

Chapter - 4

Module - 11

DATABASE AND SQL

INTRODUCTION

Data : They are raw or known facts, i.e. the data that has not been processed. For example, data of the ages of students in a class.

Information : Processed data is called information. For example, we find from the data of the age of students in a class, how many students are of age 18. The result we obtain is information.

Practically everything we do generates data. Our ancestors also needed to record such data. When there were floods, marks were made on walls in different places, to indicate the level to which the water had risen in that area. In today's complex world, with so much data being generated due to various type of financial and non-financial transactions, we need means to record this information in a meaningful manner, so that we can retrieve it later with ease.

Database

A database is a collection of data that is related in some way. A list of students enrolled for the commerce stream, the list of names, addresses and telephone numbers in a telephone directory etc are all examples of databases. A database can be compared to a filing cabinet which contains many files. The files are tables containing data. The size of a database can be very large and can have varying levels of complexity. While today most databases are maintained on computers, some of them may be manually maintained.

HIERARCHY OF DATA

Character : Consists of letters (A-Z), digits (0-9) and special characters (-, +, * etc).

Field(Column) : Is a collection of related characters. e.g. name, age etc.

Record(Row) : Is a collection of related fields. e.g. all the information of a student i.e. his name, age, address, roll number etc. is a record.

File(Table) : A collection of related records is called a file.

Database : A collection of one or more related Files is called a Database.

Note : In MySQL it is more common to call a Field as a Column, a record as a Row and a File as a Table.

DBMS

A DBMS is a collection of programs (computer based system) that enables the user to create and maintain a database. It is used to define, construct and manipulate the data in the database for various applications. It provides information storage, organization and retrieval capabilities. The DBMS also enforces necessary access restrictions and security measures in order to protect the database. Various types of control systems within the DBMS make sure that the database continues to function properly. They include - Integrity system - Security system - Concurrency control system - Recovery control system Some DBMS enable us to define "views" of the database. A view is how the database appears to the user. This enables us to show only the relevant information to different types of users and it increases security, as certain users will not be able to see data which they are not meant to see.

Advantages of DBMS

1. Redundancy is controlled.
2. Unauthorized access is restricted.
3. Provides multiple user interfaces.
4. Enforces integrity constraints.
5. Provides backup and recovery of data.

Limitations of DBMS

1. Uses extra computer time and resources.
2. It is expensive to install new systems.
3. Requires well qualified and expert personnel to operate the system.
4. If it is not designed for concurrent access to many users it will be of little use.
5. Authority of ownership rights over the data have to be fixed, so as to determine who is eligible for using the data.

RECORD RELATIONSHIPS

To make it easy for the end user who use the database for various purposes, the following record relationships are used.

One – One : Here one parent record is associated with only one child record, e.g. one doctor in a hospital will take care of one patient.

One – Many : Here one parent record is associated with many child records, e.g. one doctor in the hospital will take care of many patients.

Many – One : Here two or more parent records are associated with one child record, e.g. many doctors will be attending to a single patient.

Many – Many : Here two or more parent records are associated with two or more child records, e.g. two or more doctors will be attending to two or more patients, hence bills of different patients will be associated with different doctors.

DATABASE SYSTEM

The database and DBMS software together is called as Database system.

