**Resumen base de datos**

**Unidad 1**

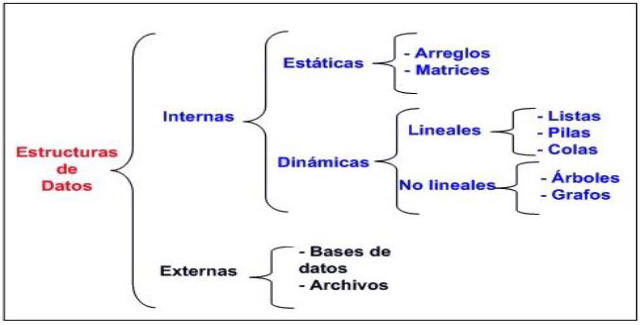
**Datos, información y conocimiento.**

Dato: Mínima unidad semántica, irrelevante por si mismo.

Información: La información es un conjunto de datos procesados que tienen significado (relevancia, propósito y contexto), son de utilidad para tomar decisiones porque reduce la incertidumbre.

Conocimiento: Es la mezcla de experiencia, valores, información y know how que sirve para adquirir nuevas experiencias e información y es útil para la acción. Se origina y aplica en la mente de los conocedores. Se encuentra en documentos pero también en rutinas organizativas, procesos, prácticas y normas.

Conclusión: Los sistemas de información, los procesos de recopilación, gestión y utilización de datos se hace cada vez mas importante para el éxito operacional. Los datos son valiosos si podemos acceder a ellos y extraer un significado, es decir son información, que al ser combinado con nuestra experiencia pueda constituirse como conocimiento para la toma de decisiones. Es imposible obtener información de datos que no están organizados, almacenados y analizados de forma efectiva.



**Sistemas de ficheros**

Definición: Un sistema basado en ficheros (de ahora en adelante SF) es un conjunto de programas informáticos que permiten al usuario almacenar, consultar y modificar datos. Los datos se almacenan en un fichero diseñado para una determinada aplicación. Cada programa define y maneja sus propios datos.

Origen: Los SF surgen al informatizar los archivadores manuales para acceder mas eficientemente a los datos almacenados. Los sistemas de ficheros tienen ciertas caracteristicas propias:

- Descentralizado: Cada departamento almacena y gestiona sus datos mediante programas de aplicación escritos especialmente para él.

- Totalmente independientes: Los programas son independientes entre departamentos para introducir datos, mantener ficheros y generar los informes que el departamento necesita.

- Duplicación de datos: Los departamentos no comparten información ni aplicaciones por lo que los datos comunes deben estar duplicados en cada departamento.

- Inconsistencia: Al haber datos duplicados un departamento puede tener un dato y otro departamento otro dato distinto ya que la actualización puede darse en uno y no en otro.

Inconvenientes

- Separación y aislamiento de datos.

- Duplicación de datos, se desperdicia capacidad de almacenamiento y al no estar sincronizados se pierde la coherencia de datos.

- Dependencia de los datos, su estructura física al tener distintos programas de aplicación dificulta cambios en la misma.

- Incompatibilidad entre formatos, cada fichero se puede utilizar con su aplicación.

- Rigidez en las consultas, no hay flexibilidad porque el fichero devuelve información concreta y no permite la utilización de otro criterio preestablecido.

**Bases de datos**

Definición: Una base de datos es un conjunto de datos que modela hechos y objetos de una parte de la realidad, sirven de soporte para aplicaciones informáticas. Los datos deben almacenarse físicamente en un soporte informático y relacionarse entre si mediante una estructura lógica. Una propiedad esencial es la “independencia de datos” es decir separar datos y aplicaciones que los manejan. Cada base de dato se diseña para satisfacer los requisitos de información de una organización. Es “una versión moderna” de los sistemas de ficheros.

