



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David



INTRODUCTION TO DATABASES

Lecture 1: Database Concepts

Today's Lecture



- ❑ What is a Database
- ❑ Why Do We Need Computerised Databases
- ❑ Historical Development of Databases
- ❑ File-Based Systems
- ❑ The Database and Database Management Systems (DBMSs)
- ❑ Advantages and Disadvantages of DBMSs

Question(s)

- What is a database?

Database – What is it?

If you had to explain what a database is to someone who has never heard of it before, how would you describe it?

<https://padlet.com/dchidambaram1/IntroDB>



Database – What is it?



"A set of information held in a computer" - Oxford English Dictionary



"One or more large structured sets of persistent data, usually associated with software to update and query the data" - Free On-Line Dictionary of Computing



A database is a structured collection of data that is organized in a way to make it easily accessible, manageable, and updateable. It is designed to efficiently store, retrieve, and manage vast amounts of information in a systematic and structured manner. Databases are a fundamental part of modern computing and play a critical role in various applications and systems.

Question(s)

- Why do we need databases?



A Database is

A collection of related pieces of data

- **Representing/capturing the information** about a real-world enterprise or part of an enterprise
- **Collected and maintained** to serve specific data management needs of the enterprise
- Activities of the enterprise are supported by the database and **continually update** the database



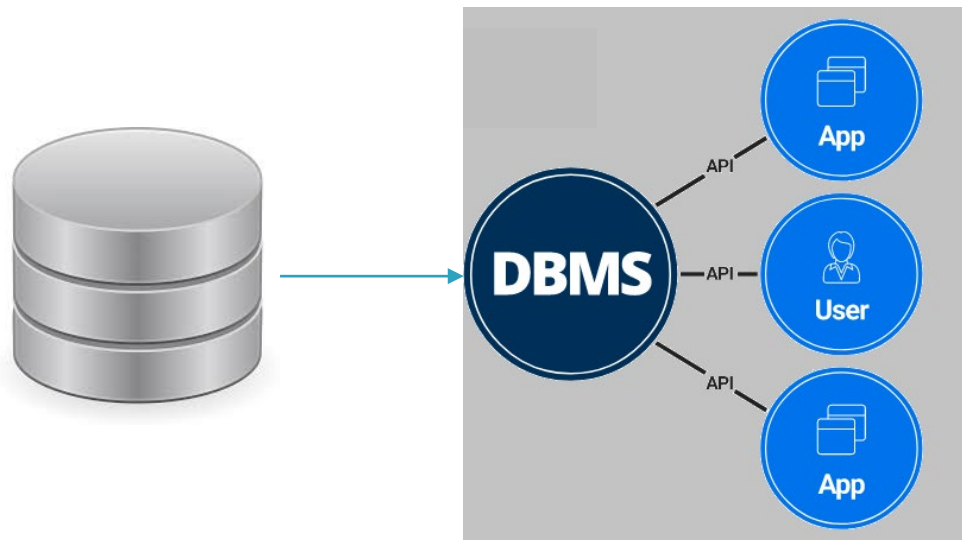
Question(s)

- What is a Database Management System (DBMS)?



What is Database Management System?

- ❑ software systems used to store, retrieve, and run queries on data
- ❑ serves as an interface between an end-user and a database, allowing users to create, read, update, and delete data in the database.



Database- Past and present



Question(s)

- Why do we need DBMSs?



Why Database Management Systems?

- Data is used everywhere.



- ▣ Personal Data
 - ▣ Product Data
 - ▣ Media Data, etc.
- ***It needs to be stored somewhere!***

- Data is valuable.



Why Database Management Systems?

“You can have data without information, but you cannot have information without data.”

Daniel Keys Moran, Computer programmer and science fiction author

“Data that is loved tends to survive.”

Kurt Bollacker, computer scientist

“We’re entering a new world in which data may be more important than software.”

Tim O’Reilly, founder, O’Reilly Media

“If somebody tortures the data enough (open or not), it will confess anything.”

