Introduction to Databases

What is a Database?

In its simplest term, a Database is an organized collection of related data. It is a structure that serves as a repository for end user data.

A database is also commonly defined as a shared collection of logically related data and its description, designed to meet the information needs of an organization.

Examples of Databases

1. When you purchase goods from a supermarket, it is likely that a database is accessed.
2. When you purchase goods using your credit card, a check is performed to ensure that you have sufficient credit left to make the purchase. This check usually involves accessing a database somewhere.
3. When you make inquiries about a vacation, your travel agent may access several databases containing vacation and flight details from around the globe.
4. Your local library probably has a database containing details of the books in the library as well as its readers, reservation etc…
5. Many of the sites on the Internet are driven by database applications like Amazon etc..

University's Database Example

All data maintained by a university about its students may be organized as a database

The purpose of the database system is to maintain all relevant data in a form that is useful to the owner for record keeping and decision making.

For a university, the database can be used to prepare lists of students taking a particular course or provide profiles of individual students. (For businesses, the database can be used to record sales and to analyze stock movement to provide information for management decision making.)

The data in the database are organized into tables; for example, a student table may contain personal and contact data on all students. The data are organized into rows and columns (similar to a spreadsheet). A database may contain several tables linked together.

Example

Student Id	Student Name	Date of Birth	Telephone
2008234	Jason Smith	20-May-1992	555-2847
2007123	Shirley Adams	17-Dec-1993	555-2908

2008687	Nigel Croft	26-Jan-1993	555-4875

Other tables which may exist in a student database are

- A course catalog with data on each course (code, title, credits).
- Courses taken which stores results for each student for each course taken (name, code, grade, year etc...)

History of Databases

The predecessor to the database system which we use is the file-based system. Although this system is largely obsolete, it is good to study the system because it will prevent us from making the same mistakes again.

What is a file-based system?

A file based system is a collection of application programs that perform services for the end-users, such as production of reports. Each program manages and defines its own data.

File-based systems were an early attempt to computerize the manual filing system that we are all familiar with. But why would we want to replace the manual filing system? Well the manual filing system works well as long as the number of items to be stored is small. The manual filing system breaks down when we have to cross-reference or process information in the different files.

The file-based system was developed in response to the needs of industry for more efficient data access. However, rather than establishing a centralized store for the company's operational data, a decentralized approach was taken, where each department stored and controlled its own data.

For example, all the data related to and needed by the Sales department will be stored in the sales department, all the data related to and needed by the HR department will be stored in the HR department and so on.

Limitations of the File-based System

1. Separation and Isolation of Data – When data is stored in separate files, it is more difficult to access data that should be available. This difficulty is compounded if we require data from more than two files.
2. Duplication of Data – Owing to the decentralized approach taken by each department, the file-based approach encouraged, if not necessitated the uncontrolled duplication of data.
3. Data Dependence – The physical structure and storage of the data files and records are defined in the application code. This means that changes to an existing structure are difficult to make.
4. Incompatible file formats – Because the structure of files is embedded in the application programs, the structures are dependent on the application programming language. For

example, the structure of a file generated by a COBOL program may be different from the structure of a file generated by a C program.

Database Approach

The database is a single possibly large repository of data that can be used simultaneously by many departments and users. Instead of disconnected files with redundant data, all data items are integrated with a minimum amount of duplication. The database is no longer owned by one department but is a shared corporate resource. The database holds not only the organization's operational data, but also a description of this data. For this reason, a database is also defined as a self-describing collection of integrated records. The description of the data is known as the **system catalog** (or **data dictionary** or **metadata** – "the data about data").

The Database Management System (DBMS)

The DBMS is a software system that enables users to define, create, maintain and control access to the database.

The DBMS is the software that interacts with the user's application programs and the database. It includes facilities for defining and creating the database tables and populating and manipulating the database for various applications. Structured Query Language (SQL) is the standard language that is used by most DBMS

The DBMS has facilities for performing the following (mainly through the use of SQL)

- Defining and creating the database tables involves creating the structure for the tables that constitute the database, specifying the type of data and the constraints for the data to be stored in the database. This is done using **Data Definition Language (DDL).**
- Populating the database involves adding data to the database.
- Manipulating the database includes functions such as querying the database to retrieve specific data, generating reports from data and updating the database to reflect changes in the database.

Example of DBMS

There are many DBMSs available; some are commercial and others are open source, for example

- MS Access is a very popular DBMS. It is a module in Microsoft Office. It has a very easy to use interface and can be used without requiring any programming. It has a number of wizards for creating tables and reports etc..

- Microsoft SQL Server is a database server developed by Microsoft. It can be used to create desktop, web-based and enterprise wide applications.

- Oracle is the popular choice among large companies for enterprise wide applications

- mySQL is an open source DBMS that can be used for small database applications as well as enterprise applications.