CHAPTER ONE

Introduction to Database Systems

Objectives of the Chapter

- ✓Describe database management system (DBMS)
- ✓Differentiate the File based verses Database approach
- ✓List out the Characteristics of the Database Approach
- ✓Define users and actors on the database scene

Introduction to Database System

- Database and database technology have a major impact on the growing use of computers.
- It is fair to say that databases play a critical role in almost all areas where computers are used, including business, e-commerce, engineering, medicine, law, education, library services more…
- To find out what database is, we have to start from data, which is the basic building block of any DBMS
- **Data** can be defined as a representation of raw facts, concepts, or instructions in a formalized manner, which should be suitable for communication, interpretation, or processing by human or electronic machine.
- Data is represented with the help of characters such as alphabets (A-Z, a-z), digits (0-9) or special characters (+,-,/,*,<,>,= etc.)

What is Data Processing?

- Data processing is the method of collecting raw data and translating it into usable information.
- It is usually performed in a step-by-step process by a team of data scientists and data engineers in an organization.
- The raw data is collected, filtered, sorted, processed, analyzed, stored, and then presented in a readable format.
- The data processing cycle consists of a series of steps where raw data (input) is fed into a process (CPU) to produce actionable insights (output).
- Each step is taken in a specific order, but the entire process is repeated in a cyclic manner.
- The first data processing cycle's output can be stored and fed as the input for the next cycle.
- Generally, there are six main steps in the data processing cycle, they are Collection, Preparation, Input, Data processing, Output and Storage.

Database System

- **Database** is a collection of related data organized in a way that data can be easily accessed, managed and updated.
- **Database** is a collection of inter-related data which helps in efficient retrieval, insertion and deletion of data from database and organizes the data in the form of tables, views, schemas, reports etc.
- For eg: Consider the names, telephone numbers, and address of the people you may know.
- Another Example also you can consider, university database organizes the data about students, faculty, and admin staff etc. which helps in efficient retrieval, insertion and deletion of data from it.
- You may have recorded these data in an indexed book or stored in a hard disk of a computer with help of some software's like MS-Excel or MS-Access.
- This collection of related data with an implicit meaning is a database.

Cont...

Generally, a database has the following properties:-

- A database represents some aspects of the real world, sometimes called the **mini world.**
  - Changes to the mini world are reflected in the database
- A database is a logically coherent collection of data with some inherent meaning.
  - A random mixture/collection of data cannot be correctly referred to as a database.
- A database is designed, built, and populated with data for a specific purpose.
  - It has an intended group of users and some fixed applications in which these users are interested.

Data Management Approaches => Development Levels

- There are three main data management Approaches (Levels);

1. Manual approach
2. Traditional file based approach and
3. Database approach

Manual Approach

- It refers to data processing that requires humans to manage and process the data throughout its existence. The data storage and retrieval is being done using **human labor**.
- **Manual data processing** utilizes **non-technological tools**, which include paper, writing utensils and physical filing cabinets. Data storage and retrieval follow the **primitive** and **traditional** way of information handling where cards and paper are used for the purpose.
- Many Files for as many event and objects as the organization has, since files are used to store information. Each of the files containing various kinds of information is labelled and stored in one ore more cabinets.
- The cabinets could be kept in safe places for security purpose based on the sensitivity of the information contained in it. **For Example**:- Cards and papers, human labor, Library catalog cards, hospitals/clinics cards, Kebele files

Manual Approach

Limitations of the Manual Approach

- Prone/liable to error
- Difficult to update, retrieve, integrate
- You have the data but it is difficult to compile the information
- Limited to small size information
- Cross referencing is difficult

❖ An alternative approach of data handling is a computerized way of dealing with the information. The computerized approach could also be either decentralized or centralized base on where the data resides in the system.

File Based Approach

- The database is stored as a collection of **files.**
- A file is organized logically as a sequence of **records.** A record is a sequence of **fields.**
- A file organization refers to the organization of the data of a file into records, blocks, and access structures;
- is basically a way of arranging the files in a storage medium like a hard disk.
- The file system organizes the files and helps in the retrieval of files when they are required. File systems consist of different files which are grouped into directories.
- The directories further contain sub directories(other folders) and files.
- The file system performs basic operations like management, file naming, giving access rules, etc
- A database can be of any size and complexity. For eg: the list of names and addresses of the students in a class may consist of only a few hundreds of records each with a simple structure.