Paolo Magrassi, former vice president, research director, Gartner

Why Database Management Systems?

□ Data is not information!

Information is data made meaningful – typically by applications. E.g.:

- Consumer applications
- Search engines
- Online banking
- Travel reservation sites
- Health care
- Libraries
- Etc.
- Registration services
- Price comparison sites



□ ***It needs to be shared!***

- Accessed by applications that can make it meaningful
- Simultaneous, multiple access to it too

Database Example

What is the purpose of having database?

What data would be stored in this database?

What will be the repercussions if there's no database maintained?



Database Example

What is the purpose of having University database?

To keep an accurate track of the academic activities of the University.

What data would be stored in this database?

Data about:

- a. students,
- b. Faculty
- c. Courses
- d. Research-laboratories,
- e. Course registration etc.

What will be the repercussions if there's no database?

Database Example

StudentID	FirstName	LastName	Department	Course	CourseProfessor	GPA
1001	John	Smith	Computer	COMP101	Dr. Johnson	3.5
1002	Jane	Doe	Biology	BIO201	Dr. Davis	3.7
1003	Michael	Johnson	Computer	COMP101	Dr. Johnson	3.2
1004	Sarah	Davis	Biology	BIO201	Dr. Davis	3.9
1005	David	Brown	Computer	COMP102	Dr. Anderson	3.1
1006	Laura	Turner	Biology	BIO202	Dr. Williams	3.8
1007	Brian	Anderson	Mathematics	MATH101	Prof. Turner	3.6
1008	Emma	White	Computer	COMP102	Dr. Anderson	3.4
1009	Mark	Taylor	Physics	PHYS101	Prof. Davis	3.2
1010	Olivia	Clark	Mathematics	MATH102	Prof. Turner	3.9

Database Example

StudentID	FirstName	LastName	Department	GPA
1001	John	Smith	Computer	3.5
1002	Jane	Doe	Biology	3.7
1003	Michael	Johnson	Computer	3.2
1004	Sarah	Davis	Biology	3.9
1005	David	Brown	Computer	3.1
1006	Laura	Turner	Biology	3.8
1007	Brian	Anderson	Mathematics	3.6
1008	Emma	White	Computer	3.4
1009	Mark	Taylor	Physics	3.2
1010	Olivia	Clark	Mathematics	3.9

Course	CourseProfessor
COMP101	Dr. Johnson
BIO201	Dr. Davis
COMP102	Dr. Anderson
BIO202	Dr. Williams
MATH101	Prof. Turner
PHYS101	Prof. Davis
MATH102	Prof. Turner

Database Example

StudentID	FirstName	LastName	Department	GPA
1001	John	Smith	Computer	3.5
1002	Jane	Doe	Biology	3.7
1003	Michael	Johnson	Computer	3.2
1004	Sarah	Davis	Biology	3.9
1005	David	Brown	Computer	3.1
1006	Laura	Turner	Biology	3.8
1007	Brian	Anderson	Mathematics	3.6
1008	Emma	White	Computer	3.4
1009	Mark	Taylor	Physics	3.2
1010	Olivia	Clark	Mathematics	3.9

Course	CourseProfessor
COMP101	Dr. Johnson
BIO201	Dr. Davis
COMP102	Dr. Anderson
BIO202	Dr. Williams
MATH101	Prof. Turner
PHYS101	Prof. Davis
MATH102	Prof. Turner

```
UPDATE Courses SET CourseProfessor = 'Prof. Anderson' WHERE Course = 'COMP101';
```

Specialty Cat Store



★★★★★

[Redacted description text]



★★★★

[Redacted description text]



★★★★★

[Redacted description text]



★★★★★★

[Redacted description text]

