7 Relational Database and SQL

7.1 INTRODUCTION

When we speak about an organization, a large amount of data is required to be processed and handled. This data handling is performed by arranging data in the form of tables and databases.

A **database** is defined as an organized collection of data (information) about an entity (something that exists) or things. It is a shared collection of related data/information used to support the activities and decision-making of a particular organization. It also allows the users to enter, access and analyze their data quickly and easily. It serves as a container which may contain various database objects. Database is **integrated** as well as **shared**. For example, all files belonging to an organization will be treated as the database of that organization. A database, therefore, is considered as a repository of stored data.

We will now discuss some components like files, tables, records, fields, etc., that are an important part of a database.

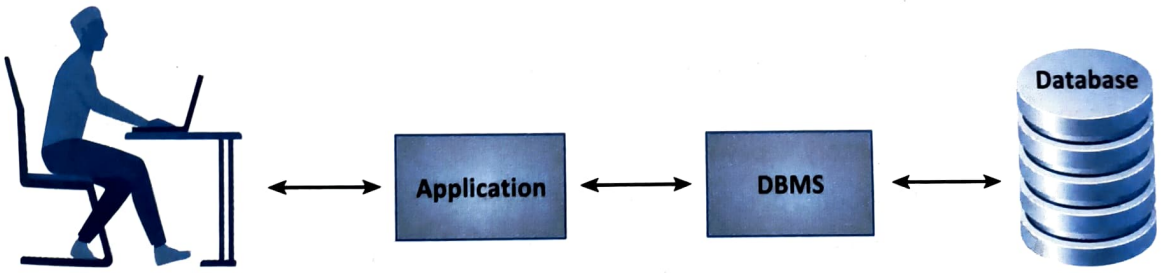

Fig. 7.1: Database and Database Management System

CTM: Database is an organized collection of interrelated data that serves many applications.

Consider the example of a "School" database. This database shall constitute tables related to student, teacher, result, etc. The data is arranged inside a database as per the file organization hierarchy as shown in Fig. 7.2.

Fig. 7.2: File Organization

- **Data/character** is the smallest unit of file organization which represents itself in the form of a bit that may be either 0 or 1. Eight bits make a byte which represents a character in a computer.
- A **field** is a set of characters which are used together to represent specific data elements. It is also termed as a data item. A specific or an individual data item within a record is known as a field. For example, Roll number, Name, Age and Marks are the fields in a student's record.
- A collection of fields is termed as a **Record**. For example, a Student record consists of the fields Roll No, Name, Age and Marks as shown in Fig. 7.3.

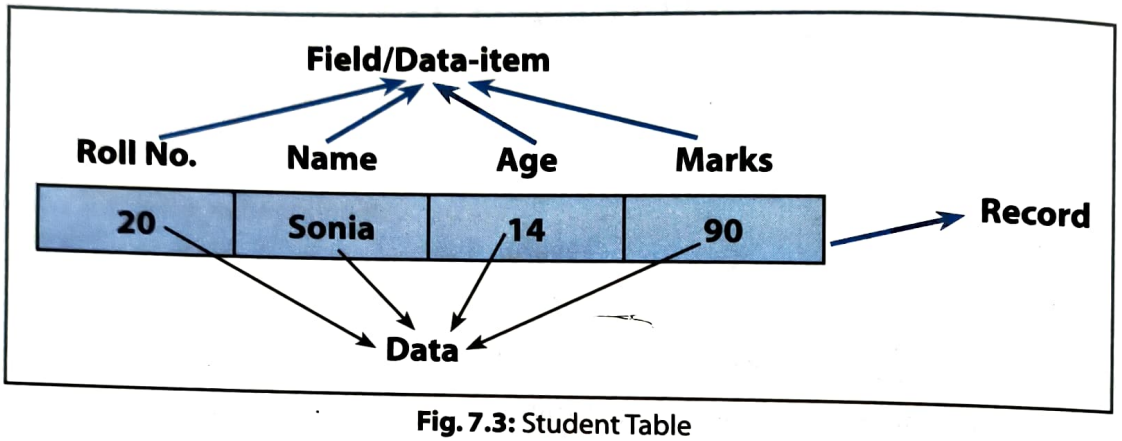

Fig. 7.3: Student Table

- A collection of logically related records is called a file. **A file is also termed as a table or a relation**. A table has rows and columns, where rows represent records or tuples and columns represent the attributes or fields. For example, the entire information about all the students (in the form of records) in a class is kept in a file or table named "student" (Fig. 7.3).
- **Database** is, therefore, a place where related information is stored and various operations can be performed on it. It is the highest unit of file organization.

