**UD1: INTRODUCCIÓN A LAS BASES DE DATOS**

**1 DATOS Y ARCHIVOS**

**1.1- La necesidad de gestionar la información**

La escritura fue la herramienta que permitió al ser humano almacenar información y realizar cálculos sobre ella. Además de permitir compartir esa información entre diferentes personas, también posibilitó que los datos se

guardaran de manera continua e incluso estuvieran disponibles para las siguientes generaciones.

**El propio nombre Informática hace referencia al hecho de ser una ciencia que trabaja con información.**

Por ello las bases de datos son una de las aplicaciones más antiguas de la informática.

**1.2- Datos e información**

**Un dato es una propiedad en crudo, es decir, sin contextualizar.**

Por ejemplo, Sánchez y 32 son datos.

Si decimos que Sánchez es mi primer apellido y que 32 son los grados centígrados de temperatura que hace ahora en la calle, esos datos pasan a ser información.

La información tiene una importancia humana, es interpretable, reconocible, presentable... El dato es simplemente el paso previo. Resumiendo: todo dato debe de ser procesado para obtener información, y es la información lo que nos interesa a los seres humanos.

Los datos se convierten en información gracias a los **metadatos: datos que**

**sirven para describir otros datos.**

**1.3- Sistemas de información**

**Un sistema se define como "un conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objeto".**

**1.3.1- Sistemas de información**

**Los sistemas que se dedican a la gestión y explotación de la información se les conoce como sistemas de información**, y está compuesto por los siguientes elementos:

* **Recursos físicos:** carpetas, documentos, equipamiento, discos, ...
* **Recursos humanos:** personal que maneja la información.
* **Protocolo:** normas que se deben cumplir al manejar la información (pensemos por ejemplo en una plantilla o modelo para un documento).

**1.3.2- Sistemas de información electrónico**

**Se trata de un sistema de información en el que el manejo de la información se realiza a través de dispositivos electrónicos (o informáticos).** Sus componentes son:

* **Datos.** Se trata de la información relevante que almacena y gestiona el sistema de información. Ejemplos de datos son: Sánchez, 12764569F, Calle Mayo 5, Azul...
* **Hardware.** Equipamiento físico que se utiliza para gestionar los datos. cada uno de los dispositivos electrónicos que permiten el funcionamiento del sistema de información: servidores, máquinas de los clientes,routers, switches, impresoras...
* **Software.** Aplicaciones informáticas que se encargan de la gestión de la base de datos y de las herramientas que facilitan su uso.
* **Recursos humanos.** Personal que maneja el sistema de información.

**2 TIPOS DE GESTIÓN MEDIANTE EL ORDENADOR**

**2.1- Sistemas de gestión de ficheros**

**2.1.1- ¿Qué es un fichero?**

**Fichero o archivo: conjunto de información relacionada, tratada como un todo y organizada de forma estructurada. Es una secuencia de dígitos binarios que organiza información relacionada con un mismo aspecto.**

Normalmente, los ficheros están formados por registros lógicos que contienen los datos relativos a un mismo elemento u objeto (por ejemplo, los datos de un alumno matriculado en un colegio). A su vez, los registros están divididos en campos que contienen los datos elementales que conforman el registro: nombre, dirección, teléfono, email,...

**2.1.2- Tipos de ficheros**

Según la forma en la que se accede a los ficheros, podemos encontrar algunos tipos, de entre los que destacamos los siguientes:

* **Secuenciales:** se caracteriza porque sus registros están almacenados de forma contigua, por lo que la única forma de acceder a él es leyendo un registro tras otro desde el principio hasta el final. Se utiliza en cintas magnéticas o CD/DVD.
* **Ficheros de acceso directo:** en este tipo se puede acceder a un registro indicando la posición del mismo, o a través de una clave que forma parte del registro como un campo más. Lo más habitual es utilizar como índice la posición física del disco donde está el inicio del registro. Así se puede localizar fácilmente sin recorrer el fichero secuencialmente.
* **Ficheros indexados:** se basa en el uso de índices, que permite el acceso totalmente libre. Estos índices son similares a los de un libro; si queremos leer un capítulo concreto, podemos consultar el índice y nos dice la página en la que comienza. Por tanto, dentro del fichero tendremos una zona de registros y una zona de índices. Usualmente, los CD/DVD funcionan de esta forma.