Purchase

Name

Address

Phone

Credit Card

AutoSave On acca5026 course work marks • Saved Search (Alt+Q) Divya Chidambaram

File Home Insert Page Layout Formulas Data Review View Help Acrobat

Clipboard Font Alignment Number Styles Cells Editing Analysis Sensitivity

Comments Share

Calibri 11 A A Wrap Text General

B I U Merge & Center

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

Sort & Filter Find & Select

Analyze Data Sensitivity

A1 Name

	A	B	C	D	E	F
1	Name	Product	Quantity	Date	Address	
2	Jenny	Cat leash	1	23-Sep	123 Main street	
3						
4						
5						
6						



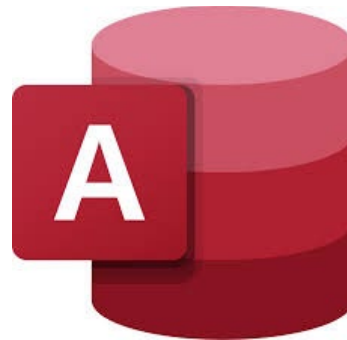
[illegible]

Popular DBMS??



Popular Database ?? DBMS??

- ❑ MySQL
- ❑ Microsoft Access
- ❑ Microsoft SQL Server
- ❑ FileMaker Pro
- ❑ Oracle Database
- ❑ dBASE.

