Course Objectives

- Introduction to very basics
- Guides through different design stages
- Familiarize with tools
- Emphasis on design stages of database

Lecture Objectives

- Some common uses of database systems.

- Characteristics of file-based systems.

- Problems with file-based approach.

- Meaning of the term database and Database Management System (DBMS).

- Database systems Approach characteristics

- Advantages and limitations of DBMS

- Components of Database systems

- Users of Database systems

Definitions

- *Data*: stored representations of meaningful objects and events or

- Referred to facts concerning objects and events that could be recorded and stored on computer media
  - Structured: numbers, text, dates
  - Unstructured: images, video, documents

- *Information*: data processed to increase knowledge in the person using the data

Files and Databases

- **File**: A collection of records or documents dealing with one organization, person, area or subject. (Rowley)
  - Manual (paper) files
  - Computer files: MS Word files, Ms Excel files etc
- Database: A collection of similar records with relationships between the records. (Rowley)

Examples of databases

- Accounts database; payroll database; UW's students database; Amazon's products database; airline reservation database

Terms and Concepts

- Database Management System -- DBMS
  - Software system used to define, create, maintain and provide controlled access to the database and repository

- ✓ Examples of DBMS

  DB2 (IBM), SQL Server (MS), Oracle, Sybase

  MySQL, Postgres,

Terms and Concepts

- Repository
  - AKA Data Dictionary
  - The place where all metadata for a particular database is stored
  - may also include information on relationships between files or tables in a particular database

Terms and Concepts

- Metadata
  - *Metadata:* data that describes the properties and context of user data
  - Data about data
    - In DBMS means all of the characteristics describing the attributes of an entity, E.G.:
      - name of attribute
      - data type of attribute
      - size of the attribute
      - format or special characteristics
  - Characteristics of files or relations
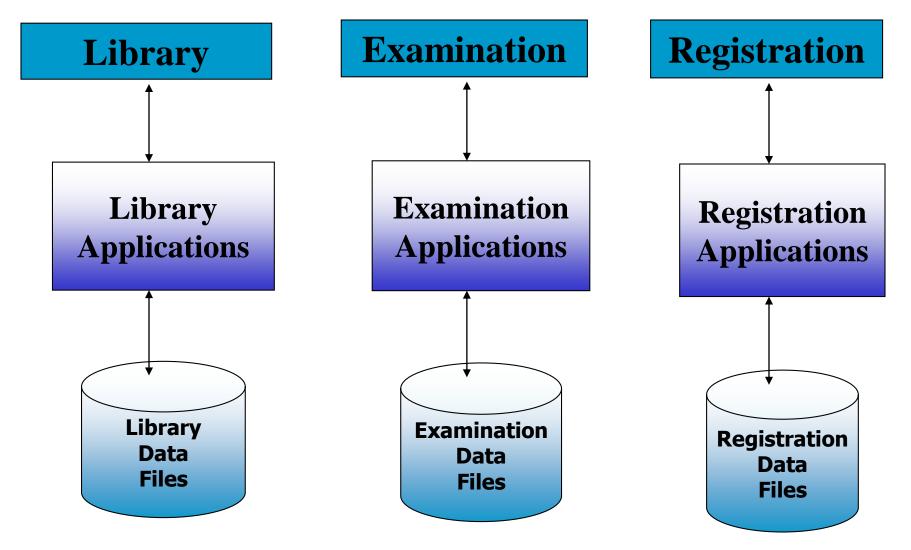   - name, content, notes, etc.

A bit of History

- Computer initially used for computational/ engineering purposes

- Commercial applications introduced File Processing System

File Processing System

❑ A collection of application programs that perform services for the end-users such as production of reports

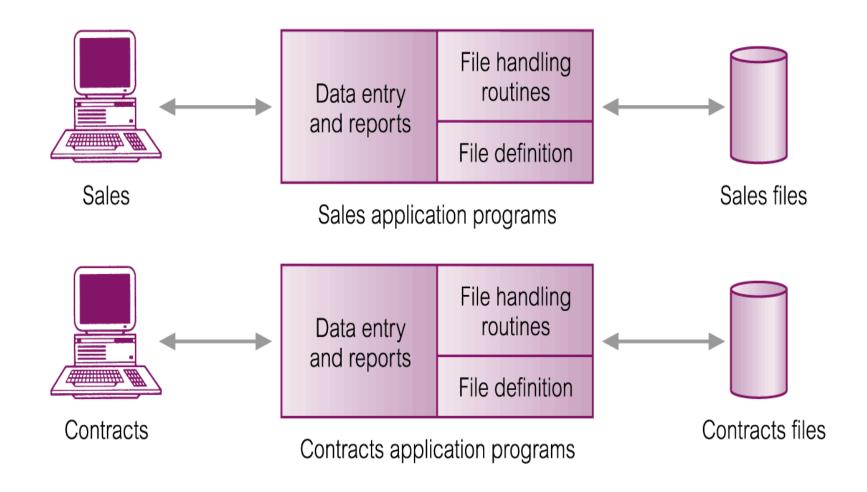❑ Each program defines and manages its own data

