

Introducción a la base de datos

1. ¿Qué es un archivo o fichero?

Un archivo o fichero informático es un conjunto de bytes que son almacenados en un dispositivo. Un archivo es identificado por un nombre y la descripción de la carpeta o directorio que lo contiene. A los archivos informáticos se les llama así porque son los equivalentes digitales de los archivos escritos en expedientes, tarjetas, libretas, papel o microfichas del entorno de oficina tradicional.

2. Tipos de ficheros.

ORGANIZACIÓN SECUENCIAL:

En un fichero con organización secuencial los registros se escriben sobre el dispositivo de almacenamiento en posiciones físicamente contiguas, sin dejar huecos entre ellos, en el mismo orden en que hayan sido introducidos.

Las ventajas de los ficheros con organización secuencial son:

- Rapidez en el acceso a un bloque de registros contiguos
- No es necesario realizar operaciones de compactación del archivo
- No se desperdicia espacio en el dispositivo de almacenamiento porque no hay huecos
- Se pueden utilizar cualquier tipo de registros: de longitud fija, variable o indefinida.

Los inconvenientes de este tipo de organización son:

- Para acceder al registro n hay que recorrer los $n-1$ registros anteriores. El acceso es secuencial.
- Para realizar una consulta hay que crear un proceso en el que se compare el valor del campo que se pretende localizar con el valor del mismo campo correspondiente a cada registro leído del fichero.
- La adición de registros se realiza a continuación del último registro ya existente. No se pueden insertar nuevos registros.
- No se pueden eliminar registros. Para eliminar un registro se marca de modo que no se muestre, pero el registro existe y ocupa espacio en el dispositivo del almacenamiento. (borrado lógico)
- Para mantener ordenado y compactado el fichero, hay que crear un fichero nuevo a partir del existente.

ORGANIZACIÓN SECUENCIAL ENCADENADA:

Son ficheros de organización secuencial gestionados mediante punteros que nos permiten tener los registros ordenados según un orden lógico diferente del orden físico en el que están grabados.

Los punteros son un campo adicional, por lo que aumentan el tamaño de los registros, y forman parte integrante de la estructura del registro, indicando cual es el siguiente o el anterior registro en secuencia lógica y no en secuencia física

Las características de los punteros son:

- Estar en cada registro en una posición fija y definida para todos los registros
- Poseer una longitud constante

ORGANIZACIÓN SECUENCIAL INDEXADA:

En esta organización los registros con los datos se graban en un fichero secuencialmente, pero se pueden recuperar con acceso directo gracias a la utilización de un fichero adicional, llamado de índices, que contiene información de la posición que ocupa cada registro en el fichero de datos.

En este tipo de ficheros se distinguen tres áreas:

Área primaria:

Es la zona donde están contenidos los registros ordenados ascendentemente por el valor de su clave. Esta área del fichero está dividida en segmentos.

Acceso directo al segmento donde se haya ubicado el registro buscado

Acceso secuencial posteriormente, a los registros del segmento, hasta localizar el registro buscado o alcanzar el final del segmento, en caso de que no se halle.

Área de índices:

Es creada por el sistema al mismo tiempo en que se almacenan los datos. En esta área los registros están formados por dos campos:

- El valor del campo clave del último registro de un bloque o segmento. Los bloques están constituidos por un número fijo de registros consecutivos.
- El segundo campo contiene la dirección de comienzo de cada uno de los segmentos en los que se halla dividida el área primaria.

Área de overflow:

Es la zona destinada a contener los registros almacenados posteriormente a la creación del fichero, por lo que no han sido incluidos en el área primaria.

El acceso a los registros se realiza mediante una consulta secuencial al área de índices para determinar el segmento donde se encuentra el registro buscado. Con el valor del segmento seleccionado, se recorren secuencialmente los registros de ese segmento.

El borrado La eliminación de los registros debe realizarse mediante marcas

La inserción de registros se hace en el área de overflow. No está permitida la instrucción de nuevo registro en el área primaria después de la creación del fichero.

En esta organización cuando el numero de registros borrados es grande, o las cadenas de desbordamientos son largas su utilización deja de ser eficiente, siendo necesario reorganizar el archivo.

ORGANIZACIÓN INDEXADA-ENCADENADA:

Este tipo de organización aprovecha lo mejor de la organización secuencial encadenada e indexada.

Se caracteriza por la utilización de punteros e índices de forma simultánea, lo que implica un aumento del espacio ocupado, pero proporciona una gran rapidez en la búsqueda de registros.

Para eliminar registros se marcan, en lugar de ser borrados físicamente.

Las adiciones se realizan sobre la zona de overflow ya que no se pueden añadir registros en el área primaria una vez creado el fichero.

