**INTRODUCCIÓN A LAS BASES DE DATOS**

1. INTRODUCCIÓN:

Actualmente cualquier organización con un sistema de información automatizado dispone de una Base de Datos (BD) para almacenar y gestionar toda la información de dicha organización. Sin embargo, antes de la aparición de las BD, esto no era así.

Vamos a describir a continuación los primitivos sistemas de almacenamiento de la información,

1.1 SISTEMA DE INFORMACIÓN:

Toda organización, necesita una infraestructura para poder desarrollar sus actividades. Esta estructura organizativa está basada en un conjunto de funciones que hay que desarrollar y que, entre otras, incluye las siguientes:

* Controlar y gestionar los recursos financieros.
* Comercializar los productos o servicios en los que la empresa basa su negocio.
* Fabricar productos o crear servicios que venden en el mercado.

Todas estas funciones tienen que coordinarse entre sí, por ello las organizaciones incluyen una infraestructura para coordinar el flujo y el registro de información necesaria para desarrollar sus actividades de acuerdo a una estrategia de negocio.

Una definición formal de Sistema de Información sería:

Un conjunto formal de procesos, que se dedican a recopilar, transformar y distribuir la información necesaria para las operaciones de una empresa y para las actividades de dirección y control correspondientes.

Los componentes que forman un SI son:

* **Información**: Son los datos que se usan y se adaptan a las personas que los necesitan y a la tecnología disponible en la organización.
* **Procedimientos**: Son las reglas y formas de trabajo que todos siguen para que la organización funcione bien. El SI ayuda a gestionar la información según estas reglas.
* **Usuarios**: Son las personas que usan el sistema, ya sea para ingresar, manejar o utilizar la información para hacer su trabajo.
* **Recursos**: Son las herramientas y equipos, como computadoras y redes, que permiten almacenar, procesar y compartir la información. Es la parte más visible del SI.

1.2 SISTEMAS DE INFORMACIÓN AUTOMATIZADO

Las organizaciones han ido incorporando nuevas tecnologías desde el siglo XX hasta la actualidad para mejorar el rendimiento y la eficacia de los SI. Cuando el SI emplea las TIC en el procesamiento, comunicación y almacenamiento de la información se denomina Sistema de Información Automatizado (SIA).

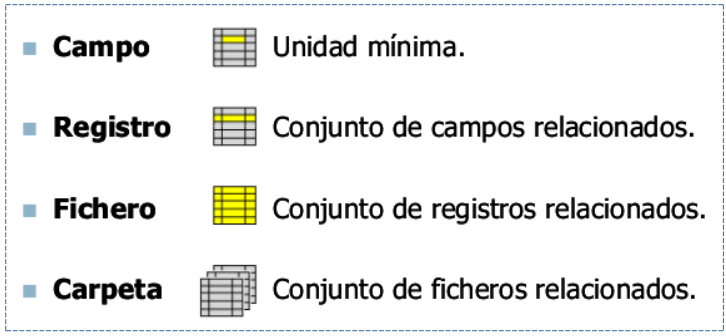
1.2.1 SIA basado en ficheros

Si analizamos la situación pasada de algunas organizaciones nos encontramos con sistemas de información compuestos por diversas aplicaciones software para gestionar la información de la organización. Cada aplicación realiza su tarea y las almacenan en ficheros dentro de las unidades de almacenamiento del ordenador. Este tipo de SI se denomina SI orientado a procesos. Los datos que gestionaba una organización se almacenaban en un conjunto de ficheros.

Cada fichero era generado por la propia aplicación que gestionaba los datos almacenados en el propio fichero, es decir, nos encontramos una fuerte dependencia de los datos con la aplicación que los manipula y gestiona.

1.3 TIPOS DE ARCHIVOS

Si un SIA utiliza varias aplicaciones que emplean información de diversas fuentes, entonces era necesario varios archivos independientes que se almacenaban en una carpeta o directorio. Los ficheros podían ser de distinto tipo en función de la organización de los registros dentro de cada fichero, cada uno de ellos con ventajas e inconvenientes.



