بوليتكنك البحرين

Bahrain Polytechnic

Database Systems

Topic 1: Introduction to Database Systems

Learning Objectives

Learning Objectives:

- Understand the role of data as a resource in an organisation
- Provide definitions for: Data, Database and DBMS
- Describe the database system environment
- Understand the motivation for the database approach
- State the objectives of the database approach
- Explain what data independence means
- Be aware of a database transaction
- Be aware of database backup and recovery
- Explain how a database improves the security of data
- Understand how a database inter-relates data through a Data Model
- Describe the major data models being: Relational, Hierarchical and Network

References:

- Rob, P. & Coronel, C., Database Systems, 8th Edition, Chpt 1, Chpt 2 - Sections 2.1 - 2.5

Activity - Introduction to databases

For discussion: (5 minutes)

- What is data?
- What is information?
- What is a database?

Data

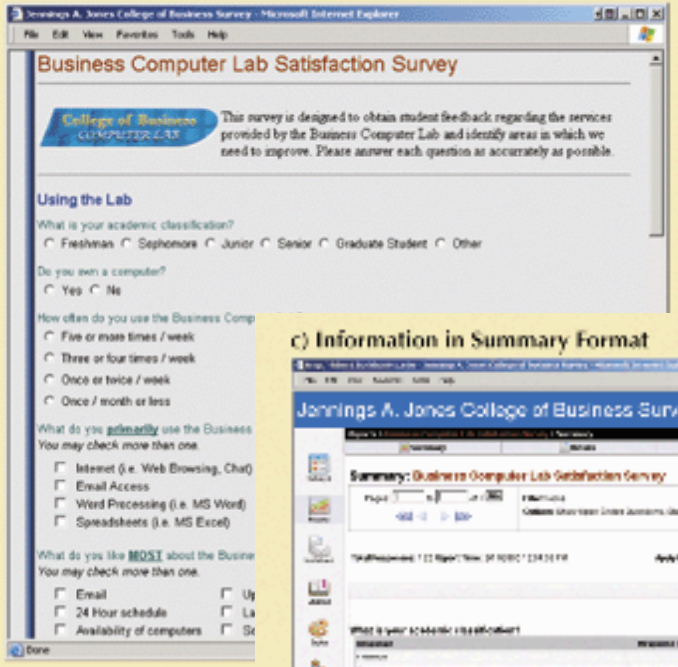
- Data refers to raw facts. Facts concerning people, places, events, or other objects/concepts
- Data by itself is useless unless it is somehow processed
- Data that is processed to reveal meaning becomes information
- A lack of data leads to inadequate information and thus ill-informed decisions
- Data represents a valuable corporate resource which needs adequate integrity and security controls
- Data management is a discipline that focuses on the proper generation, storage and retrieval of data

Activity

- In pairs, discuss the difference between data and information and think of 4 examples of each. (5 minutes)
- We will talk about this using the board

