**Database System**

# **Audience**

This Course will especially help computer science graduates in understanding the basic-to-advanced concepts related to Database Management Systems.

# **Prerequisites**

A good understanding of basic computer concepts such as primary memory, secondary memory, and data structures and algorithms.

# **Course Objectives**

* To be able to understand the underlying concepts of database, and database management system (DBMS)
* To introduce students to the concepts of relational data model
* Analysis and design of database application or information system
* Experience with SQL
* Implementation of database using SQL
* Learn to work with Microsoft SQL Server 2008 environment

# **Lecture 01**

**Overview of Lecture**

o Introduction to the course

o Database definitions

o Importance of databases

o Introduction to File Processing Systems

o Advantages of the Database Approach

**Database definitions:**

Definition 1: A shared collection of logically related data, designed to meet the information needs of multiple users in an organization. The term database is often erroneously referred to as a synonym for a “database management system (DBMS)”. They are not equivalent and it will be explained in the next section.

Definition 2: A collection of data: part numbers, product codes, customer information, etc. It usually refers to data organized and stored on a computer that can be searched and retrieved by a computer program.

Definition 3: A data structure that stores metadata, i.e. data about data. More generally we can say an organized collection of information.

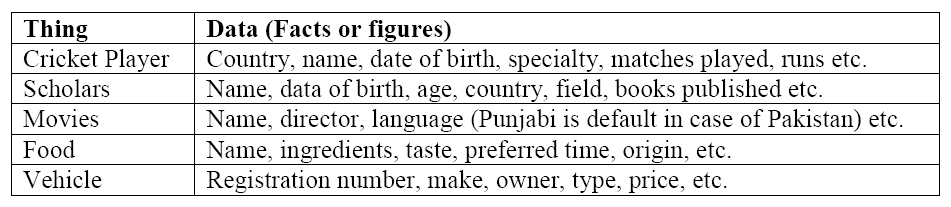
Definition 4: A collection of information organized and presented to serve a specific purpose. (A telephone book is a common database.) A computerized database is an updated, organized file of machine readable information that is rapidly searched and retrieved by computer.

Definition 5: An organized collection of information in computerized format.

Definition 6: A collection of related information about a subject organized in a useful manner that provides a base or foundation for procedures such as retrieving information, drawing conclusions, and making decisions.

Definition 7: A Computerized representation of any organizations flow of information and storage of data.

Important thing that you should be very clear about is the difference between database and the database management system (DBMS). See, the database is the collection of data about anything, could be anything. Like cricket teams, students, busses, movies, personalities, stars, seas, buildings, furniture, lab equipment, hobbies, hotels, pets, countries, and many more anything about which you want to store data. What we mean by data; simply the facts or figures. Following table shows the things and the data that we may want to store about them:



The database management system (DBMS), on the other hand is the software or tool that is used to manage the database and its users. A DBMS consist of different components or subsystem that we will study about later. Each subsystem or component of the DBMS performs different function(s), so a DBMS is collection of different programs but they all work jointly to manage the data stored in the database and its users.

**Importance of the Databases**

Databases are important; why? Traditionally computer applications are divided into commercial and scientific (or engineering) ones. Scientific applications involve more computations, that is, different type of calculations that vary from simple to very complex.

Today such applications exist, like in the fields of space, nuclear, medicine that take hours or days of computations on even computers of the modern age. On the other hand, the applications that are termed as commercial or business applications do not involve much computations, rather minor computation but mainly they perform the input/output operations. That is, these applications mainly store the data in the computer storage, then access and present it to the users in different formats (also termed as data processing) for example, banks, shopping, production, utilities billing, customer services and many others. As is clear from the example systems mentioned, the commercial applications exist in the day to day life and are related directly with the lives of common people. In order to manage the commercial applications more efficiently databases are the ultimate choice because efficient management of data is the sole objective of the databases. So such applications are being managed by databases even in a developing country like Pakistan, yet to talk about the developed countries. This way databases are related directly or indirectly almost every person in society.

Databases are not only being used in the commercial applications rather today many of the scientific/engineering application are also using databases less or more. Databases are concern of the effectively latter form of applications are more Commercial applications involve the goal of this course is to present an in-depth introduction to databases, with an emphasis on how to organize information in the database and to maintain it and retrieve it efficiently, that is, how to design a database and use it effectively.

**Databases and Traditional File Processing Systems**

Traditional file processing system or simple file processing system refers to the first computer-based approach of handling the commercial or business applications. That is why it is also called a replacement of the manual file system. Before the use computers, the data in the offices or business was maintained in the files (well in that perspective some offices may still be considered in the pre-computer age). Obviously, it was laborious, time consuming, inefficient, especially in case of large organizations.

Computers, initially designed for the engineering purposes were though of as blessing, since they helped efficient management but file processing environment simply transformed manual file work to computers. So processing became very fast and efficient, but as file processing systems were used, their problems were also realized and some of them were very severe.

It is not necessary that we understand the working of the file processing environment for the understanding of the database and its working. However, a comparison between the characteristics of the two definitely helps to understand the advantages of the databases and their working approach. That is why the characteristics of the traditional file processing system environment have been discussed briefly here.

