

4. Introducción a las Bases de Datos

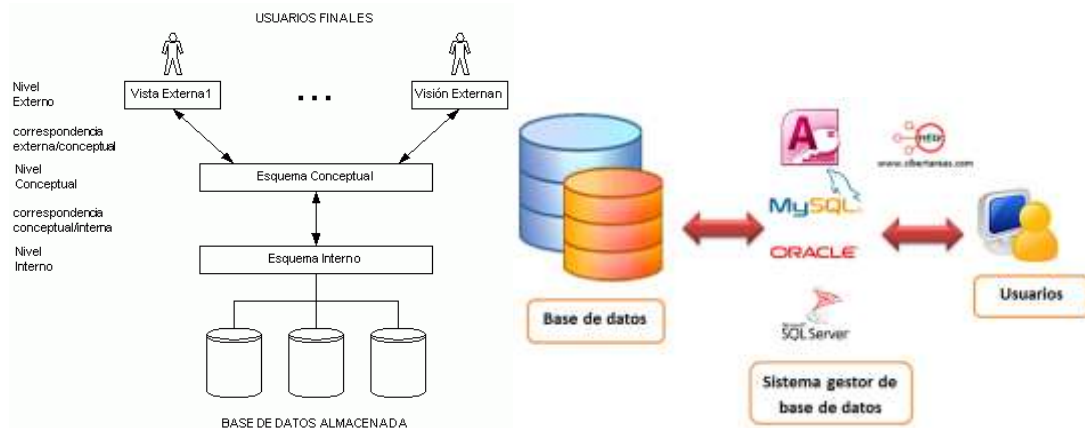
Por lo general, una eficiente administración de datos requiere el uso de una base de datos. Una base de datos es una estructura computarizada compartida e integrada que guarda un conjunto de:

- ❖ **Datos del usuario final**, son datos sin elaborar que son de interés para el usuario final.
- ❖ **Metadatos**, o datos a cerca de datos, por medio de los cuales los datos del usuario son integrados y manejados. Los metadatos dan información sobre los datos (nombre, tipo de datos, si puede dejarse vacío...)

Para implementar una base de datos y administrar su contenido, se requiere un sistema de administración de bases de datos (DBMS), que sirve como intermediario entre el usuario y la BD.

4.1. Sistemas Gestores de Bases de Datos

Un **Sistema Gestor de Base de Datos (SGBD)** es el software que permite gestionar bases de datos, ocultando la física de la misma y permitiendo manejarla desde un nivel más conceptual. Dicho software permite separar las aplicaciones (los programas) de los datos; de modo que los programas negocian con el SGBD el acceso a los datos.

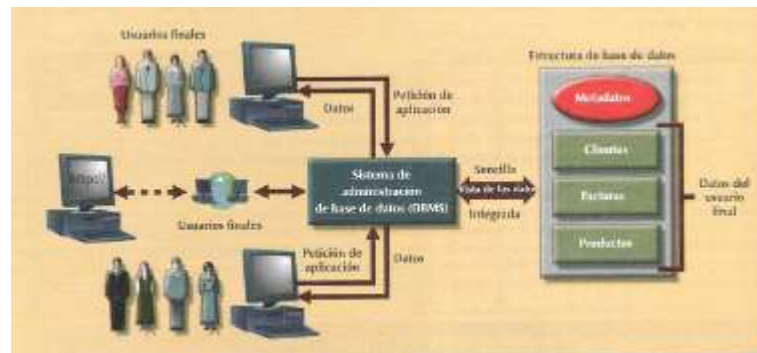


4.2. Función y Ventajas de SGBD

El **SGBD** sirve como intermediario entre el usuario final y la BD. La estructura de la BD está guardada como un conjunto de archivos y la única forma de tener acceso a los datos de estos archivos es a través del SGBD.

El SGBD:

- ❖ Presenta al usuario final una sola vista integrada de los datos de la BD.
- ❖ Recibe todas las peticiones de aplicación y las traduce en las complejas operaciones requeridas para cumplirlas.
- ❖ Oculta gran parte de la complejidad interna de la BD de los programas de aplicación a los usuarios.
- ❖ El programa de aplicación puede ser escrito usando un LP como .NET, Java, C o PHP entre otros.



