Chapter 01 : Introduction to the database

1. ***Data, Databases, and DBMS**:*

   The first step is to understand some terminologies:

   What is data? It is any kind of information, such as, for example: here is a person named Jean. It is also a relationship between pieces of information: Jean teaches databases. Relationships of this kind define structures. A database is a set, generally large, of such information, stored permanently. Hence the definition:

***Definition 1.1**: A database is a large set of structured information stored on a permanent medium.*

Why we use databases instead of directly using files?

The direct use of files raises significant issues:

- ***Difficulty of data access:** In practice, for each access, even the simplest, a program would need to be written.*
- ***Lack of security:** If every programmer can directly access the files, it becomes impossible to guarantee the security and integrity of the data.*
- ***No concurrency control:** In an environment where multiple users access the same files, concurrency issues arise.*

Hence the use of software responsible for managing the files that make up a database, handling protection and security features, and providing the various types of interfaces necessary for data access. This software (the DBMS for DataBase Management System) is very complex and forms the main topic of this course.

In particular, one of the main tasks of the DBMS is to hide from the user the complex and tedious details related to file management. Hence the definition:

***Definition 1.2:** A Database Management System (DBMS) is high-level software that allows the manipulation of information stored in a database.*

The architecture of a DBMS consists of three levels:

a) ***External Level (or User Level)**:*

This level corresponds to the view of the data that each user or application has. It defines how data is presented to the end user, without worrying about how it is actually stored. It allows adapting data to the specific needs of users.

b) ***Conceptual Level**:*

This level represents the global structure of the database. It describes how data is organized without focusing on its physical storage. It includes the relationships between different entities and the associated constraints, while hiding implementation details.

The Data Definition Language (DDL) is used to specify the conceptual schema:

- *Definition and naming of object categories,*
- *Definition and naming of relationships between objects,*
- *Expression of constraints on the data.*

c) ***Internal Level (or Physical Level)**:*

This level describes how data is actually stored on the physical medium (hard drive, SSD, etc.). It includes technical details such as file management, indexing, and how data is optimized for performance and quick access.

This three-level architecture ensures data independence, meaning a change at the physical level does not affect the logical and external levels, and vice versa. This facilitates the management, security, and maintenance of databases.

2. What should we know to use a DBMS?

The use of a DBMS requires understanding (and thus knowing how to use) the following functionalities:

- *Defining the data schema using the DBMS's data models.*
- *Operations on data: searching, updating, etc.*
- *Sharing data among multiple users (Transaction mechanisms).*
- *Optimizing performance by adjusting the physical organization of the data. This aspect is more related to administration of the database.*

***2.1 Definition of the data schema**: A schema is simply the description of the data contained in the database. This description follows a data model that provides tools for describing (structures, constraints, and operations). In fact, in a DBMS, there are several models, more or less abstract, of the same objects, for example:*

- ***The conceptual model**: the description of the information system.*
- ***The logical model**: the interface with the DBMS, for example relational model where data are represented by relation.*
- ***The physical model***: This model concerns the physical management of storage space.

In this course, we focus on the example of the most common conceptual model: the Entity/Relationship model. It is essentially, a very abstract description that offers the following advantages:

- *Analysis of the real world.*
- *Design of the information system.*
- *Communication between different actors in the organization.*

2.2 The problems to be addressed by the DBMS :

Managing a database poses complex problems. This management is handled by Database Management Systems (DBMS).

- ***Organization of Data**: The DBMS organizes data into permanent tables stored on disk; it creates mechanisms to ensure quick access to the data; it informs users about these structures.*
- ***Data Management**: The DBMS ensures the consistent evolution of data; it checks that constraints (e.g., uniqueness, referential integrity between tables, etc.) are respected.*

- ***Access to Data****: The DBMS allows access to data both by occasional users and by data processing programs.*
- ***Access Control****: The DBMS ensures that only authorized users and software can access and modify the data.*
- ***Managing Concurrent Access****: Multiple users must be able to access the same data simultaneously. The DBMS must be able to:*

  - ✓ *Handle conflicts during concurrent access for updates.*
  - ✓ *Provide a rollback mechanism if modifications in progress need to be undone.*
  - ✓ *Ensure a consistent view of the data if one user is querying and another is updating.*

The goal*: To avoid deadlocks while preventing chaotic modifications.*