ROMPEMEGATON BASES DE DATOS

**Archivos**: conjunto de registros homogéneos que contienen datos heterogéneos.

**Registros**: conjunto de campos relacionados que contienen datos elementales.

**Campo**: unidad mínima de información de un registro.

Clasificación de archivos

* por ciclo de vida:
* persistentes: se almacenan en memoria secundaria y se quedan ahí hasta que se borren.
* temporales: pueden estar en memoria principal o secundaria, y tienen un tiempo de vida determinado.
* según la información:
* de programa: almacenan instrucciones.
* de datos: almacenan datos.

Organización de los archivos (número de alternativas posibles para organizar casi ilimitado):

**Archivo secuencial**: no queda posición sin usar y tampoco se puede acceder directamente, se accede en orden en el que fueron escritos, del primero al último de uno en uno.

* alta: el nuevo registro se agrega al final del archivo.
* baja: no existe la baja física, se marcan los registros con un valor testigo.
* modificaciones: se puede reescribir.

**Archivo indexado**: se asocia a la clave de un registro su dirección relativa dentro del fichero a través de otra tabla (índice). El área de datos almacena los registros y está direccionada por el área de índices que incluye una clave de registro y la dirección de almacenamiento.

* alta: graba el registro en el área de datos y se guarda la clave con su dirección en el área de índices.
* baja: tiene baja física, se elimina el registro del área de índices y se cambia el puntero.
* modificaciones: busca el registro en el área de índices, obtiene la dirección, se ubica en el área de datos y modifica.

**Archivo secuencial indexado**: se organiza por bloques de registros donde los índices apuntan a cada bloque. Tiene tres zonas: el área de datos que se divide en bloques y almacena los registros con tamaño fijo, el área overflow que almacena los registros cuando se llena la anterior, y el área de índices que guarda la primera clave del área de datos y su dirección de memoria.

**Archivo directo**: rápido acceso a los registros haciendo referencia a su posición relativa, por lo que se dimensiona antes.

Estructura en el manejo de datos a través de bases de datos:

Datos -> Archivos -> Tratamientos -> Resultados.

**Sistema de Bases de datos**: consiste en alguna colección de datos, de redundancia mínima, persistentes e independientes usados por una organización determinada. Es integrada por varios archivos de forma que se elimina total/parcial la redundancia de datos y es compartida porque cada parte individual es accedida por un conjunto de usuarios.

El diseño de base de datos se hace en tres fases:

* diseño conceptual: representación total y abstracta de los datos que componen la Base. (DER).
* diseño lógico: transforma el diseño anterior al modelo de datos del Sistema de Gestión de Bases de Datos.
* diseño físico: implementa de forma eficiente el diseño anterior, depende por completo del SGBD y el equipo donde se haga.

Ventajas de las BD: muy eficiente en la recuperación, codificación y entrada de datos, mayor coherencia y valor informativo en resultados, acceso más rápido y sencillo de los usuarios finales.

Desventajas de las BD: costosa en equipos, ausencia de estándares, necesita personal especializado, hay desfase entre teoría y práctica.

Un sistema de bases de datos está compuesto por la BD, SGBD/motores de BD, programas de aplicación, conjunto de usuarios finales, hardware y programas utilitarios.

La arquitectura general es de tres vistas:

* vista externa: visión particular de usuarios de la Base.
* vista conceptual: representación total y abstracta de los datos.
* vista interna: es el almacenamiento físico de la Base.

Las funciones del Administrador de Bases de Datos (persona/s encargada del control general):

* decide el contenido de la BD.
* decide la estructura de almacenamiento y estrategia de acceso.
* se vincula con los usuarios.
* seguridad.
* define estrategia de respaldo y recuperación.
* controla desempeño y responde a cambios.

El **Sistema de Gestión de Bases de Datos (SGBD)** es una colección de datos interrelacionados y un conjunto de programas para acceder a ellos, su objetivo es proporcionar un entorno eficiente para ser utilizado al extraer y almacenar información. Suministra lenguaje, herramientas y funciones.

Las funciones del SGBD:

* de definición: provee un lenguaje de definición de datos y debe definir las estructuras de datos a los tres niveles, al interno define espacio reservado, modo de representación de los datos, a los externo y conceptual da instrumentos para definir entidades, atributos, etc.
* de manipulación: permite a los usuarios buscar, eliminar o modificar los datos según normas de seguridad.
* de utilización: reúne interfaces que necesitan los usuarios para comunicarse con la BD y da una serie de procedimientos para el administrador.

El SGBD tiene que proveer distintos tipos de lenguajes según las operaciones que deba realizar, los hay de definición y de manipulación:

* DDL (data definition language): son lenguajes internos dedicados a la definición de la BD, son palabras reservadas como CREATE, ALTER, DROP.
* DML (data manipulation language): es significativo a nivel usuario, permite interactuar directamente con la manipulación de datos, son palabras reservadas como INSERT, UPDATE, DELETE, SELECT.

el SQL (structured query language) como lenguaje de manipulación de datos tiene la propiedad dual de actuar como huésped o autocontenido.

Para definir el diseño conceptual de un BD usamos el DER (diagrama entidad relación): rectángulos para entidades, rombos para relaciones, cardinalidad, chupetín para atributo, doble chupetín para atributo multivaluado, doble rectángulo para entidad débil, recuadro sobre entidad para agregación, generalización: ( T/P , E/S) explicado con ejemplos abajo

Vamos a poner el ejemplo de la entidad Empleado, de la cual heredan Administrador y Técnico. Si T es total, quiere decir que, si tengo un empleado, este es administrador o técnico, alguna de las 2 tiene que ser, no puede ser ninguna; mientras que P nos dice que puede no ser ni administrador ni técnico. S de superpuesta quiere decir que un empleado puede ser Admin y técnico a la vez, mientras que E de exclusiva es uno o es otro, pero no puede ser ambos.

La teoría de normalización permite detectar y corregir errores en las tablas, existen 6 formas solo damos 3. La normalización es el proceso en que se transforman datos complejos en un conjunto de datos más chicos, simples, estables y fácil de mantener. Ayuda a mejorar el esquema lógico, que con ciertas restricciones evita duplicidad de datos. Minimiza redundancia, mantenimiento de datos e impacto de futuras anomalías.

1FN primera formal normal: cada atributo de una tabla tiene un valor atómico (simple)

2FN segunda forma normal: está en 1FN, cada atributo no clave depende de la clave completa y no de parte de ella.

3FN tercera forma normal: está en 3FN, todos los atributos no claves son independientes entre sí.

**Clave (key)**: es un atributo o conjunto de atributos que identifican a un registro.

Tipos de claves:

* clave primaria (PK): atributo/conjunto de atributos que identifican unívocamente a una tupla.
* clave candidata: cumple requisitos para ser PK, tiene potencial.
* clave secundaria: únicas o no, permiten al sistema de gestión ordenar una tabla por otro criterio que no sea el establecido en la PK.
* claves foráneas (FK): atributo dentro de la tabla que apunta a un registro único en otra tabla.