

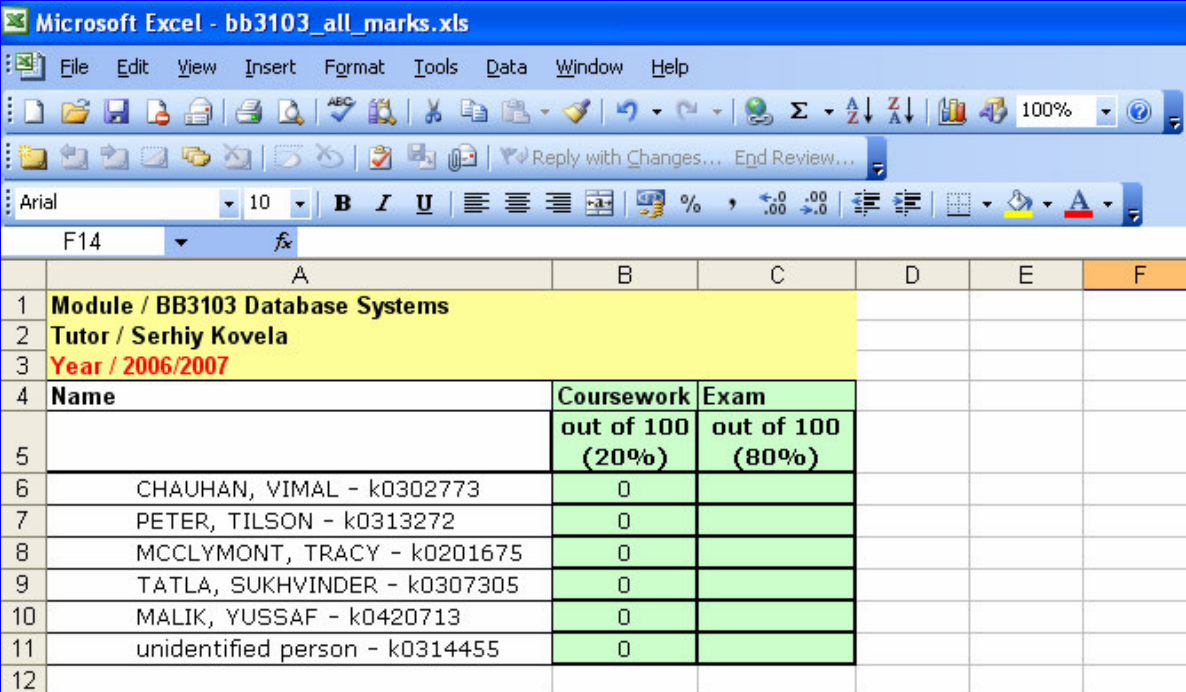
Introduction to Databases

Objectives

- ◆ **Most common types of digital information processing – enter file-based systems**
- ◆ **Why database systems came around**
- ◆ **DBMS environment: components / roles**
- ◆ **Is it all worth it?**
- ◆ **DBMS timeline**

File-Based Systems as we know them

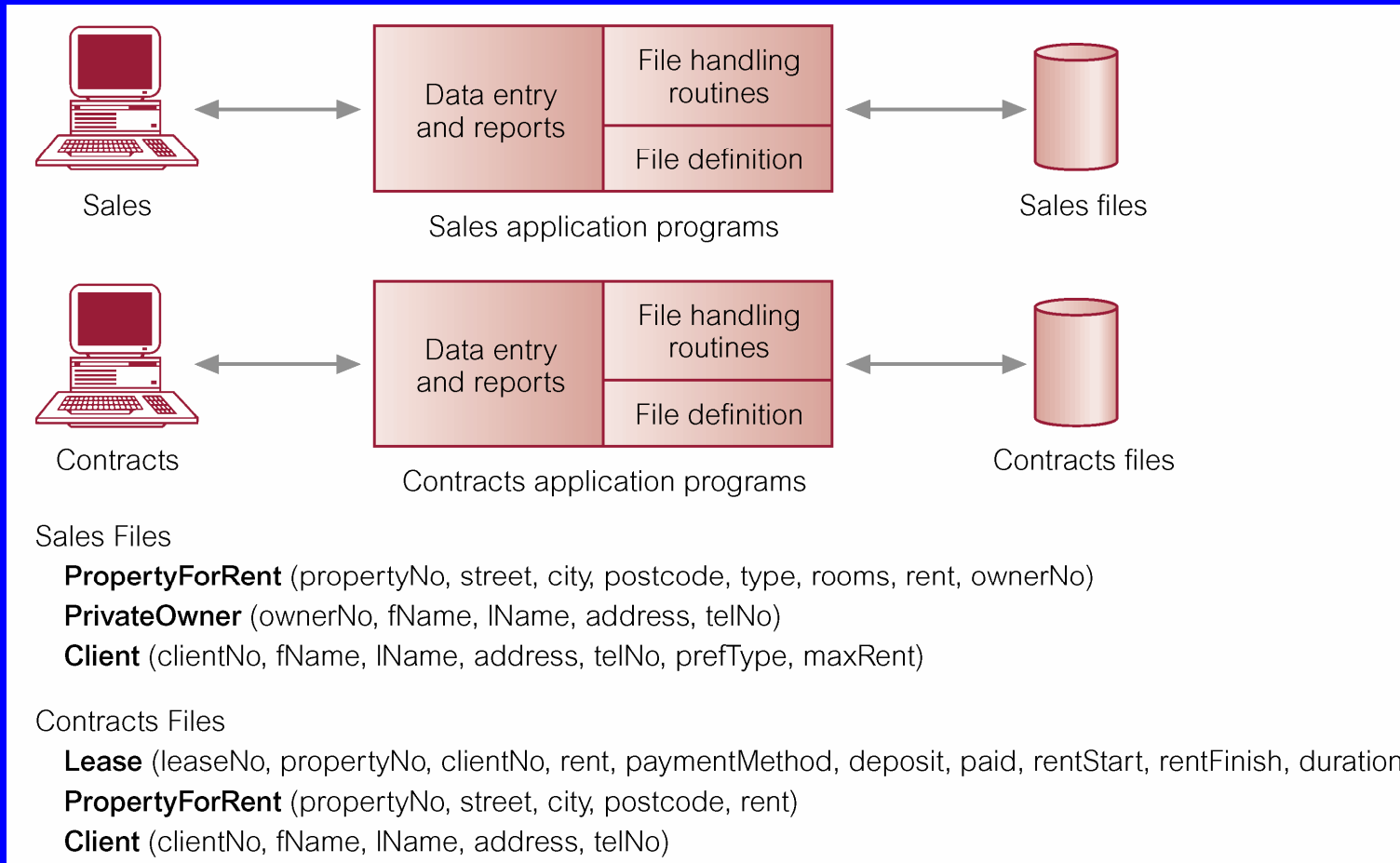
- ◆ Collection of application programs that perform services for the end users (e.g. MS Excel?)
- ◆ Each program defines and manages its own data