Una base de datos se puede percibir como un gran almacén de datos que se define y se crea una sola vez y que puede ser utilizada al mismo tiempo por distintos usuarios. En una BD los datos se integran con una mínima cantidad de duplicidad (idealmente 0), es común a toda la organización (a diferencia del fichero que es para el departamento particular). No solo almacena los datos si no que también almacena una descripción de dichos datos, esta descripción normalmente es llamada “metadatos”, se almacena en el diccionario de datos y permite la independencia de datos lógica-física. Representan algún aspecto del mundo real, también denominado dominio y está diseñada, construida y poblada con datos para un propósito especifico.

**DBMS (Data base management system o sistema de gestión de base de datos SGBD)**

El sistema de gestion de base de datos es una aplicación que permite a los usuarios “definir, crear, mantener, consultar y controlar el acceso” a base de datos.  
  
- Definir: Especificar tipos de datos, estructuras, relaciones y restricciones.

- Crear: Generar la estructura definida, poblar datos iniciales.

- Mantener: realizar acciones de actualización, inserción y eliminación.

- Consultar: Recuperar información según criterios de selección.

- Controlar el acceso: Garantizar que cada tipo de acción es posible solamente realizada por grupos de usuarios autorizados.

En el modelo seguido con los sistemas de BD se da una implementación interna de un objeto y una especificación externa separada. Los usuarios solo ven la especificación externa y no deben preocuparse como se implementa internamente el objeto. La ventaja de este modelo es conocido como “abstracción de datos”, esto es que se puede cambiar una implementación interna de un objeto sin afectar a sus usuarios ya que la especificación externa no se ve alterada.

Componentes o servicios de los DBMS

DDL (Lenguaje de definición de datos): Define los elementos de datos, estructuras, relaciones y validaciones a nivel externo.

DML (Lenguaje de manipulación de datos): Permite buscar, añadir suprimir y modificar datos.

SQL (Lenguaje de consulta estandar): Funcionalidad para cargar archivos, realizar copias de seguridad, gestion de concurrencia y estadisticas de utilización.

Ventajas de los DBM

- Control de concurrencia

- Control de integridad de información. Consistencia de datos.

- Copias de seguridad y plan de mantenimiento.

- Optimizador de consultas.

El SGBD proporciona acceso controlado a la BD mediante:

- Un sistema de seguridad de modo que los usuarios no autorizados no puedan acceder a la BD.

- Un sistema de integridad que mantiene la integridad y consistencia de los datos.

- Un sistema de control de concurrencia que permite el acceso compartido a la BD.

- Un sistema de control de recuperación que restablece la BD tras producirse un fallo de hardware o software.

- Un diccionario de datos o catalogo, accesible por el usuario que contiene la descripción de la BD.

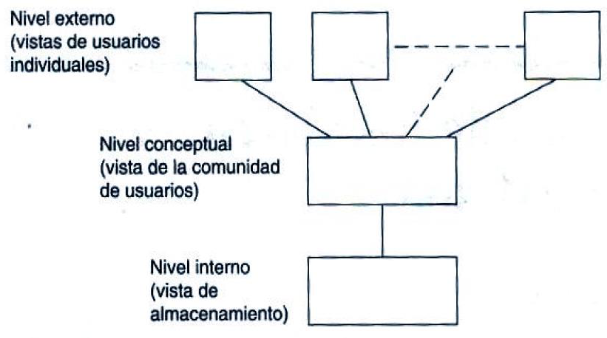
**Arquitectura de un BDMS (SGBD)**

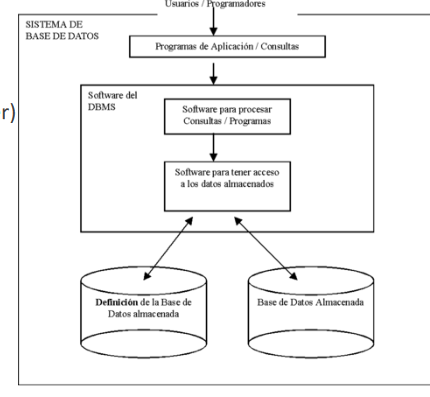
La arquitectura ANSI/SPARC se divide en tres niveles, interno, conceptual y externo.

