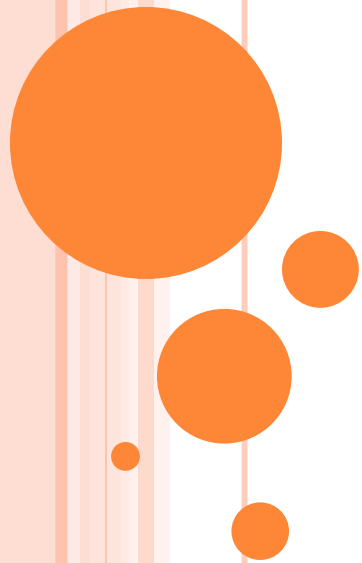


BASE DE DATOS



HISTORIA DE LAS B.D.



Tuvieron sus orígenes en 1960 - 1962, cuando se empezaron a usar las máquinas que codificaban la información en tarjetas perforadas por medio de agujeros. Las bases de datos se crean con el objetivo de almacenar grandes cantidades de datos que antes se almacenaba en libros, lo que era lento, costoso y complejo (cualquier actualización a realizar, había que hacerla en cada uno de los libros en los que apareciera dicha información a modificar).

Las primeras bases de datos manejaban ficheros que eran almacenados en tarjetas o soportes magnéticos. Cuando los ordenadores evolucionan, aparecen las cintas y los discos, a la vez que las máquinas son dotadas de mucha más potencia y facilidad de manipulación, es por tanto en ese momento cuando las bases de datos comienzan a ser realmente útiles.



En 1970 se convoca una Conferencia de Lenguajes de Programación y se establece un modelo llamado CODASYL (Modelo para el tratamiento de bases de datos que fue publicado por E. Cod en 1970. Cod, propuso una forma de organizar las bases de datos mediante un modelo matemático lógico.

Una vez creado este modelo se crea un modelo estándar de actuación.



¿QUÉ ES UNA BASE DE DATOS?

Abreviada B.D.D

1-Un conjunto de información almacenada en memoria auxiliar que permite acceso directo y un conjunto de programas que manipulan esos datos

2-Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso

B.D.D. EN LA VIDA COTIDIANA

Las B.D.D se usan tan ampliamente que se pueden encontrar en organizaciones de todos los tamaños e incluso en su hogar.

Ejemplos:

- Cuando compras objetos por internet.**
- Cuando pagas con tarjeta de crédito.**
- Al hacer las compras del supermercado.**



TIPOS DE BASE DE DATOS



Las bases de datos pueden clasificarse de varias maneras, de acuerdo al contexto que se esté manejando, o la utilidad de la misma.



□ **Según la variabilidad de los datos almacenados:**

- * Bases de datos estáticas
- * Bases de datos dinámicas



BASES DE DATOS ESTÁTICAS

- Éstas son bases de datos de sólo lectura, utilizadas primordialmente para almacenar datos históricos que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo, realizar proyecciones y tomar decisiones.



BASES DE DATOS DINÁMICAS

- Éstas son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización, borrado y adición de datos, además de las operaciones fundamentales de consulta. Un ejemplo de esto puede ser la base de datos utilizada en un sistema de información de una tienda de abarrotes, una farmacia, un videoclub.



□ Según el contenido:

- * Bases de datos bibliográficas.
- * Bases de datos de texto completo.
- * Bases de datos o "bibliotecas" de información química o biológica.



BASES DE DATOS BIBLIOGRÁFICAS

Un registro típico de una base de datos bibliográfica contiene información sobre el autor, fecha de publicación, editorial, título, edición, de una determinada publicación, etc.



BASES DE DATOS DE TEXTO COMPLETO

Almacenan las fuentes primarias, como por ejemplo, todo el contenido de todas las ediciones de una colección de revistas científicas.



