Unit 1
Database and Database Users

- Introduction
- Characteristics of the Database Approach
- Actors on the Scene
- Workers behind the Scene
- Advantages of Using the DBMS Approach

Introduction

- **Data** is defined as the known facts that can be recorded and have an implicit meaning.
- A **database** is the collection of related persistent data and contains information relevant to an enterprise. The database is also called the repository or container for a collection of data files. For example, **university database** for maintaining information about *students*, *courses* and *grades* in university.
- A **database management system** (DBMS) is a computerized system that enables users to create and maintain a database. For example, MySQL, Oracle, PostgreSQL, MariaDB, Microsoft SQL Server etc.

Introduction

- The DBMS is a general-purpose software system that facilitates the processes of ***defining***, ***constructing***, ***manipulating***, ***sharing***, ***protecting***, *and* ***maintaining*** the databases.
- **Defining** a database involves specifying the data types, structures, and constraints of the data. The database definition information is also stored by the DBMS in the form of a database catalog or dictionary called ***meta-data***.
- **Constructing** databases is the process of storing the data on some storage medium that is controlled by the DBMS.
- **Manipulating** a database includes functions like querying the database to retrieve specific data, updating the database to reflect changes, and generating reports from the data.

Introduction

- **Sharing** a database allows multiple users and programs to access the database simultaneously.
- **Protection** includes system protection against hardware or software malfunction (or crashes) and security protection against unauthorized or malicious access.
- DBMS must be able to **maintain** the database system by allowing the system to evolve as requirements change over time.
- An **application program** accesses the database by sending queries or requests for data to the DBMS. A **query** typically causes some data to be retrieved. A **transaction** may cause some data to be read and some data to be written into the database.

Introduction

- The database and DBMS software together is called **database system**.
- Databases play a critical role in almost all areas including business, electronic commerce, social media, engineering, medicine, genetics, law, education, and library science.
- Users of a DBMS system can create a database and perform the following basic operations on the database:
  - Adding new, empty data files to the database.
  - Inserting data into existing data files.
  - Retrieving data from existing data files.
  - Changing data in existing data files.
  - Deleting data from existing data files.
  - Removing existing data files from the database.