Otros tipos de ficheros serían: secuenciales indexados (también conocidos como parcialmente indexados), ficheros de acceso calculado o hash…

**2.1.3- Sistemas de gestión de ficheros**

La idea es que los datos se almacenan en ficheros y se crean aplicaciones (cuyo código posee la empresa que crea dichas aplicaciones) para acceder a los ficheros. Cada aplicación organiza los datos en los ficheros como le parece mejor y si incorporamos aplicaciones nuevas, estas usarán sus propios ficheros.

Cada aplicación almacena y utiliza sus propios datos de forma un tanto caótica. **La ventaja de este sistema (la única ventaja), es que los procesos son independientes por lo que la modificación de uno no afecta al resto.**

**Inconvenientes:**

* **Programación de aplicaciones compleja.** Ya que los programadores se deben de encargar de lo que tiene que hacer la aplicación y además de estructurar los datos en disco.
* **Datos redundantes.** Ya que se repiten continuamente. Podría, por ejemplo, ocurrir que una segunda aplicación utilice datos personales, que resulta que ya estaban almacenados en los ficheros de una primera aplicación, pero como ambas son independientes, los datos se repetirán.
* **Datos inconsistentes.** En relación con el problema anterior, ya que un proceso cambia sus datos y no los del resto. Por lo que la misma información puede tener distintos valores según qué aplicación acceda a él.
* **Difícil acceso a los datos.** Cada vez que se requiera una consulta no prevista inicialmente, hay que modificar el código de las aplicaciones o incluso crear una nueva aplicación. Esto hace imposible pensar en nuevas consultas e instantáneamente obtener sus resultados; inviable para aplicaciones que requieren grandes capacidades de consultas y análisis de datos.
* **Coste de almacenamiento elevado.** Al almacenarse varias veces el mismo dato, se requiere más espacio en los discos. Además, las aplicaciones también ocupan mucho al tener que pensar en todas las posibles consultas sobre los datos que la organización precisa.
* **Dependencia de los datos a nivel físico.** Para poder saber cómo se almacenan los datos, es decir qué estructura se utiliza de los mismos, necesitamos ver el código de la aplicación; es decir el código y los datos no son independientes.
* **Dificultad para el acceso simultáneo a los datos.** El acceso simultáneo requiere que varios usuarios la puedan acceder a la misma información. Con este tipo de sistemas es extremadamente difícil conseguir esta capacidad.
* **Dificultad para administrar la seguridad del sistema.** Ya que cada aplicación se crea independientemente. Es, por tanto, muy difícil establecer criterios de seguridad uniformes. Es decir, los permisos que cada usuario tiene sobre los datos, se establecen de forma muy confusa (y nada uniforme ya que cada aplicación puede variarla seguridad).

**2.2- Sistemas de Bases de Datos**

**Una base de datos es una colección de datos relacionados lógicamente entre sí, con una definición y descripción comunes y que están estructurados de una determinada manera. Es un conjunto estructurado de datos que representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos eficientemente por parte de varias aplicaciones y usuarios.**

La base de datos no sólo contiene los datos de la organización, también almacena una **descripción de dichos datos**. Esta descripción es lo que se denomina **metadatos**, se almacena en el diccionario de datos o catálogo y es lo que permite que exista independencia de datos lógica-física.

En este tipo de sistemas, los datos se centralizan en una **base de datos** común a todas las aplicaciones. Un software llamado **Sistema Gestor de Bases de Datos (SGBD)** es el que realmente accede a los datos y se encarga de gestionarlos. Las aplicaciones que crean los programadores, no acceden directamente a los datos, de modo que la base de datos es común para todas las aplicaciones.