BASES DE DATOS O "BIBLIOTECAS" DE INFORMACIÓN QUÍMICA O BIOLÓGICA

Son bases de datos que almacenan diferentes tipos de información proveniente de la química, las ciencias de la vida o médicas. Se pueden considerar en varios subtipos:

- * Las que almacenan secuencias de nucleótidos o proteínas.
- * Las bases de datos de rutas metabólicas.
- * Bases de datos de estructura, comprende los registros de datos experimentales sobre estructuras 3D de biomoléculas.
- * Bases de datos clínicas.
- * Bases de datos bibliográficas (biológicas, químicas, médicas y de otros campos): Medline.



MODELOS DE BASE DE DATOS.



MODELOS DE LAS B.D.D.

Es una descripción de algo conocido como contenedor de datos (algo en donde se guarda la información), así como de los métodos para almacenar y recuperar información de esos contenedores.

Los modelos de datos son abstracciones que permiten la implementación de un sistema eficiente de base de datos, por lo general son algoritmos y conceptos matemáticos.



BASES DE DATOS JERÁRQUICAS

Estas son bases de datos que almacenan su información en una estructura jerárquica.

Son especialmente útiles en el caso de aplicaciones que manejan un gran volumen de información y datos muy compartidos permitiendo crear estructuras estables y de gran rendimiento.

Su principal limitación de este modelo es su incapacidad de representar eficientemente la redundancia de datos



BASE DE DATOS DE RED

Su diferencia fundamental hacia la B.D.D jerárquica es la modificación del concepto de *nodo*, en este tipo de base de dato se permite que un mismo nodo tenga varios padres.

Fue una gran mejora con respecto al modelo jerárquico, ya que ofrecía una solución eficiente al problema de redundancia de datos; pero, aun así, la dificultad que significa administrar la información en este modelo hizo que su principal usuario sean programadores.

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



BASES DE DATOS RELACIONALES

En este modelo, el lugar y la forma en que se almacenen los datos no tienen relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar para un usuario esporádico de la base de datos. La información puede ser recuperada o almacenada mediante consultas que ofrecen una amplia flexibilidad y poder para administrar la información.

El lenguaje más habitual para construir las consultas a bases de datos relacionales es SQL



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, trata de almacenar en la base de datos los objetos completos.

Esta es una base de datos que incorpora todos los conceptos importantes del paradigma de objetos:

- Encapsulación
- Herencia
- Polimorfismo

Una operación se especifica en dos partes. La interfaz (o signatura) y la implementación (o método)



BASES DE DATOS DOCUMENTALES

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

Tesaurus es un sistema de índices optimizado para este tipo de bases de datos.



BASES DE DATOS DEDUCTIVAS

También llamadas bases de datos lógicas.

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.



ESTRUCTURA DE UNA BASE DE DATOS.



- La estructura de una base de datos hace referencia a los tipos de datos, los vínculos o relaciones y las restricciones que deben cumplir esos datos (integridad de datos y redundancia de datos).
- La estructura de una base de datos es diseñada o descripta empleando algún tipo de modelo de datos.



- Un ejemplo a modo de descripción de la estructura de una base de datos puede ser:

ALUMNO:

Numero de alumno (entero de 6 números)

Nombre (cadena de 30 caracteres)

Apellido (cadena de 30 caracteres)

Año de nacimiento (entero de 4 números)

Especialidad (entero de 3 números).



INTEGRIDAD Y SEGURIDAD DE DATOS



INTEGRIDAD DE DATOS

- Integridad: conjunto de seguridades que son utilizadas para mantener los datos correctos.
- Fuente de Error: estas fuentes de error se origina si el programa de entrada de datos no esta validado. Ej: fallas de hardware, actualizaciones incompletas, defectos del software, inserción de datos no válidos, errores humanos.
- Una técnica que usa la BD de una entrada de datos no válida es la validación.



- Validación: es proteger los datos, validar los datos en la entrada de datos. Existen tipos de validaciones:
- Tipo de Dato: es si se define un campo como carácter o char y no puede ingresar números enteros.
- Valor de Dato: si se define un valor entero se puede especificar un rango y no se puede pasar de ese valor.
- Valores Claves / No Nulos: asegura registros únicos y cuyos valores no sean nulos.
- Integridad Referencial: asegura a la BD, que no exista registros hijos sin sus registros padres correspondientes.