The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - bb3103_all_marks.xls". The spreadsheet contains a table with the following data:

	A	B	C	D	E	F
1	Module / BB3103 Database Systems					
2	Tutor / Serhiy Kovela					
3	Year / 2006/2007					
4	Name	Coursework	Exam			
5		out of 100 (20%)	out of 100 (80%)			
6	CHAUHAN, VIMAL - k0302773	0				
7	PETER, TILSON - k0313272	0				
8	MCCLYMONT, TRACY - k0201675	0				
9	TATLA, SUKHVINDER - k0307305	0				
10	MALIK, YUSSAF - k0420713	0				
11	unidentified person - k0314455	0				
12						

File-Based Processing



Limitations of File-Based Approach

- ◆ **Separation and isolation** of data
 - Each program maintains its own set of data.
 - 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.

Limitations of File-Based Approach

- ◆ **Data dependence**

- File structure is defined in the program code.

- ◆ **Incompatible file formats**

- Programs are written in different languages, and so cannot easily access each other's files.

- ◆ **Fixed Queries/Proliferation of application programs**

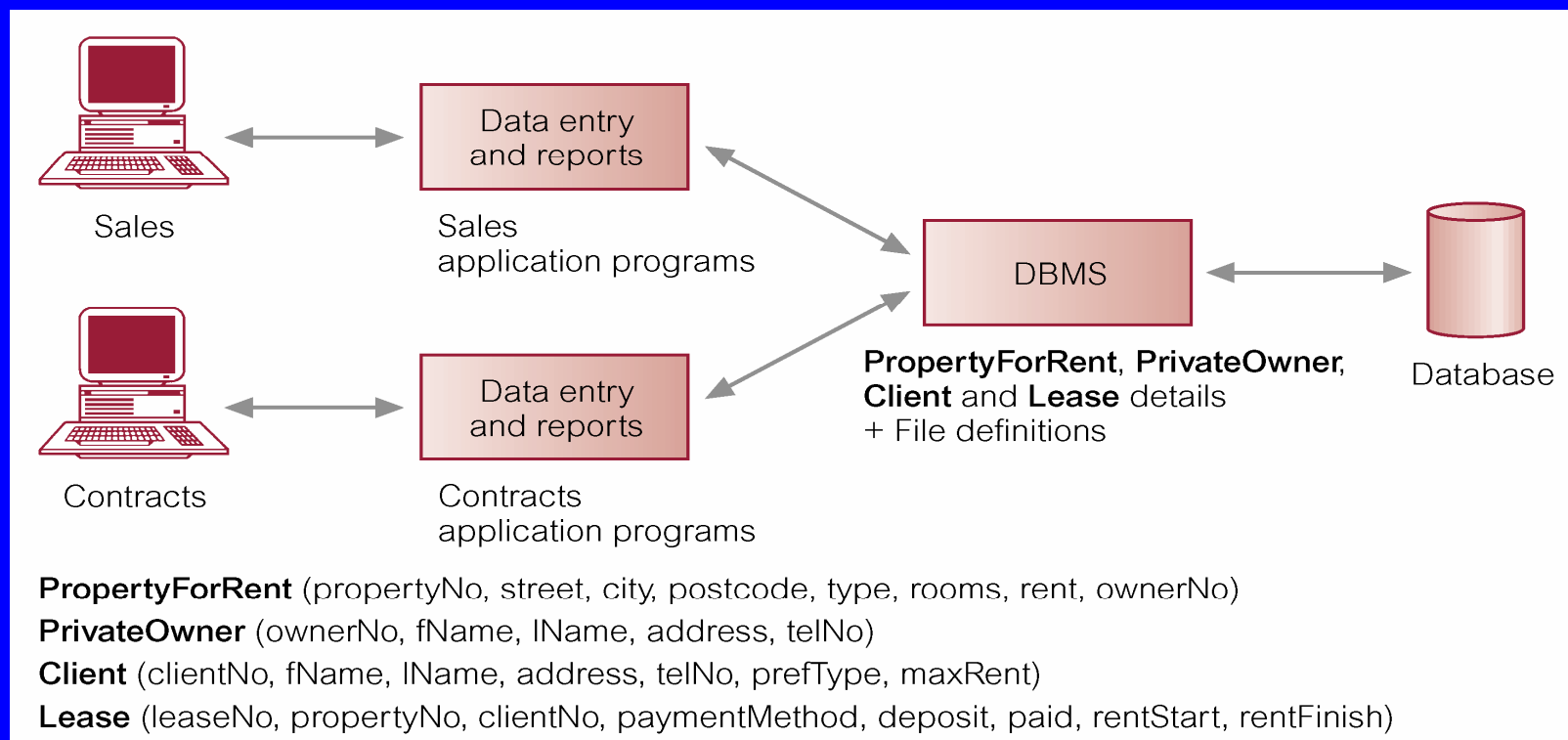
- Programs are written to satisfy particular functions.
- Any new requirement needs a new program.

