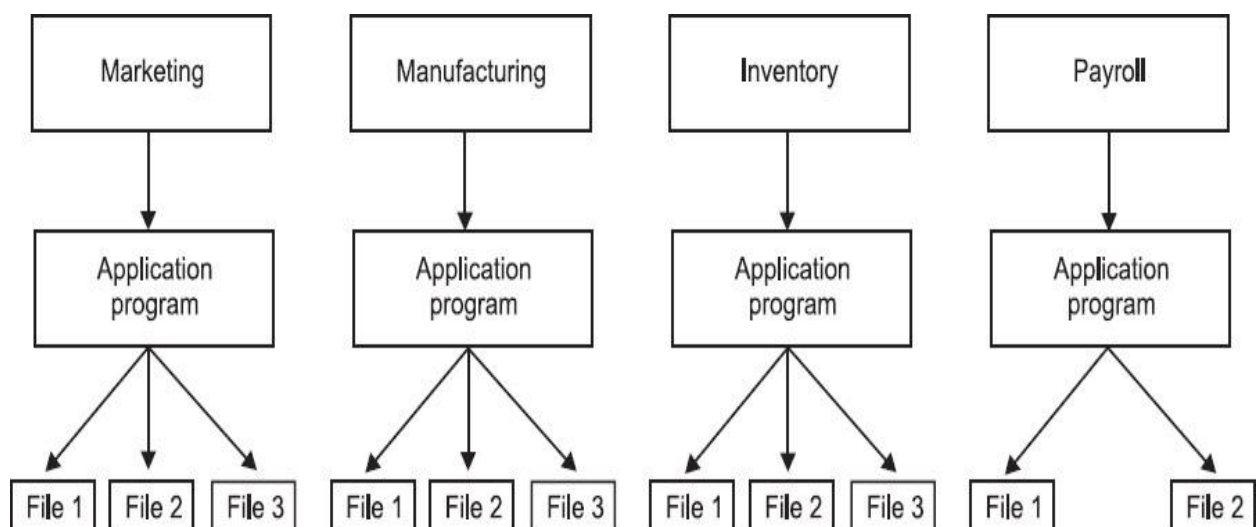


COMPUTER AND COMMUNICATION TECHNOLOGY

INTRODUCTION TO COMPUTER FILES AND DATABASES

FILE SYSTEM VERSUS DATABASE SYSTEMS

- Conventionally, the data were stored and processed using traditional file processing systems.
 - In these traditional file systems, each file is independent of other file, and data in different files can be integrated only by writing individual program for each application.
 - The data and the application programs that use the data are so arranged that any change to the data requires modifying all the programs that uses the data.
 - This is because each file is hard-coded with specific information like data type, data size etc.
 - Some time it is even not possible to identify all the programs using that data and is identified on a trial-and-error basis.
-
- A file processing system of an organization is shown in figure below.
 - All functional areas in the organization creates, processes and disseminates its own files.
 - The files such as inventory and payroll generate separate files and do not communicate with each other.



- No doubt such an organization was simple to operate and had better local control but the data of the organization is dispersed throughout the functional sub-systems.
- These days, databases are preferred because of many disadvantages of traditional file systems.

Disadvantages of Traditional File System

- A traditional file system has the following disadvantages.

1) Data Redundancy:

- Since each application has its own data file, the same data may have to be recorded and stored in many files.
- For example, personal file and payroll file, both contain data on employee name, designation etc.
- The result is unnecessary duplicate or redundant data items.
- This redundancy requires additional or higher storage space, costs extra time and money, and requires additional efforts to keep all files up to-date.

2) Data Inconsistency:

- Data redundancy leads to data inconsistency especially when data is to be updated.
- Data inconsistency occurs due to the same data items that appear in more than one file do not get updated simultaneously in each and every file.
- For example, an employee is promoted from Clerk to Superintendent and the same is immediately updated in the payroll file may not necessarily be updated in provident fund file.
- This results in two different designations of an employee at the same time.
- Over the period of time, such discrepancies degrade the quality of information contain in the data file that affects the accuracy of reports.

3) Lack of Data Integration:

- Since independent data file exists, users face difficulty in getting information on any ad hoc query that requires accessing the data stored in many files.
- In such a case complicated programs have to be developed to retrieve data from every file or the users have to manually collect the required information.

4) Program Dependence:

- The reports produced by the file processing system are program dependent, which means if any change in the format or structure of data and records in the file is to be made, the programs have to be modified correspondingly.
- Also, a new program will have to be developed to produce a new report.

5) Data Dependence:

- The Applications/programs in file processing system are data dependent i.e., the file organization, its physical location and retrieval from the storage media are dictated by the requirements of the particular application.
- For example, in payroll application, the file may be organised on employee records sorted on their last name, which implies that accessing of any employee's record has to be through the last name only.

6) Limited Data Sharing:

- There is limited data sharing possibilities with the traditional file system.
- Each application has its own private files and users have little choice to share the data outside their own applications.
- Complex programs required to be written to obtain data from several incompatible files.