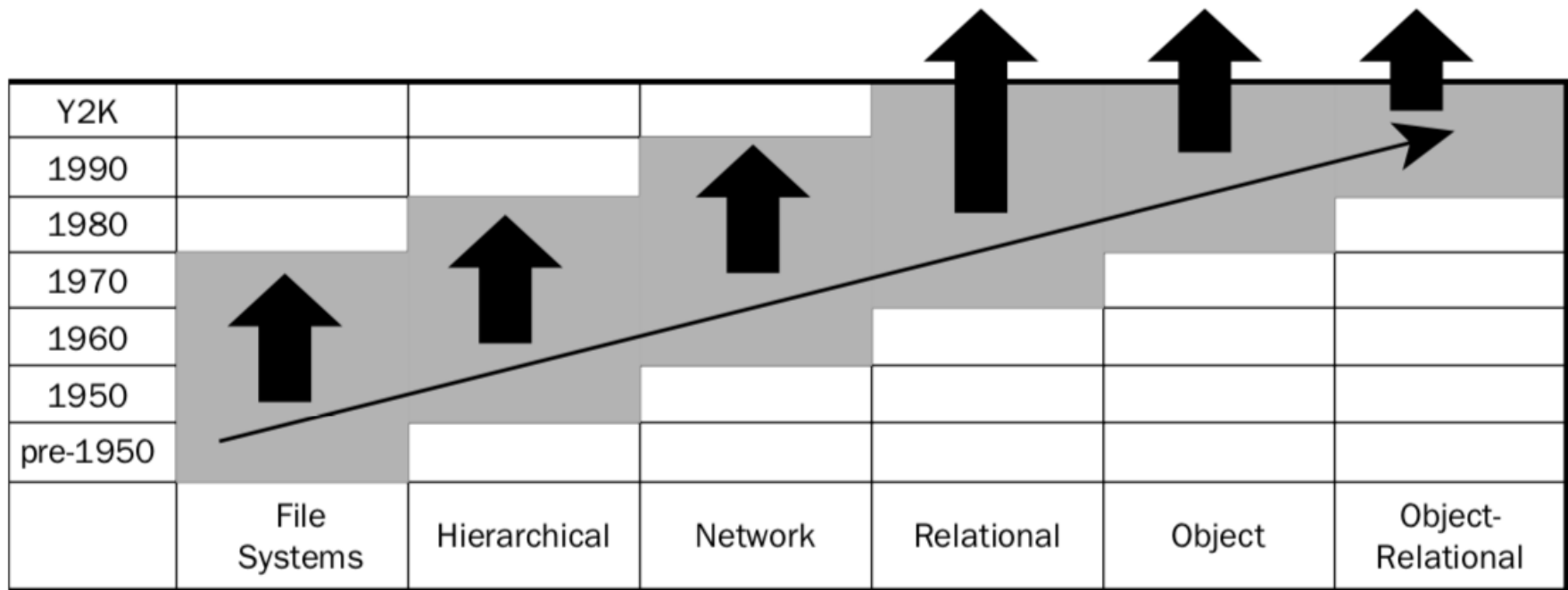


Historical Development of Databases

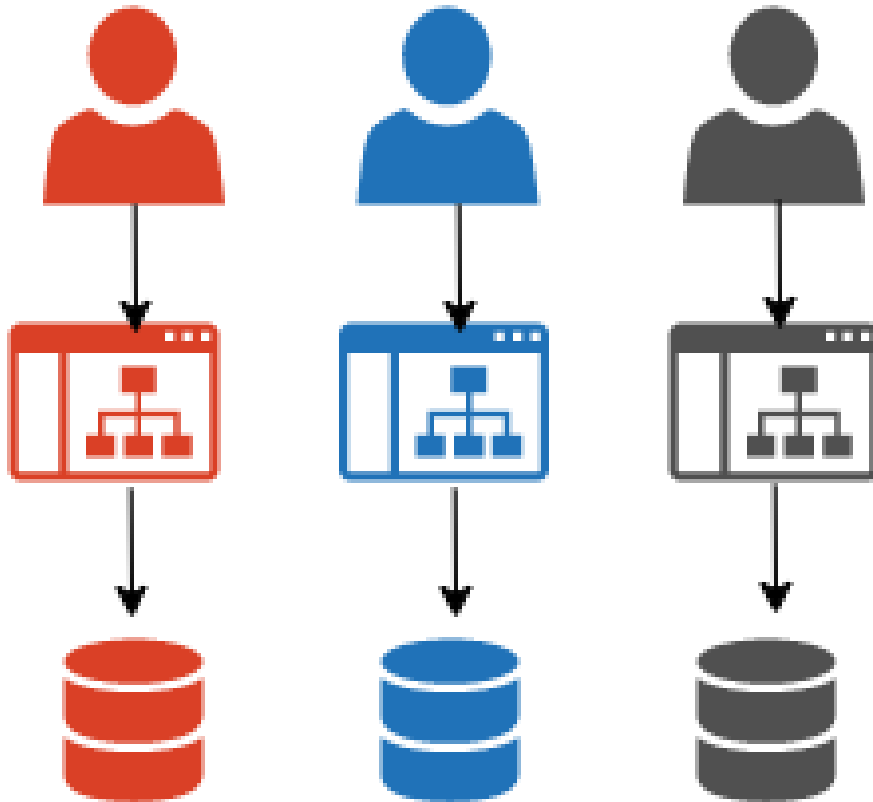
TIMEFRAME	DEVELOPMENT	COMMENTS
1960s (onwards)	File-based systems	Precursor to the database system. Decentralized approach: each department stored and controlled its own data.
Mid-1960s	Hierarchical and network data models	Represents first-generation DBMSs. Main hierarchical system is IMS from IBM and the main network system is IDMS/R from Computer Associates. Lacked data independence and required complex programs to be developed to process the data.
1970	Relational model proposed	Publication of E. F. Codd's seminal paper "A relational model of data for large shared data banks," which addresses the weaknesses of first-generation systems.
1970s	Prototype RDBMSs developed	During this period, two main prototypes emerged: the Ingres project at the University of California at Berkeley (started in 1970) and the System R project at IBM's San José Research Laboratory in California (started in 1974), which led to the development of SQL.
1976	ER model proposed	Publication of Chen's paper "The Entity-Relationship model—Toward a unified view of data." ER modeling becomes a significant component in methodologies for database design.
1979	Commercial RDBMSs appear	Commercial RDBMSs like Oracle, Ingres, and DB2 appear. These represent the second generation of DBMSs.
1987	ISO SQL standard	SQL is standardized by the ISO (International Standards Organization). There are subsequent releases of the standard in 1989, 1992 (SQL2), 1999 (SQL:1999), 2003 (SQL:2003), 2008 (SQL:2008), and 2011 (SQL:2011).
1990s	OODBMS and ORDBMSs appear	This period initially sees the emergence of OODBMSs and later ORDBMSs (Oracle 8, with object features released in 1997).
1990s	Data warehousing systems appear	This period also see releases from the major DBMS vendors of data warehousing systems and thereafter data mining products.
Mid-1990s	Web-database integration	The first Internet database applications appear. DBMS vendors and third-party vendors recognize the significance of the Internet and support web-database integration.
1998	XML	XML 1.0 ratified by the W3C. XML becomes integrated with DBMS products and native XML databases are developed.

