Introduction to Databases

Objectives

- After studying this lesson you should be able to:
  - Describe the characteristics of a file-based system and discuss its shortcomings
  - Describe the components of, functions of and people involved with a DBMS
  - Distinguish between the three levels of a database architecture
  - Understand the nature and importance of data independence

Data and Information

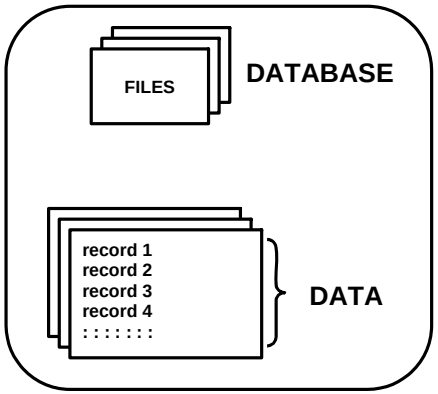
- **Data** are raw facts concerning things such as people, objects or events ("Raw" indicates that facts and figures have not yet been processed to reveal their meaning).
  - Examples: Employee Number, Sales Amount.
- **Information** is data that have been processed and presented in a form suitable for human interpretation, often with the purpose of revealing trends or patterns
  - Examples: Sales per quarter, Sales by employee etc.

Data Management

- Data Management is a discipline that focuses on the proper generation, storage and retrieval of data.
- Efficient data management requires the use of a computer-based database system..

Database Concept

- A **Database System** is composed of a shared collection of logically related data and a description of this data (sometimes called metadata), designed to meet the information needs of an organization



Conventional File Processing System

- Typically consist of a set of application programs that performed various tasks, such as the production of reports.
- Each program would define and mange its own data.
- As business became more complex, this approach developed a number of shortcomings and limitations

Basic File Terminology

- Data
  - Raw facts that have little meaning unless they have been organized in some logical manner.
- Field
  - A character or group of characters that has a specific meaning. A field might be defined as telephone number, name, date and so on.

Basic File Terminology

- Record
  - A logically connected set of one or more fields that describe a person, place or thing.
- File
  - A collection of related records. For examples student records, vendor records.

Limitations of Conventional File Processing System

- Uncontrolled Redundancy
- Inconsistent Data
- Inflexibility
- Limited Data Sharing
- Poor Enforcement of Standards
- Excessive Program Maintenance

Uncontrolled Redundancy

- Each application has it own files, this is an approach that inevitably leads to high levels of data redundancy.
- Valuable storage space is waste
- The same data will have to be input several times to update all occurrences of data item
- Inconsistencies in the various versions

Inconsistent Data

- Changes to data must be made simultaneously (and correctly) to each of the files that contain same data item.
- Inconsistencies in stored data are one of the most common sources of errors.
- Inconsistent documents and reports.
- Undermine the confidence of users in the integrity of the information system.

Inflexibility

- It cannot easily respond to request for a new or redesigned documents and reports.
- It cannot readily satisfy demands for information in a new format.

Limited Data Sharing

- Each application has its own private files, and users have little opportunity to share the data outside of their own applications.
- One consequence of this is that the same data may have to be entered several times.
- Another consequence is that in developing new applications, the designer often cannot

Poor Enforcement of Standards

- Standards are required for data names, formats and access restrictions.
- Two types of inconsistencies may result from poor enforcement of standards: synonyms and homonyms.
  - A **Synonym** results when two different names are used for the same data item
    - E.g. student number and registration number
  - A **Homonym** is a single name that is used for two different data items
    - E.g. The term balance might be used to designate a checking account or savings account balance in different departments

Extensive Program Maintenance

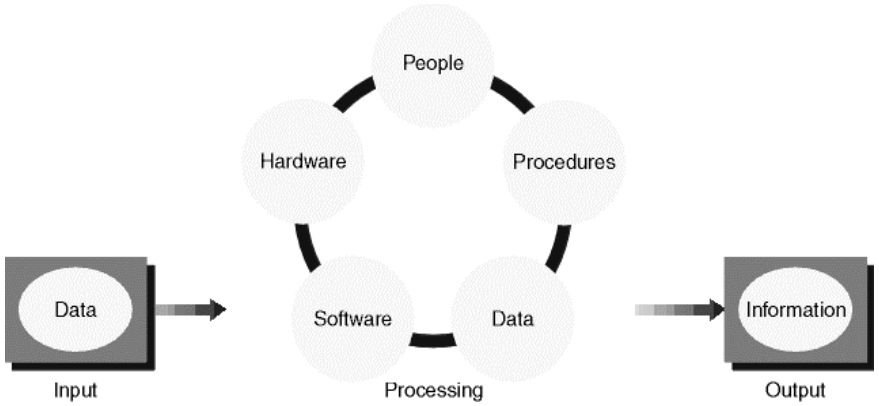
- The descriptions of files, records and data items are embedded within individual application programs.
- Any modification to a data file (change in data name format or method of access) requires that the program also be modified.

Database System Environment

- The term database system refers to an organization of components that define and regulate the collection, storage, management and use of data within a database environment.

Database System

- An Database System has five key components: hardware, software, data, processes, and people.



Database System – Hardware

- Hardware
  - Identifies all the system's physical devices. This can be anything from a single PC to a full network of computers supported by one or more servers.

Database System – Software

- Software
  - Refers to the collection of programs used within the database system. It takes three types of software to make the database system function fully:
    - Operating Systems
    - DBMS Software
    - Applications Programs and Utilities

Database System – Software

- Operating Systems Software
  - Manages all hardware components and makes it possible for all other software to run on the computers.
- DBMS Software
  - Manages the database within the database system. Examples include Microsoft's Access and SQL Server, Oracle and IBM's DB2.
- Application Program and Utilities
  - Used to access and manipulate the data in the DBMS. Application programs are used to access the data and generate reports, tabulations and other information to facilitate decision-making.