Tener un SGBD entre el usuario final y las aplicaciones ofrece algunas **ventajas** como:

- ❖ **Los datos se comparten mejor.**
- ❖ **Mejor seguridad de datos.**
- ❖ **Mejor integridad de datos.**
- ❖ **Mínima inconsistencia de datos**, hay inconsistencia de datos cuando diferentes versiones de éstos aparecen en lugares diferentes.
- ❖ **Mejora acceso a datos**, el SGBD hace posible producir respuestas rápidas de las consultas.

Todas estas ventajas permiten una **mejora en la toma de decisiones** y **aumento en la productividad**.

4.3. Tipos de Bases de Datos

Un SGBD puede soportar numerosos tipos de bases de datos. Las BD se pueden clasificar según el número de usuarios, ubicación(es) de la BD y el tipo y grado de uso esperado.

Desde el punto de vista del número de usuario.

- ❖ **BD de un solo usuario o BD de escritorio**, sólo soporta un usuario a la vez.
- ❖ **BD de múltiples usuarios**, soporta a varios usuarios de forma simultánea:
 - Si número de usuarios no excede de 50 se denomina **BD de grupo de trabajo**.
 - Más de 50 usuarios se denomina **BD empresarial**.
 -
- ❖ Desde el punto de vista de la ubicación de la BD:
 - ❖ **BD centralizada**, una BD que soporte datos ubicados en un solo sitio.
 - ❖ **BD distribuida**, una BD que soporte datos en diferentes lugares.
- ❖ Desde el punto de vista del modo en que se usarán:
 - ❖ **BD operacional, de transacciones o de producción**, son BD diseñadas para soportar operaciones diarias que deben ser registradas con toda precisión y de inmediato, por ejemplo, operaciones de venta de un producto o pagos y compras de materias primas.
 - ❖ **Almacén de datos**, se concentra en guardar datos empleados para generar información requerida.
- ❖ Desde el punto de vista del grado en que están estructurados los datos:
 - ❖ **BD con datos no estructurados**, los datos están almacenados sin refinar, es decir, en el formato en el que se recogieron.
 - ❖ **BD con datos estructurados**, son el resultado de tomar datos no estructurados y formatearlos.

- ❖ **BD con datos semiestructurados**, datos sin estructurar al que se le ha aplicado algún tipo de formato, por ejemplo las BD XML.

Producto	Nº de usuarios		Ubicación de los datos		Uso de datos		XML
	Un usuario	Multiusuario	Centralizada	Distribuida	Operacional	Almacen de datos	
MS Access	X		X		X		
MS SQL Server	X	X	X	X	X	X	X
IBM DB2	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X
Oracle RDBMS	X	X	X	X	X	X	X

5. ¿Por qué es importante el diseño de BD?

El **diseño de la BD** se refiere a las actividades que se concentran en el diseño de la estructura de la BD que se usará para guardar y administrar datos del usuario final. Una BD que satisfaga todas las necesidades del usuario no es tarea fácil, de hecho, el diseño de la BD es un aspecto tan importante que aprender buenas técnicas para el diseño nos llevará la mayor parte del curso. Un SGBD funcionará mal con una BD mal diseñada.

6. Evolución del procesamiento de datos de un Sistema de Archivos

Entender qué es una BD, qué hace y la forma apropiada de usarla puede ser más fácil de aclarar si se considera lo que no es una BD.

6.1. Sistemas de Archivos Manuales

Históricamente, una organización proponía para guardar los datos, sistemas manuales hechos con papel y lápiz. Los papeles estaban organizados en carpetas de archivos. Mientras el conjunto de datos era relativamente pequeño y los financieros tenían pocas necesidades de informar, el sistema cumplía con su función como depósito de datos. No obstante, cuando las organizaciones crecieron y las necesidades de informar se hicieron más complejas, dar seguimiento a datos con un sistema manual se hizo más difícil. Por tanto, las empresas buscaron ayuda en la tecnología de las computadoras.

6.2. Sistemas de Archivos Computarizados

Elaborar informes a partir de sistemas manuales era lento y tedioso. Por tanto un especialista en procesamiento de datos (**DP**) era contratado para crear un sistema computarizado que rastrearía datos y elaboraría los informes solicitados.