Table 7.1: Use of Database in Real-life Applications

Application	Database to maintain data about
1. Banking	Customer information, Account details, Loan details, Transaction details, etc.
2. Crop Loan	Kisan credit card data, Farmer's personal data, Land area and Cultivation data, Loan history, Repayment data, etc.
3. Inventory Management	Product details, Customer information, Order details, Delivery data, etc.
4. Organization Resource Management	Employee records, Salary details, Department information, Branch locations, etc.
5. Online Shopping	Item description, User login details, User preference details, etc.

These databases are generally managed by a special software known as **Database Management System (DBMS)**.

7.2 DATABASE MANAGEMENT SYSTEM (DBMS)

Database Management Systems are specially-designed applications that connect the user and program, and store data in an organized manner. The purpose of DBMS software is to allow the user to create, modify and control a database.

A DBMS stores data in such a manner that it becomes easier and highly efficient to retrieve, manipulate and produce information. Thus, a **DBMS** is an electronic or computerized record-keeping system. It maintains the various pieces of information in an integrated and summarized form instead of keeping them in separate independent files.

Examples of Database Management Systems are MS-Access, MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle, SAP, dBase, FoxPro, etc.

Few customized DBMSs are computerized library systems, automated teller machines, flight reservation systems, computerized parts, inventory systems, etc.

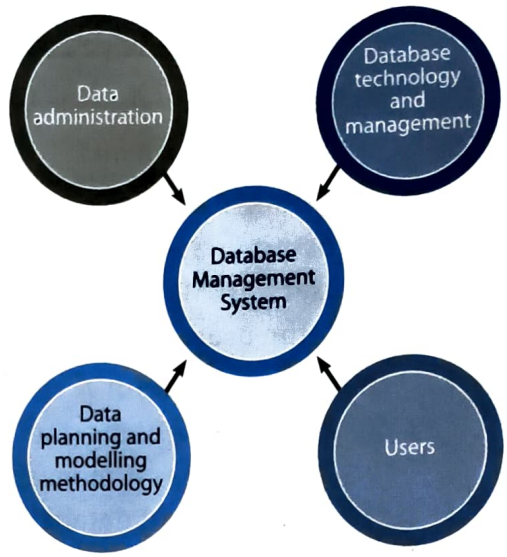

Fig. 7.4: Role of Database Management System

A DBMS gives us tools to:

- Store data in a structured way.
- Query the database (*i.e.*, ask questions about the data).
- Sort and manipulate the data in the database.
- Validate the data entered and check for inconsistencies.
- Produce flexible reports, both on screen and on paper, that make it easy to comprehend the information stored in the database.

Also, it maintains data consistency in the case of multiple users.

CTM: A DBMS is a general purpose software system that facilitates the process of defining, constructing and manipulating databases for various applications.

7.2.1 Need for DBMS

The database system is used to eliminate the problems of data redundancy and data inconsistency. It does not maintain separate files for different applications. Rather, it works on the centrally maintained database, which means that data is kept at one place and all applications that require the data may refer to this database. Whenever any file gets updated, the updated version of the file is available to all applications using the database system, as shown in Fig. 7.5. So, data redundancy and data inconsistency are controlled to a large extent.

However, at times, there might be data redundancy due to some technical requirements in business applications. In such cases, we are required to maintain same data for different files but this is not recommended.