SEGURIDAD DE LOS DATOS

- Se presentan cuando no es posible establecer claves de acceso y resguardo en forma uniforme para todo el sistema, facilitando así el acceso a intrusos.
- La seguridad de los datos se puede definir en las siguientes aspectos:
- Objeto a asegurar: el primer objeto a asegurar son los objetos, programas y finalmente al esquema.
- Codificación de Claves: el DBMS provee la seguridad de los Login (usuario y password).
- Control de Acceso: se especifican seguridades contra accesos indicados orientado a personas no autorizada.



GESTORES DE BASES DE DATOS



GESTORES DE BASE DE DATOS

- Son un tipo de software muy específico, dedicado a servir de interfaz entre la base de datos, el usuario y las aplicaciones que la utilizan. Los hechos general de los sistemas de gestor de base de datos es el de manejar de manera clara, sencilla y ordenada un conjunto de datos que posteriormente se convertirán en información relevante para una organización.
- Los sistemas de gestores de base de datos mas usados son: ORACLE, D2 de IBM, SQL SERVER, MYSQL, POSTGRESQL.



OBJETIVOS

- **Abstracción de la información:** Ahorran a los usuarios detalles acerca del almacenamiento físico de los datos.
- **Independencia:** La independencia de los datos consiste en la capacidad de modificar el esquema (físico o lógico) de una base de datos sin tener que realizar cambios en las aplicaciones que se sirven de ella.
- **Seguridad:** La información almacenada en una base de datos puede llegar a tener un gran valor. Los SGBD deben garantizar que esta información se encuentra segura de permisos a usuarios y grupos de usuarios, que permiten otorgar diversas categorías de permisos.

- **Manejo de transacciones:** Una transacción es un programa que se ejecuta como una sola operación. Esto quiere decir que luego de una ejecución en la que se produce una falla es el mismo que se obtendría si el programa no se hubiera ejecutado. Los SGBD proveen mecanismos para programar las modificaciones de los datos de una forma mucho más simple que si no se dispusiera de ellos.
- **Tiempo de respuesta:** Lógicamente, es deseable minimizar el tiempo que el SGBD tarda en darnos la información solicitada y en almacenar los cambios realizados.



VENTAJAS

- Proveen facilidades para la manipulación de grandes volúmenes de datos.
- Simplifican la programación de equipos de consistencia.
- Manejando las políticas de respaldo adecuadas, garantizan que los cambios de la base serán siempre consistentes sin importar si hay errores correctamente.



- Organizan los datos con un impacto mínimo en el código de los programas.
- Bajan drásticamente los tiempos de desarrollo y aumentan la calidad del sistema desarrollado si son bien explotados por los desarrolladores.
- Usualmente, proveen interfaces y lenguajes de consulta que simplifican la recuperación de los datos.



DESVENTAJAS

- Típicamente, es necesario disponer de una o más personas que administren de la base de datos, en la misma forma en que suele ser necesario en instalaciones de cierto porte disponer de una o más personas que administren los sistemas operativos. Esto puede llegar a incrementar los costos de operación en una empresa.
- Si se tienen muy pocos datos que son usados por un único usuario por vez y no hay que realizar consultas complejas sobre los datos, entonces es posible que sea mejor usar una planilla de cálculo.



- Complejidad, los software muy complejos y las personas que vayan a usarlo deben tener conocimiento de las funcionalidades del mismo para poder aprovecharlo al máximo.
- Tamaño, la complejidad y la gran cantidad de funciones que tienen hacen que sea un software de gran tamaño, que requiere de gran cantidad de memoria para poder correr.
- Coste del hardware adicional: los requisitos de hardware para correr un SGBD por lo general son relativamente altos, por lo que estos equipos pueden llegar a costar gran cantidad de dinero.