Para acceder a un registro se busca en el área de índices la dirección de inicio del bloque de registros a la que pertenece el registro buscado.

Si se desea consultar todo el fichero, el último registro de cada bloque del área de overflow tiene un puntero al primer registro del área siguiente y así se continuará la búsqueda hasta acabar de leer el fichero.

Los nuevos registros se insertan y quedan enlazados entre sí mediante punteros conservando el orden lógico que marca la clave o índice principal.

ORGANIZACIÓN RELATIVA DIRECTA:

La organización relativa está basada en la independencia entre el orden en el que se dan de alta los registros y la posición en que se graban en el soporte.

Hay dos tipos de organización relativa: directa y aleatoria.

En el caso de la organización relativa directa se emplean claves numéricas,

Las ventajas de este tipo de organización de ficheros son:

- Acceso directo a los registros. No se necesita un algoritmo de transformación

- Permite realizar operaciones de escritura y lectura simultáneamente
- El acceso a los datos se realiza de dos formas diferentes:
 - Directamente, mediante la clave del registro
 - Secuencialmente, a partir del primer registro almacenado en el fichero, por lo que son muy rápidos en el tratamiento individual de registros.

Los inconvenientes de este tipo de organización de ficheros son:

- Al realizar un acceso secuencial, en una consulta sobre todos los registros del fichero hay que recorrer todas las direcciones aunque estén vacías.
- Deja gran cantidad de posiciones libres de memoria dentro del fichero (huecos), debido a que las claves de los registros pueden indicar posiciones de almacenamiento no contiguas, lo que implica una falta de aprovechamiento del soporte del almacenamiento respecto al número real de registros almacenados.
- Se producen colisiones, ya que puede existir más de un registro con la misma clave. Esto causa errores.

ORGANIZACIÓN RELATIVA ALEATORIA:

Son ficheros con organización relativa y clave alfanumérica o bien numérica pero que se debe transformar obteniéndose un valor numérico entero que facilite la correspondencia directa entre la clave y la dirección de memoria.

El valor de la clave debe estar en relación con la capacidad máxima del soporte físico

La dirección del almacenamiento de un registro dentro del dispositivo se obtiene de su clave de la siguiente forma:

- clave alfanumérica se aplican algoritmos de transformación para obtener valores enteros positivos.
- Si la clave es numérica se aplica un algoritmo que permita obtener un rango de valores comprendidos en el intervalo de valores de las direcciones de memoria disponibles, de modo que existe una relación directa entre la dirección lógica (clave) y la dirección física (memoria).

El algoritmo de transformación o hashing debe cumplir las siguientes condiciones:

- Que sea fácil de aplicar, estableciendo una relación directa entre dirección lógica y dirección física
- Que deje el mínimo número de huecos posible, maximizando el espacio disponible en el dispositivo de almacenamiento.

- Que las claves de registros diferentes nos den direcciones diferentes. Producir el menor número de registros que con distintas claves creen las mismas direcciones de almacenamiento

Las ventajas de este tipo de organización de ficheros son:•Acceso inmediato a los registros mediante su clave

- No es necesario ordenar el fichero
- Se pueden realizar operaciones de escritura y lectura a la vez
- Son muy rápidos en el tratamiento individual de registros.
- Se pueden realizar accesos secuenciales.

Los inconvenientes de este tipo de organización de ficheros son:

- Las consultas sobre todo el fichero son lentas
- El fichero contiene gran cantidad de huecos o espacios libres
- El algoritmo para la conversión de las claves y el algoritmo necesario para el almacenamiento de sinónimos han de ser creados de modo que dejen el menor número de huecos libres y se genere el menor número de sinónimos.

3. Define brevemente los conceptos de registro y campo.

Un registro (también llamado fila o tupla) representa un objeto único de datos implícitamente estructurados en una tabla. En términos simples, una tabla de una base de datos puede imaginarse formada de *filas* y *columnas* (campos o atributos).

Un campo es un espacio de almacenamiento para un dato en particular.¹

En las bases de datos, un campo es la mínima unidad de información a la que se puede acceder; un campo o un conjunto de ellos forman un registro, donde pueden existir campos en blanco, siendo este un error del sistema operativo. Aquel campo que posee un dato único para una repetición de entidad, puede servir para la búsqueda de una entidad específica.

En las hojas de cálculo los campos son llamados "celdas"

4. ¿Qué es una base de datos?

Las **bases de datos** son un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso, la mayoría de las bases de datos están en formato digital, siendo este un componente electrónico, por tanto se

ha desarrollado y se ofrece un amplio rango de soluciones al problema del almacenamiento de datos.