Cont…

- Traditionally, data accessed through computers has been stored on different storage media in the form of individual files. File based systems were an early attempt to computerize the manual filing system. Files proved to be quite satisfactory as long as computerization was limited to few application areas.
- However, the **number of users became increased,** especially with the introduction of internet and online transaction systems, the file systems gave rise to **many serious problems**.
- A collection of application programs perform services for the end-users. In such systems, every application program that provides service to end users define and manage its own data .
- Such systems have number of programs for each of the different applications in the organization. Since every application defines and manages its own data, the system is subjected to serious **data duplication problem**. File, in traditional file based approach, is a collection of records which contains logically related data.

Problems with File Systems

A large number of files gave rise to the following problems.

- ✓ Files involve a high level of redundancy in data; i.e., same data item being stored at many d/t places.
- ✓ Individual files are not adaptable to rapid changes, especially with respect to the way the data items are structured within the file.
- ✓ Data stored in a file system is purely dependent of the physical medium used.
- ✓ In file system, only pre-defined questions can be answered.
- ✓ Modifications in one program may require modifications in other programs, which interface with this program.

Problems with File Systems

A large number of files gave rise to the following problems.

- ✓ **Data inconsistency:** If the data is kept in different files, there could be problems when an item of data needs updating, as it will need to be updated in all the relevant files; if this is not done, the data will be inconsistent, and this could lead to errors.
- ✓ **Difficult to implement data security:** Data is stored in different files by different application programs. This makes it difficult and expensive to implement organization-wide security procedures on the data.

Database Approach

- Database is a shared collection of logically related data designed to meet the information needs of an organization. Since it is a shared corporate resource, the database is integrated with minimum amount of or no duplication. It is a collection of logically related data where these logically related data comprises entities, attributes, relationships, and business rules of an organization's information.
- The purpose of a database is to store information and to allow users to retrieve and update that information on demand and it designed once and used simultaneously by many users.
- DBMS is a collection of programs that enables users to create and maintain a database.
- The DBMS is a **general purpose software system** that facilitates the process of defining, constructing, manipulating and sharing database among various users and applications.
- Defining a database involves specifying the data type, structures and constraints/restrictions of the data to be stored in the database.

Data Base Management System (DBMS)..

- Constructing the database is the process of storing the data on some storage medium that is controlled by the DBMS.
- **Manipulating a database** includes functions such as **querying** the database to **retrieve specific data, updating** the database to **reflect changes** in the mini world, and **generating reports** from the data.
- Other important functions provided by the DBMS include **protecting** the database and **maintaining** it over a long period of time.
- Here are some examples of popular DBMS used these days: MySql, MS SQL Server, PostgreSQL, IBM DB2, oracle, Amazon SimpleDB (cloud based) etc.
- DBMS provides proper **security measures** for protecting the data from unauthorized access and also provides mechanisms for data recovery and data backup.
- In Database Management System the data can be **fetched by SQL queries** and **relational algebra**.

Difference between File System and DBMS

Basis	File System	DBMS
Structure	The file system is software that manages and organizes the files in a storage medium within a computer.	DBMS is software for managing the database.
Data Redundancy	Redundant data can be present in a file system.	In DBMS there is no redundant data.
Backup and Recovery	It doesn't provide backup and recovery of data if it is lost.	It provides backup and recovery of data even if it is lost.
Query processing	There is no efficient query processing in the file system.	Efficient query processing is there in DBMS.
Consistency	There is less data consistency in the file system.	There is more data consistency because of the process of normalization.
Complexity	It is less complex as compared to DBMS.	It has more complexity in handling as compared to the file system.
Security Constraints	File systems provide less security in comparison to DBMS.	DBMS has more security mechanisms as compared to file systems.
Cost	It is less expensive than DBMS.	It has a comparatively higher cost than a file system.
Data Independence	There is no data independence.	In DBMS data independence exists.
User Access	Only one user can access data at a time.	Multiple users can access data at a time.

Why Do We Need a Database?

