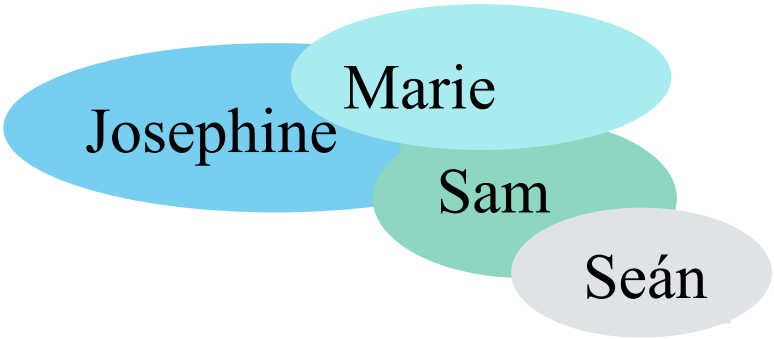
The SAP DB Database System

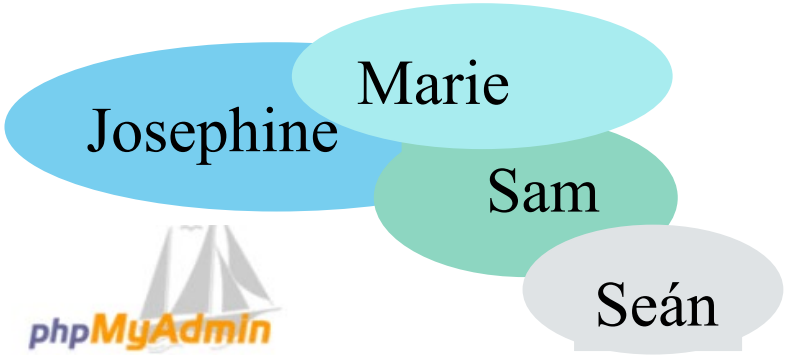
MySQL

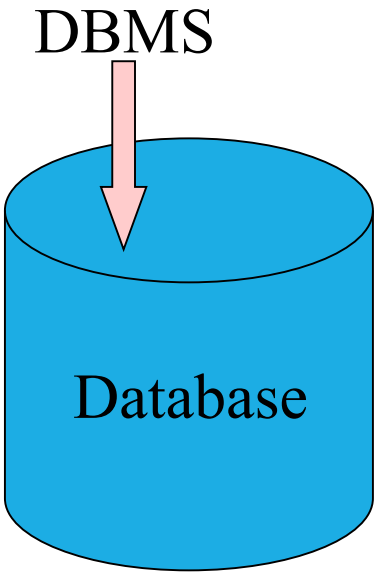
WE WILL USE MYSQL

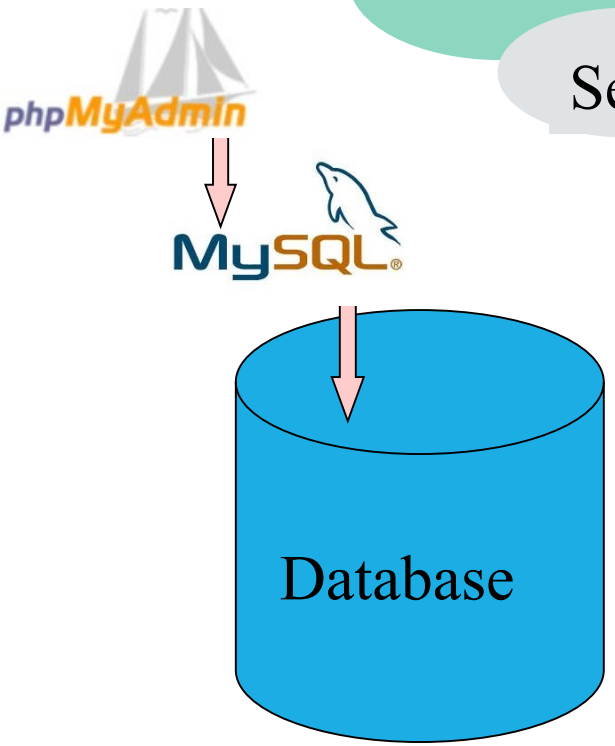
It is available on the CS servers – you do not need to install it

IMPORTANT DISTINCTIONS: The Database; the DBMS; the GUI of the DBMS; and the user(s)









We will use phpMyAdmin and (a reduced featured phpMyAdmin) Adminer

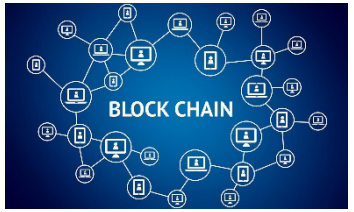
These are also available on the IT/CS servers – you do not need to install the – you access it using a browser:

OTHER APPROACHES (NON DBMS)

NoSQL databases:

- Non relational
- Examples: MongoDB, Redis, Apache Cassandra, SAP Hanna, Couchbase, etc.