Data vs. Information

a) Initial Survey Screen

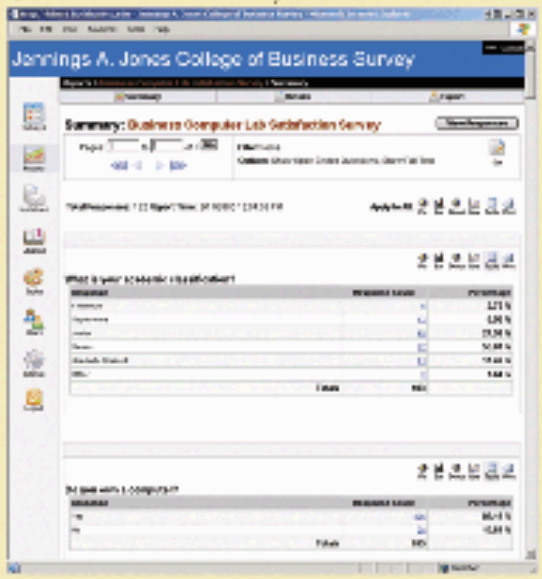


b) Raw Data

	A	B	C	D	E	F	G	H	I	J
1	AcadClas	OwnComputer	HowOftenUseLab	PrimarilyU	PrimarilyU	PrimarilyU	PrimarilyU	PrimarilyU	PrimarilyU	PrimarilyU
2	Gra	N	5W	0	1	1	1	1	1	1
3	Sen	Y	5W	1	0	0	0	1	1	1
4	Sen	Y	1W	0	0	0	1	0	0	0
5	Sen	Y	1W	1	0	0	0	1	0	0
6	Sen	Y	3W	0	0	0	0	1	0	0
7	Gra	N	5W	0	0	0	0	1	1	1
8	Sen	Y	1W	1	0	0	0	1	0	0
9	Sen	Y	3W	1	0	1	0	1	1	0
10	Sen	Y	1W	0	0	0	1	0	1	0
11	Sen	Y	5W	0	1	0	0	0	0	0
12	Jun	Y	1W	1	0	0	0	1	0	1
13	Sen	N	5W	1	0	0	1	1	1	1
14	Jun	Y	1W	0	0	0	1	0	0	0
15	Sen	Y	5W	0	0	0	1	0	0	0
16	Jun	Y	1M	0	1	0	0	0	0	1
17	Sen	Y	1W	0	0	1	0	0	0	0
18	Sen	Y	1W	0	0	1	1	0	0	0
19	Gra	N	5W							
20	Gra	Y	1M							
21	Gra	Y	5W							
22	Sen	N	3W							
23	Jun	Y	1W							
24	Sen	Y	3W							
25	Jun	Y	1W							
26	Jun	Y	1W							
27	Sen	Y	1M							
28	Sen	Y	5W							
29	Gra	Y	1M							
30	Gra	Y	5W							
31	Jun	N	5W							
32	Jun	Y	3W							
33	Gra	Y	5W							

SurveyRawData

c) Information in Summary Format



d) Information in Graphic Format

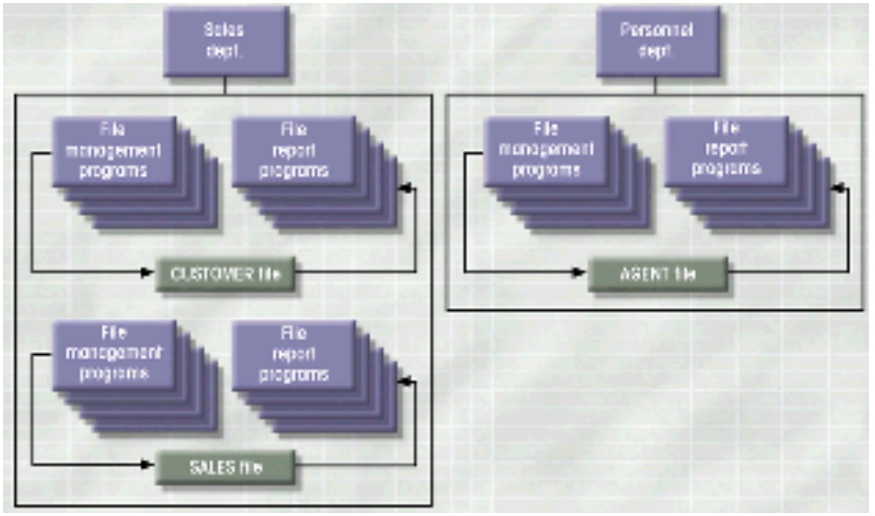
Traditional File Systems

Databases are often contrasted to traditional file systems though the latter are now rarely used. But as the problems that existed were the impetus for the development of the "Database Concept" it is worth noting some of these.

- Problems:
  - Requires extensive programming in third-generation language (3GL)
  - Data and structural dependence
  - Data redundancy

Traditional File Systems

- Requests for information (reports) required a DP specialist to write programs for the department that required the report
- File systems developed to address needs
- Data was organized in the files according to expected use - led to islands of information



- To retrieve data required extensive programming in third-generation language (3GL), this was time consuming and made ad hoc queries impossible
- Often the same data was stored in many different locations, eg. agent details occurred in both the CUSTOMER and AGENT files

Traditional File Systems

- Example of an insurance company file – *can you see the problems here?*

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	$100.00	05-Apr-2006
Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	$250.00	16-Jun-2006
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	$150.00	29-Jan-2007
Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	$300.00	14-Oct-2006
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	$100.00	28-Dec-2006
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	$850.00	22-Sep-2006
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	$120.00	25-Mar-2006
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	$250.00	17-Jul-2006
Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	$100.00	03-Dec-2006
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	$500.00	14-Mar-2006

C_NAME = Customer name
C_PHONE = Customer phone
C_ADDRESS = Customer address
C_ZIP = Customer zip code

A_NAME = Agent name
A_PHONE = Agent phone
TP = Insurance type
AMT = Insurance policy amount, in thousands of $
REN = Insurance renewal date

Traditional File Systems

- Example of storing student records in different departments in a University, can you see the problems here?