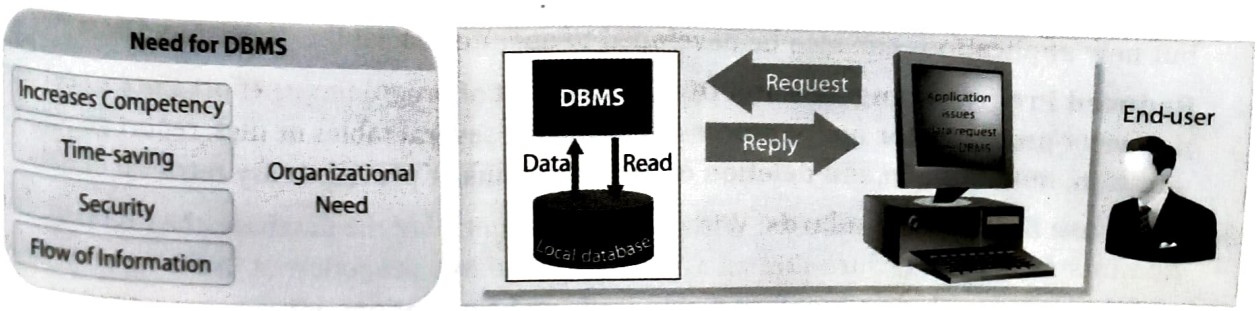

Fig. 7.5: A Centralized Database System

7.2.2 Components of a Database System

The various components of a database system are described in Fig. 7.6.

Let us discuss these components.

1. **Users:** Users can be of varied types, usually a DB administrator, System or Application developers and end-users. DBMS provides the following critical services to the user:
   (a) **Database Creation:** A DBMS helps the user in creating and defining the required data or, in turn, a database. It manages and organizes the required data and databases.
   (b) **Database Maintenance:** It helps in maintenance of data and database by addition, deletion, modification and regular updation of the tables and its records.
   (c) **Database Processing:** A DBMS performs one of the major tasks of query processing—it processes the queries or the information requirement of users and retrieves necessary information from the database.

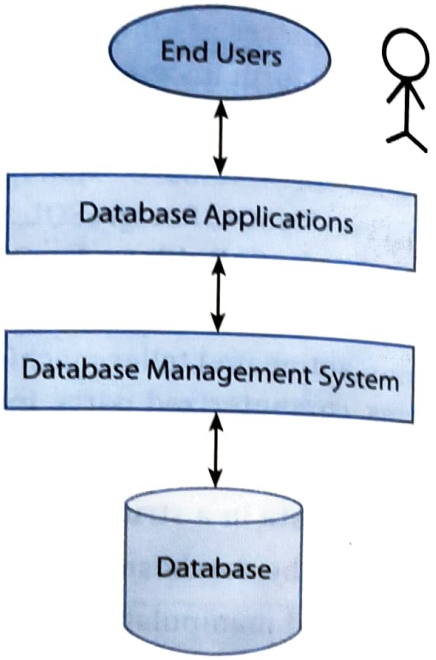

Fig. 7.6: Components of a Database System

2. **Database Application:** Database application may be Personal, Departmental, Enterprise and Internal. It may be general-purpose or customized as per the needs of a user.
3. **DBMS:** Software that allows users to define, create, access and manage database(s) is termed as a DBMS. For example, MySQL, Oracle, etc.
4. **Database:** It is a collection of logically related data.

7.2.3 Advantages of a DBMS

Apart from providing various salient features described above, a DBMS has several advantages over traditional data processing techniques.

1. **Elimination of Data Redundancy:** Duplication of data leads to wastage in storage space. A DBMS eliminates data redundancy (duplication of data) by integrating the files so that multiple copies of the same data are not stored.
2. **Data Consistency:** A DBMS provides data consistency to a large extent as the changes made at one place are reflected at all other places or to all the users.
3. **Sharing of Data:** By using a DBMS, not only can existing applications share data in the database, but new applications can also be developed to operate against the same stored data.
4. **Reduced Programming Effort:** A DBMS saves a lot of programming effort since a user need not write programs for query processing involving several tables or files, report generation, addition, modification and deletion of data, etc. Thus, it provides easy retrieval of data.
5. **Database Enforces Standards:** With centralized control of the database, the DBA (Database Administrator) can ensure that all applicable standards are followed in the representation of data, *i.e.*, format, documentation standards and conventions, etc.