File Processing Systems

Library	Examination	Registration

Library Applications	Examination Applications	Registration Applications

Library Data Files	Examination Data Files	Registration Data Files

Program and Data Interdependence

File Processing Systems

Library
Reg_Number
Name
Father Name
Books Issued
Fine

Examination
Reg_Number
Name
Address
Class
Semester
Grade

Registration
Reg_Number
Name
Father Name
Phone
Address
Class

Files Based Processing

Problems/Disadvantages of File based systems

Data Redundancy and Inconsistency:

- Data redundancy: The presence of duplicate data in multiple data files so that the same data are stored in more than one place or location

- Data inconsistency: The same attribute may have different values.

Problems/Disadvantages of File based systems

Program-data dependence:

- The coupling of data stored in files and the specific programs required to update and maintain those files such that changes in programs require changes to the data

Lack of flexibility:

- A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad-hoc reports or respond to unanticipated information requirements in a timely fashion.

Problems/Disadvantages of File based systems

Poor security:

- Because there is little control or management of data, management will have no knowledge of who is accessing or even making changes to the organization's data.

Limitation of data sharing and availability:

- Information cannot flow freely across different functional areas or different parts of the organization. Users find different values of the same piece of information in two different systems, and hence they may not use these systems because they cannot trust the accuracy of the data.
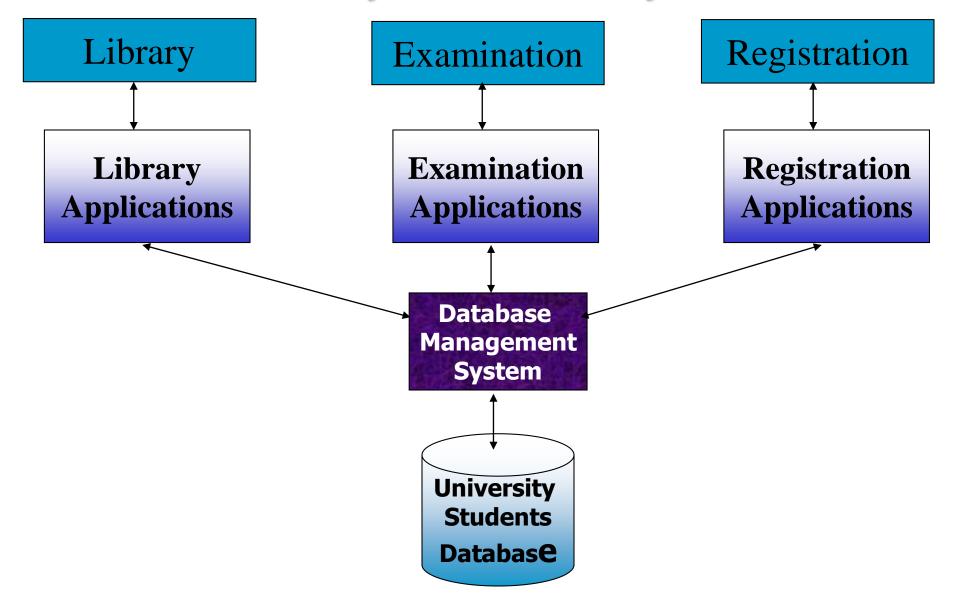
Problems/Disadvantages of File based systems

- **Incompatible file formats**
  - Programs are written in different languages, and so cannot easily access each others files.

- **Fixed Queries/Proliferation of application programs**
  - Programs are written to satisfy particular functions. Any new requirement needs a new program.
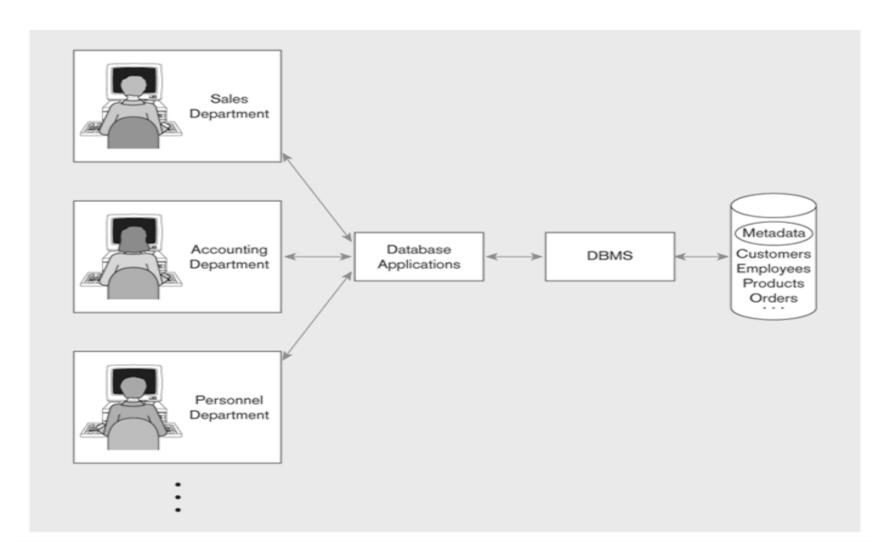
SOLUTION:
The DATABASE Approach