**Advantages of Databases**

It will be helpful to reiterate our database definition here, that is, database is a shared collection of logically related data, designed to meet the information needs of multiple users in an organization. A typical database system environment is shown in the figure below:



The figure shows different subsystem or applications in an educational institution, like library system, examination system, and registration system. There are separate, different application programs for every application or subsystem. However, the data for all applications is stored at the same place in the database and all application programs, relevant data and users are being managed by the DBMS. This is a typical database system environment and it introduces the following advantages:

o Data Sharing

The data for different applications or subsystems is placed at the same place. This introduces the major benefit of data sharing. That is, data that is common among different applications need not to be stored repeatedly, as was the case in the file processing environment. For example, all three systems of an educational institution shown in figure 3 need to store the data about students. The example data can be seen from figure 2. Now the data like registration number, name, address, and father name that is common among different applications is being stored repeatedly in the file processing system environment, where as it is being stored just once in database system environment and is being shared by all applications. The interesting thing is that the individual applications do not know that the data is being shared and they do not need to. Each application gets the impression as if the data is being for stored for it. This brings the advantage of saving the storage along with others discussed later.

o Data Independence

Data and programs are independent of each other, so change is once has no or minimum effect on other. Data and its structure is stored in the database where as application programs manipulating this data are stored separately, the change in one does not unnecessarily effect other.

o Controlled Redundancy

Means that we do not need to duplicate data unnecessarily; we do duplicate data in the databases, however, this duplication is deliberate and controlled.

o Better Data Integrity

Very important feature; means the validity of the data being entered in the database.

Since the data is being placed at a central place and being managed by the DBMS, so it provides a very conducive to check or ensure that the data being entered into the database is actually valid. Integrity of data is very important, since all the processing and the information produced in return are based on the data. Now if the data entered is not valid, how can we be sure that the processing in the database is correct and the results or the information produced is valid? The businesses make decisions on the basis of information produced from the database and the wrong information leads to wrong decisions, and business collapse. In the database system environment, DBMS provides many features to ensure the data integrity, hence provides more reliable data processing environment.

**Difference between Data and Information**

Data is the collection of raw facts collected from any specific environment for a specific purpose. Data in itself does not show anything about its environment, so to get desired types of results from the data we transform it into information by applying certain processing on it. Once we have processed data using different methods data is converted into meaningful form and that form of the Data is called information

Example:



If we consider the data in the above figure without the titles or the labels associated with the data (EmpName, age, salary) then it is not much useful. However, after attaching these labels it brings some meanings to us, this meaningfulness is further increased when we associate some other labels, like the company name and the department name etc. So this is a very simple example of processing that we can do on the data to make it information.

Once we have clear idea of what data and information is we proceed with another term knows as “schema” Schema is a repository or structure to express the format and other different information about data and database, as we can see from the database definition “Database is a self-describing collection of interrelated records.” The word self-describing means that the data storage and retrieval mechanism and its format is described in the database, Actual place where these definitions and descriptions are performed is database schema.

o Database Application:

Database Application is a program or group of programs which is used for performing certain operations on the data stored in the database. These operations may contain insertion of data into a database or extracting some data from the database based on a certain condition, updating data in the database, producing the data as output on any device such as Screen, disk or printer.

o Database Management Systems:

Database management system is software of collection of small programs to perform certain operation on data and manage the data.

Two basic operations performed by the DBMS are:

* Management of Data in the Database
* Management of Users associated with the database.

Management of the data means to specify that how data will be stored, structured and accessed in the database. Management of database users means to manage the users in such a way that they can perform any desired operations on the database. DBMS also ensures that a user cannot perform any operation for which he is not allowed. And also an authorized user is not allowed to perform any action which is restricted to that user.

In General DBMS is a collection of Programs performing all necessary actions associated to a database.

**Users of Database Systems:**

o Application Programmers

o End Users

Naïve

Sophisticated

o Application programmers:

This category of database users contains those people who create different types of database application programs that we have seen earlier. Application programmers design the application according to the needs of the other users of the database in a certain environment. Application programmers are skilled people who have clear idea of the structure of the database and know clearly about the needs of the organizations.

o End Users:

Second category of the Database users are the end users, this group of users contains the people who use the database application programs developed by the Application programmers. This category further contains two types of users

Naïve Users

Sophisticated Users

Naïve Users

This category of users is that category that simply uses the application database programs created by the programmers. This group has no interaction with other parts of their database and only use the programs meant for them. They have not to worry about the further working of the database.

Sophisticated Users:

This type of users has some additional rights over the Naïve users, which means that they can access the data stored in the database any of their desired way. They can access data using the application programs as well as other ways of accessing data. Although this type of users has more rights to access data, but these users have to take more responsibility and they need to be aware of the database structure. Moreover such users should be skilled enough to be able to get data from database with making and damage or loss to the data in database.

o Database Administrators (DBA):

This class of database users is the most technical class of db users. They need to have the knowledge of how to design and manage the database use as well as to manage the data in the database. DBA is a very responsible position in an organization. He is responsible for proper working of the database and DBMS, has the responsibility of making proper database backups and make necessary actions for recovering the database in case of a database crash. To fulfill the requirements of a DBA position a DBA needs vast experience and very elegant technical skills.

**Typical Components of a Database Environment:**

Different typical components of a database environment are shown in the figures below; they describe graphically the role of different types of users.



Database is used to store data and DBMS uses mechanisms to get data from the database.



Application programs talk to DBMS and ask for the data required.



Database designers design (for large organizations) the database and install the DBMS for use by the users of the database in any specific organization.



Once Database has been installed and is functioning properly in a production environment of an organization the Database Administrator takes over the charge and performs specific DBA related activities including:

* Database maintenance.
* Database Backup.
* Grant of rights to database users.
* Monitoring of Running Jobs
* Managing Print jobs
* Ensuring quality of Service to all users.



* Database administrator can interact with the database designer during database design phase so that he has a clear idea of the database structure for easy reference in future.
* This helps DBA perform different tasks related to the database structure.
* DBA also interacts with the application programmers during the application development process and provides his services for better design of applications.
* End users also interact with the system using application programs and other tools as specified in the description above.