De esta forma, hay, al menos, dos capas a la hora de acceder a los datos. Las aplicaciones se abstraen sobre la forma de acceder a los datos, dejando ese problema al SGBD. Así se pueden concentrar exclusivamente en la tarea de conseguir una interfaz de acceso a los datos para los usuarios.

**Ventajas:**

* **Independencia de los datos y los programas.** Esto permite modificar los datos sin modificar el código de las aplicaciones y viceversa.
* **Menor redundancia.** Este modelo no requiere que los datos se repitan para cada aplicación que los requiera, en su lugar se diseñan los datos de forma independiente a las aplicaciones. Los programadores de aplicaciones deberán conocer la estructura creada para los datos y la forma en la que deben acceder a ellos.
* **Integridad de los datos.** Al estar centralizados, es más difícil que haya datos incoherentes. Es decir, que una aplicación muestre información distinta al resto de aplicaciones, ya que los datos son los mismos para todas.
* **Mayor seguridad en los datos.** El SGBD es el encargado de la seguridad y se puede centrar en ella de forma independiente a las aplicaciones. Como las aplicaciones deben atravesar la capa del SGBD para llegar a los datos, no se podrán saltarla seguridad.
* **Visiones distintas según el usuario.** Nuevamente, centralizarlos datos facilita crear políticas que permitan que los usuarios vean la información de la base de datos de forma distinta.
* **Datos más documentados.** Las bases de datos tienen mucho mejor gestionados los metadatos, que permiten describirla información de la base de datos y que pueden ser consultados por las aplicaciones.
* **Acceso a los datos más eficiente.** Esta forma de organizarlos datos produce un resultado más óptimo en rendimiento ya que los sistemas gestores centralizan el acceso pudiendo ejecutar políticas diferentes en función de la demanda.
* **Menor espacio de almacenamiento.** Puesto que hay muy poca redundancia.
* **Acceso simultáneo a los datos.** Nuevamente el SGBD tiene más capacidad de conseguir esto. Cuando hay varias aplicaciones que intentan acceder a los datos en los sistemas orientados a los ficheros, compiten por los datos y es fácil el bloqueo mutuo. En el caso de los sistemas orientados a bases de datos, toda petición pasa la capa del SGBD y esto permite evitarlos bloqueos.

**Desventajas**

* **Instalación costosa.** El control y administración de bases de datos requiere de un software y hardware poderoso.
* **Requiere personal cualificado.** Debido a la dificultad de manejo de este tipo de sistemas.
* **Implantación larga y difícil.** En relación a los puntos anteriores. La adaptación del personal y del equipamiento es mucho más complicada y lleva bastante tiempo.
* **Ausencia de estándares totales**. Lo cual significa una excesiva dependencia hacia los sistemas comerciales del mercado. Aunque, hoy en día, hay un funcionamiento base y un lenguaje de gestión (SQL) que desde hace tiempo se considera estándar (al menos en las bases de datos relacionales).

**3 ¿CÓMO FUNCIONA UN SISTEMA GESTOR DE BASES DE DATOS?**

**3.1- Funciones de un SBGD**

**3.1.1- A nivel de definición o descripción**

Permite al diseñador de la base de datos crear las estructuras apropiadas para integrar adecuadamente los datos. Aunque la función de definición lo que hace es gestionar los metadatos. **Los metadatos son la estructura de la que dispone el sistema de base de datos para documentar cada dato. Los metadatos también son datos que se almacenan en la propia base de datos; pero su finalidad es describir los datos.**

**Un metadato nos permite saber a qué información real se refiere cada dato.** Por ejemplo: Sánchez, Rodríguez y Crespo son datos. Pero Primer Apellido es un metadato que, si está correctamente creado, nos permite determinar que Sánchez, Rodríguez y Crespo son primeros apellidos.

**Los metadatos son la base de la creación de las bases de datos.**

Un lenguaje conocido como **lenguaje de descripción datos** o **DDL**, es el que permite realizar la función de definición en las bases de datos.