General office	Library	Hostel	Account office
RollNo	RollNo	RollNo	RollNo
Name	Name	Name	Name
Class	Class	Class	Class
Father_Name	Address	Father_Name	Address
Date of birth	Date_of _birth	Date_of_birth	Phone_No
Address	Phone_No	Address	Fee
Phone_No	No_of_books_issued	Phone_No	Installments
Previous_Record	Fine	Mess_Bill	Discount
Attendance	etc.	Room No	Balance
Marks		etc.	Total
etc.			etc.

Traditional File Systems

- Data Dependence
  - Changes in file's data characteristics requires modification of data access programs
  - Must tell program what to do and how
  - Makes file systems cumbersome from a programming and data management views
- Structural Dependence
  - Change in file structure requires modification of related programs
- Data Redundancy
  - Different and conflicting versions of same data
  - Results of uncontrolled data redundancy
    - > Data anomalies
      - Modification, Insertion, Deletion

Traditional File Systems

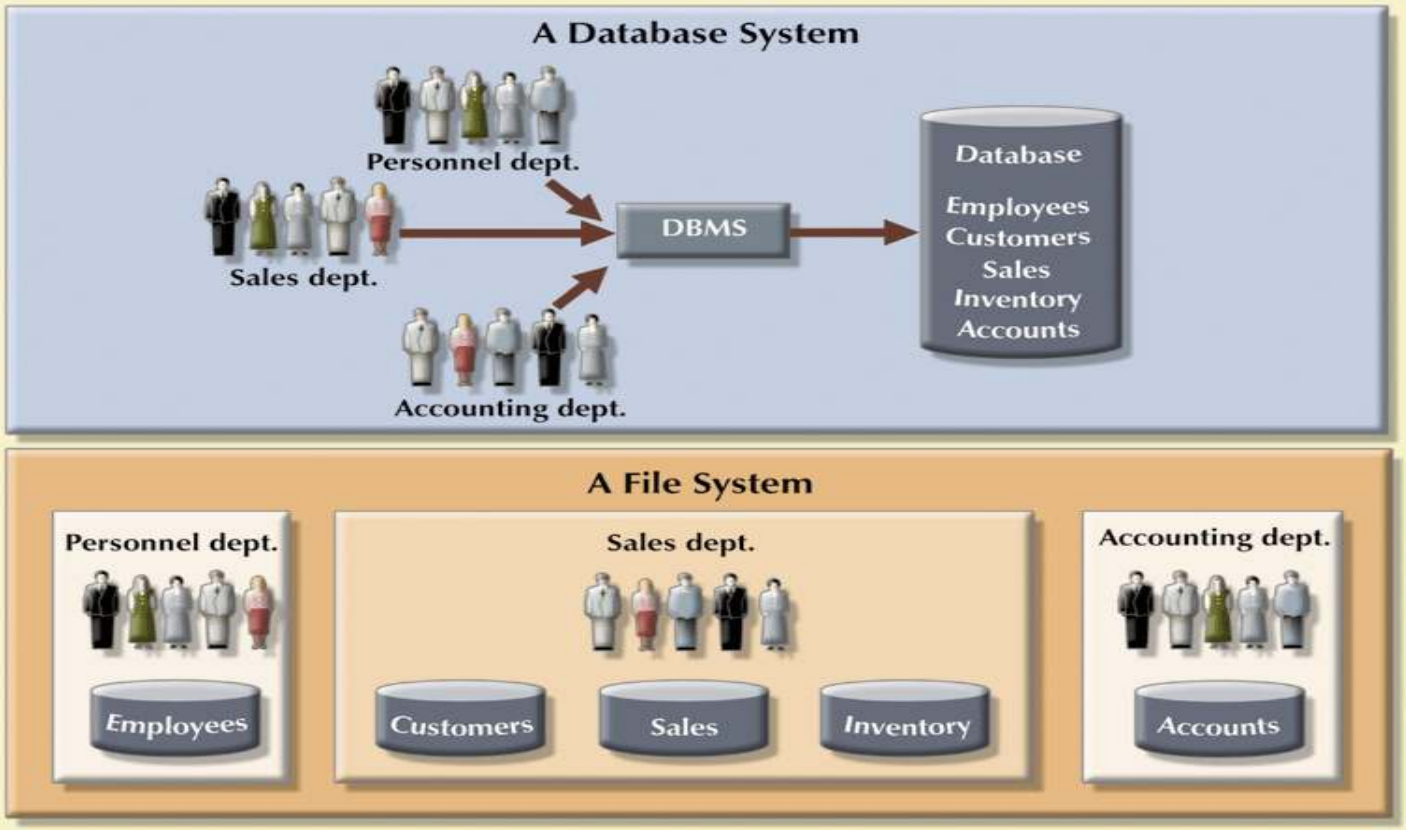
- Data Redundancy (cont)
  - Data inconsistency
    - > Lack of data integrity
- File Terminology
  - Field
    - > group of characters with specific meaning
  - Record
    - > logically connected fields that describe a person, place, or thing
  - File
    - > collection of related records