Figure 1.10 Historical development of database systems.

History of Databases Summarised

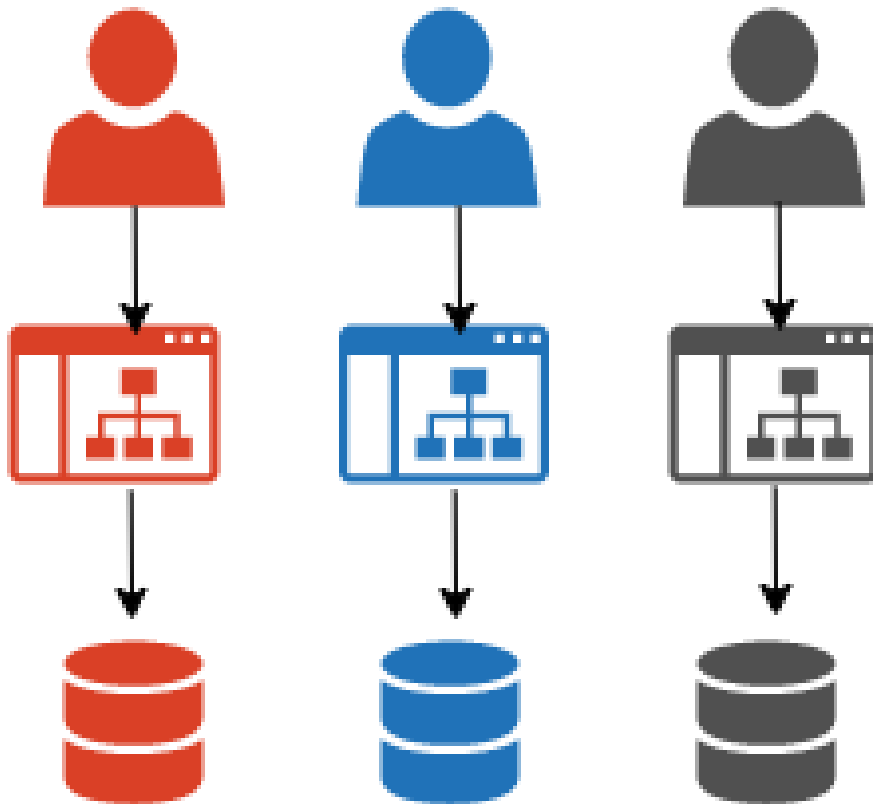


File-Based Systems



- All application was tailor-made
- Applications store (& persist) their data in files
- Each file has its own format
- ⑩ Program must know format
- ⑩ Any other program using file has to know format

File-Based Systems



- All application was tailor-made
- Applications store (& persist) their data in files
- Each file has its own format
- ⑩ Program must know format
- ⑩ Any other program using file has to know format
- ⑩ **Issues?**

