Chapter 1
Introduction to Database Management System

1.1 Data :

In simple words data can be <mark>facts related to any object</mark> in consideration. For example our name, age, height, weight, etc are some data related to us. A picture , image , file , pdf etc can also be considered data.

Data is <mark>rarely useful in its raw form</mark>. For example, in a banking application, data is the whole collection of bank account numbers; bank customers' names, addresses, and ages; bank transactions and so on. However, when data is arranged relationally, it then becomes information, which is much more useful to users.

Data are logically organized into:

1. Bits

    A bit is the smallest unit of data representation (value of a bit may be a 0 or 1)

2. Fields

    A data field represents an attribute (a characteristic or quality) of some entity (object, person, place, or event)

3. Records

    A record consists of fields, with each field describing an attribute of the entity.

4. Files

    A group of related records

5. Databases

    Collection of logically related records or files.

1.2 Database :

A database is a <mark>collection of data that is saved and organized to allow easy retrieval when needed</mark>. It is the <mark>collection of schemas, tables, queries, reports, views, and other objects</mark>. In order to maintain and access the database we need a DBMS (database management system).

Databases consists of tables that include groups of related data fields that are known as records. <mark>Databases are not limited to only computers</mark>; in fact, a phone book is an example of a database.

Some of the <mark>examples of database applications are telephone directory system, Computerized Library, Inventory System etc</mark>.

<mark>Early databases were relatively "flat"</mark> which means they were limited to simple rows and columns, like a spreadsheet.

1.2.1 Types of Database:

1. **Relational database** :-

    A tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. It <mark>stores data in tables with relationships to other tables.</mark>

2. **Distributed database** :-

    A distributed database is one that can be dispersed or <mark>replicated among different point</mark> in a network.

3. **Object oriented programming database** :-

In this database, the ==data is defined in object classes== and sub classes.

1.3 Database Management System:

A database-management system (DBMS) is a ==computer-software application that interacts with end-users, other applications, and the database itself to capture and analyze data==.

A database management system (DBMS) is ==system software for creating and managing databases==. The DBMS provides users and programmers with a systematic way to ==create, retrieve, update and manage data==.

==Examples of DBMS are MS-Sql Server, MySQL, MS-Access, dBase, Oracle etc.==
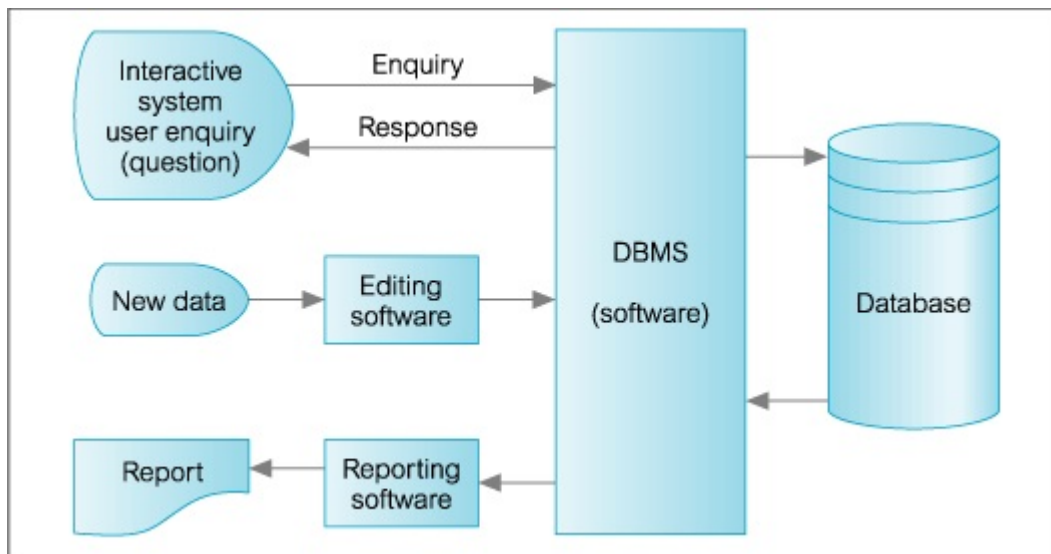
1.3.1 Application of DBMS

✔   Banking:

   For customer and their account information.

✔   Airlines:

   For reservation and schedules information.