**3.1.2- A nivel de manipulación**

Permite cambiar y consultar los datos de la base de datos. Se realiza mediante un **lenguaje de modificación de datos** o **DML**. Mediante este lenguaje se puede:

* Añadir datos
* Eliminar datos
* Modificar datos
* Consultar datos

La función de consulta se realiza con un **lenguaje de consulta de datos** o **DQL (Data Query Language)**.

**3.1.3- A nivel de control**

Mediante esta función los administradores poseen mecanismos para proteger los datos. Es la función encargada de establecer los permisos de acceso a los elementos que forman parte de la base de datos.

El lenguaje que implementa esta función es el **lenguaje de control de datos** o **DCL**.

**3.2- Utilidades de un SBGD**

**Una base de datos sería el conjunto de los datos y metadatos, mientras que el sistema gestor de base de datos sería el conjunto de herramientas que nos van a permitir definir, manipular y controlar nuestra base de datos.**

**Los SGBD proporcionan un conjunto coordinado de programas, procedimientos y lenguajes que permiten a los distintos usuarios realizar sus tareas habituales con los datos, garantizando además la seguridad de los mismos.**

Herramientas que proporcionan los SBGD:

* **Herramientas para la creación y especificación del diccionario de datos.** El diccionario de datos es la estructura de la base de datos que almacena los metadatos. Es decir el diccionario de datos contiene la descripción de todos los datos de la base de datos.
* **Herramientas para administrar y crear la estructura física de la base de datos.** El SGBD proporciona herramientas para especificar la forma en la que se almacenarán los datos en la computadora (o computadoras) que alojen la base de datos. Estas herramientas nos permitirán diseñar una forma de almacenamiento centrada en optimizar el acceso a los datos.
* **Herramientas para la manipulación de los datos.** Nos permitirán añadir, modificar, suprimir o consultar datos (función de manipulación) de la forma más sencilla posible.
* **Herramientas de recuperación en caso de desastre.** Si ocurre un mal funcionamiento del sistema, un fallo en la alimentación del sistema, errores de red, etc. En ese caso los buenos SGBD poseen y proporcionan mecanismos para que se recupere la máxima información posible y se asegure su integridad.
* **Herramientas para la creación y restablecimiento de copias de seguridad.** Es una de las tareas fundamentales, ya que permite recuperarla información en caso de pérdida de datos.
* **Herramientas para la gestión de la comunicación de la base de datos.** Encargadas de configurar el hardware y software de conexión a la red. Así como los mecanismos necesarios para configurar adecuadamente el software que se encarga de recibir y comunicarlas peticiones de los clientes.
* **Herramientas para la creación de aplicaciones de usuario.** Es decir, herramientas para los programadores de aplicaciones, los cuales crean el software con el que los usuarios accederán de forma cómoda a la base de datos.
* **Herramientas de instalación y configuración de la base de datos.**
* **Herramientas para la exportación e importación de datos a o desde otros sistemas.**
* **Herramientas para gestionar la seguridad.** Permiten establecer privilegios y permisos diferentes para los usuarios, así como impedir el acceso no deseado (función de control).

**3.3- Niveles de abstracción de una base de datos**

En cualquier software siempre hay dos puntos de vista: la que tienen los usuarios finales, llamada visión o **nivel externo**; y la que tienen los creadores del software, llamada **nivel interno**.

En el caso de una base de datos hay más niveles, más formas diferentes de observar la base de datos. Cada nivel se corresponde con los diferentes usuarios que pueden manejar la base de datos. A cada nivel se le llama nivel de abstracción, que nos permite abstraernos de detalles concretos para entender la base de datos sin más complejidad de la necesaria.

