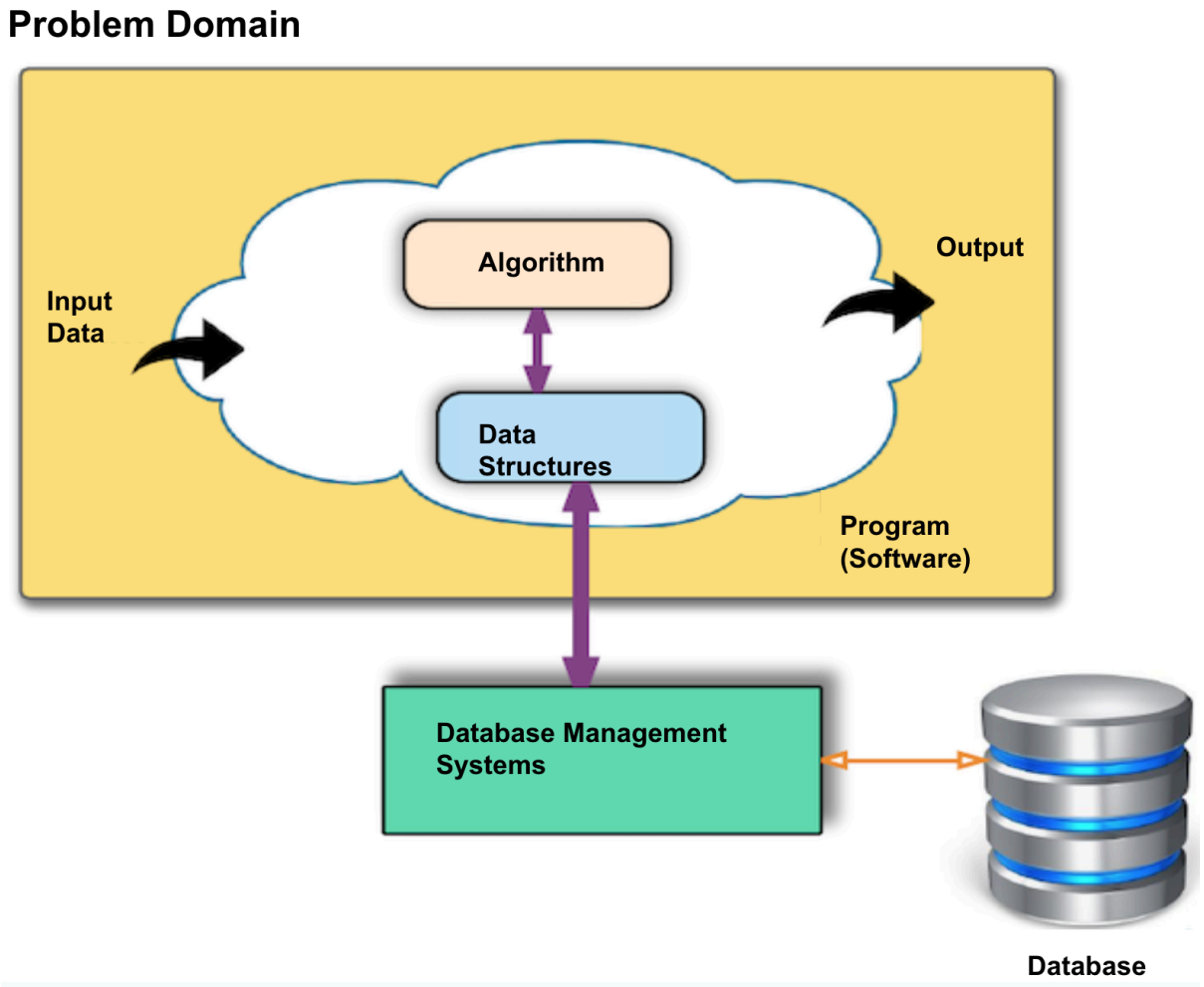
Module 1: Database Systems

Introduction to database systems

The Big Picture



A collection of related data stored in files is called a **database**.

Data Growth and Its Management

- **China's giant telescope**, which is the size of thirty football fields, collects **38 gigabytes of data per second**. (2020)
- **Google processes an average of 40,000 searches per second** (3.5 billion daily on average, with a total of 5 billion searches per day). Search results are quickly delivered to users. (2018)
- **Facebook has 2 billion users**, with **1.5 billion active daily users**. (2018)
- **Every minute**, the following activities occur:
    - **4,146,600 YouTube videos are watched**
    - **456,000 tweets are posted**
    - **46,740 photos are uploaded to Instagram**
    - **510,000 comments are added to Facebook** (2018)

Storing, managing, and quickly accessing such massive amounts of data requires the use of **databases**.