LENGUAJES DE PROGRAMACIÓN



LENGUAJES DE PROGRAMACIÓN

DELPHI

- Delphi es un entorno de Programación visual orientado a objetos para desarrollo rápido de aplicaciones de propósito general, incluyendo aplicaciones cliente / servidor. Desarrollo de bases de datos multinivel dimensionable, auténtica capacidad de reutilización orientada a objetos y compilador de código original de alto rendimiento.

COBOL

- El Cobol es un lenguaje de alto nivel compilado y orientado hacia la gestión de empresas y de ahí viene su nombre. Lenguaje común orientado hacia los negocios. Aunque existen diferentes versiones es el más estándar existiendo en leves diferencias entre una y otra versión.



PASCAL

- El Pascal es un lenguaje compilado de orientación general pseudocientífica, se trata de un lenguaje de finales de la década de los 70 que surgió como alternativa al Basic y al Fortran. Al igual que el Cobol se trata de un lenguaje estructurado

LENGUAJE C

- El Lenguaje C es un lenguaje de nivel medio, es decir, sin ser un lenguaje de alto nivel como COBOL, BASIC o Pascal, tampoco es un Lenguaje Ensamblador.

HTML

- HTML (Hyper Text Markup Language) es un lenguaje sencillo que permite describir hipertexto, es decir, texto presentado de forma estructurada, con enlaces (hyperlinks) que conducen a otros documentos o fuentes de información relacionadas, y con inserciones multimedia (gráficos, sonido...).

BASE DE DATOS PARALELAS



BASES DE DATOS PARALELAS

- Hace 15 años los sistemas paralelos de base de datos han estado casi descartados incluso por algunos de sus mas firmes defensores.
- Actualmente están comercializados con éxito por prácticamente todos los fabricantes de bases de datos.



BASE DE DATOS PARALELAS

- Un SGBD que se ejecuta sobre múltiples procesadores y discos que han sido diseñados para ejecutar operaciones en paralelo, cuando sea posible, con el propósito de mejorar el rendimiento.



- Los sistemas paralelos de base de datos constan de varios procesadores y varios discos conectados a través de una red de interconexión de alta velocidad. Para medir el rendimiento de los sistemas de base de datos existen 2 medidas principales:

- La productividad

- El tiempo de respuesta



- **La productividad** (throughput) que se entiende como el número de tareas que pueden completarse en un intervalo de tiempo determinado.
- **El tiempo de respuesta** (response time) que es la cantidad de tiempo que necesita para completar una única tarea a partir del momento en que se envíe.



ARQUITECTURA

Existen cuatro arquitecturas de sistemas paralelos:

- **De memoria compartida:** Todos los procesadores comparten una memoria común.
- **De discos compartidos:** Todos los procesadores comparten un conjunto de discos común.
- **Sin compartimiento:** Los procesadores no comparten ni memoria ni disco.
- **Jerárquica:** Este modelo es un híbrido de las arquitecturas anteriores.



FACTORES NEGATIVOS DEL PARALELISMO

Estos pueden atenuar tanto la ganancia de velocidad como la ampliabilidad:

- **Costes de inicio.** El inicio de un único proceso lleva asociado un coste de inicio.
- **Interferencia.** Como los procesos que se ejecutan en un sistema paralelo acceden con frecuencia a recursos compartidos, pueden sufrir un cierto retardo como consecuencia de la esta.
- **Sesgo.** Al dividir cada tarea en un cierto número de pasos paralelos se reduce el tamaño del paso medio. Normalmente es difícil dividir una tarea en partes exactamente iguales, entonces se dice que la forma de distribución de los tamaños es sesgada.



PARALELISMO DE E/S

- Las relaciones se dividen entre los discos disponibles para que se pueda realizar la recuperación de datos mas rápidamente. Hay 3 técnicas de división:
 - División por turno rotatorio
 - División por asociación
 - División por rangos



- **Turno rotatorio:** asegura una distribución de las tuplas entre los discos, por ello, cada disco tiene aproximadamente el mismo número de tuplas que los demás.
- **División por asociación:** Se escogen varios atributos del esquema de la relación y se designan como atributos de división. Se escoge una función de asociación cuyo rango es $\{0,1,\dots,n-1\}$.
- **División por rangos:** Se distribuye rangos continuos de valores de los atributos a cada disco.