The various reasons for which we require databases are:

- ✓ **Size of Data**: The small amount of data storing into spreadsheet is fine, however it might turns into a large amount of data then Spreadsheet solution will not work.
- ✓ **Ease of Updating Data**: Multiple peoples cannot access the same file on same time.
- ✓ **To manage large chunks of data:** For instance: if your size of data increases into thousands of records, it will simply create a problem of speed.
- ✓ **Security of data:** Anyone can easily get access to file and can make changes to it. With databases you have security groups and privileges, you set to restrict access.
- ✓ **Data integrity:** In databases, you can be assured of accuracy and consistency of data due to the built in integrity checks and access controls.
- ✓ **Accuracy:** refers to error-free records that can be used as a reliable source of information.

Why Do We Need a Database?

The various reasons for which we require databases are:

- ✓ **Less redundancy** – DBMS follows the rules of normalization. Normalization is a mathematically rich and scientific process that reduces data redundancy.
- ✓ **Consistency** – is a state where every relation in a database remains consistent. There exist methods and techniques, which can detect attempt of leaving database in inconsistent state.
- ✓ **Query Language** – DBMS is equipped with query language, which makes it more efficient to retrieve and manipulate data. A user can apply as many and as different filtering options as required to retrieve a set of data.

Why Do We Need a Database?

The various reasons for which we require databases are:

- ✓ **ACID Properties** – DBMS follows the concepts of **A**tomicity, **C**onsistency, **I**solation, and **D**urability (ACID). These concepts are applied on transactions, and help the database to stay healthy in multi-transactional environments and in case of failure.
- ✓ **Multiuser and Concurrent Access** – DBMS supports multi-user environment and allows them to access and manipulate data in parallel. Though there are restrictions on transactions when users attempt to handle the same data item, but users are always unaware of them.
- ✓ **Multiple views** – DBMS offers multiple views for different users. A user who is in the Sales department will have a different view of database than a person working in the Production department (Hierarchy).

Characteristics of the Database Approach

- In the database approach, a single repository of data is maintained. That is, defined once and then accessed by various users.
- In file systems, each application is free to name data elements independently.
- In contrast, in a database, the names and labels of data are defined once and used repeatedly by queries, transactions, and applications.

The main characteristics of the database approach against the file processing approach are the following:

- ✓ Self- describing nature
- ✓ Separation between Programs and Data, and Data Abstraction
- ✓ Support of Multiple Views of the Data
- ✓ Sharing of Data and Multiuser Transaction Processing

Self- describing nature

- A fundamental characteristic of the database approach is that, the database system 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 structure of each file, 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 primary database.
- A general-purpose DBMS software package is not written for a specific database application.

Self- describing nature

- Therefore, it must refer to the catalog to know the structure of the files in a specific database, such as type and format of data it will access.
- The DBMS software must work with any number of *database applications*- for example, a university database, a banking database, or a company database- as long as the database definition is stored in the catalog.
- In traditional file processing, data definition is typically part of the application program themselves. Hence, these programs are constrained to work with only one *specific database*, whose structure is declared in application programs.

Separation between Programs and Data, and Data Abstraction

- In traditional file system, the structure of data files is 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, DBMS access programs do not require such changes in most cases.
- The structure of data files is stored in the DBMS catalog separately from the access programs. This property is known as **program-data independence.**
- DBMS provides users with a **conceptual representation** of data that does not include many of the details of how the data is stored or how the operations are implemented.
- This characteristic is called as **data abstraction.**
- Many other details of file storage organization such as the access path specified on a file can be hidden from database users by the DBMS.

Support of Multiple Views of the Data

- A database typically has many users, each of whom may require different perspective or view of the database.
- A multiuser DBMS, as its name implies, must allow multiple users to access the database at the same time.
- The users will have variety of distinct applications which defines multiple views

Sharing of Data and Multiuser Transaction Processing

- The DBMS must include **concurrency control** software to ensure a controlled, correct updation since several users may try to update the same data in a multiuser DBMS.
- For example, when several reservation clerks try to assign a seat on an airline flight, the DBMS should ensure that each seat can be accessed by only one clerk at a time for assignment to a passenger.
- These types of applications are generally called **online transaction processing (OLTP)** applications.
- 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.

DATABASE DEVELOPMENT LIFE CYCLE

- The major steps in database design are:

- **Planning** is identifying information gap in an organization and propose a database solution to solve the problem.