6. **Improved Data Integrity:** Data integrity refers to the validity and consistency of stored data. For example, the system itself checks for the correct information to be entered by the user in the correct format. It consists of various constraints.
7. **Privacy and Security:** Data security refers to protection of data against accidental or intentional disclosure to unauthorized persons. Since there is centralized control, the data is protected.
8. **Economical:** Combining all the organization's operational data into one database and creating a set of applications that work on this single source of data can result in cost savings. The overall maintenance cost of data is reduced.
9. **Improved Backup and Recovery System:** A database system provides facilities for recovery from hardware or software failures.
10. **Meeting Enterprise Requirements than Individual Requirements:** Since many types of users with varying levels of technical knowledge use a database, a DBMS should provide a variety of user interfaces.

CTM: The repetition (duplication) of same data at multiple places in a database is known as **data redundancy**.

7.3 DBMS MODELS

Data models define how the logical structure of a database is modelled. A data model is an integrated collection of conceptual tools that can be used to describe the structure of the database along with the appropriate data types, relationships and constraints required to be applied on the data.

Data models are used to implement abstraction in a DBMS. They are a communication tool. Data models define how data is connected and how it is processed and stored inside the system. They organize data for various users. A data model should be able to give best data representation and should possess the following desirable characteristics:

1. Data models should be presented graphically using diagrams and symbols.
2. Data representation in a data model should have no data redundancy.
3. A data model should be made available and shared by various applications.
4. Data represented should be consistent, stable and valid in all aspects.

7.3.1 Types of Data Models

Data models are categorized into three different categories:

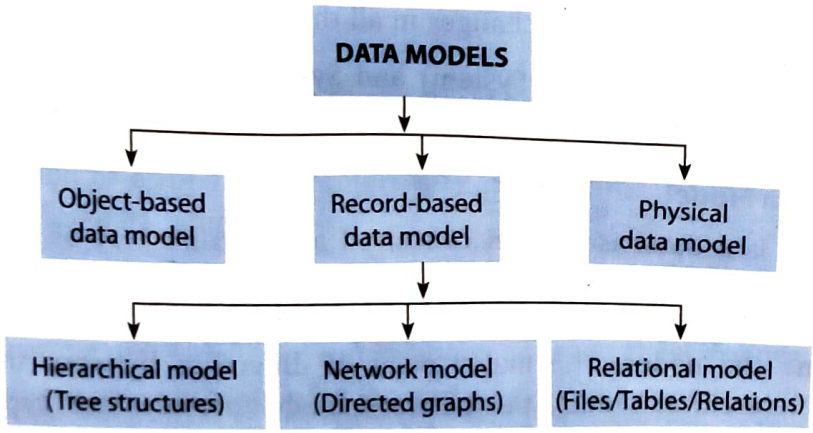

Fig. 7.7: Classification of Data Models

A relational model consists of a collection of tables, each of which is assigned a unique name, *i.e.*, it represents the database as a collection of relations or tables. **Characteristics** of relational database are:

1. Data is conceptually represented as an orderly arrangement of data into rows and columns, termed as a relation.
2. Each relation is represented as a table.
3. Columns described in a table are the attributes that belong to an entity which is modelled as a table.
4. Every row in a table represents a single entity.
5. All the values in a relation are scalar, *i.e.*, at any given row or column position, there is one and only one value.
6. All operations are performed on the entire relation and the result is an entire relation.

Advantages of a relational model are as follows:

1. A relational model provides structural independence by using independent tables.
2. Changes made in the table structure do not affect the data access or other application programs.
3. It is represented in the form of tables; so, it is simple and easier to understand.
4. Tabular view also provides easier database design, use, implementation and management.
5. Built-in query support based on SQL is provided by RDBMS (Relational Database Management System).
6. Data organization and manipulation is easy, flexible and faster.
7. Powerful structure designing and processing capabilities of RDBMS isolate the end-user from physical-level details and, thus, improve implementation and management simplicity.
8. Mathematical operations can be successfully carried out using RDBMS.