Database System – People

- This component includes all users of the database system.
  - **Systems Administrators** oversee the database system's general operations.
  - **Database Administrators**, also known as DBA, and manage the DBMS's use functionality.
  - **Database Designers** design the database structure.
  - **Systems Analyst and Programmers** design and implement the application programs. They design and create the data entry screens, reports and procedures to access and manipulate the data.
  - **End Users** are the people who use the application programs.

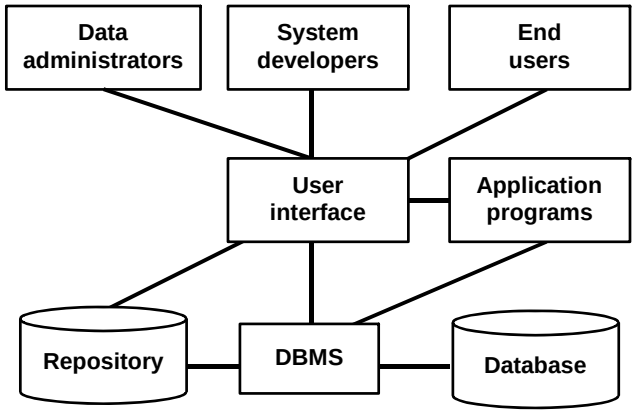
Database System – Procedures

- The instructions and rules that govern the design and use of the database system.
  - These could include: how to log on, how to use a particular feature of the database, how to backup data and what to do in the event of a failure, either of hardware or software.

Database System – Data

- Collection of facts and figures stored in the database from which information is generated.

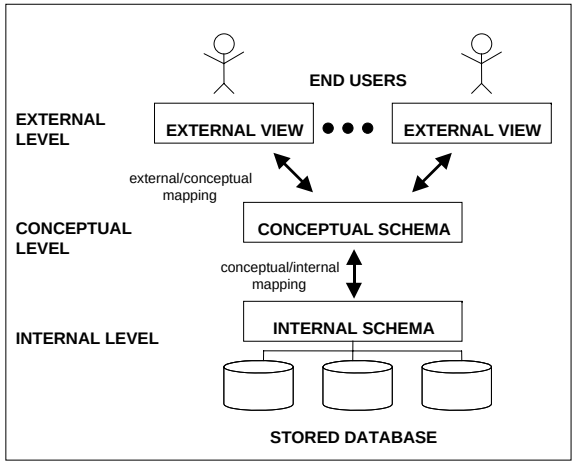
Components of the Database Environment



Components of Database System

- Different databases can be structured in different ways, most will have the following recognizable components:
  - **Physical Database**: stores the actual data
  - **File Manager**: manages disk storage and data structure
  - **Database Manager:** interfaces between the low level data and request/application programs
  - **Query Processor:** translates requests into low level instructions
  - **DDL (Data Definition Language) Compiler**: converts DDL statements into sets of tables
  - **Data Dictionary:** stores the "meta data" about the structure of the database

The Three Levels Architecture



Three Levels Architecture

- The objective of the three-level architecture is to separate a user's view of the database from the way that the data is physically represented.
- In this architecture, schemas can be defined at the following three levels:
  - The Internal Level
  - The Conceptual Level
  - The External or View Level

External Level

- This is concerned with **Views** of individual users or groups of users
- There can be many distinct external views
- Such a view will includes only those data concepts in which the user is interested
- This enables the DBA to limit access and ensure security

Conceptual Level

- This describes the **Logical Structure** of the entire database
- It hides the physical storage structures and concentrates on describing the data concepts
- It includes all the data and their relationships
- It is constructed using one of a number of possible basic data models (Examples of these are: Relational, Hierarchical, Network or Object-Oriented).

Internal Level

- This is the **Physical Storage Structure** of the database
- It describes how the data is held and manipulated
- It is described by an internal schema which includes such details as data types, addressing modes, compression and encryption techniques.
- It contains occurrences of internal/stored records

Reasons for Three Levels Architecture

- Each user should be able to access the same data, but may need to have a different customized view of the data.
- Users should not have to deal directly with physical database storage details, such as indexing or hashing.
- DBA should be able to change the database storage structures without affecting the users' views.
- The internal structure of the database should be unaffected by changes to the physical aspects of storage.
- The DBA should be able to change the conceptual or global structure of the database without affecting all users.

Schemas, Mappings and Instances

- The overall description of the database is called the **Database Schema**.
- The data in the database at any particular point in time is called a **Database Instance**. The schema is sometimes called the intension of the database, while an instance is called an extension (or state) of the database.

Three Level Architecture for Schema

- There are three different types of schema in the database and these are defined according to the levels of abstraction of the three-level architecture
  - External schema (also called subschema)
  - Conceptual schema
  - Internal schema

Mapping

- Three schemas are only descriptions of data
- If the request is a database retrieval, the database extracted from the stored database must be reformatted to match the user's external view before it is presented to the user.
- The process of transforming requests and results between levels are called **Mapping**.

Conceptual/Internal Mapping

- The conceptual schema is related to internal schema
- Enables DBMS to find the actual record or combination of records in physical storage that constitutes a logical record
- Allows any differences in entity names, attribute names, data types to be resolved

External/Conceptual Mapping

- Each external schema is related to the conceptual schema
- Enables DBMS to map names in the user’s view onto the relevant part of the conceptual schema