Issue #1: Duplication of Data / Data Redundancy

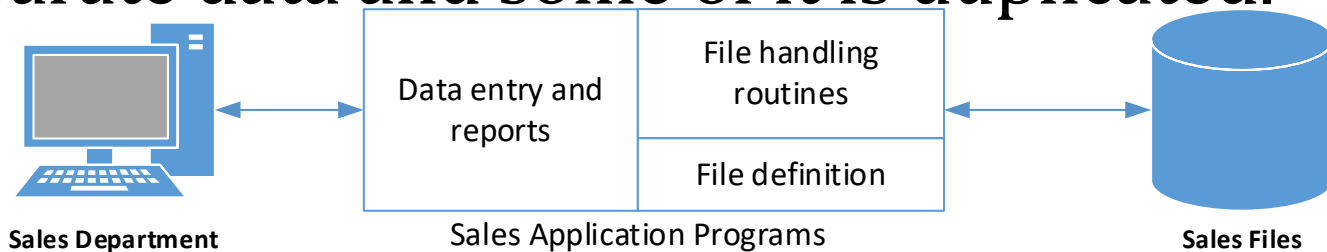
Issue #2: Incomplete Updates (falling out of sync)

Limitations of File-Based Approach

- ❑ **Separation and isolation of data**
 - ▣ Each program maintains its own set of data.
 - ▣ Thus, users of one program may be unaware of potentially useful data held by other programs.
- ❑ **Duplication of data**
 - ▣ Same data is held by different programs.
 - ▣ Wasted space and potentially different values and/or different formats for the same item.
 - ▣ Duplication can lead to loss of data integrity.

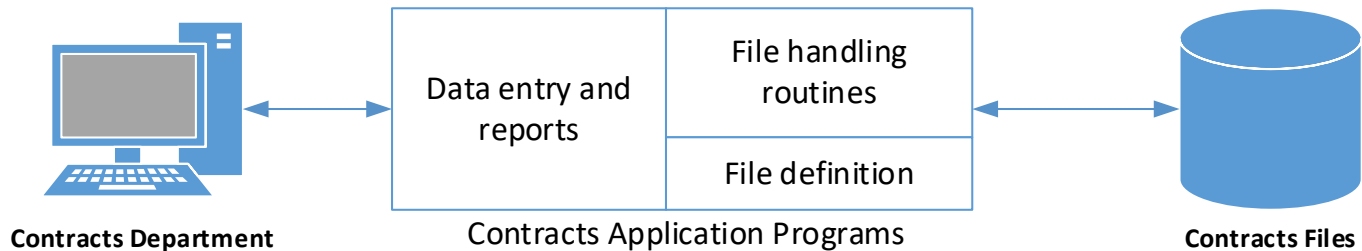
Limitations of File-Based Approach

- Sales and Contracts each have their own separate data and some of it is duplicated.



Sales Files

- PropertyForRent (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)
- PrivateOwner (ownerNo, fName, lName, address, telNo)
- Client (clientNo, fName, lName, address, telNo, prefType, maxRent)



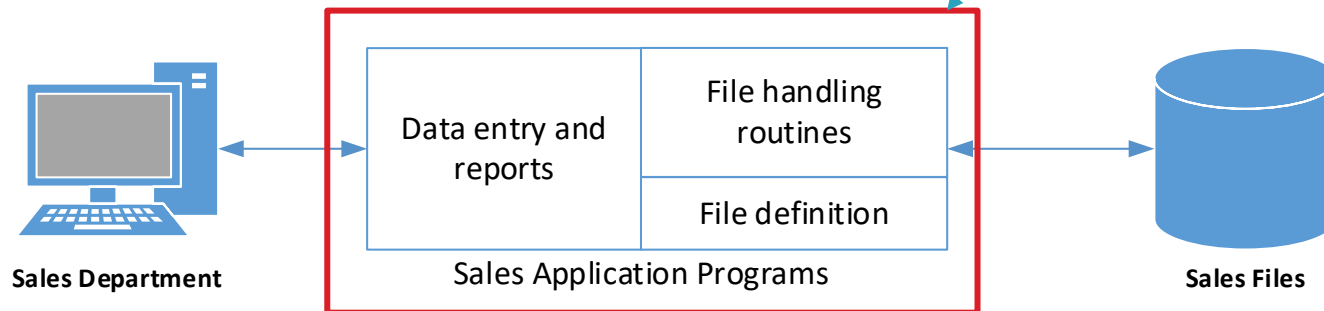
Contracts Files

- Lease (leaseNo, propertyNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish)
- PropertyForRent (propertyNo, street, city, postcode, rent)
- Client (clientNo, fName, lName, address, telNo)

Limitations of File-Based Approach

□ Data dependence

- ▣ The file structure is defined in the program code, making changes to an existing structure difficult.