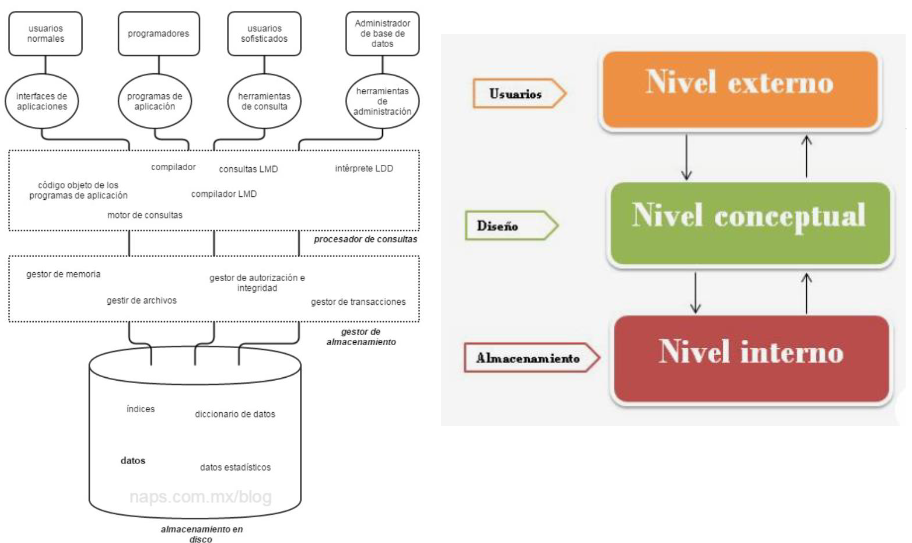
- Nivel interno: También conocido como el nivel físico es el que está mas cerca del almacenamiento físico, es decir tiene que ver con la forma en que los datos están almacenados físicamente.

- Nivel externo: También conocido como el lógico del usuario, es el mas próximo al usuario, tiene que ver con la forma en que los usuarios individuales ven los datos.

- Nivel conceptual: También conocido como el lógico de la comunidad o solamente nivel lógico es un nivel de indirección entre los otros dos.







**Clasificación de un SGBD**

Según el modelo lógico:  
  
- Jerárquico

- En red

- Relacional

- Orientado a objetos

Según número de usuarios:

- Monousuario

- Multiusuario

Según numero de sitios:

- Centralizado

- Distribuido

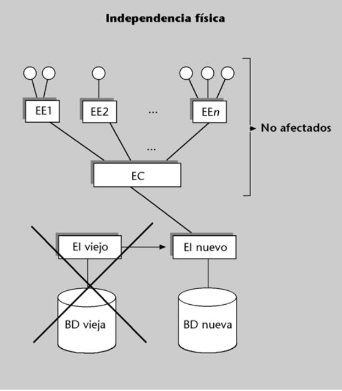
Independencia en los datos

La arquitectura de tres niveles proporciona dos tipos de independencia:

- Física

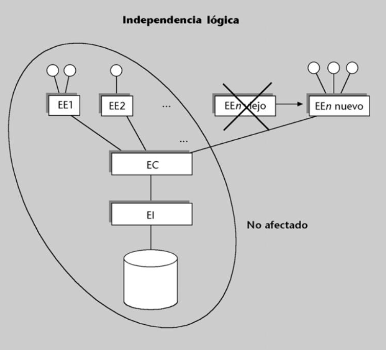
- Lógica

Hay independencia física cuando los cambios en la organización física de la BD no afectan al mundo exterior, es decir los programas usuarios o los usuarios directos. Es decir, los cambios en el esquema interno no afectan al esquema conceptual ni externo.



Si hay independencia física de los datos lo único que variará al cambiar el esquema interno son las correspondencias entre el esquema conceptual y el interno. La mayoría de los cambios en el esquema interno obligaran a rehacer la BD real es decir la física.

Hay independencia lógica cuando los usuarios no se ven afectados por el cambio en el nivel lógico.



**Funciones del DBA (Administrador de base de datos, ABD)**

Tiene tres tareas:

- Crear el esquema de la BD.

- Definir la estructura y el método de acceso.

- Modificar el esquema y la organización física.

