# Introduction To Database Systems

Imagine trying to operate a business without knowing who your customers are, what products you are selling, who is working for you, who owes you money, and whom you owe money. All businesses have to keep this type of data and much more; and just as importantly, they must have those data available to decision makers when they need them. It can be argued that the ultimate purpose of all having these data is to help businesses use information as an organizational resource. These data could vary from a few megabytes on just one or two topics to terabytes covering hundreds of topics within the business's internal and external environment. How can these businesses process this much data? How can they store it all, and then quickly retrieve just the facts that decision makers want to know, just when they want to know it? The answer is that they use databases.

## Data and Information

In an organization, the data is the most basic resource. To run the organization efficiently, the proper organization and management of data is essential. The term data can be defined as known facts that could be recorded and stored on Computer Media. It is also defined as raw facts from which the required information is produced. Data and information are closely related and are often used interchangeably. Information is nothing but refined data. In other way, we can say, information is processed, organized or summarized data. According to Burch *et. al.,* "Information is data that have been put into a meaningful and useful content and communicated to a recipient who uses it to made decisions".

In these days, there is no lack of data, but there is lack of quality information. The quality information means information that is accurate, timely and relevant, which have these three major key attributes.

1. **Accuracy:** It means that the information is free from errors, and it clearly and accurately reflects the meaning of data on which it is based. It also means it is free from bias and conveys an accurate picture to the recipient.
2. **Timeliness:** It means that the recipients receive the information when they need it and within the required time frame.
3. **Relevancy:** It means the usefulness of the piece of information for the corresponding persons. It is a very subjective matter. Some information that is relevant for one person might not be relevant for another and vice versa *e.g.,* the price of printer is irrelevant for a person who wants to purchase computer.

**Database**

A database is an organized collection of data, so that it can be easily accessed and managed with controlled redundancy to serve one or more applications in an optimal way. The data are stored in such a way that they are independent of the programs used by the people for accessing the data. The approach used in adding the new data, modifying and retrieving the existing data from the database is common and controlled one. It is also defined as a collection of logically related data stored together that is designed to meet information requirements of an organization. The example of a database is a telephone directory that contains names, addresses and telephone numbers of the people stored in the computer storage.

**File System**

Conventionally, the data were stored and processed using file 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 uses 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. Sometimes, it is even not possible to identify all the programs using that data and is identified on a trial-and-error basis. 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.

**Disadvantages of File System**

A 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 contained in the data file that affects the accuracy of reports.

3. **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.
4. **Lack of security and limited data sharing:** Sharing data among multiple geographically dispersed users introduce a lot of security risks. Security and data-sharing features are difficult to programs and are, therefore, often omitted in a file system environment. For instance, features such as the ability to lock out parts of files or parts of the system itself, and other measures designed to safeguard data confidentiality. Even when an attempt is made to improve system and data security, the security devices tend to be limited in scope and effectiveness.

**Advantages of Database Systems**

The database systems provide the following advantages over the file system:

1. **Reduced redundancy:** In a 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 updates 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 reduced.
2. **Data consistency:** The problem of updating multiple files in 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. **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 authentical 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.