Enter the ‘Database’

- ◆ **Shared collection of logically related data** (and a description of this data), designed to meet the information needs of an organization.
- ◆ Logically related data comprises **entities, attributes, and relationships** of an organization’s information.
- ◆ **System catalog** (metadata) provides description of data to enable program–data independence.

Database Management System (DBMS)

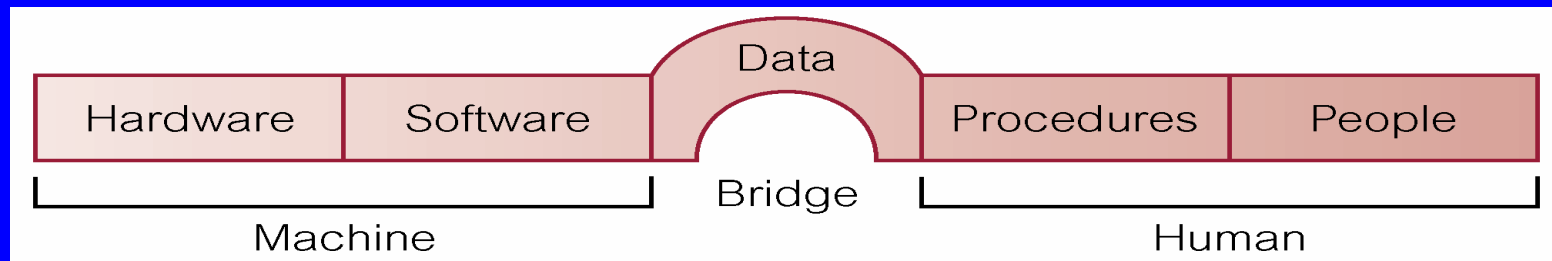
- ◆ A software system that enables users to define, create, and maintain the database and that provides controlled access to this database.



Database Approach

- ◆ **Data definition language (DDL).**
 - Permits specification of data types, structures and any data constraints. All specifications are stored in the database.
- ◆ **Data manipulation language (DML).**
 - General enquiry facility (query language) of the data.
- ◆ **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.

DBMS Environment: components



◆ Hardware

- Can range from a PC to a network of computers.

◆ Software

- DBMS, operating system, network software (if necessary) and also the application programs.

◆ Data

- Used by the organization and a description of this data called the schema.

◆ Procedures

- Instructions and rules that should be applied to the design and use of the database and DBMS.

◆ People

DBMS Environment: roles

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

Advantages of DBMSs

- ◆ **Control of data redundancy**
- ◆ **Improved data integrity / security**
- ◆ **Data consistency / more information from the same amount of data**
- ◆ **Improved backup and recovery services**
- ◆ **Sharing of data / increased concurrency / improved data accessibility and responsiveness**
- ◆ **Enforcement of standards / economy of scale**
- ◆ **Improved maintenance through data independence**

Disadvantages of DBMSs

- ◆ **Complexity**
- ◆ **Size**
- ◆ **Cost of DBMS / additional hardware**
- ◆ **Cost of conversion**
- ◆ **Performance**
- ◆ **Higher impact of a failure**

History of Database Systems

◆ First-generation

- Hierarchical and Network

◆ Second generation

- Relational

◆ Third generation

- Object Relational
- Object-Oriented