Database System

- Consists of related data stored in a single repository
- Provides advantages over file system management approach
  - Eliminates inconsistency, data anomalies, data dependency, and structural dependency problems
  - Stores data structures, relationships, and access paths
- The centralised control of data means that for many applications the data already exists
- The data is no longer related by application programs, but by the structure defined in the database

Database vs. File Systems

FIGURE 1.6 Contrasting database and file systems



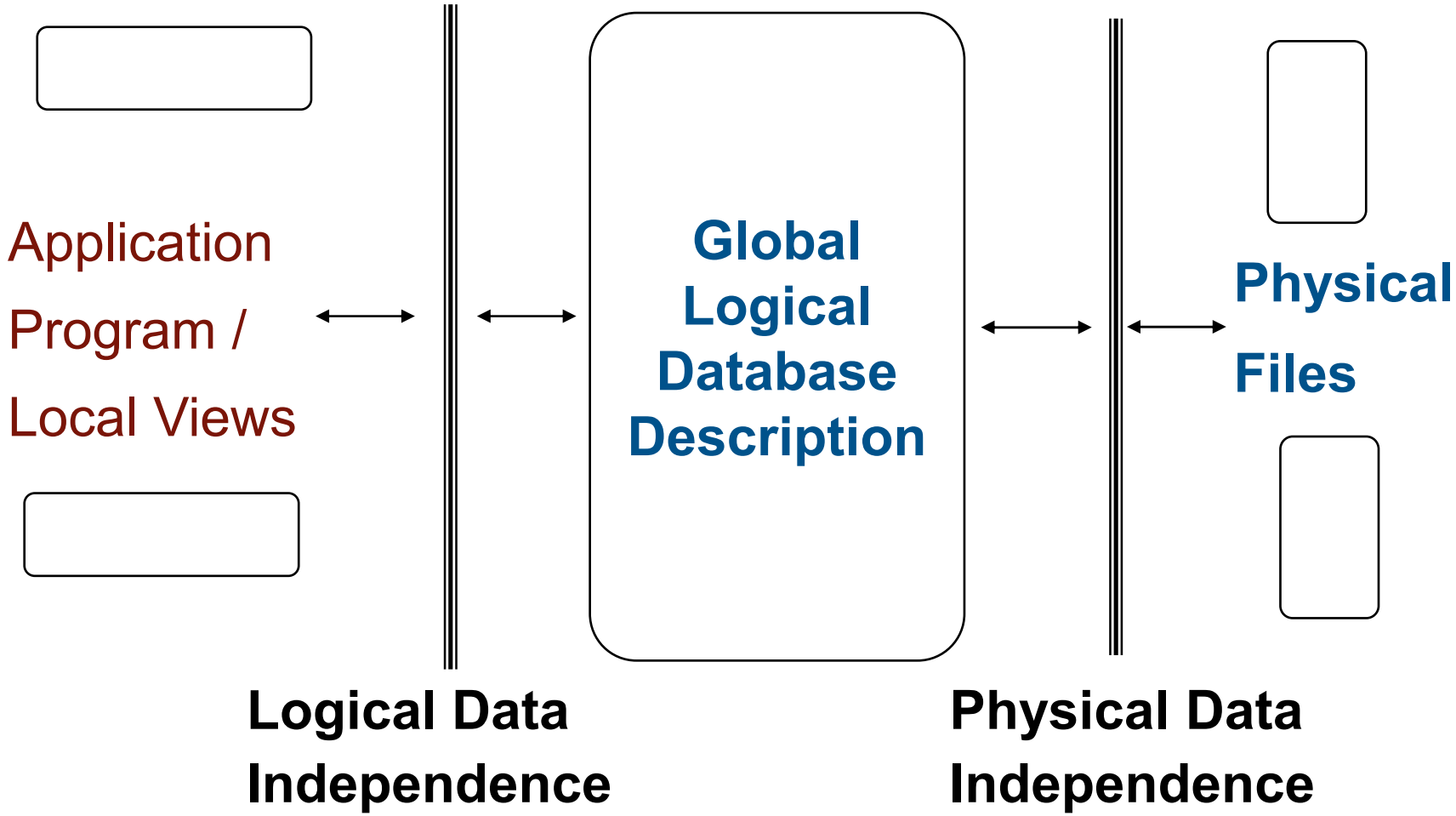
Objectives of Database Technology

- Data Independence
- Minimal Data Redundancy
- Increased Data Sharing
- Improved Data Quality
- Improved Security of Data
- Improved Access to Data
- Reduced Program Maintenance
- Inter-relate data thru the model

Data Independence

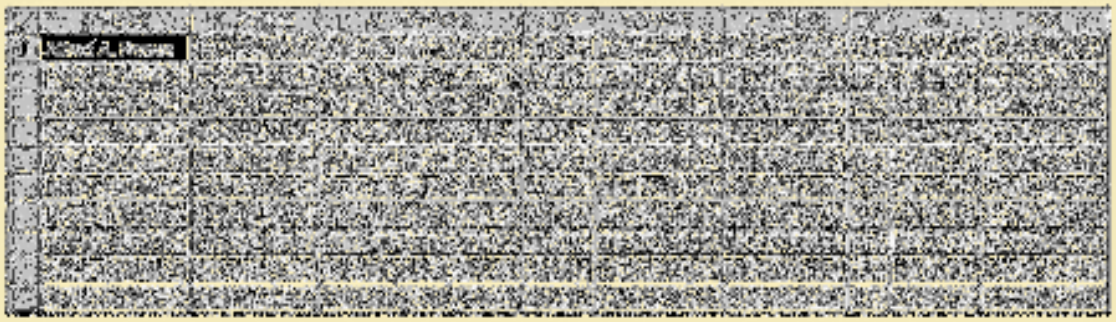
- Is the property of being able to change the logical or physical structure of data without requiring changes to application programs that manipulate that data
- Results in a more complex structure
- Data is stored independently of the programs
- Data descriptions are not embedded in application programs.
- Can the database structure be changed with no impact on application programs?
- Role of the database catalog or dictionary (the metadata)

Logical vs Physical Data Independence



Minimal Data Redundancy

Returning to slide 8:



- Minimise the duplication of data, separate into two 'piles' of data
  - CUSTOMER data:
    - c_no, c_name, c_address, c_zip, tp, amt, ren, a_code,
  - AGENT data:
    - a_no, a_name, a_phone
    - Note agent details now only appears once for each agent

Sharing of Data – the problem

- The DBMS should:
  - support multiple concurrent users of the same data and
  - ensure that the data remains consistent at all times
- To avoid concurrency problems, transactions must be made logically serial
  - One common technique used is record locking
  - That is, a transaction can lock a record, preventing update by another transaction, until the update has completed.

Data Integrity

- Validation or integrity rules should be defined and automatically invoked at run time by the DBMS
- Significant variation exists among DBMSs in the level of support for data integrity.
- ANSI/ISO suggest that 100% of all enterprise rules should be held in the database, and specifically none in application programs.
- An area of significant development during the 1990s.