5. Diferencias entre un sistema convencional de archivos y un Sistema Gestor de Base de Datos.

Los sistemas en base a archivo se caracterizaron por el manejo de los datos en función y del área a la que apoyaban, así como del personal que desarrolló esos sistemas.

Una manera de mantener información en un computador es hacerlo mediante un sistema de procesamiento de archivos típico o tradicional, que permitirá tener a los archivos estructurados y organizados, y poder realizar operaciones con ellos. Este sistema de archivos se mantiene mediante un sistema operativo convencional. Antes de la llegada de los sistemas de gestión de bases de datos (SGBD), las organizaciones normalmente han almacenado la información usando estos sistemas, pero mantener la información en estos sistemas de archivos tiene una serie de inconvenientes importantes.

Sistema de archivos:

1. Manejo de archivos en texto plano
2. Búsqueda de tres tipos: secuencial, aleatorio, indexado.
3. No existen las actualizaciones, se simulan dando de baja el registro y dando de alta el nuevo con la modificación
4. Tienen independencia lógica y física la mayoría de ellos.
5. Cuenta con redundancia no controlada y errores de referencias.
6. Sus componentes son: archivos, registros, campos.

Sistemas de bases de datos:

1. Existe redundancia controlada e integridad referencia.
2. Implementa mayor seguridad.
3. Existen actualizaciones directas.
4. No tienen independencia ni lógica ni física.
5. Sus componentes son Hw, Sw, BD y Usuarios.

6. Componentes de un Sistema Gestor de Base de Datos.

El **motor de la base de datos** acepta peticiones lógicas de los otros subsistemas del SGBD, las convierte en su equivalente físico y accede a la base de datos y diccionario de datos en el dispositivo de almacenamiento.

El **subsistema de definición de datos** ayuda a crear y mantener el diccionario de datos y define la estructura del fichero que soporta la base de datos.

El **subsistema de manipulación** de datos ayuda al usuario a añadir, cambiar y borrar información de la base de datos y la consulta para extraer información.

El **subsistema de generación de aplicaciones** contiene utilidades para ayudar a los usuarios en el desarrollo de aplicaciones. Usualmente proporciona pantallas de entrada de datos, lenguajes de programación e interfaces.

El **subsistema de administración** ayuda a gestionar la base de datos ofreciendo funcionalidades como almacenamiento y recuperación, gestión de la seguridad, optimización de preguntas, control de concurrencia y gestión de cambios.

7. Tipos de usuarios de las bases de datos, funciones y características de cada uno de ellos.

Programador de aplicaciones: son aquellos profesionales en informática que interactúan con el sistema a través del DML (Lenguaje de Manipulación de Datos), los cuales se encuentran en un lenguaje de programación (Pascal, Cobol, etc.) Es el encargado de escribir programas de aplicación que usen Bases de Datos.

Usuario Final: accede a la base de datos desde un equipo en el cual puede utilizar lenguaje de consulta generado como parte del sistema o acude a un programa de aplicación suministrado por un programador.

Administrador de Bases de Datos: es el encargado del control general del sistema.

Programador de Aplicaciones.

Usuario sofisticado: interactúa con el sistema sin escribir programas. Generan consultas en un lenguaje de bases de datos.

Usuario Especializado: algunos usuarios sofisticados desarrollan aplicaciones de bases de datos especializadas. Entre estas aplicaciones se encuentran los sistemas de diseño asistido por computador.

Usuarios ingenuos: es el usuario final que utiliza bases de datos sin saberlo, para él es totalmente transparente como se generan las consultas de la información.

Actores en el escenario: personas que su trabajo depende del uso constante una base de datos.

DataBase Administrators(DBA): administran 2 recursos 1. la base de datos y 2. es el SGBD y el software con el relacionado. El Administrador de Base de Datos (DBA) es quien autoriza el acceso a la base de datos, vigilar el uso y adquirir hardware y software necesarios para su uso. También es el responsable de velar por la seguridad y lentitud en el sistema.

Diseñador de Base de Datos: es el encargado de estructurar la arquitectura para representar y almacenar los datos. Él debe atender a los usuarios de Bases de Datos para comprender sus necesidades presentando un diseño que de respuesta a sus necesidades.

Usuarios Finales: son quienes requieren acceso a la base de datos para generar consultas e informes. Hay varios usuarios finales como son:

- **Usuarios finales esporádicos:** acceden de vez en cuando pero esto no significa que siempre requieran la misma información.
- **Usuarios finales simples o paramétricos:** su función gira en torno a consultas y actualizaciones de la base de datos. Todos estamos acostumbrados a tratar con estos usuarios, como los cajeros bancarios al revisar los saldos, al generar retiros y depósitos.
- **Usuarios finales avanzados:** estos son ingenieros, analistas de negocios, científicos, son quienes conocen los recursos del SGBD para satisfacer requerimientos complejos.
- **Usuarios Autónomos:** utilizan bases de datos personalizadas basadas en programas comerciales que cuentan con interfaces de fácil uso.