✔   Telecommunications:

   Maintaining the bills, calls and other useful information.

✔   Government:

   For Taxes, Budgets, and Census.

✔   Sales:

   For inventory, Customers information, purchase and sales information.

✔   Universities:

   For maintaining student records, course registration and grades.

1.3.2 Advantages of DBMS

1. **Minimal Data Redundancy**:

   Data Redundancy means that the same data field appears in different files and format i.e. <mark>Duplication of data</mark>. So the waste of storage is reduced.

2. **Sharing of Data**:

   The data sharing feature of DBMS allows the <mark>data to be shared among the existing applications and newly developed application</mark> without having to create any additional stored data.

3. **Increased Security:**

   The DBMS <mark>creates a security system that enforces user security and data privacy</mark> within the database.

4. **Better Service to Users**:

   The DBMS allow users who don't know any programming language to <mark>interact with the data more easily.</mark>

5. **Features of Backup and Recovery:**

   Backup is a task of creating a copy of original data so that the <mark>data can be recovered in case of data loss.</mark>

1.3.3 Disadvantages of DBMS

1. **Cost of Hardware and Software:**

   A processor with <mark>high speed of data processing and memory of large size is required</mark> to run the DBMS software .

   DBMS <mark>software is also very costly.</mark>

2. **Database Failures:**

   In most of the organizations, <mark>all data is Centralized</mark>. If database is corrupted due to power failure or it is corrupted on the storage media, then our valuable data may be lost or whole system stops.

3. **Complexity of Backup and Recovery:**

   Since data is centralized, <mark>adequate backup of the data is necessary</mark> so that in case of failure data can be recovered. So backup problem is also the drawback.

4. **Appointing Technical Staff:**

   The <mark>trained technical persons </mark>such as database administrator and application programmers etc are required to handle the DBMS. We have to pay high salaries to these persons. Therefore, the system cost increases.

5. **Larger File Size:**

   The <mark>DBMS itself is large program</mark> so additional memory is required.

1.3.4 Components of DBMS

A DBMS is an intermediate link between the physical database, computer and operating system and on the other hand users.

There are 5 components in database system environment and they are

1. **Hardware**:

The hardware is the actual computer system for keeping and accessing the database. The DBMS hardware consists of secondary storage device like hard disk on which the database physically resides, input output devices etc.

2. **Software**:

The software is the actual DBMS. A DBMS is software that provides interface between the user and physical database for accessing the data. All requests from users for access to the database are handled by DMBS.

3. **Data:**

Data stored in database includes numerical data, non numerical data such as characters, logical values and complex data such as images. In addition to these operational data, the database contains meta-data, the data of data.
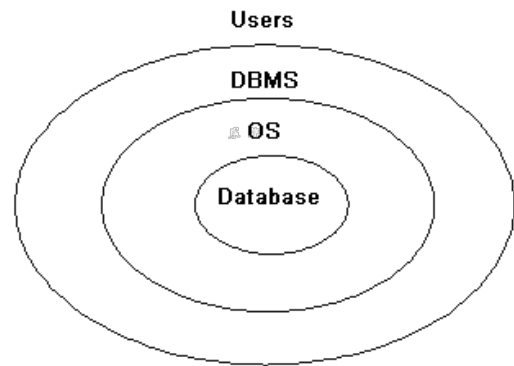
4. **Users:**

There are a number of users who can access or retrieve data on demand using the application and interfaces provided by DBMS.

5. **Procedures:**

Procedure refers to the rules that govern the design and use of the database. There must be a documented rule for the database users on how to use or run the system. Some of them may be

    a.   Logon to DBMS

    b.   Use Particular DBMS facility or application program

    c.   Make backup copies of Database

    d.   Handle hardware and software failure

1.4 File Environment

The traditional filing system is a method of storing and arranging computer files and the information in the file.

For this method, a file name is assigned to a file in order to secure storage location in the computer memory or data storage devices such as hard disks or CD-ROMs are used to maintain the physical location of the files. By this file name a file can be further accessed. Whether the file system has an underlying storage device or not, file systems typically have directories which associate file names with files, usually by connecting the file name to an index in a file allocation table.

A new concept for file management is the concept of a database-based file system

1.4.1 Advantages of traditional file based system

✗ No need of external storage

✗ No need of a highly technical person to handle the database.

