Topic 02: Types of data and Data types

Types of Data:

In data management, there are three main types of data: **Structured Data**, **Unstructured Data**, and **Semi-Structured Data**. These classifications help in organizing and analyzing data based on its format and structure.

1. Structured Data

Structured data is highly organized and easily stored in traditional relational databases or spreadsheets. It follows a predefined model (e.g., rows and columns in a table) and is easily searchable.

- **Characteristics**:
  - Data is in a fixed format (e.g., tables, rows, columns).
  - Easily stored, queried, and processed.
  - Can be represented in relational databases (RDBMS) like SQL.
- **Examples**:
  - **Databases**: Customer information, transaction records.
  - **Data Formats**: CSV, Excel sheets, SQL databases.

2. Unstructured Data

Unstructured data does not have a predefined data model. It is not stored in an easily searchable format, making it more challenging to process and analyze. It often requires advanced techniques such as natural language processing (NLP) or machine learning for analysis.

- **Characteristics**:
  - No fixed structure or organization.
  - It can be text-heavy or multimedia content.
  - Requires processing to derive meaningful information.
- **Examples**:
  - **Text**: Emails, social media posts, articles, blogs.
  - **Multimedia**: Images, videos, audio files, raw logs.

3. Semi-Structured Data

Semi-structured data does not fit neatly into a table format, but it contains some structure, such as tags or metadata. It provides flexibility but still allows for some organization and easier analysis compared to unstructured data.

- **Characteristics**:
  - Contains elements like tags, labels, or other markers for structure.
  - Can be stored in formats like XML, JSON, or NoSQL databases.
- **Examples**:
  - **Data Formats**: XML, JSON, NoSQL databases (e.g., MongoDB).
  - **Data**: Web logs, metadata, configuration files.

Data Types in SQL:

SQL supports various data types, allowing users to define the type of data stored in each column of a database table. These data types can be broadly categorized into several categories, such as numerical types, text types, date/time types, and Boolean types.

1. String/Text Data Types

These data types are used for storing textual or string-based data.

- **CHAR**: Fixed-length string (e.g., CHAR(10) for a string of exactly 10 characters).
- **VARCHAR**: Variable-length string (e.g., VARCHAR(255) for strings of up to 255 characters).
- **TEXT**: Used for larger texts (can store long-form text, such as paragraphs or large documents).

2. Numerical Data Types

These data types are used for storing numbers, either whole numbers or decimals.

- **INT** (INTEGER): Used for whole numbers (e.g., INT can store values like 1, -500, 1000).
- **BIGINT**: A larger integer for storing very large numbers.
- **DECIMAL** or **NUMERIC**: Fixed-point numbers for precise storage of numbers with decimals (e.g., DECIMAL(10, 2) can store numbers like 12345678.99).
- **FLOAT** or **REAL**: Approximate floating-point numbers (e.g., FLOAT can store values like 3.14, 0.0001, 150.555).
- **DOUBLE**: Double-precision floating-point numbers for higher precision.

3. Date/Time Data Types

These are used for storing date and time-related data.

- **DATE**: Stores the date in the format YYYY-MM-DD (e.g., 2024-12-22).
- **TIME**: Stores the time in the format HH:MM:SS (e.g., 14:30:00).
- **DATETIME** or **TIMESTAMP**: Stores both the date and time (e.g., 2024-12-22 14:30:00).
- **YEAR**: Stores the year as a four-digit value (e.g., 2024).

4. Boolean Data Type

A Boolean data type stores logical values, which are typically represented as TRUE or FALSE. In SQL, this is often represented as BOOLEAN or using BIT (with 0 as false and 1 as true).

- **BOOLEAN** or **BIT**: Stores truth values, typically TRUE (1) or FALSE (0).

Example of Data Types in SQL:

Here's how you might define a table in SQL with various data types:

```
CREATE TABLE Employee (
    EmployeeID INT PRIMARY KEY,             -- Integer for employee ID
    FirstName VARCHAR(50),                  -- Variable-length string for first name
    LastName VARCHAR(50),                   -- Variable-length string for last name
    BirthDate DATE,                         -- Date type for birth date
    Salary DECIMAL(10, 2),                  -- Decimal for salary
    HireDate DATETIME,                      -- Datetime for hire date
    IsActive BOOLEAN                        -- Boolean for active status (TRUE or FALSE)
);
```

Summary of Key Data Types:

Data Type	Category	Description
CHAR(n)	String/Text	Fixed-length string.
VARCHAR(n)	String/Text	Variable-length string.
TEXT	String/Text	Large variable-length text data.
INT	Numeric (Whole Number)	Integer (whole numbers).
DECIMAL	Numeric (Decimal)	Fixed-point number for precise values.
FLOAT/REAL	Numeric (Decimal)	Approximate floating-point number.
DATE	Date/Time	Date without time.
TIME	Date/Time	Time of day without date.
DATETIME	Date/Time	Date and time together.
BOOLEAN	Boolean	Logical true/false value (1/0).

This table structure helps organize data in SQL databases and ensures efficient data processing and querying.