Los tipos de archivo según su organización más habituales son los siguientes:

1.3.1 Archivo secuencial

Consiste en un fichero en el que los registros se añaden uno detrás de otro en secuencia. Un archivo secuencial es la forma más simple de almacenar y recuperar registros de un archivo. Siempre se almacenan los registros al final del archivo, después del último. Los registros tienen la misma estructura y por tanto todos son del mismo tamaño.

Las operaciones de actualización son muy costosas. Para ello se usan marcas de registro de borrado, pero las actualizaciones son muy costosas ya que hay que crear un archivo para copiar todos los registros más el modificado.

Entre sus ventajas están:

* Inserciones rápidas y eficiente uso del espacio en disco.
* Acceso rápido cuando se leen todos los registros.
* Fácil de gestionar.

Desventajas:

* Lento si se necesita acceder a registros en orden diferente.
* Ineficiente en dispositivos de acceso aleatorio.
* Modificar registros es lento y vulnerable a fallos del sistema.

1.3.2 Archivo relativo

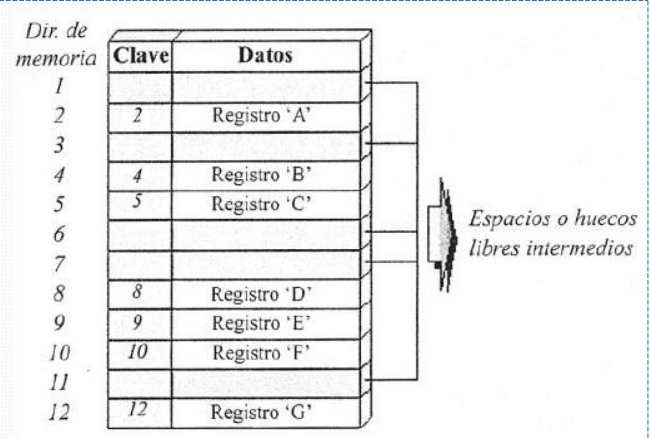
En un archivo con organización relativa el orden de inserción del registro es independiente de la posición en la que se almacena el registro en el archivo. Cuando se crea el archivo se dimensiona para albergar un número limitado de registros, por tanto tienen un tamaño definido que siempre es el mismo. Todo el espacio disponible para el archivo se divide en posiciones numeradas cada una de las cuales es capaz de albergar un registro. Las celdas están numeradas desde el 0 y en orden creciente.

Ventajas de los archivos relativos:

* Acceso rápido a cualquier registro usando el campo clave
* Actualización directa sin archivos temporales
* Operaciones rápidas de inserción
* Modificación y borrado.

Desventajas:

* Desperdicio de espacio si el archivo está sobredimensionado
* Riesgo de colisiones en las direcciones de clave
* Acceso secuencial limitado
* Uso de un solo campo clave.

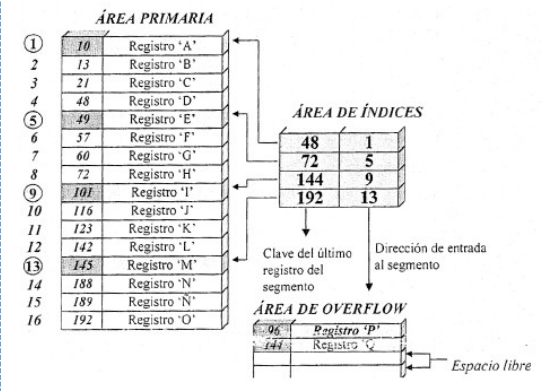


1.3.3 Archivo secuencial indexado

Es una variante de la organización secuencial que permite acceso directo o aleatorio a los registros, además de su acceso secuencial. De nuevo los registros tienen la misma estructura y consecuentemente el mismo tamaño. Sin embargo, en el diseño del registro se indica un campo o combinación de campos que permite identificar a cada registro de forma única, es decir, no hay dos registros con el mismo valor de campo identificador. A este campo se le llama clave.