1.4.2 Disadvantages of traditional file based system
 ✗ Provides less security.
 ✗ Redundancy is more.

1.5 Schema and Instance

Database changes over a time as information is inserted or deleted. The collection of information stored in database at particular moments is called the instance of database. The overall design of database is called the database schema. So schema is the logical structure of database.
There are three types of schema.

a) **Physical schema:**
This describes how the relations described in the conceptual schema are actually stored on secondary storage device such as disk.

b) **Conceptual/logical schema:**
This schema hides the details of physical storage structures and concentrates on describing the entities data types, relationship and constraints.

c) **External schema:**
This schema which is at view level allows data access to be customized at the level of individual users   or groups of users. i.e. describes the different views of database.

1.6 Data Abstraction in DBMS

Database systems are made-up of complex data structures. To ease the user interaction with database, the developers hide internal irrelevant details from users. This process of hiding irrelevant details from user is called data abstraction.

We have three levels of abstraction:

1. **Physical level**: This is the lowest level of data abstraction. It describes how data is actually stored in database. You can get the complex data structure details at this level.

2. **Logical level**: This is the middle level of 3-level data abstraction architecture. It describes what data is stored in database.

3. **View level**: Highest level of data abstraction. This level describes the user interaction with database system.
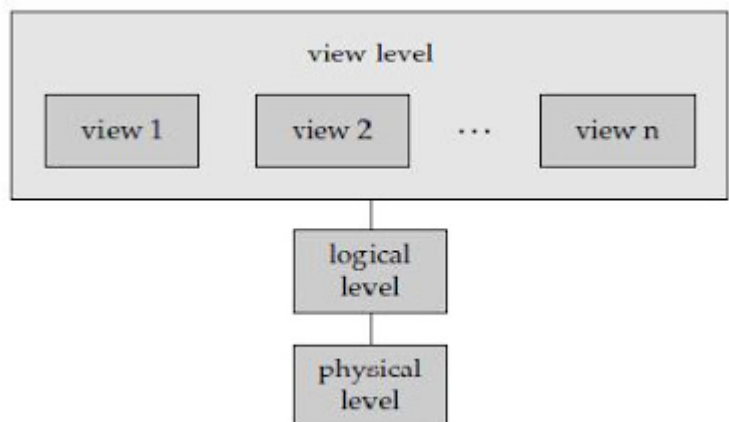
Example:

Let's say we are storing customer information in a customer table.

At **physical level** these records can be described as blocks of storage (bytes, gigabytes etc.) in memory. These details are often hidden from the programmers.

At the **logical level** these records can be described as fields and attributes along with their data types, their relationship among each other can be logically implemented.

At **view level**, user just interact with system with the help of GUI and enter the details at the screen, they are not aware of how the data is stored and what data is stored; such details are hidden from them.

1.7 Data Independence in DBMS

A database system normally contains a lot of data in addition to users' data. For example, it stores data about data, known as metadata, to locate and retrieve data easily. It is rather difficult to modify or update a set of metadata once it is stored in the database. But as a DBMS expands, it needs to change over time to satisfy the requirements of the users. If the entire data is dependent, it would become a tedious and highly complex job.

1.7.1 Types of Data Independence:
There are two types of data independence:

a. Physical Data Independence

Physical data independence is the power to change the physical data without impacting the schema or logical data.

For example, in case we want to change or upgrade the storage system itself – suppose we want to replace hard-disks with SSD – it should not have any impact on the logical data or schemas.
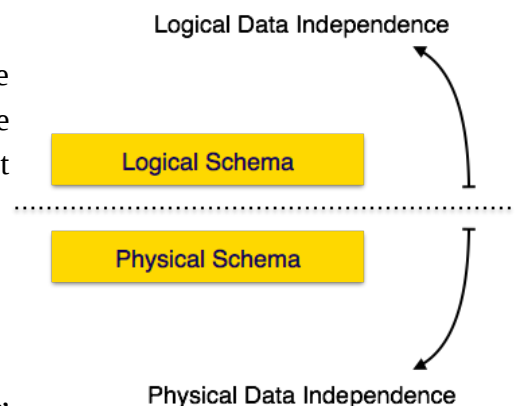
b. Logical Data Independence

Logical data independence is a kind of mechanism, which changes itself from actual data stored on the disk. If we do some changes on table format, it should not change the data residing on the disk.