PARALELISMO ENTRE CONSULTAS

- Se ejecutan en paralelo entre si diferentes consultas o transacciones. La productividad de transacciones puede aumentarse con esta forma de paralelismo.
- Es la forma mas sencilla de paralelismo que se permite en los sistemas de bases de datos.



PARALELISMO EN CONSULTAS

- Se refiere a la ejecución en paralelo de una única consulta en varios procesadores y discos. Hay 2 tipos de paralelismo en consultas:
 - Paralelismo en operaciones
 - Paralelismo entre operaciones



- **Paralelismo en operaciones:** se utiliza para ejecutar operaciones relacionales, como la ordenación, selección, proyección, y reunión.
- **Paralelismo entre operaciones:** se puede acelerar el procesamiento de consultas ejecutando en paralelo las diferentes operaciones de las expresiones de las consultas.



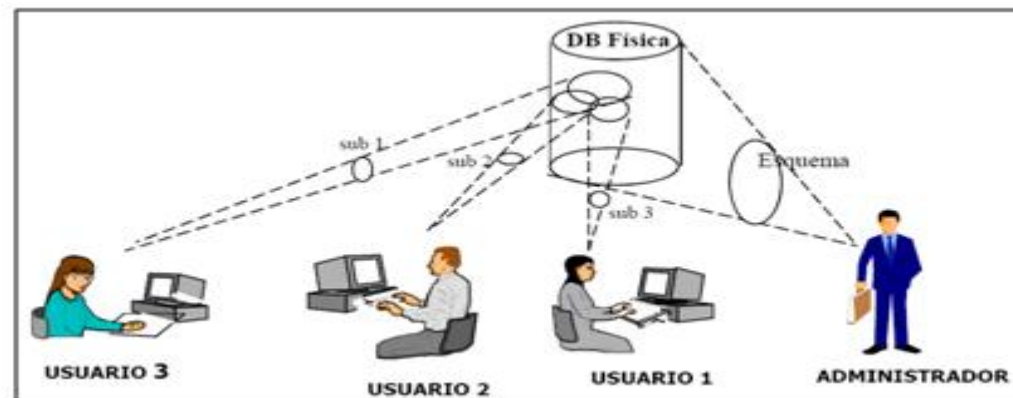
BASE DE DATOS DISTRIBUIDAS



BASE DE DATOS DISTRIBUIDA

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas entre diferentes sitios interconectados por una red de comunicaciones, los cuales tienen la capacidad de procesamiento autónomo lo cual indica que puede realizar operaciones locales o distribuidas. Un sistema de Bases de Datos Distribuida (SBDD) es un sistema en el cual múltiples sitios de bases de datos están ligados por un sistema de comunicaciones de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si los datos estuvieran siendo accedidos de forma local.

En un sistema distribuido de bases de datos se almacenan en varias computadoras.



VENTAJAS

- Refleja una estructura organizacional - los fragmentos de la base de datos se ubican en los departamentos a los que tienen relación.
- Autonomía local - un departamento puede controlar los datos que le pertenecen.
- Disponibilidad - un fallo en una parte del sistema solo afectará a un fragmento, en lugar de a toda la base de datos.



- Rendimiento - los datos generalmente se ubican cerca del sitio con mayor demanda, también los sistemas trabajan en paralelo, lo cual permite balancear la carga en los servidores.
- Economía - es más barato crear una red de muchas computadoras pequeñas, que tener una sola computadora muy poderosa.
- Modularidad - se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).