7) Poor Data Control:

- There was no centralised control at the data element level, hence a traditional file system is decentralised in nature.
- It could be possible that the data field may have multiple names defined by the different departments of an organization and depending on the file it was in.
- This situation leads to different meaning of a data field in different context or same meaning for different fields. This causes poor data control.

8) Problem of Security:

- It is very difficult to enforce security checks and access rights in a traditional file system, since application programs are added in an adhoc manner.

9) Data Manipulation Capability is Inadequate:

- The data manipulation capability is very limited in traditional file systems since they do not provide strong relationships between data in different files.

10) Needs Excessive Programming:

- An excessive programming effort was needed to develop a new application program due to very high interdependence between program and data in a file system.
- Each new application requires that the developers start from the scratch by designing new file formats and descriptions and then write the file access logic for each new file.

Database Systems or Database System Environment

- The DBMS software together with the Database is called a database system.
- In other words, it can be defined as an organization of components that define and regulate the collection, storage, management and use of data in a database.
- Furthermore, it is a system whose overall purpose is to record and maintain information. A database system consists of four major components:

Data:

- The whole data in the system is stored in a single database.
- This data in the database are both shared and integrated.
- Sharing of data means individual pieces of data in the database is shared among different users and every user can access the same piece of data but may be for different purposes.
- Integration of data means the database can be function of several distinct files with redundancy controlled among the files.

Hardware:

- The hardware consists of the secondary storage devices like disks, drums and so on, where the database resides together with other devices.

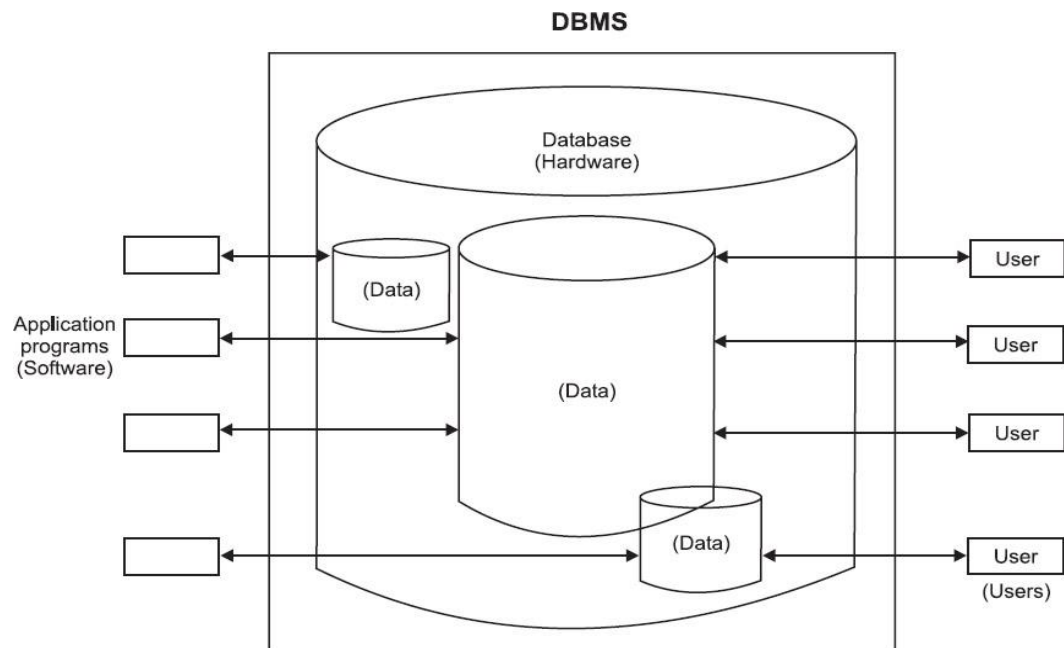
- There are two types of hardware: The first one, i.e., processor and main memory that supports in running the DBMS.
- The second one is the secondary storage devices, i.e., hard disk, magnetic disk etc., that are used to hold the stored data.

Software:

- A layer or interface of software exists between the physical database and the users.
- This layer is called the DBMS.
- All requests from the users to access the database are handled by the DBMS.
- Thus, the DBMS shields the database users from hardware details.
- Furthermore, the DBMS provides the other facilities like accessing and updating the data in the files and adding and deleting files itself.

Users:

- The users are the people interacting with the database system in any way.
- There are four types of users interacting with the database systems.
- These are Application Programmers, online users, end users or naive users and finally the Database Administrator (DBA).



Advantages of Database Systems (DBMS's)

- The Database Systems provide the following advantages over the traditional file system.

1) Controlled redundancy:

- In a traditional file system, each application program has its own data, which causes duplication of common data items in more than one file.
- This duplication/redundancy requires multiple updations for a single transaction and wastes a lot of storage space.
- We cannot eliminate all redundancy due to technical reasons. But in a database, this duplication can be carefully controlled, that means the database system is aware of the redundancy and it assumes the responsibility for propagating updates.