□ Incompatible file formats

- ▣ Structure of files are dependent on the application's programming language.
- ▣ Application programs written in different languages cannot easily access each other's files.

Limitations of File-Based Approach

- ❑ **Fixed queries or proliferation of application code**
 - ▣ Program's are written for particular functions with fixed reports and queries - no support for ad hoc/dynamic queries.
 - ▣ New requirements meant ***new*** application program code.

Access Ad Hoc Query

DDA_Student

- *
 - DDA
 - First_Name
 - Last_Name
 - Post_Code
 - Email_Student_personal
 - Email_work
 - Employment_Status
 - College_Study_ID
 - Follow_Up
 - Application_Link
 - ALO_ID
 - Welsh_Speaker
 - LinkedIn_Follower
 - Status
 - EmployerID
 - StudentID
 - Fundina_Stream

ALO

- *
 - ALO_ID
 - First_Name
 - Last_Name
 - Email

Field:	First_Name	Last_Name	Email_work	ALO_ID	First_Name	StudentID	Start_Date
Table:	DDA_Student	DDA_Student	DDA_Student	DDA_Student	ALO	DDA_Student	DDA_Student
Sort:							
Show:	☒	☒	☒	☒	☒	☒	☒
Criteria:					"Stephen"		
or:							

SQL Generated from Ad Hoc Query

```
1  SELECT DDA_Student.First_Name, DDA_Student.Last_Name,  
2         DDA_Student.Email_work, DDA_Student.ALO_ID,  
3         ALO.First_Name, DDA_Student.StudentID,  
4         DDA_Student.Start_Date  
5  FROM ALO INNER JOIN DDA_Student  
6         ON ALO.ALO_ID = DDA_Student.ALO_ID  
7  WHERE (((ALO.First_Name)="Stephen"));
```

Limitations of File-Based Approach

- ❑ In an Estate Management file-based application, different files may exist for: each property, each sale/rent, each buyer, each renter, and one for each member of staff.
- ❑ Imagine trying to answer the following queries using all these separate files:
 - ▣ What three-bedroom properties are available with a large garden and garage?
 - ▣ What two-bed properties do you have for rent within three miles of the University?

Database Approach

- All the limitations of the file-based approach can be attributed to two factors:
 - ▣ The definition of the data was embedded in application programs, rather than being stored separately and independently.
 - ▣ There is no control over access and manipulation of data beyond that imposed by application programs.
- As a result, the **Database** and the **Database Management System** emerged.

Database(Recap)

- The database is also defined as *a self-describing collection of integrated records* because it stores:
 - ▣ the organisation's operational data,
 - ▣ and also a description of this data.
- It is this self-describing nature of databases that provides **data independence**, which file-based systems lacked.
- More on data independence in the next lecture.

Database Management System (DBMS) Recap

- A DBMS is software system that enables users to define, create, maintain and control access to the database.
 - ▣ Access to the database is through the DBMS.
 - ▣ It sits between application programs and the backend database.
 - ▣ The DBMS is the software that interacts with users (individuals on workstations and users' application programs) and the database.
- Example DBMSs include PostgreSQL, Access, Oracle, MySQL, MS SQL Server, etc.

Database Approach

- ❑ **Controlled access to database may include:**
 - ❑ **A security system.**
 - ❑ **An integrity system.**
 - ❑ **A concurrency control system.**
 - ❑ **A recovery control system.**
 - ❑ **A user-accessible catalog.**
- ❑ **A view mechanism.**
 - ❑ **Provides users with only the data they want or need to use.**

Advantages and Disadvantages of DBMSs

- ❑ The DBMS has many advantages, but also some disadvantages.
- ❑ The advantages and disadvantages are listed on the following slides.
- ❑ **Please read** the accompanying notes for explanations.
 - ▣ The advantages and disadvantages highlight the differences between the file-based approach and database approach.
 - ▣ The advantages further highlight why the database approach is better than the file-based approach.