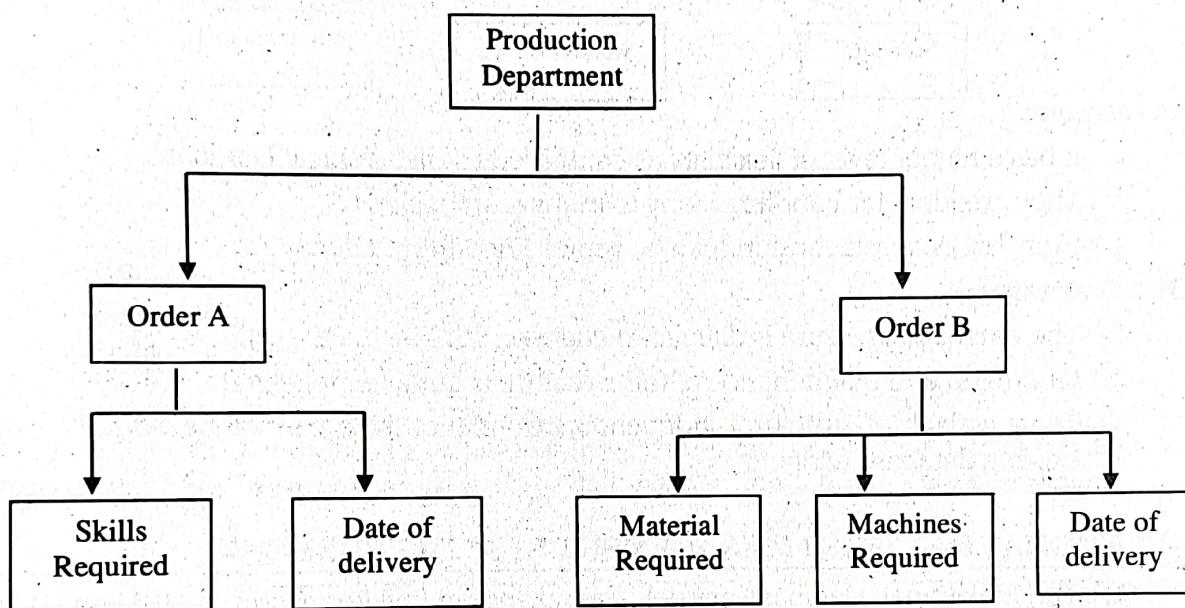
NON RELATIONAL DATABASE SYSTEMS

They are popularly categorized according to the data structures they present to the user. The Non-relational or pre-relational models are Hierarchical and Network database systems.

Hierarchical model

- a) In this model the different records are related through inverted tree like structures. A parent record can have many child records, but a child record can have only one parent record.
- b) It uses the one : one and one : many relationships.
- c) In a hierarchical database, records contain groups of parent/child relationships. For example, under Production Department (parent) would be stored the orders A and B received by the department (child), under that would be the Skills required for the order A and the date of delivery (child), and for order B the material required, machines required and the date of delivery (child), another child would be the employees assigned to each order.
- d) Although this model has been widely used, it is often considered not suitable for many application areas as its structure is inflexible and is not suitable for complex relationships.

Example of Hierarchical Database



Advantages

1. The structure is simple to construct and operate.
2. As the batch operations like payroll are done at maximum speed they permit day to day structured operations rapidly.

Disadvantages

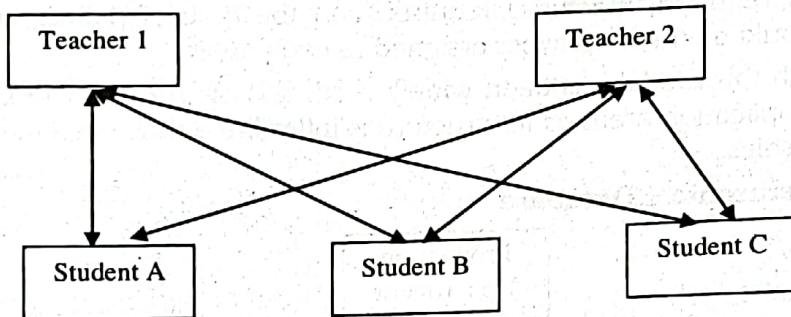
1. Many to many relationships cannot be handled in this model.
2. Its structure is quite rigid and hence adding a field is very complicated often resulting in making a new database.
3. Deleting a parent record is complex as the entire branch of child records associated with it has to be first deleted.
4. Ad hoc queries are not supported.

Network model

- a) In this model a parent record can have many child records and a child record can have many parent records.
- b) It uses the many : many relationships.
- c) The records are physically linked through linked lists.

- d) Hence this model addresses some of the drawbacks of the Hierarchical model. It supports more complex queries and relationships than are possible in the Hierarchical model.
- e) However, the Network model has its limitations, as it is very complex. While accessing the database the users have to be familiar with the structure and keep a track of where they have reached and how they got there.
- f) If any change is to be made to the structure, it would affect the application that interacts with the database.

Example of Network Database



Advantages

1. It has a higher level of flexibility as compared to Hierarchical Database.
2. They are ideal for handling many to many relationships.
3. Complex data relationships can be represented more effectively.

Disadvantages

1. The database structure is difficult to change.
2. Operation and maintenance of this structure is quite complicated.
3. There is lack of structural independence, relationships cannot be added to the existing database easily.

