**LESSON 1 - DATABASE CONCEPTS**

1. **WHAT IS A DATABASE?**

**1.1. DEFINITION.**

A database is a collection of data that are logically interrelated. The data are structured according to a model of databases that reflects the relationships and constraints of these data in the real world. The description and definition of the data are stored in the same database.

The treatments that are performed with the data must preserve the integrity and security of the database.

* 1. **ADVANTAGES OF DATABASE SYSTEMS WITH RESPECT TO THE FILE SYSTEMS.**
* **Insulation between programs and data.**

The traditional files systems are called "process-oriented Systems", since the data are stored in files designed specifically for each application, so that if a datum is used by more than one application, it appears repeated in two or more files. In addition, if you modify the treatment that performs a process on the data, it is necessary to restructure the files. Therefore, in these systems data are dependent on the treatments carried out on them.

The database systems are called "data-oriented Systems", since the data is structured according to a model that reflects the characteristics they have in the real world, and that model allows any process to treat the data. In these systems, the data are independent of the treatments carried out on them. (Integrated Information Systems).

* **Centralized data description. Self-describing data.**

The definition of the data is stored in the same database, so the treatments do not have to specify the definition of data. In addition, this offers a better and more standardized documentation of information.

* **Efficiency in the structure of data.**

As the data are structured according to a model, data redundancies are not produced. That is to say, a same datum does not appear in two or more files. Three problems are avoided:

* On the one hand, there's no wasted space in memory by storing the same information several times.
* Process saves time by not having to modify the data in several files.
* It avoids the corruption of the database since modifying a datum we would update all files in which that appears, and if not changed in any of the files, the information in the database would be inconsistent.
* **Reduction of disk storage space.**

The database systems optimize the space required to store data on disk, up to 50% more than traditional file systems.

* **Higher level equipment.**

The database systems offer more powerful and simpler tools for the manipulation of data. The database systems are integrated into fourth generation development environments. The file systems are third generation development environments.

1. **COMPONENTS OF A DATABASE SYSTEM**

* **The Data**

Is the set of structured data stored on a magnetic or optical media for direct access. This data set is what is often referred to as databases.

* **The Metabase**

Is information about the data in the database. It is essential so that the self-documented system may provide to users, the administrator and Database Management System (DBMS), the information they need about the data stored and its behavior.

The main elements of the Metabase are:

* **Catalog or Data Dictionary**: Logical Description (that is, from the point of view of the user) of each of the datum: type, file in which it is located, relationships between files, view, etc...
* **Data directory**: Physical Description that allows to move from the external to the internal representation.
* **The Security system**: It is on two levels.
* **Logical Security**: Control of users’ access to the data.
* **Physical Security**: consists of recording in some files called LOG or diary, operations performed on the database with the aim of being able to undo them in case of errors.
* **The System Management Software (logical).**

Is the set of programs, procedures, languages etc, capable of performing all the treatments required for the operation of the system.

* **The DBMS**: is the core of a database system. Allows the user to create, update and retrieve the data contained in the database. That is to say, it interprets user commands and executes them physically.
* The data definition language (**DDL**) and data manipulation language (**DML**) (for example SQL contains instructions of definition and manipulation).
* **Administration Tools:** database administration tools that facilitate the maintenance and security of the database. These programs perform tasks such as: modification of the size of the files, obtaining the usage statistics, loading files, obtaining copies, recovery after a crash, etc.
* **Database administrator (DBA)**.

Is the person whose role is to ensure the quality of data and allows their use, designing the database so that the design of its logical structure and physical implementation responds as well as possible to the needs of different users.

1. **CHARACTERISTICS OF THE DATABASE MANAGEMENT SYSTEMS.**

The database systems provide mechanisms to protect data against failures that alter and corrupt them. (This is the main difference with traditional systems of files).

The Failures that can affect the data are:

* Physical: Failures in the physical RAM or disk, system crashes, power interruption,
* Logical: Failures in the programs, operating system or the DBMS.
* Human errors: intentionally or unintentionally.

To protect the data, the DBMS has to have three features: Security, Integrity and Confidentiality.

**Security**

It is retrieving a **consistent state** of the data when faced with failures that destroy then wholly or partially.

A very common cause of the inconsistency of the database is when a logical operation translates into several actions in the database, in such a way that the operation only makes sense when you perform all actions. (Example of the two files that store an invoice: Header and lines). This problem is solved by the DBMS with transactions.

A **transaction** is a sequence of operations that have to be executed atomically, that is to say, *or performs all operations that make up the transaction or none*. When it is necessary to undo the actions of an unfinished transaction the system is based on the dietary or Log file which is a file in which the system is recording all operations that are performed on the database.

Another type of failure that can occur is one that affects memory, in this case the information would be recovered from a backup that restores the state of the data on a certain date.

**Integrity**

The aim of the integrity is to ensure that the data contained in the database are correct. To do this the DBMS must be able to detect and avoid the incorrect operations (those that introduce inconsistencies in the database).

There are two types of operations that can affect the integrity of the data:

* The operations that violate the rules of the integrity. (Existence and Referential).
* The interference by concurrent accesses.

In the first case the integrity constraints are defined at the time of the design of the database so that the DBMS do not allow operations that violate them.

In the case of concurrent control to data, the system solves it by the mechanism of locks. (Example of movie tickets for several terminals. Also problem of deadly embrace).

**Confidentiality**

The Confidentiality refers to the protection of data against access by unauthorized users. Decisions on which users should be allowed to perform operations on objects are company policy decisions, are not technical issues.

To make access control possible, it is necessary that users of the system must be identified by a login and password, and privileges, permissions and authorizations have to be registered in the database.

Systems that require great safety, use cryptography systems in such a way that the data are recorded on the basis of encrypted data, being necessary to know the key to decrypt them.

Finally, another mechanism to maintain security in the confidentiality is to facilitate the auditory, recording all user access to data in some auditor files, being able to detect unauthorized access *a posteriori*, and consequently to improve the safety system.

1. **The CLIENT-SERVER Architecture.**

A database system has a structure composed of two parts:

* A server, also called the backend.
* A set of clients or frontends.

The server offers services to perform physical operations on the data, that is to say that the server is the own S.G.B.D.

The clients are applications executed around the DBMS which perform multiple operations, but when access to data is required they communicate with the server that serves the requirements and returns the results.

There are several topologies for deploying the client-server architecture:

* The client and server are running on the same machine. The server and clients are running on different machines. This architecture is called "distributed" process that requires a communications network.
* There are several servers and each of them contain part of the database. To this architecture is called "distributed database".
* Finally, there is the client/server architecture to three levels to be used for World Wide Web applications both on intranets and the Internet. Here the server is installed on a machine, but the client portion is divided into two levels: a machine application server and a client machine. (Most of the application process is carried out in a centralized manner on the application server, leaving for the client machines the part of the application that acts as a user interface).

1. **Data models.**

The aim or goal of any information system is to represent all the information of the real world necessary for the fulfilment of its purposes, using an **abstraction** (to obtain invoices, stocks of warehouse, etc...)

This abstraction is carried out using a data model that is a set of methods and rules that indicate which information is stored and how to manipulate data.

There are two types of models:

* Conceptual model, which is a representation of reality and is not committed to any computing environment. Entity-relationship model of Chen.
* Logical models, which determine the criteria for the storage and data manipulation operations within a type of computing environment.
* The DBMS are based on a logical model specific.

Historically the logical models of data are:

* Hierarchical model: IMS from IBM.
* Web model: Codasyl.
* Relational Model: Oracle, Informix, DB II, Ingres, VisualDbase, Access...
* Object Oriented Model: Oracle.