Introduction

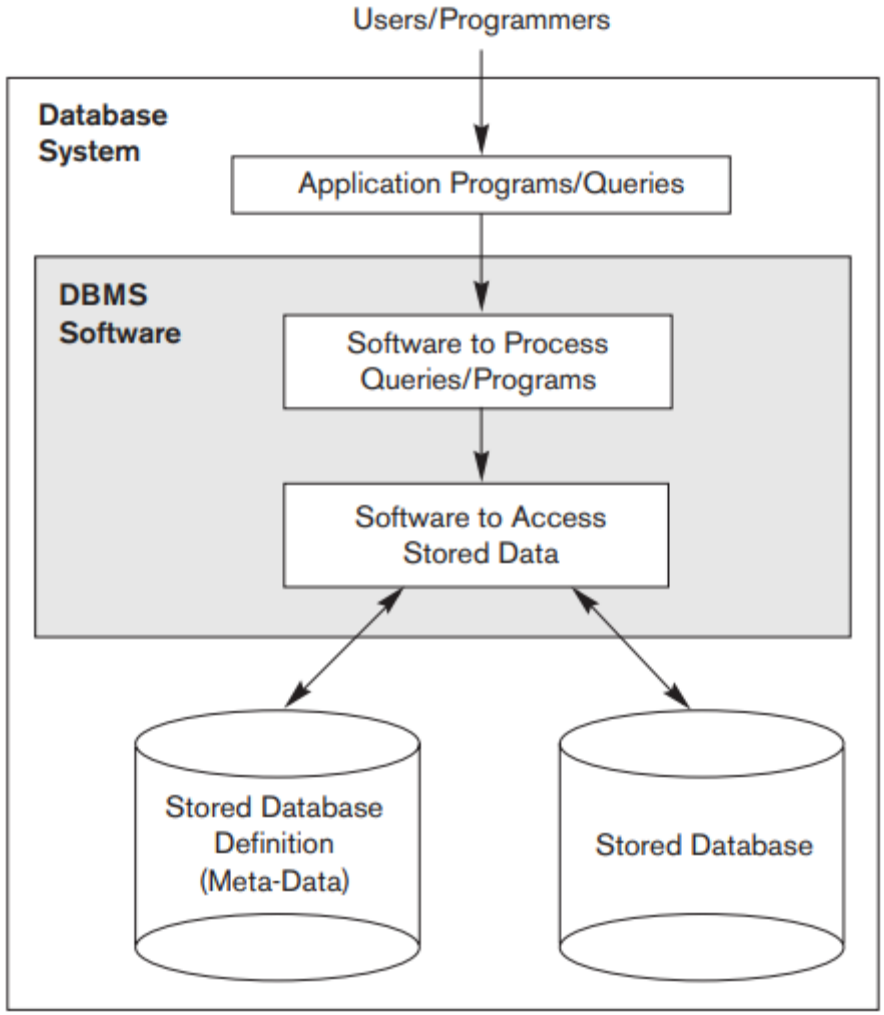

Fig: A simplified database system environment

Characteristics of Database Approach

- A number of characteristics distinguish the database approach from the much older approach of writing customized programs to access data stored in files.
- In traditional **file processing**, each user defines and implements the files needed for a specific software application as part of programming the application. This redundancy in defining and storing data results in wasted storage space and in redundant efforts to maintain common up-to-date data.
- In the **database approach**, a single repository maintains data that is defined once and then accessed by various users repeatedly through queries, transactions, and application programs.

Characteristics of Database Approach

- The main characteristics of the database approach versus the file-processing approach are the following:
  - Self-describing nature of a database system
  - Insulation between programs and data, and data abstraction
  - Support of multiple views of the data
  - Sharing of data and multiuser transaction processing
- **Self-describing nature of a database system:**
  - DBMS contains not only the database itself but also a complete definition or description of the database structure and constraints.
  - This definition is stored in the DBMS catalog, which contains information such as the structure of each file, the type and storage format of each data item, and various constraints on the data.
  - The information stored in the catalog is called **meta-data**, and it describes the structure of the database.

Characteristics of Database Approach

- **Insulation between programs and data, and data abstraction:**
  - Traditional file processing allows structure of data files to be embedded in the application programs, so any changes to the structure of a file may require changing all programs that access that file. By contrast, in database processing, the structure of data files is stored in the DBMS catalog separately from the access programs. We call this property **program-data independence**.
  - Some database management systems allow user application programs to operate on the data by invoking operations through their names and arguments, regardless of how the operations are implemented. This is termed **program-operation independence**.
  - The characteristic that allows program-data independence and program-operation independence is called **data abstraction**. A DBMS provides conceptual representation of data that hides details of how the data is stored or how the operations are implemented.

Characteristics of Database Approach

- **Support multiple views of the data:**
  - A database typically has many types of users, each of whom may require a different perspective or view of the database.
  - A view may be a subset of the database or it may contain virtual data that is derived from the database files but is not explicitly stored.
  - A multiuser DBMS whose users have a variety of distinct applications must provide facilities for defining multiple views.

Characteristics of Database Approach

- **Sharing of data and multiuser transaction processing:**
  - A multiuser DBMS allows multiple users to access the database at the same time.
  - The DBMS must include **concurrency control** software to ensure that several users trying to update the same data do so in a controlled manner so that the result of the updates is correct.
  - A fundamental role of multiuser DBMS software is to ensure that concurrent transactions operate correctly and efficiently.
  - A **transaction** is an executing program or process that includes one or more database accesses, such as reading or updating of database records. Each transaction is supposed to execute a logically correct database access if executed in its entirety without interference from other transactions. The DBMS must enforce several transaction properties such as *atomicity*, *consistency*, *isolation*, and *durability*.