Security

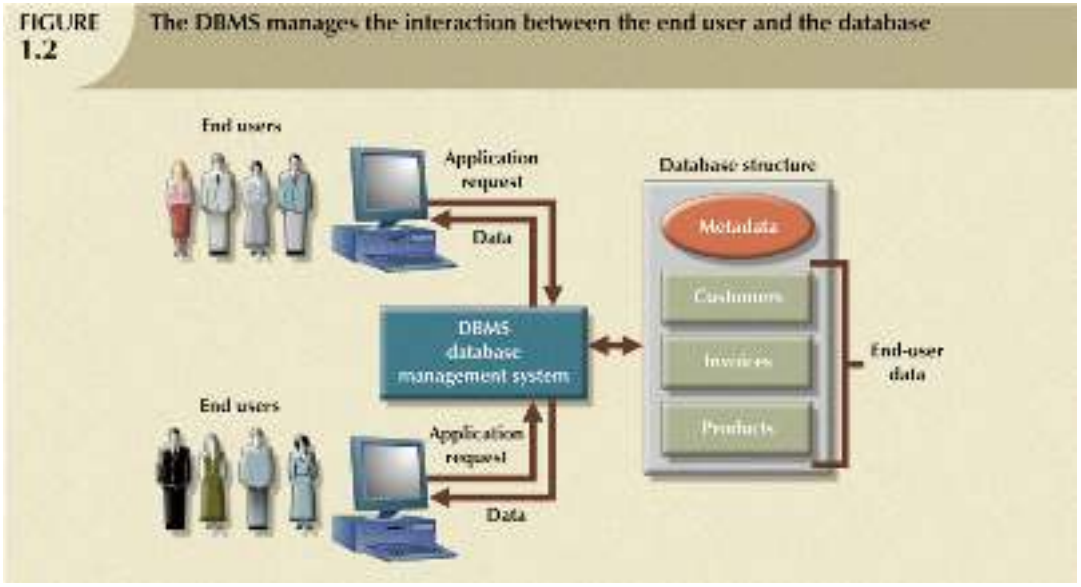
- Protecting data against accidental or intentional use by unauthorised users
- Each user requires identification with a user-id and password
- Users can be restricted in the data they can see and what actions they can perform on that data.
- The DBMS encrypts and decrypts data as it is stored and retrieved.
- **Views** are often used to limit users access to data

Database and DBMS

- So…a **database** is a shared integrated computer structure that houses a collection of:
  - **end user data** eg. product details, orders placed for these products
  - **metadata**, or data about data, through which the data is integrated and managed
- A **database management system** (DBMS) is

  — a collection of programs that manages the database structure and controls access to the data stored in the database

  — includes a query language



DBMS Functions

- Data dictionary management
  - stores the definitions of data and their relationships (metadata) in a data dictionary; any changes made are automatically recorded in the data dictionary.
- Data storage management
  - creates and manages the complex structures required for data storage.
- Data transformation and presentation
  - transforms entered data to conform to the data structures that are required to store the data

DBMS Functions

- Security management
  - creates a security system and enforces security within that system.
- Multi-user access control
  - creates complex structures that allow multiple-user access to the data.
- Backup and recovery management
  - performs backup and data recovery procedures to ensure data safety.

DBMS Functions

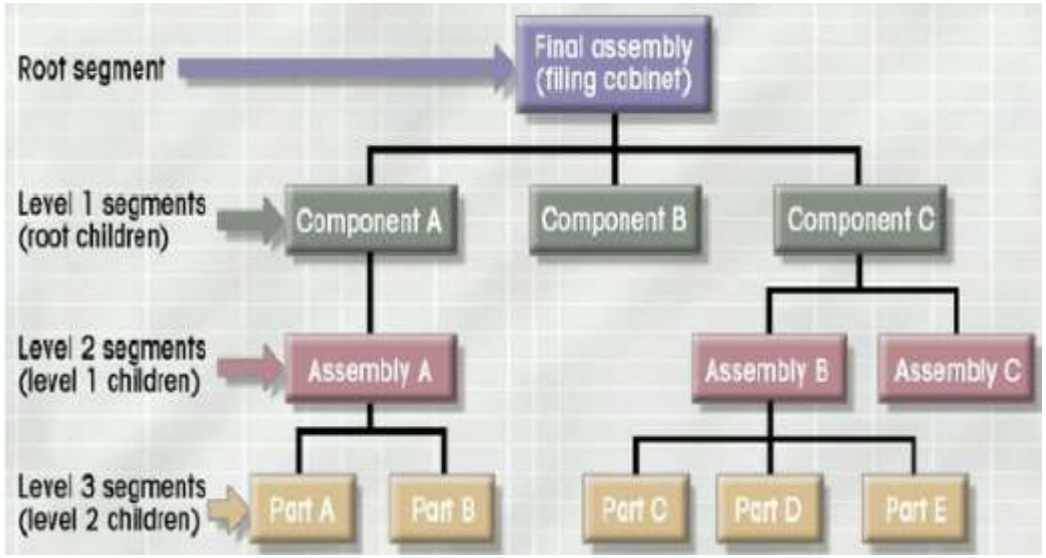
- Data integrity management
  - promotes and enforces integrity rules to eliminate data integrity problems
- Database language and application programming interfaces
  - provides access to the data via utility programs and from programming languages interfaces.
- Database communication interfaces
  - provides end-user access to data within a computer network environment.