Blockchain (Distributed database)



XML databases (Extensible Markup Language)

- Data stored in XML format – structure and links provided by XML
- Query language like SQL - XQuery
- Examples: BaseX, eXistDB, Sedna

DISADVANTAGES OF DBMS APPROACH

- Strict schema and multiple tables/relations
- Complexity
- Size
- Cost of DBMS
- Additional Hardware costs
- Cost of conversion
- Performance
- Higher impact of a failure

So, … why learn about relational DBMS?

90% of industry/enterprise/business applications are STILL Relational DBMS or Relational DBMS with extensions (e.g. OO Relational).

Majority of industry applications require:

- Correctness
- Completeness
- Efficiency (Complex optimisation techniques and complex Indexing structures).

Relational DBMS provide this.

DBMS USERS

Administrators (DBA) - accounts, passwords, privileges. Requiring constant vigilance!

System Analysts (what's required to solve a problem?; what's the business need?).

Designers (ER diagrams; mapping ER diagrams to tables).

Application programmers (creating tables; adding data; creating queries).

End users (all of us).

DATABASE ABSTRACTION

Refers to the hiding of the details of data storage that are not needed by most database users

- The aim is to separate user's views of the database from the way it is physically represented.
- 3 ways in which data can be described:

  External: user's view.

  Conceptual: logical structure as seen by DBA.

  Internal: DBMS and OS view of data.

DATABASE SCHEMAS AND INSTANCES

Similar to types and variables in programming languages.

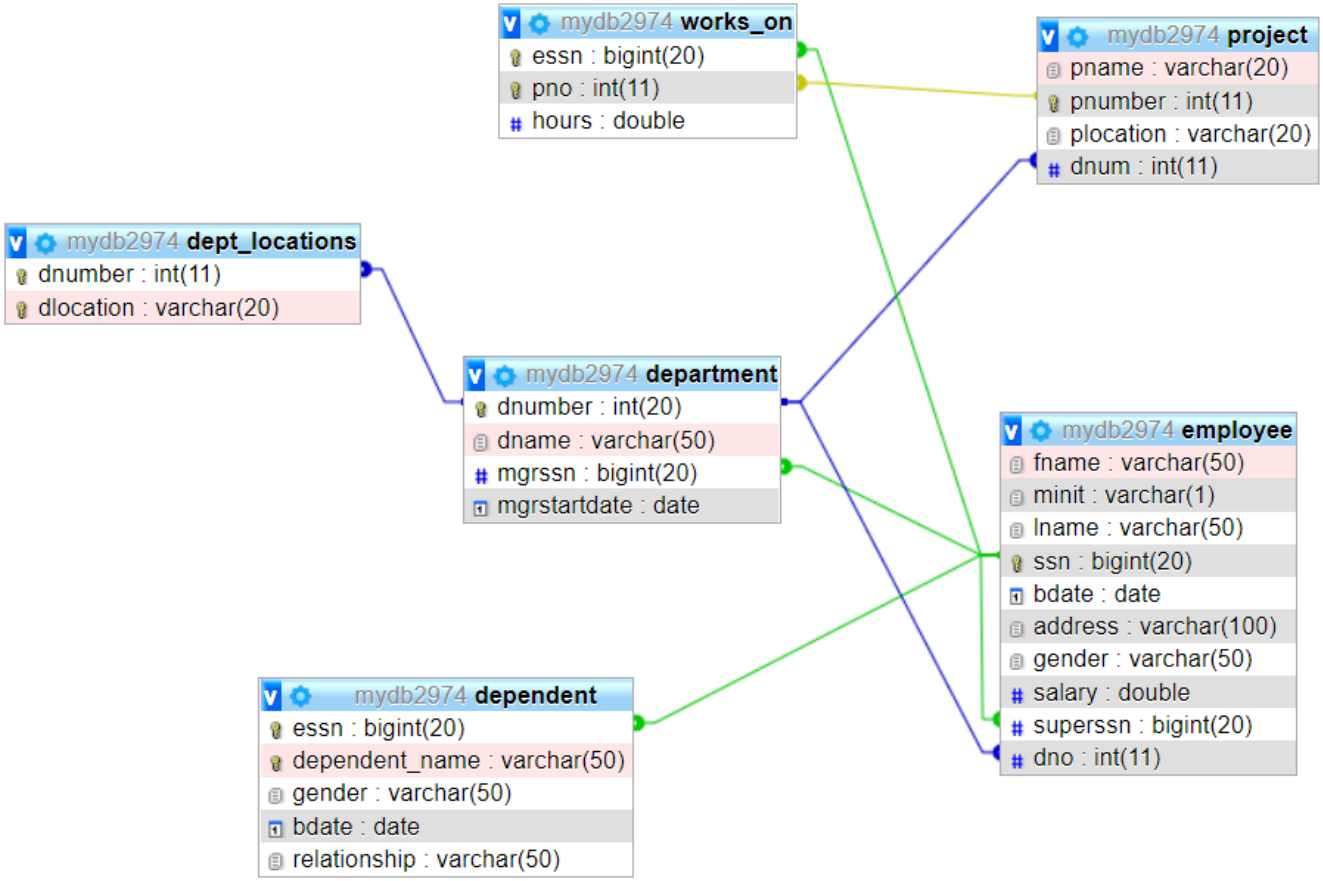
Schema: the logical structure of a database.