RELATIONAL DATABASE MANAGEMENT SYSTEMS(RDBMS)

1. The Relational Database model was proposed by Mr. Codd of IBM in 1970. However, the first commercial system appeared only in 1981-82.
2. In this model, unlike the Hierarchical and Network models, there are no physical links.
3. All the data is kept in tables which contain rows and columns.
4. The data in two tables are linked through the columns and not through physical links. Hence there is no reason to traverse pointers as in a Non-Relational Model.
5. Relational databases have become a predominant choice for the storage of information in new databases used for financial records, manufacturing and logistical information, personnel data and much more.
6. Relational databases have often replaced hierarchical databases and network databases because they are easier to understand and use, even though they are much less efficient.
7. As computer processing power has increased, the inefficiencies of relational databases, which made them impractical in earlier times, have been outweighed by their ease of use.
8. However, relational databases have been challenged by Object Databases, which were introduced in an attempt to address the object-relational impedance mismatch in relational database, and XML databases.

Although still used in some organizations the Hierarchical and Network Databases are no longer in wide use. The Relational model is most widely implemented model in modern business systems and it is the foundation for SQL.

Note : There are 12 rules for a Relational DBMS, also called as Codd's Rule. A Relational Database Management System that satisfies at least six of these rules is acknowledged as a fully operational RDBMS.

Consider a sample Inventory database given below, which consists of the tables of Stock and Supplier. As the supp_id column is the same in both tables we can obtain information from both tables at the same time.

Stock Table			
stock_id	supp_id	stock_name	clstock
1138	001	Dove Soap	245
1145	002	Colgate	153
1167	003	Nirma	110

Supplier table		
supp_id	Supp_name	Phone_no
001	Manu & Co.	9821066723
002	Raju & Co.	9912456023
003	Hiral & Co.	9983457120

PRIMARY KEY

A column in the table that uniquely identifies each row in that table is called the primary key. In the above example stock_id in the stock table can be the primary key.

CANDIDATE KEY

In a table there could be more than one column that uniquely identifies a row in the table. Then such columns are called candidate keys.

From among the candidate keys, we can choose one to be the primary key, or a combination of them to be the primary key. In the above stock table stock_id and supp_id are candidate keys. Any one of them can be selected as the primary key.

FOREIGN KEY

It is a column in one table, whose value matches with the primary key in another table. A foreign key establishes a relationship, or constraint, between two tables.

In the above tables, stock_id is the primary key in the stock table and supp_id is the primary key in the supplier table. The supp_id column in the stock table is the foreign key, as it is associated with the supp_id key which is the primary key in the supplier table.

STRUCTURED QUERY LANGUAGE(SQL)

Introduction

SQL is a computer language used to manage and interact with data kept in a relational database. It is designed not only to organize and simplify the method of obtaining information from the database, but also to reorganize data within the database.

SQL is easy to use and has been categorized as a fourth generation language (4GL).

SQL a Non-Procedural Query Language

SQL is called a non-procedural Query language, as the user only gives the request of "what is to be obtained from the database (table), without having to worry how that information will be obtained". In procedural languages, such as Java, C etc. one is in complete control over what the program does, as the language defines both the results required as well the process or method to be followed to obtain the same. In SQL on the other hand, it is left to the component of the database engine, to decide the best way to obtain the required information from the table. The SQL statements merely define the input and output, but not the procedure.

Hence with SQL we cannot write complete applications.

SQL Statements

They can be broadly categorized into the following :

1. Data Definition Language(DDL)

This language is used by the designers and programmers of the database to indicate the content and the structure of the database. It is used to indicate the physical structure of the database i.e. field names, their types etc. and also the record relationships. E.g. CREATE, ALTER, and DROP statements.

2. Data Manipulation Language (DML)

This language is used primarily for data manipulation and processing. It involves retrieving the data, arranging the data, deleting the data and displaying the data etc. A user queries the database and receives the required reports. e.g. SELECT, UPDATE, INSERT, DELETE statements

3. Data Control Language(DCL)

This is used for controlling the data and access to the database. It is used to address security issues and restrict the access to the database. E.g. GRANT, REVOKE statements.

4. Transaction Control Statements.

These statements are used to handle transactions. They are **essentially a group of DML statements executed together** as one unit. They are used to begin, end and rollback transactions. E.g. COMMIT and ROLLBACK statements.

QUESTIONS

Answer the following :

1. What is the difference between Data and Information?
2. What is the hierarchy of data?
3. What is a Database Management System?
4. What are the advantages and disadvantages of a DBMS?
5. What are the various record relationships?
6. What is the Hierarchical Database model?
7. What is the Network Database model?
8. What is the Relational Database model?
9. Why SQL is called a non procedural language?
10. What are different SQL statements?
11. Explain Primary key, Candidate key and Foreign key.