Advantages of DBMSs



- ❑ Control of data redundancy
- ❑ Data consistency
- ❑ More information from the same amount of data
- ❑ Sharing of data
- ❑ Improved data integrity
- ❑ Improved security
- ❑ Enforcement of standards
- ❑ Economy of scale

Advantages of DBMSs



- ❑ Balance conflicting requirements
- ❑ Improved data accessibility and responsiveness
- ❑ Increased productivity
- ❑ Improved maintenance through data independence
- ❑ Increased concurrency
- ❑ Improved backup and recovery services

Disadvantages of DBMSs

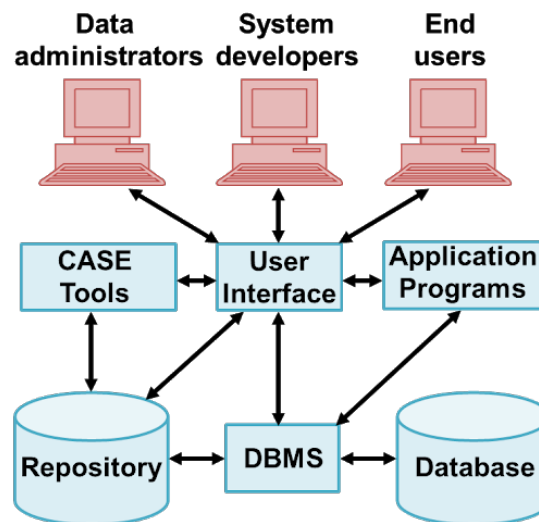
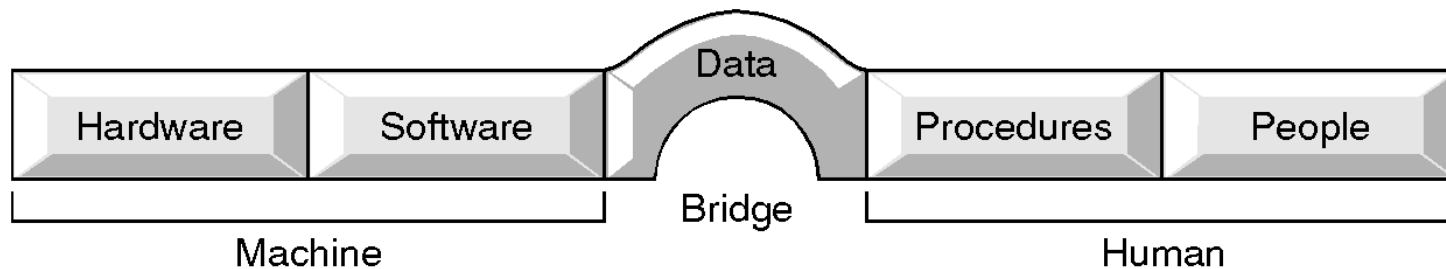


- ❑ Complexity
- ❑ Size
- ❑ Cost of DBMS
- ❑ Additional hardware costs
- ❑ Cost of conversion
- ❑ Performance
- ❑ Greater impact of a failure

Roles in the Database Environment

- ❑ **Data Administrator (DA)**
- ❑ **Database Administrator (DBA)**
- ❑ **Database Designers (Logical and Physical)**
- ❑ **Application Programmers**
- ❑ **End Users (native and sophisticated)**

Components of DBMS Environment



Source: http://cdn.wagmob.com/subject/G124/html/introduction00to00dbms_1.html



Further Reading :

Chapter 1: Introduction to Databases

- ❑ Database Systems: A Practical Approach to Design, Implementation, and Management Sixth edition by • Thomas Connolly • Carolyn Begg