Actors on the Scene

- For a small personal database, one person typically defines, constructs, and manipulates the database.
- However, in large organizations, many people are involved in the design, use, and maintenance of a large database with hundreds or thousands of users.
- Actors on the scene are the people whose jobs involve the day-to-day use of a large database. These people design, use, and administer a database.
- **Database Administrators (DBA):**
  - Database administrators oversee and manage resources such as database, DBMS, and related software.
  - Administering these resources is the responsibility of the database administrator (DBA).

Actors on the Scene

- The DBA is responsible for authorizing access to the database, coordinating and monitoring its use, and acquiring software and hardware resources as needed.
- The DBA is accountable for problems such as security breaches and poor system response time.

- **Database Designers:**
  - Database designers are responsible for identifying the data to be stored in the database and for choosing appropriate structures to represent and store this data. These tasks are mostly undertaken before the database is actually implemented and populated with data.
  - It is the responsibility of database designers to communicate with all prospective database users and groups in order to understand their requirements and to create a design and develop views that meet these requirements.

Actors on the Scene

- **End Users:**
  - End users are the people whose jobs require access to the database for querying, updating, and generating reports. There are several categories of end users as given below.
  - **Casual end users** occasionally access the database, but they may need different information each time. They use a sophisticated database query interface to specify their requests and are typically middle- or high-level managers or other occasional browsers.
  - **Naive or parametric end users** make up a sizable portion of database end users. Their main job function revolves around constantly querying and updating the database, using standard types of queries and updates – called canned transactions – that have been carefully programmed and tested. For example, bank customers and tellers, reservation agents or customers for airlines, hotels, and car rental companies etc.

Actors on the Scene

- **Sophisticated end users** include engineers, scientists, business analysts, and others who thoroughly familiarize themselves with the facilities of the DBMS in order to implement their own applications to meet their complex requirements.
- **Standalone users** maintain personal databases by using ready-made program packages that provide easy-to-use menu-based or graphics-based interfaces.

- **System Analysts and Application Programmers (Software Developers or Software Engineers):**
  - System analysts determine requirements of end users, especially naive and parametric end users, and develop specifications for standard canned transactions that meet these requirements.
  - Application programmers implement these specifications as programs; then they test, debug, document, and maintain these canned transactions.

Workers behind the Scene

- These people are associated with the design, development, and operation of the DBMS software and system environment.
- These persons are typically not interested in the database content itself. They include the following categories.
- **DBMS system designers and implementers:**
  - These people design and implement the DBMS modules and interfaces as a software package.
  - A DBMS is a very complex software system that consists of many components, or modules such as query processing, controlling concurrency etc.
  - The DBMS must interface with other system software, such as the operating system and compilers for various programming languages.

Workers behind the Scene

- **Tool developers:**
  - These people design and implement tools – the software packages that facilitate database modeling and design, database system design, and improved performance.
  - Tools are optional packages that are often purchased separately. They include packages for database design, performance monitoring, natural language or graphical interfaces, prototyping, simulation, and test data generation.
  - In many cases, independent software vendors develop and market these tools.
- **Operators and maintenance personnel (system administration personnel):**
  - These people are responsible for the actual running and maintenance of the hardware and software environment for the database system.

Advantages of Using DBMS Approach

- Some of the advantages of using a DBMS approach are explained below.
- **Controlling redundancy:**
  - The database is said to be redundant if the same information is duplicated in several places. For example, the address and telephone number of a particular customer may appear in a file that consists of saving-account records and in a file that consists of checking-account records.
  - Redundancy leads to serval problems. First, it causes duplicate effort during database update. Second it wastes storage space. Third it may cause data inconsistency.
  - Using **data normalization**, we can reduce data redundancy.

Advantages of Using DBMS Approach

- **Restricting Unauthorized Access:**
  - Most users will not be authorized to access all information in the database.
  - Some users may only be permitted to retrieve data, whereas others are allowed to retrieve and update.
  - A DBMS should provide a **security and authorization subsystem**, which the DBA uses to create accounts and to specify account restrictions.
  - We can also apply similar controls to the DBMS software. For example, only the DBA's staff may be allowed to use certain privileged software, such as the software for creating new accounts.