The Use of Databases in Various Sectors

Today, **databases** are widely used across many industries and institutions, including:

- **Finance**
- **Education**
- **Transportation**
- **Logistics**
- **Communication**
- **Media**
- **Healthcare**
- **Information Technology**
- **Manufacturing**

Data and Information

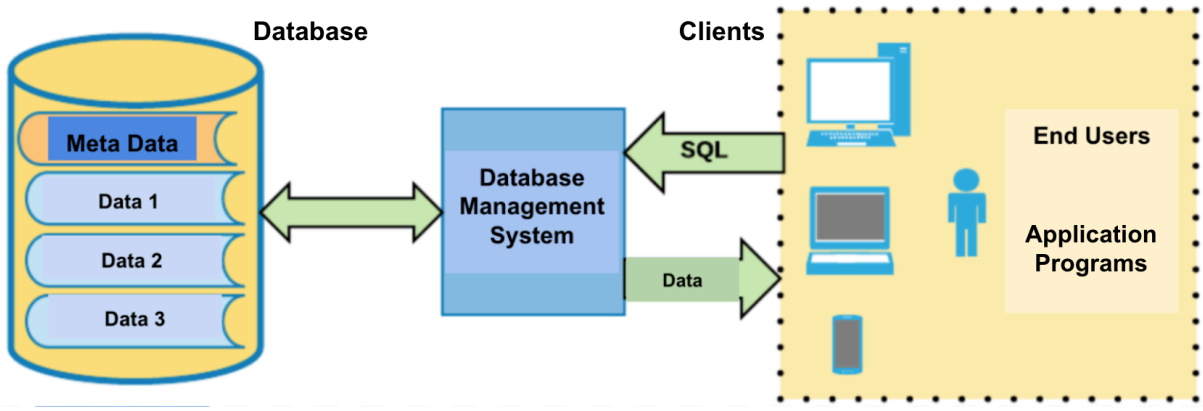
- Raw facts that have not been processed or given meaning are called data.

- When data is processed, it becomes information.
- Information is used to represent the meaning of data.
- Accurate, relevant, and timely information is highly effective in decision-making processes.
- Making the right decisions is crucial for the survival of organizations.
- Data management is one of the most fundamental activities in organizations.
- Data management is the discipline that deals with the proper generation, storage, and access to data.

Database System

Database System = Database + DBMS + Clients

- **Database = Raw Data + Metadata** (Relationships + Data Characteristics).
- Software that manages the database structure and allows access to the data is called a **Database Management System (DBMS)**.



Database System Environment

Hardware

- Servers, workstations, network infrastructure, storage devices, RAID systems, etc.

Software

- **Operating Systems**

- **Database Management Systems (DBMS):** PostgreSQL, MySQL, DB2, Oracle, MSSQL, Cassandra, MongoDB, Redis, etc.
- **Application programs and utility tools**

People

- System administrator, database administrator (DBA), database designer, application developer, end user

Data

- In the center of everything.

Essential Features of a Database Management System (DBMS)

- **Data Integration:** Ensures efficient and redundancy-free storage of data.
- **Data Integrity:** Guarantees that data remains accurate and consistent. Constraints such as key constraints and integrity rules help prevent inconsistencies.
- **Data Security:** Protects data from system failures and cyberattacks while maintaining its integrity through transactions, RAID systems, recovery mechanisms, and advanced authorization structures.
- **Data Abstraction:** Provides users with a logical model that simplifies data interaction, making complex physical data structures easier to understand and manage. Consider the following example comparing data abstraction in a traditional file system and a DBMS.

  File system

```
String file = "Students.csv";
try {
    FileReader fileReader = new FileReader(file);
    BufferedReader bufferedReader = new BufferedReader(fileReader);

    String satir = null;
    while ((satir = bufferedReader.readLine()) != null) {
        System.out.println(satir);
    }
    bufferedReader.close();
} catch (IOException e) {
    e.printStackTrace();
}
```

  Relational DBMS (SQL)

```
Select * From Students;
```

Some well-known Relational Database Management Systems (RDBMS)

- PostgreSQL – An advanced open-source RDBMS with strong support for SQL and NoSQL features, known for data integrity and extensibility.
- MySQL – An open-source RDBMS known for its speed and reliability, widely used in web applications and data-driven platforms.
- Oracle Database – A commercial RDBMS used in enterprise applications, offering powerful performance, security, and scalability features.
- Microsoft SQL Server – A robust RDBMS from Microsoft, widely used in business applications and integrated with Windows-based environments.
- SQLite – A lightweight, embedded RDBMS often used in mobile applications and local storage without requiring a separate server.
- IBM Db2 – A high-performance RDBMS designed for large-scale enterprise applications, offering AI-driven optimizations.
- MariaDB – A MySQL-compatible open-source database known for enhanced performance, security, and scalability.

References

- Carlos Coronel, Steven Morris, and Peter Rob, Database Systems: Design, Implementation, and Management, Cengage Learning.
- http://www.digitalinformationworld.com/2015/02/fascinating-social-networking-stats-2015.html
- https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#230ac18d60ba
- Apache Spark Tutorial | Spark Tutorial for Beginners | Apache Spark Training | Edureka
    - https://www.youtube.com/watch?v=9mELEARcxJo