2) Data consistency:

- The problem of updating multiple files in traditional file system leads to inaccurate data as different files may contain different information of the same data item at a given point of time.
- This causes incorrect or contradictory information to its users. In database systems, this problem of inconsistent data is automatically solved by controlling the redundancy.

3) Program data independence:

- The traditional file systems are generally data dependent, which implies that the data organization and access strategies are dictated by the needs of the specific application and the application programs are developed accordingly.
- However, the database systems provide an independence between the file system and application program, that allows for changes at one level of the data without affecting others.
- This property of database systems allow to change data without changing the application programs that process the data.

4) Sharing of data:

- In database systems, the data is centrally controlled and can be shared by all authorized users.
- The sharing of data means not only the existing applications programs can also share the data in the database but new application programs can be developed to operate on the existing data.

- Furthermore, the requirements of the new application programs may be satisfied without creating any new file.

5) Enforcement of standards:

- In database systems, data being stored at one central place, standards can easily be enforced by the DBA.
- This ensures standardised data formats to facilitate data transfers between systems.
- Applicable standards might include any or all of the following—departmental, installation, organizational, industry, corporate, national or international.

6) Improved data integrity:

- Data integrity means that the data contained in the database is both accurate and consistent.
- The centralized control property allow adequate checks can be incorporated to provide data integrity.
- One integrity check that should be incorporated in the database is to ensure that if there is a reference to certain object, that object must exist.

7) Improved security:

- Database security means protecting the data contained in the database from unauthorised users.
- The DBA ensures that proper access procedures are followed, including proper authenticational schemes for access to the DBMS and additional checks before permitting access to sensitive data.
- The level of security could be different for various types of data and operations.

8) Data access is efficient:

- The database system utilizes different sophisticated techniques to access the stored data very efficiently.

9) Conflicting requirements can be balanced:

- The DBA resolves the conflicting requirements of various users and applications by knowing the overall requirements of the organization.

- The DBA can structure the system to provide an overall service that is best for the organization.

10) Improved backup and recovery facility:

- Through its backup and recovery subsystem, the database system provides the facilities for recovering from hardware or software failures.
- The recovery subsystem of the database system ensures that the database is restored to the state it was in before the program started executing, in case of system crash.

11) Minimal program maintenance:

- In a traditional file system, the application programs with the description of data and the logic for accessing the data are built individually.
- Thus, changes to the data formats or access methods results in the need to modify the application programs. Therefore, high maintenance effort are required.
- These are reduced to minimal in database systems due to independence of data and application programs.

12) Data quality is high:

- The quality of data in database systems are very high as compared to traditional file systems. This is possible due to the presence of tools and processes in the database system.

13) Good data accessibility and responsiveness:

- The database systems provide query languages or report writers that allow the users to ask ad hoc queries to obtain the needed information immediately, without the requirement to write application programs (as in case of file system), that access the information from the database.
- This is possible due to integration in database systems.

14) Concurrency control:

- The database systems are designed to manage simultaneous (concurrent) access of the database by many users.
- They also prevents any loss of information or loss of integrity due to these concurrent accesses.

15) Economical to scale:

- In database systems, the operational data of an organization is stored in a central database.
- The application programs that work on this data can be built with very less cost as compared to traditional file system.
- This reduces overall costs of operation and management of the database that leads to an economical scaling.

16) Increased programmer productivity:

- The database system provides many standard functions that the programmer would generally have to write in file system.
- The availability of these functions allow the programmers to concentrate on the specific functionality required by the users without worrying about the implementation details.
- This increases the overall productivity of the programmer and also reduces the development time and cost.

Disadvantages of Database Systems

- In contrast to many advantages of the database systems, there are some disadvantages as well.
- The disadvantages of a database system are as follows:

1) Complexity increases:

- The data structure may become more complex because of the centralised database supporting many applications in an organization.
- This may lead to difficulties in its management and may require professionals for management.

2) Requirement of more disk space:

- The wide functionality and more complexity increase the size of DBMS.
- Thus, it requires much more space to store and run than the traditional file system.

3) Additional cost of hardware:

- The cost of database system's installation is much more. It depends on environment and functionality, size of the hardware and maintenance costs of hardware.

4) Cost of conversion:

- The cost of conversion from old file-system to new database system is very high.
- In some cases the cost of conversion is so high that the cost of DBMS and extra hardware becomes insignificant.
- It also includes the cost of training manpower and hiring the specialized manpower to convert and run the system.

5) Need of additional and specialized manpower:

- Any organization having database systems, need to be hire and train its manpower on regular basis to design and implement databases and to provide database administration services.

6) Need for backup and recovery:

- For a database system to be accurate and available all times, a procedure is required to be developed and used for providing backup copies to all its users when damage occurs.

7) Organizational conflict:

- A centralised and shared database system requires a consensus on data definitions and ownership as well as responsibilities for accurate data maintenance.

8) More installational and management cost:

- The big and complete database systems are more costly. They require trained manpower to operate the system and has additional annual maintenance and support costs.