Database Models

- Collection of logical constructs used to represent data structure and relationships within the database
  - Conceptual models: logical nature of data representation
  - Implementation/Physical models: emphasis on how the data are represented in the database
- Conceptual Models consists of
  - Context Model
  - Key Model
  - Fully attributed Model
- Implementation/Physical Models consists of
  - Hierarchical
  - Network
  - Relational

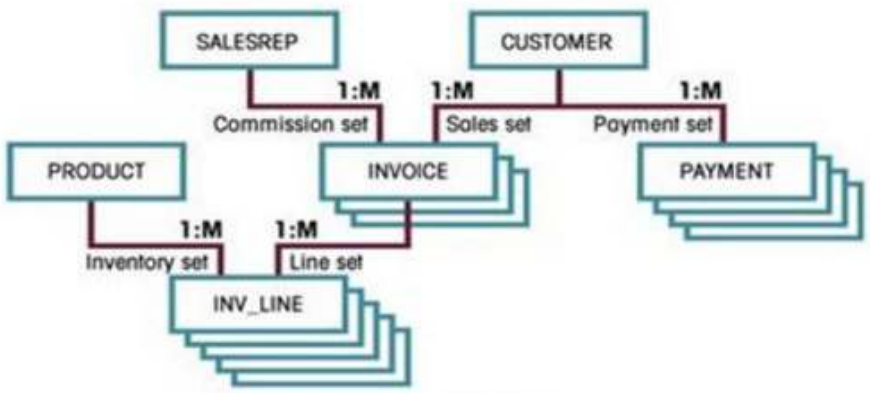
Implementation Models

- Hierarchical
  - Logically represented by an upside down tree
    - Each parent can have many children
    - Each child has only one parent



Implementation Models

- Network
  - Each record can have multiple parents
    - Composed of sets
      - Each set has owner record and member record
      - Member may have several owners



Network data model

Relational Database Model

- **Perceived by user** as a collection of tables for data storage
- A table is a series of row/column intersections
- Tables related by sharing common entity characteristics
  - *This semester we will be concentrating on the Relational database model (the basis for most current DBMS)*

Relational Database Model

- Advantages
  - Structural independence
  - Improved conceptual simplicity
  - Easier database design, implementation, management, and use
  - Ad hoc query capability with SQL
  - Powerful database management system
- Disadvantages
  - Substantial hardware and system software overhead
  - Conceptual simplicity allows untrained people to potentially generate poor designs
  - Ease of development may promote 'islands of information' as departments 'go their own way'

Relational Database Terminology

A relational database represents a collection of relations

- **EMPLOYEE relation: EMPLOYEE (emp_num, emp_lname, emp_fname, emp_initial, emp_hiredate, job_code)**

Table name: EMPLOYEE

	EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
▶	101	News	John	G	08-Nov-00	502
	102	Senior	David	H	12-Jul-89	501
	103	Arbough	June	E	01-Dec-97	503
	104	Ramoras	Anne	K	15-Nov-88	501
	105	Johnson	Alice	K	01-Feb-94	502
	106	Smithfield	William		22-Jun-05	500
	107	Alonzo	Maria	D	10-Oct-94	500
	108	Washington	Ralph	B	22-Aug-89	501
	109	Smith	Larry	W	18-Jul-99	501
	110	Olenko	Gerald	A	11-Dec-96	505
	111	Wabash	Geoff	B	04-Apr-89	506
	112	Smithson	Darlene	M	23-Oct-95	507
	113	Joenbrood	Delbert	K	15-Nov-94	508
	114	Jones	Annelise		20-Aug-91	508
	115	Bawangi	Travis	B	25-Jan-90	501
	116	Pratt	Gerald	L	05-Mar-95	510
	117	Williamson	Angie	H	19-Jun-94	509
	118	Frommer	James	J	04-Jan-06	510

Row or TUPLE

Column or ATTRIBUTE

Tables, as above, are used as pictorial representations of relations although there are some problems with this approach – see next week

Summary

- Information is derived from data, which are usually stored in a database
- A DBMS is software that implements and manages a database
- Databases were developed to address the weaknesses of file systems
- A DBMS
  - presents to the user a single data repository that promotes data sharing
  - Enforces data integrity, eliminates redundancy and promotes data security