Un archivo secuencial indexado tiene tres áreas:

* Área de datos: Almacena los registros en orden según la clave.
* Área índice: Tabla que asocia claves con direcciones de registros.
* Área de excedentes: Guarda registros que no caben en el área de datos, manteniendo el orden mediante punteros.



Ventajas de los archivos secuenciales indexados:

* Permiten acceso secuencial y directo a los registros usando la clave.
* Los registros se pueden actualizar en el mismo archivo, sin necesidad de copias.

Desventajas:

* Ocupan más espacio en disco por el uso del área de índices.
* El tiempo de acceso puede aumentar con muchas inserciones.
* No funcionan en dispositivos de almacenamiento no direccionable.

1.4 INCONVENIENTES DE LAS BD BASADAS EN FICHEROS

Independientemente del tipo de archivo que utilizara, la BD formada por un conjunto de ficheros tenía una fuerte dependencia de la aplicación que lo gestiona. Esto supone una serie de inconvenientes:

* **La redundancia de datos:** ocurre cuando varias aplicaciones almacenan la misma información en diferentes archivos, lo que desperdicia recursos y genera diferencias en los resultados. Esto sucede porque cada archivo solo puede ser accedido por una aplicación a la vez, bloqueando a las demás.
* **Aumento del espacio de almacenamiento**: La redundancia de datos produce una ocupación inútil de memoria secundaria.
* **Degradación del rendimiento:** Aumentan los tiempos de procesamiento por la repetición de controles en diferentes archivos.
* **Inconsistencia de la información:** Surgen incongruencias al no actualizar simultáneamente los datos en múltiples ficheros.
* **Cambio de estructura de datos:** Modificar la estructura de un archivo requiere cambiar el programa, ya que los datos están integrados en él.
* **Falta de protección:** Los datos son vulnerables a accesos no autorizados, ya que basta con tener el fichero para acceder a ellos.

Otra cuestión importante respecto a este tipo de BD era que los ficheros no estaban relacionados entre sí, y los datos almacenados representaban las relaciones existentes en la información mediante referencias simbólicas y/o físicas.

Surge la necesidad de una gestión más eficiente de los datos, lo que lleva al uso de bases de datos donde la información se recoge y almacena una sola vez, sin depender de las aplicaciones que la utilizan. Este enfoque transforma radicalmente el tratamiento de datos, organizándolos en un conjunto estructurado que no está diseñado para una aplicación específica, sino que busca satisfacer las diversas necesidades de información de toda la organización a lo largo del tiempo. Este tipo de Sistema de Información se denomina SIA orientado a datos.

2. LAS BASES DE DATOS

Una base de datos es un conjunto de archivos relacionados que almacena los datos y su representación para un sistema de información de una organización. Estos datos están sujetos a ciertas restricciones y se organizan de manera específica para facilitar un acceso y actualización rápidos y sencillos.

Existe una independencia de los datos respecto a las aplicaciones.

2.1 VENTAJAS DE LAS BD

* Independencia de los datos respecto a las aplicaciones y viceversa
* Coherencia de los resultados: Como la información sólo se recoge y almacena una vez, en las aplicaciones se utilizan siempre los mismos datos, por lo que los resultados son coherentes.
* Mejor disponibilidad de los datos para el conjunto de los usuarios: Cada usuario ya no es propietario de los datos, puesto que están compartidos por el conjunto de aplicaciones.
* Mayor valor informativo al representar no solo los datos, sino también las relaciones entre ellos, de modo que la información total es más valiosa que la simple suma de sus partes.
* Mejor documentación de la información
* Mayor eficiencia en la recogida, validación e introducción de los datos en el sistema.
* Reducción del espacio de almacenamiento

2.2 CARACTERÍSTICAS DE LAS BD

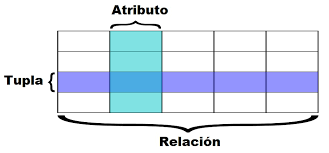
Para que una base de datos (BD) se considere válida, debe cumplir con varias características:

* Versatilidad: Debe permitir que diferentes programas generen registros lógicos a partir de la información existente.
* Desempeño: Necesita asegurar tiempos de respuesta adecuados y permitir accesos simultáneos a los datos por distintas aplicaciones.
* Mínima redundancia: Aunque se busca evitar la redundancia de datos, es necesario para representar relaciones entre entidades, lo que puede provocar inconsistencia e incremento en el almacenamiento.
* Integridad: Los datos deben ser veraces y no modificados anómalamente. Se deben establecer procedimientos para verificar que los valores cumplan con los requerimientos.
* Seguridad y privacidad: La BD debe proteger los datos de pérdidas por fallos del sistema y de accesos no autorizados. Debe incluir medidas contra la destrucción de datos por el entorno y procedimientos para la recuperación de información.

2.3 TIPOS DE BASE DE DATOS

**RELACIONAL:**

Son las más comunes. En este tipo de BD, la información se organiza en relaciones, que son conjuntos de datos llamados tuplas. Cada tupla tiene atributos, que son elementos simples de información. Se suele visualizar cada relación como una tabla, donde las filas representan las tuplas y las columnas representan los atributos.



**DISTRIBUIDA:**

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas en diferentes localizaciones físicas e interconectados por una red de comunicaciones.

**ORIENTADA A OBJETO:**

Las bases de datos orientadas a objetos son un tipo de sistema de gestión de bases de datos que utiliza conceptos de programación orientada a objetos. En lugar de almacenar datos en tablas como en las bases de datos relacionales, estas bases de datos almacenan datos como objetos, que pueden contener tanto atributos (datos) como métodos (funciones) para manipular esos datos.

**JERARQUICA:**

Una base de datos jerárquica es un tipo de sistema de gestión de bases de datos que organiza la información en una estructura de árbol, donde los datos se relacionan entre sí en una jerarquía. Esto significa que los datos están organizados en niveles, y se accede a ellos siguiendo la jerarquía.

2.4 NIVELES DE BASES DE DATOS

Dependiendo de quién accede a una BD, esta le presenta la información de una forma que pueda entender, interpretar y manejar. Existen, pues tres visiones de los datos en una BD en función del tipo de usuario que la maneja y que permite describir la información de una BD en tres niveles:

EXTERNA:

* Es la que tienen los usuarios finales
* Es una visión parcial de los datos que necesita en un momento dado y es diferente a la de otro usuario
* Estas visiones las proporcionan las aplicaciones que acceden a las BD

CONCEPTUAL:

* Es una representación del problema de gestión de información tal y como se presenta en el mundo real
* Se definen las entidades, sus propiedades y las relaciones entre las entidades
* Es una representación abstracta del problema independiente de como se trata la información, como se almacena físicamente y de las visiones externas.
* La visión conceptual no cambia si no cambia la definición del problema

FÍSICA:

* Representación de como la información de la BD se almacena en un dispositivo físico de almacenamiento
* Se describen las estructuras de datos u organizaciones físicas, dispositivos, ficheros, tipos de datos, etc
* Estas estructuras físicas hacen la BD entendible para el sistema informático que las contiene.

Estas tres visiones, en realidad suponen tres representaciones del mismo problema a diferentes niveles. En cada nivel se describen aquellos objetos que pueden ser entendidos por los usuarios de ese nivel de la base de datos.



La descripción de datos a diferentes niveles garantiza la independencia de los datos, permitiendo que se modifiquen la organización física y la representación conceptual sin afectar a los programas de aplicación. Esta independencia incluye la posibilidad de cambiar visiones externas según nuevos requerimientos sin modificar las descripciones de datos. Aunque en muchas bases de datos esta independencia no se logra por completo, se puede mejorar al representar una misma descripción conceptual de varias formas físicas y al mantener el número y disposición de ítems independientes de su representación conceptual. El nivel conceptual es el más importante, ya que incluye todas las entidades, atributos y relaciones del problema. A partir de este, se puede hablar de un **nivel lógico**, que aplica reglas y restricciones para determinar cómo se tratará la información.