1.8 Concept of DDL, DML, DCL

1.8.1 Data Defination Language (DDL)

DDL or Data Definition Language actually consists of the SQL commands that can be used to define the database schema. It simply deals with descriptions of the database schema and is used to create and modify the structure of database objects in database.

Examples of DDL commands:

- CREATE

is used to create the database or its objects (like table, index, function, views, store procedure and triggers).

- DROP

  is used to delete objects from the database.

- ALTER

  is used to alter the structure of the database.

- TRUNCATE

  is used to remove all records from a table, including all spaces allocated for the records are removed.

- COMMENT

  is used to add comments to the data dictionary.

- RENAME

  is used to rename an object existing in the database.

1.8.2 Data Manipulating Language (DML)

The SQL commands that deals with the manipulation of data present in database belong to DML or Data Manipulation Language and this includes most of the SQL statements.

Examples of DML:

- **SELECT** – is used to retrieve data from the a database.
- **INSERT** – is used to insert data into a table.
- **UPDATE** – is used to update existing data within a table.
- **DELETE** – is used to delete records from a database table.

1.8.3 Data Manipulating Language (DML)

DCL includes commands such as COMMIT, GRANT and REVOKE which mainly deals with the rights, permissions and other controls of the database system.

Examples of DCL commands:

- **GRANT**-gives user's access privileges to database.
- **REVOKE**-withdraw user's access privileges given by using the GRANT command.
- **COMMIT:** Save work done.
- **ROLLBACK:** Restore database to original since the last commit

1.9 Database Users

Users are differentiated by the way they expect to interact with the system. Altogether there are five types of database users.

1. **Application Programmer:**

They are ==computer professionals who write application program.== They can choose many tools to develop user interface. Examples in School Management system include program that displays student information, their obtained marks etc.

2. **Sophisticated Users**:

They ==interact with the system without writing program. They form their request in a database query language.==

3. Specialized Users:

They are ==sophisticated users who write specialized database application==. Expert System, CAD, System that stores data with complex data type (audio, video , graphics) are some of the examples of such application.

4. Native Users:

They ==invoke one of the application programs that have been written previously==. E.g.: bank teller

5. **DA**

Data Administrator is the ==person who has the central responsibility for an organizations data.==

The responsibilities include:

- ✗ ==Establishing the policies and specific procedures== for collecting, validating, sharing, and inventorying data to be stored in databases and for making information accessible to the members of the organization and, possibly, to persons outside of it.

- ✗ Data administration is a policy making function and the DA should have access to senior corporate management.

6. **DBA**

Database Administrator is a ==specialist responsible for maintaining standards== for the development, maintenance, and security of an organization's databases.

Responsibilities include:

- ✗ **Schema Definition**:

  The ==DBA creates the original data structure using DDL== which under compilation results to a set of tables that are permanently stored in data dictionary.

- ✗ **Granting of authorization of data access**:

  The ==DBA is responsible to regulate which part of database various users are authorized to access.==

- ✗ **Installation of new software:**

  It is primarily job of the DBA to ==install new versions of DBMS software, application software and other software related to DBMS administration==. It is important that the DBA test the new software before it is moved into productive environment.

- ✗ **Security Administration:**

One of the main responsibilities of DBA is to <mark>monitor and administer DBMS security.</mark> This involves adding and removing users, auditing and checking security problems.

✗ **Data analysis:**

The DBA will frequently be called on to <mark>analyze the data stored in the database and to make recommendations relation to performance and efficiency of the data storage.</mark>

1.9 Data Dictionary

A *data dictionary* is a <mark>software module and database containing descriptions and definitions concerning the structure, data elements, interrelationships, and other characteristics of an organization's database.</mark>

1.10 Information Management System

Information Management System (IMS) is a <mark>general term for software designed to facilitate the storage, organization and retrieval of information.</mark> IMS gives the business manager the information that they need to make decision. IMS provide a variety of information to managers. Periodic Scheduled Report, Exception report, Demand report and response report etc are the part of IMS in business enterprise.

Assignment 1

Q. List the major dates in the history of digital database.

Q. What are the difference between Traditional File Based System and Modern System.

Q. All the Pokhara University Board Exam Questions From this Chapter.