- Autoriza el acceso a la base de datos.

- Da mantenimiento de rutina a la BD, es decir, back ups, verificación de espacio libre y supervisión del rendimiento del sistema.

**Diferencia entre Sistema de archivos y DBMS**

****

**Unidad 2**

**Bases de datos relacionales**

Fueron propuestas por Edgar F Codd en 1970, se organizan mediante tablas. Podríamos definir a una BD relacional como un conjunto de tablas que almacena datos donde las tablas están relacionadas entre sí.

Las tablas se organizan en registros (filas) que recoge los datos asociados por ej a un individuo y campos (columnas) que son las partes en la que se desglosa cada registro, por ende la tabla es el conjunto de todos los registros junto con su definición de campos.

**Diseño de una base de datos**

El diseño de base de datos consta de tres fases:

Diseño conceptual: Se parte de una “especificación de requisitos” que describe los datos que debe recoger la BD y las prestaciones a ofrecer, de la cual se obtiene un “esquema conceptual” que es una descripción en lenguaje natural/gráfico de la estructura. Para ello se emplea un “modelo conceptual” que es el lenguaje natural o gráfico que se utiliza para describir un esquema conceptual. Como modelo conceptual utilizaremos el “modelo entidad-interrelación”.

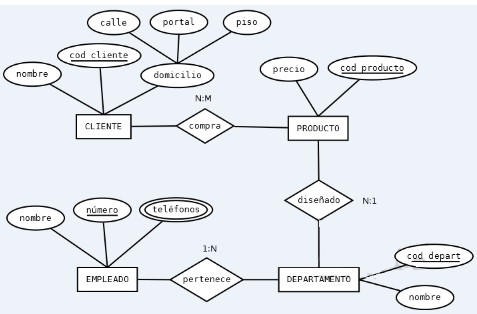
El diseño lógico: Se parte del “esquema conceptual” para obtener el “esquema lógico” es decir la descripción de la estructura de la BD según el modelo de datos elegido para implementarla. Para ello se emplea un “modelo lógico” que es el lenguaje empleado para describir el “esquema lógico”, esto es independiente del SGBD pero no del modelo de datos.

Como modelo lógico utilizaremos el “modelo relacional”.

El diseño físico: Parte del “esquema lógico” para obtener el “esquema físico” que es la descripción de la implementación de la BD según el SGBD. Se detalla la estructura física de almacenamiento de datos según el ordenador, métodos de acceso a dichos datos y el modelo de seguridad. El objetivo es determinar las “estructuras de almacenamiento” y “estructuras de acceso” para que las aplicaciones que accedan a la BD obtengan un buen rendimiento.

Como modelo físico utilizaremos el “SGBD o DBMS”.

MER (modelo entidad relacion)



Permite realizar el diseño conceptual de la BD, es una representación lingüística y grafica de los objetos que forman parte del mundo real.

Conceptos básicos de este modelo son:

- Entidades: Lugar, persona, cosa, concepto, suceso, real o abstracto de interés del cual queremos almacenar información y es distinguible de los demás. Tienen:

- Tipo: Estructura generica compartida por un conjunto de entidades.

- Ocurrencia: realizaciones concretas de este tipo de entidad.

- Conjunto: Grupo de entidades del mismo tipo.

Pueden ser entidades fuertes que son aquellas que tienen existencia por si misma, o débiles que dependen de otra entidad para existir. Todas las entidades están definidas por sus atributos, que pueden ser atributos simples o atómicos, que son aquello que son indivisibles y atributos compuestos que pueden llegar a formar una jerarquía, a su vez los atributos pueden ser mono-valuados, es decir tienen un solo valor para una ocurrencia en particular o multivaluados que tienen un conjunto de valores para una misma concurrencia. También tenemos dos divisiones mas en “elementales y derivados”, donde los elementales no se puede calcular en base a otro atributo y el derivado o calculado que se puede determinar según otros valores y atributos “opcionales u obligatorios” donde un atributo obligatorio no puede tomar un valor nulo y los opcionales si pueden.

- Relaciones

- Atributos

- Claves