* **Nivel físico.** Nos permite saber la forma en la que está almacenada la base de datos. Por ejemplo en qué discos duros, qué archivos utiliza, de qué tipo son los archivos, bajo qué sistema operativo,... Este nivel es el que está más cercano a la visión de la base de datos que posee la computadora, por lo que es absolutamente dependiente del hardware y el software (especialmente del Sistema Operativo).
* **Nivel interno.** Un poco más cercano a la visión que tenemos las personas. Permite observar la base de datos como un conjunto de estructuras que relacionan la información humana con la información digital. A este nivel no se depende del hardware concreto que tengamos; es decir, no se habla de discos, servidores, archivos,... sino de las estructuras que disponemos en nuestro SGBD en particular para organizar los datos.
* **Nivel conceptual.** Es el nivel de mayor abstracción y el más importante. Se trata de una visión organizativa de los datos independiente tanto del hardware como del software que tengamos. Es el plano o modelo general de la base de datos y a este nivel es al que trabajan las o los analistas y diseñadores cuando crean el primer esquema de la base de datos. En ningún momento queda influido por el SGBD en particular que usemos.
* **Nivel externo.** Se trata de la visión de los datos que poseen los usuarios y usuarias finales de la base de datos. Esa visión es la que obtienen a través de las aplicaciones. Las aplicaciones creadas por los desarrolladores abstraen la realidad conceptual de modo que el usuario no conoce las relaciones entre los datos, como tampoco conoce dónde realmente se están almacenando los datos. Es la forma en la que cualquier persona desea manejar una base de datos a través de formularios, informes, listas,...

Las dos columnas que dividen algunos de estos niveles reflejan dos fronteras a tener en cuenta:

* **Independencia Física (Nivel físico - interno).** La da la barrera entre el esquema físico y el interno e indica que el esquema interno es independiente del hardware concreto que usemos.
* **Independencia Lógica (Nivel interno - conceptual).** Los esquemas de los niveles conceptual y externo son independientes del software concreto de base de datos que usemos; no dependen en absoluto de él. Por ello estos esquemas nos valdrían para cualquier SGBD que utilicemos.

**3.4- Diferentes tipos de usuarios de una base de datos**

Los **informáticos**, lógicamente, son los profesionales que definen y preparan la base de datos. Pueden ser:

* **Directivos/as.** Organizadores y coordinadores del proyecto a desarrollar y máximos responsables del mismo..
* **Analistas.** Son los encargados de controlar el desarrollo de la base de datos aprobada por la dirección. Dirigen a los desarrolladores y operadores. Normalmente son, además, los diseñadores de la base de datos: es decir, crean el esquema conceptual de la misma.
* **Administradores/as de las bases de datos.** Encargados de crear el esquema interno de la base de datos. También gestionan el correcto funcionamiento del SGBD. Sus tareas incluyen la planificación de copia de seguridad, gestión de usuarios y permisos, optimización del rendimiento, monitorización de problemas y creación de los objetos de la base de datos.
* **Desarrolladores/as o programadores/as.** Encargados de la realización de las aplicaciones de usuario para que estos accedan a la base de datos.
* **Equipo de mantenimiento.** También se les llama operadores. Encargados de dar soporte a los usuarios en el trabajo diario (suelen incorporar además tareas administrativas como la creación de copias de seguridad por ejemplo o el arreglo de problemas de red por ejemplo).

Por otro lado, estarían los **usuarios**:

* **Expertos/as.** Realizan operaciones avanzadas sobre la base de datos. Normalmente conocen el lenguaje de manipulación de datos (DML) para acceder a la base de datos. Son usuarios, por lo tanto, con conocimientos informáticos que se encargan en las empresas de los clientes de algunas acciones más complejas sobre la base de datos que las que realizan los usuarios habituales.
* **Habituales.** Utilizan las aplicaciones creadas por los desarrolladores para consultar y actualizar los datos. Son los que trabajan en la empresa a diario con estas herramientas y el objetivo fundamental de todo el desarrollo de la base de datos.
* **Ocasionales.** Son usuarios que utilizan un acceso mínimo a la base de datos a través de una aplicación que permite consultar ciertos datos. Serían por ejemplo los usuarios que consultan el horario de trenes a través de Internet. Aunque se les llama ocasionales son el núcleo del trabajo con la base de datos ya que son los que más la utilizan (ya que son sus usuarios más numerosos) y son, por ejemplo, los que visitan la base de datos para realizar compras o para informarse del negocio representado en la base de datos.