The **limitations** of a relational model are:

1. RDBMS incurs hardware and system software overheads.
2. The size of database becomes very large.

7.4 RELATIONAL DATABASE

A **relational database** is a type of database that stores and provides access to data points that are related to one another. Relational databases are based on the relational model, an intuitive, straightforward way of representing data in tables. In a relational database, each row in the table is a record with a unique ID called the key. The columns hold attributes of the entity and each record usually has a value for each attribute, making it easy to establish the relationships among data points.

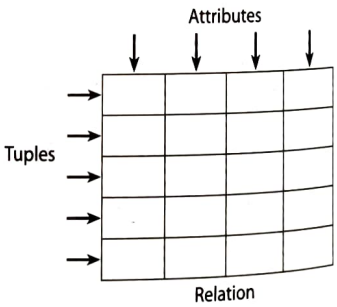

Fig. 7.10: Tuples and Attributes

Often, data in a relational database is organized into tables.

Basic Terminologies related to a Relational Database:

1. **Entity:** An entity is something that exists and about which we can store some information. It is an object which can be distinctly identified. For example, student entity, employee entity, item entity, etc. Entity becomes the name of the table.

2. **Attribute:** In a relational table, an attribute is a set of values of a particular type. The term attribute is also used to represent a column. A table consists of several records (row); each record can be broken into several smaller entities known as fields or attributes or columns. A set of attributes defines the characteristics or properties of an entity. In the given table, Student relation (Fig. 7.11) consists of four fields or attributes—Roll number, Name, Address and Gender.
3. **Tuple:** Each row in a table is known as tuple. It is also called a row/record. A single entry in a table is called a record or row. A record in a table represents a set of related data. For example, the Student table given below has 10 records.

   Each tuple (row) in a relation (table) corresponds to data of a real world entity (for example, Student, Employee and Attendance)
4. **Cardinality of Relation:** It is the number of records or tuples in the relation. Thus, the cardinality of Student relation is 10.
5. **Degree of Relation:** Number of columns or attributes is known as degree of a relation. Thus, the degree of Student relation is 4.
6. **Domain of Relation:** It defines the kind of data represented by the attribute. It is the set of all possible values that an attribute may contain. For example, in the given table Student, domain for the field Gender is two since it can have either 'M' or 'F' as the possible and available values that it may contain.
7. **Body of the Relation:** It consists of an unordered set of 0 or more tuples.

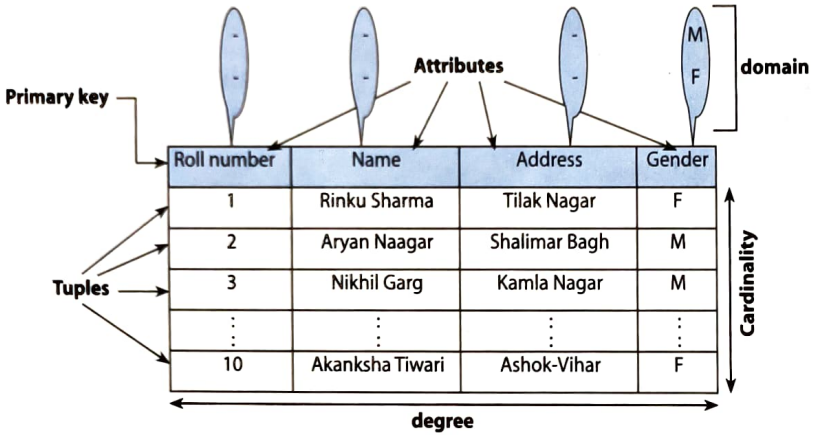

Fig. 7.11: Relational Model (Student Relation)

Therefore, with reference to Fig. 7.11, in the given Student relation:

- There are 10 tuples (*i.e.*, cardinality=10) and 4 attributes (*i.e.*, degree=4).
- Roll number, Name, Address and Gender are the attribute names.
- The first tuple contains the values (1, "Rinku Sharma","Tilak Nagar",'F').
- The domain of the attribute Gender is (M,F).

7.5 DATABASE KEYS

Keys are a very important part of relational database. They allow us to identify an attribute or a set of attributes on the basis of which a table is identified. They are used to establish and identify relation among two or more tables. They also ensure that each record can be uniquely identified by a combination of one or more fields within a table.