e.g., database consists of information about students, courses, and lecturers and how the data is related.

Instance: the actual content of the database at some point in time

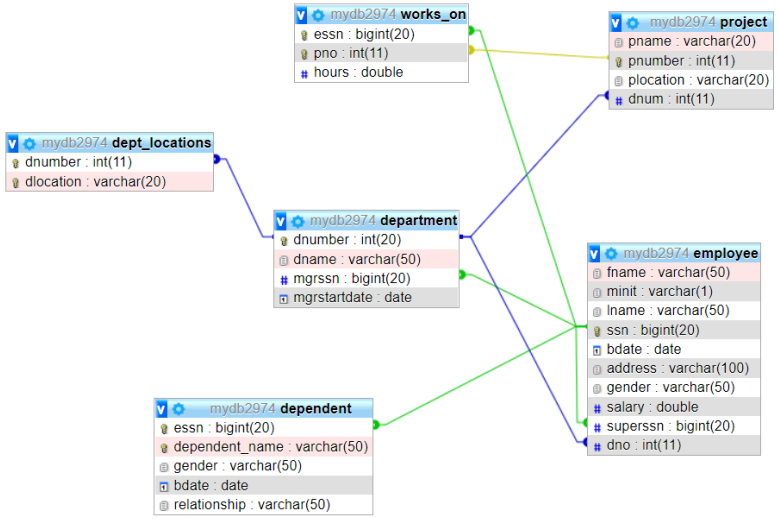
e.g., actual information on 1500 lecturers, 18,000 students, 3000 modules, etc.

A FIRST DATABASE (FROM E&N)
The "company" database



OUR NOTATION

case **not** significant;
spaces not allowed



DATABASE SCHEMA

employee(fname, minit, lname, ssn, bdate, address, gender, salary, superssn, dno)

department(dname, dnumber, mgrssn, mgrstartdate)

dept_locations(dnumber, dlocation)

project(pname, pnumber, plocation, dnum)

works_on(essn, pno, hours)

dependent(essn, dependent_name, gender, bdate, relationship)

YOUR OWN SUMMARY AND REFLECTIONS

What is your take-home message from today's lecture?

MY SUMMARY OF IMPORTANT CONCEPTS …

The traditional (computer) file processing approach, relational and non-relational are all approaches to storing and retrieving structured data.

Differ in a number of ways:

- Domain they are suited to
- Importance of correctness, consistency and efficiency to that domain (which is strength of DBMS).
- Cost

- Who are the different DBMS users and what do they do?
- What is DBMS abstraction and why is it important?
- What is the difference between a Schemas and Instances

NEXT WEEK: GETTING YOUR OWN DATABASE SET-UP

- Setting up an account on the CS server and accessing a mySQL database.
- Looking at the sample company database in detail.
- Defining the relational model in more detail.