8. Organización de una base de datos.

Una base de datos se organiza en campos, registros, archivos. Un campo es una pieza de información ejemplo: el campo nombre o edad. Un registro es un conjunto de campos ejemplo: registro de los alumnos de un colegio. Un archivo es una colección de registros ejemplo: los archivos del área de una empresa "X". Otro ejemplo es un conjunto de documentos de identidad cada uno tiene campos que son: nombre, apellido paterno, apellido materno, fecha de nacimiento, etc. y todo este conjunto de campos conforman un registro, este con otros registros conforman un documento de identidad. El cual te sirve para realizar consultas ya sean tu dirección, tu cumpleaños o tu estado civil.

9. Modelos de bases de datos.

Bases de datos jerárquicas

Las B.D.D. jerárquicas almacenan su información en una estructura jerárquica, similar a un árbol donde un nodo es un padre que puede varios hijos

Base de datos de red

Este a diferencia de la anterior permite que un nodo tenga varios padres. Ofrece una solución al problema de redundancia de datos, aun así, la dificultad que significa administrar la información en una base de datos de red ha significado que sea un modelo utilizado en su mayoría por programadores más que por usuarios finales.

Bases de datos transaccionales

Son bases de datos cuyo único fin es el envío y recepción de datos a grandes velocidades, estas bases son muy poco comunes y están dirigidas por lo general al entorno de análisis de calidad, datos de producción e industrial, es importante entender que su fin único es recolectar y recuperar los datos a la mayor velocidad posible, por lo tanto la redundancia y duplicación de información no es un problema como con las demás bases de datos, por lo general para poderlas aprovechar al máximo permiten algún tipo de conectividad a bases de datos relacionales.

Bases de datos relacionales

Éste es el modelo utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente.

Bases de datos multidimensionales

Son bases de datos ideadas para desarrollar aplicaciones muy concretas, como creación de Cubos OLAP. Básicamente no se diferencian demasiado de las bases de datos relacionales

Bases de datos orientadas a objetos

Este modelo, bastante reciente, y propio de los modelos informáticos orientados a objetos, trata de almacenar en la base de datos los objetos completos

na base de datos orientada a objetos es una base de datos que incorpora todos los conceptos importantes del paradigma de objetos:

- Encapsulación - Propiedad que permite ocultar la información al resto de los objetos, impidiendo así accesos incorrectos o conflictos.
- Herencia - Propiedad a través de la cual los objetos heredan comportamiento dentro de una jerarquía de clases.
- Polimorfismo - Propiedad de una operación mediante la cual puede ser aplicada a distintos tipos de objetos.

Bases de datos documentales

Permiten la indexación a texto completo, y en líneas generales realizar búsquedas más potentes.

Bases de datos deductivas

Un sistema de base de datos deductiva, es un sistema de base de datos pero con la diferencia de que permite hacer deducciones a través de inferencias. Se basa principalmente en reglas y hechos que son almacenados en la base de datos. Las bases de datos deductivas son también llamadas bases de datos lógicas, a raíz de que se basa en lógica matemática.

.

10. Tipos de archivos según su función en el tiempo

Existen dos tipos de ficheros: Los permanentes y los temporales.

Los Permanentes:

ARCHIVOS MAESTROS. Se llaman también archivos de situación. Contienen la información actualizada de los datos que pueden cambiar. Reflejan el estado actual de la aplicación o de algún aspecto o parte concreta de la misma.

ARCHIVOS CONSTANTES. Contienen los datos fijos de la aplicación o aquellos otros que no suelen cambiar apenas. Se suelen utilizar como archivos de consulta.

ARCHIVOS HISTÓRICOS. Contienen los datos acumulados a lo largo del tiempo de los archivos que han experimentado actualizaciones. Se emplean para efectuar estadísticas, para posibles consultas o para reconstrucción de archivos que han sufrido deterioros.

Temporales:

ARCHIVOS DE MOVIMIENTO O DE TRANSACCIONES, Contienen la información necesaria para actualizar los archivos maestros con los datos o resultados de las operaciones que se realizan. Una vez terminada cada operación, la información que contienen pierde su valor y utilidad, y puede ser destruida.

ARCHIVOS DE MANIOBRA O INTERMEDIOS. Se utilizan para almacenar datos o resultados auxiliares que han de ser utilizados por otros procesos o programas. Su vida depende del tiempo que dure la operación en la que intervienen.