- **Analysis** is concentrates more on fact finding about the problem or the opportunity. It identify Feasibility analysis, requirement determination and selection of best design

- **Design**
  - ✓ **Conceptual Design**: concise description of the data, data type, relationship between data and constraints on the data
  - ✓ **Logical Design**: a higher level conceptual abstraction with selected specific data model to implement the data structure.
  - ✓ **Physical Design**: physical implementation of the upper level design of the database

DATABASE DEVELOPMENT LIFE CYCLE

- ✓**Implementation**: the testing and deployment of the designed database for use.
- ✓**Operation and Support**: administering and maintaining the operation of the database system and providing support to users.

Actors on the Scene

- In large organizations, many people are involved in the design, use, and maintenance of a large database with hundreds of users.
- The people whose job involve the day-to-day use of a large database: called as **actors on the scene,** and are classified as follows:-
  - ✓ Database Administrator (DBA)
  - ✓ Database designers
  - ✓ End user
  - ✓ System Analysts and Application Programmers (software engineers)

Database Administrator (DBA)

- The DBA is responsible for authorizing access to the database, coordinating and monitoring its use and acquiring software and hardware resources as needed.
- In a database environment, the primary resource is the database itself, and the secondary resource is the responsibility of the DBA.
- Administrators maintain the DBMS and are responsible for administrating the database.
- They are responsible to look after its usage and by whom it should be used.
- They create access profiles for users and apply limitations to maintain isolation and force security.
- Administrators also look after DBMS resources like system license, required tools, and other software and hardware related maintenance.

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.
- Responsible to define the content, the structure, the constraints, and functions or transactions against the database. They must communicate with the end-users and understand their needs.
- They keep a close watch on what data should be kept and in what format.
- They identify and design the whole set of entities, relations, constraints, and views.
- Identifies the data to be stored and choose the appropriate structures to represent and store the data.
- Database Designers are the users who design the structure of data base which includes tables, indexes, views, constraints, triggers, stored procedures.
- He/she controls what data must be stored and how the data items to be related.

System Analysts

- **System analysts** determine the requirements of end users, especially naive users, and develop specifications for canned transactions that meet these requirements.
- It determines the user requirement and how the user wants to view the database.
- The application programmer implements these specifications as programs; code, test, debug, document and maintain the application program.
- Determines the interface on how to retrieve, insert, update and delete data in the database.
- The application could use any high level programming language according to the availability, the facility and the required service.
- Application programmers implement these specifications as programs; then they test, debug, document, and maintain these canned transactions.

End Users

- **End Users:** Workers, whose job requires accessing the database frequently for various purpose. There are different group of users in this category.
- End users are the people whose jobs require access to the database for querying, updating, and generating reports; the database primarily exists for their use. There are several categories of end users:
- They use the data for queries, reports and some of them update the database content.

Cont….

There are several categories of end users:

- **Casual end user:** they access the database occasionally, but they may need different information each time.
- **Naive or parametric end user:** 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.
- **Sophisticated end user**: include engineers, scientists, business analysts who try to learn most of the DBMS facilities in order to achieve their complex requirements.
- **Standalone users**: maintain personal database by using ready-made program packages that provide easy-to use-menu-based or graphical-based interfaces.

Database Users – Actors Behind The Scene

- **System Designers and Implementers**: Design and implement DBMS packages in the form of modules and interfaces and test and debug them.
- The DBMS must interface with applications, language compilers, operating system components, etc.
- **Tool Developers**: Design and implement software systems called tools for modeling and designing databases, performance monitoring, prototyping, test data generation, user interface creation, simulation etc. that facilitate building of applications and allow using database effectively.
- **Operators and Maintenance Personnel**: They manage the actual running and maintenance of the database system hardware and software environment.

When not to Use a DBMS

- In spite of the advantages of using a DBMS, there are a few situations in which such a system may involve unnecessary overhead costs, as that would not be incurred in traditional file processing.

The overhead costs of using a DBMS are due to the following:

- ✓ High initial investment in hardware, software, and training.
- ✓ Overhead for providing security, concurrency control, recovery, and integrity functions.
- ✓ Complexity increases,
- ✓ Requirement of more disk space
- ✓ Cost of conversion
- ✓ Need of additional and specialized manpower
- ✓ Need for backup and recovery
- ✓ More installation and management cost

Thank you!!