Advantages of Using DBMS Approach

- **Providing Persistent Storage for Program Objects:**
  - Databases can be used to provide persistent storage for program objects and data structures.
  - Programming languages typically have complex data structures, such as structs or class definitions.
  - Object-oriented database systems are compatible with programming languages such as C++ and Java, and the DBMS software provide persistent storage for objects and data structures.
  - Object-oriented database systems typically offer data structure compatibility with one or more object-oriented programming languages.

Advantages of Using DBMS Approach

- **Providing Storage Structures and Search Techniques for Efficient Query Processing:**
  - Database systems must provide capabilities for efficiently executing queries and updates.
  - The DBMS must provide specialized data structures and search techniques to speed up disk search for the desired records. Auxiliary files called indexes are often used for this purpose.
  - The query processing and optimization module of the DBMS is responsible for choosing an efficient query execution plan for each query based on the existing storage structures.
- **Providing Backup and Recovery:**
  - A DBMS must provide facilities for recovering from hardware or software failures.
  - The backup and recovery subsystem of the DBMS is responsible for recovery.

Advantages of Using DBMS Approach

- **Providing Multiuser Interfaces:**
  - Because many types of users use a database, a DBMS should provide a variety of user interfaces.
  - These include apps for mobile users, query languages for casual users, programming language interfaces for application programmers, forms and command codes for parametric users, and menu-driven interfaces and natural language interfaces for standalone users.
- **Representing Complex Relationships among Data:**
  - A database may include numerous varieties of data that are interrelated in many ways.
  - A DBMS must have the capability to represent a variety of complex relationships among the data, to define new relationships as they arise, and to retrieve and update related data easily and efficiently.

Advantages of Using DBMS Approach

- **Enforcing Integrity Constraints:**
  - A DBMS should provide capabilities for defining and enforcing these constraints that must hold for the data.
  - The simplest type of integrity constraint involves specifying a data type for each data item.
  - A more complex type of constraint that frequently occurs involves specifying that a record in one file must be related to records in other files. This is known as a referential integrity constraint.
  - Another type of constraint specifies uniqueness on data item values. This is known as a key or uniqueness constraint.
  - It is the responsibility of the database designers to identify integrity constraints during database design.
  - Some constraints can be specified to the DBMS and automatically enforced. Other constraints may have to be checked by update programs or at the time of data entry.

Advantages of Using DBMS Approach

- **Permitting Inferencing and Actions Using Rules and Triggers:**
  - Some database systems provide capabilities for defining deduction rules for inferencing new information from the stored database facts. Such systems are called deductive database systems.
  - In relational database systems, it is possible to associate triggers with tables. A trigger is a form of a rule activated by updates to the table, which results in performing some additional operations to some other tables, sending messages, and so on.
  - More powerful functionality is provided by active database systems, which provide active rules that can automatically initiate actions when certain events and conditions occur.

Advantages of Using DBMS Approach

- **Additional Implications of Using Database Approach:**
  - Few additional implications of using the database approach that can benefit most organizations are given below:
  - **Potential for Enforcing Standards:** Database approach permits the DBA to define and enforce standards among database users. Standards can be defined for names and formats of data elements, display formats, report structures, terminology, and so on.
  - **Reduced Application Development Time:** Developing a new application – such as the retrieval of certain data from the database for printing a new report – takes very little time. Once a database is up and running, substantially less time is generally required to create new applications using DBMS facilities.
  - **Flexibility:** DBMSs allow certain types of evolutionary changes to the structure of the database without affecting the stored data and the existing application programs.

Advantages of Using DBMS Approach

- **Availability of Up-to-Date Information:** As soon as one user's update is applied to the database, all other users can immediately see this update.
- **Economies of Scale:** The DBMS approach permits consolidation of data and applications, thus reducing the amount of wasteful overlap and redundancies. This enables the whole organization to invest in more powerful processors, storage devices, or networking gear, rather than having each department purchase its own (lower performance) equipment. This reduces overall costs of operation and management.