- Central repository of shared data

- Data is managed by a controlling agent

- Stored in a standardized, convenient form

This requires a
Database and Database Management System (DBMS)

University Database Systems

Database approach at Pine Valley Furniture Company

DBMS manages data resources like an operating system manages hardware resources

Main Characteristics of the Database Approach

- **Self-describing nature of a database system:**
  - A DBMS **catalog** stores the description of a particular database (e.g. data structures, types, and constraints)
  - The description is called **meta-data**.
  - This allows the DBMS software to work with different database applications.
- **Insulation between programs and data:**
  - Called **program-data independence**.
  - Allows changing data structures and storage organization without having to change the DBMS access programs.

Main Characteristics of the Database Approach (continued)

- **Data Abstraction:**
  - A **data model** is used to hide storage details and present the users with a conceptual view of the database.
  - Programs refer to the data model constructs rather than data storage details

- **Support of multiple views of the data:**
  - Each user may see a different view of the database, which describes **only** the data of interest to that user.

Main Characteristics of the Database Approach (continued)

- **Sharing of data and multi-user transaction processing:**
  - Allowing a set of **concurrent users** to retrieve from and to update the database.
  - *Concurrency control* within the DBMS guarantees that each **transaction** is correctly executed or aborted
  - *Recovery* subsystem ensures each completed transaction has its effect permanently recorded in the database
  - **OLTP** (Online Transaction Processing) is a major part of database applications. This allows hundreds of concurrent transactions to execute per second.

Advantages of Using the Database Approach

- Controlling redundancy in data storage and in development and maintenance efforts.

- Sharing of data among multiple users.

- Restricting unauthorized access to data.

- Enforcing integrity constraints on the database.

- Providing Storage Structures (e.g. indexes) for efficient Query Processing

- Providing multiple interfaces to different classes of users (Views).

Advantages of Using the Database Approach (continued)

- Providing backup and recovery services.
- Potential for enforcing standards:
  - This is very crucial for the success of database applications in large organizations. **Standards** refer to data item names, display formats, screens, report structures, meta-data (description of data), Web page layouts, etc.
- Flexibility to change data structures:
  - Database structure may evolve as new requirements are defined.

Additional Implications of Using the Database Approach

- Availability of current information:
  - Extremely important for on-line transaction systems such as airline, hotel, car reservations.

- Economies of scale:
  - Wasteful overlap of resources and personnel can be avoided by consolidating data and applications across departments.

Additional Implications of Using the Database Approach

- Reduced application development time:
  - Incremental time to add each new application is reduced.

Disadvantages of DBMSs

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

Components of DBMS Environment

- ## 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.

Components of DBMS Environment

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

- **People**

Tutorial Question

Discuss the roles of the following personnel in the

database environment:

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

Historical Development of Database Technology

- Early Database Applications:
  - The Hierarchical and Network Models were introduced in mid 1960s and dominated during the seventies.
  - A bulk of the worldwide database processing still occurs using these models, particularly, the hierarchical model.
- Relational Model based Systems:
  - Relational model was originally introduced in 1970, was heavily researched and experimented within IBM Research and several universities.
  - Relational DBMS Products emerged in the early 1980s.

Historical Development of Database Technology

- Object-oriented and emerging applications:
  - Object-Oriented Database Management Systems (OODBMSs) were introduced in late 1980s and early 1990s to cater to the need of complex data processing in CAD and other applications.
    - Their use has not taken off much.
  - Many relational DBMSs have incorporated object database concepts, leading to a new category called *object-relational* DBMSs (ORDBMSs)
  - *Extended relational* systems add further capabilities (e.g. for multimedia data, XML, and other data types)

Historical Development of Database Technology (continued)

- Data on the Web and E-commerce Applications:
  - Web contains data in HTML (Hypertext markup language) with links among pages.

  - This has given rise to a new set of applications and E-commerce is using new standards like XML (eXtended Markup Language).

  - Script programming languages such as PHP and JavaScript allow generation of dynamic Web pages that are partially generated from a database

    - Also allow database updates through Web pages

Extending Database Capabilities

- New functionality is being added to DBMSs in the following areas:
  - Scientific Applications
  - XML (eXtensible Markup Language)
  - Image Storage and Management
  - Audio and Video Data Management
  - Data Warehousing and Data Mining
  - Spatial Data Management
  - Time Series and Historical Data Management

- The above gives rise to *new research and development* in incorporating new data types, complex data structures, new operations and storage and indexing schemes in database systems.