Data vs. Information

Data	Information
▪ Raw facts ▪ Raw data - Not yet been processed to reveal the meaning ▪ Building blocks of information ▪ **Data management** ▪ Generation, storage, and retrieval of data	▪ Produced by processing data ▪ Reveals the meaning of data ▪ Enables **knowledge** creation ▪ Should be accurate, relevant, and timely to enable good decision making

Types of Databases

- **Single-user database**: Supports one user at a time
  - **Desktop database**: Runs on PC Local Store DB
- **Multiuser database**: Supports multiple users at the same time
  - **Workgroup databases**: Supports a small number of users or a specific department
  - **Enterprise database**: Supports many users across many departments

Role of the DBMS

- Intermediary between the user and the database
- Enables data to be shared Scalable
- Presents the end user with an integrated view of the data
- Receives and translates application requests into operations required to fulfill the requests Robust
- Hides database's internal complexity from the application programs and users
  - Abstraction

Types of Databases

- **Centralized database**: Data is located at a single site Trading Platform (Live Data)
- **Distributed database**: Data is distributed across different sites Hadoop, Spark
- **Cloud database**: Created and maintained using cloud data services that provide defined performance measures for the database
  AWS

Advantages of the DBMS

- Better data integration and less data inconsistency
  - **Data inconsistency**: Different versions of the same data appear in different places
- Increased end-user productivity
- Improved:
  - Data sharing
  - Data security
  - Data access
  - Decision making
    - **Data quality**: Promoting accuracy, validity, and timeliness of data

Types of Databases

- **General-purpose databases**: Contains a wide variety of data used in multiple disciplines
- **Discipline-specific databases**: Contains data focused on specific subject areas

Types of Databases

- **Operational database**: Designed to support a company's day-to-day operations
- **Analytical database**: Stores historical data and business metrics used exclusively for tactical or strategic decision making
  - **Data warehouse**: Stores data in a format optimized for decision support

Types of Databases

- **Online analytical processing (OLAP)**
  - Enable retrieving, processing, and modeling data from the data warehouse
- **Business intelligence**: Captures and processes business data to generate information that support decision making

Types of Data[illegible]s

- **Unstructured data**: It exists in their original state
- **Structured data**: It results from formatting
  - Structure is applied based on type of processing to be performed
- **Semistructured data**: Processed to some extent
- **Extensible Markup Language (XML)**
  - Represents data elements in textual format

Data Redundancy Implications

- Poor data security
- Data inconsistency
- Increased likelihood of data-entry errors when complex entries are made in different files
- **Data anomaly**: Develops when not all of the required changes in the redundant data are made successfully

Table 1.2 - Basic File Terminology

FILE SYSTEM

TERM	DEFINITION
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data have little meaning unless they have been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Types of Data Anomaly

- Update Anomalies
- Insertion Anomalies
- Deletion Anomalies

Problems with File System Data Processing

- Lengthy development times
- Difficulty of getting quick answers
- Complex system administration
- Lack of security and limited data sharing
- Extensive programming

Disadvantages of Database Systems

- Increased costs
- Management complexity
- Maintaining currency
- Vendor dependence
- Frequent upgrade/replacement cycles

Data Redundancy

- Unnecessarily storing same data at different places
- **Islands of information**: Scattered data locations
  - Increases the probability of having different versions of the same data