DESVENTAJAS

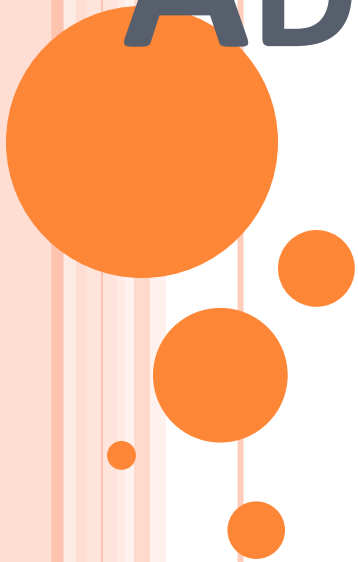
- Complejidad - Se debe asegurar que la base de datos sea transparente, se debe lidiar con varios sistemas diferentes que pueden presentar dificultades únicas.
- Seguridad - se debe trabajar en la seguridad de la infraestructura así como cada uno de los sistemas.
- Integridad - Se vuelve difícil mantener la integridad, aplicar las reglas de integridad a través de la red puede ser muy caro en términos de transmisión de datos.



- Falta de experiencia - las bases de datos distribuidas son un campo relativamente nuevo y poco común por lo cual no existe mucho personal con experiencia o conocimientos adecuados.
- Diseño de la base de datos se vuelve más complejo, el diseño de una base de datos distribuida debe considerar la fragmentación, replicación y ubicación de los fragmentos en sitios específicos.



USUARIOS Y ADMINISTRADORES



USUARIOS DE BASE DE DATOS

- Las personas que trabajan con una base de datos se pueden catalogar como usuarios o administradores de base de datos.



Hay 4 tipos de usuarios:

- Usuarios normales
- Programadores de aplicaciones
- Usuarios sofisticados
- Usuarios especializados



USUARIOS NORMALES

- Son usuarios no sofisticados que interactúan con el sistema mediante la innovación de alguno de los programas de aplicación permanente que se ha escrito previamente.
- Los usuarios normales pueden también leer informes generados de la base de datos.



PROGRAMADORES DE APLICACIONES

- Son profesionales informáticos que escriben programas de aplicación.
- Los programadores de aplicaciones pueden elegir entre muchas herramientas para desarrollar interfaces de usuario.



- Las herramientas de desarrollo rápido de aplicaciones (DRA) permiten al programador construir formularios e informes sin construir un programa.
- Hay lenguajes llamados “*lenguajes de cuarta generación*” que incluyen características especiales para facilitar la generación de formularios y presentación de datos en la pantalla.



USUARIOS SOFISTICADOS

- Interactúan con el sistema sin programas escritos. En su lugar, ellos forman sus consultas en un lenguaje de consulta de base de datos.



USUARIOS ESPECIALIZADOS

- Son usuarios sofisticados que escriben aplicaciones de base de datos especializadas que no son adecuadas en el marco de procesamiento de datos tradicional.



Entre estas aplicaciones están:

- Sistemas de diseño asistido por computador
- Sistemas de bases de conocimientos y sistemas expertos
- Sistemas que almacenan los datos con tipos de datos complejos
- Sistemas de modelado de entorno



ADMINISTRADOR DE BASE DE DATOS

- Es la persona que llevan un control centralizado sobre el sistema, tanto de los datos como de los programas que acceden a esos datos.



Las funciones del administrador de base de datos incluyen las siguientes:

- **Definición del esquema:** el administrador crea el esquema original de la base de datos escribiendo un conjunto de instrucciones de definición de datos
- **Definición de la estructura y del método de acceso.**



- **Modificación del esquema y de la organización:** los administradores realizan cambio en el esquema y en la organización para reflejar las necesidades cambiantes de la organización o para alterarla y mejorar el rendimiento.



- **Concesión de autorización para el acceso a los datos:** permite al administrador determinar a que parte de la base de datos puede acceder cada usuario .
- **Mantenimiento rutinario:** algunos ejemplos de actividades rutinarias de mantenimiento del administrador son:



- Copia de seguridad periódica de la base de datos, para prevenir la pérdida de datos en caso de desastres.
- Asegurarse de que haya suficiente espacio libre en disco para las operaciones normales y aumentar el espacio según sea necesario.
- Supervisión de los trabajos que se ejecuten en la base de datos