C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	100.00	05-Apr-2010
Leone K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	250.00	16-Jun-2010
Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	150.00	29-Jan-2011
Paul F. Cowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	300.00	14-Oct-2010
Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	100.00	28-Dec-2010
Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	850.00	22-Sep-2010
James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	120.00	25-Mar-2011
George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	250.00	17-Jul-2010
Anne G. Farniss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	100.00	03-Dec-2010
Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	500.00	14-Mar-2011

C_NAME	= Nombre del cliente	A_NAME	= Nombre del agente
C_PHONE	= Teléfono del cliente	A_PHONE	= Teléfono del agente
C_ADDRESS	= Dirección del cliente	TP	= Tipo de seguro
C_ZIP	= Código postal del cliente	AMT	= Cantidad de póliza de seguro, en miles de dólares
		REN	= Fecha de renovación de seguro

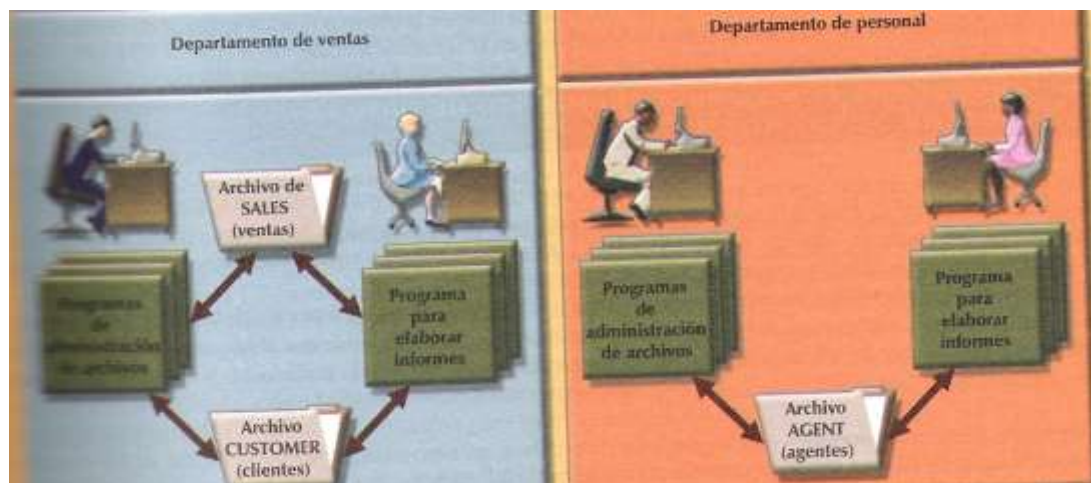
Contenido archivo de CLIENTE

La descripción de archivos computarizados requiere un vocabulario específico: **datos**, **campo**, **registro** y **archivo**. Cuando los usuarios deseaban datos del archivo, hacían una petición al DP éste diseñaba un programa para recuperar los datos que entregaba en forma de informe impreso al usuario.

A_NAME	A_PHONE	A_ADDRESS	ZIP	HIRED	YTD_PAY	YTD_FIT	YTD_FICA	YTD_SLS	DEP
Alex B. Alby	713-228-1249	123 Toll, Nash, TN	37119	01-Nov-2000	26566.24	6641.56	2125.30	132737.75	3
Leah F. Hahn	615-882-1244	334 Main, Fox, KY	25246	23-May-1986	32213.78	8053.44	2577.10	138967.35	0
John T. Okon	615-123-5589	452 Elm, New, TN	36155	15-Jun-2005	23198.29	5799.57	1855.86	127093.45	2

A_NAME	= Nombre del agente	YTD_PAY	= Pago acumulado
A_PHONE	= Teléfono del agente	YTD_FIT	= Impuesto federal acumulado pagado
A_ADDRESS	= Dirección del agente	YTD_FICA	= Impuestos de Seguro Social acumulados pagados
ZIP	= Código postal del agente	YTD_SLS	= Ventas acumuladas
HIRED	= Fecha de contratación del agente	DEP	= Número de empleados

Contenido archivo de AGENTE



Sistema sencillo de archivos

Este método para organizar y manejar datos fue una mejora definitiva y fue utilizado durante más de dos décadas. A medida que más y más archivos computarizados se elaboraron, los problemas y limitaciones con este tipo de sistema de archivos se hicieron evidentes. Algunos de ellos son:

- ❖ **Largos tiempos de desarrollo**, la tarea más sencilla requería un extenso programa lo que dificultaba obtener respuestas rápidas.
- ❖ **Administración de un sistema complejo**. La administración de un sistema se hace más difícil a medida que se amplía el número de archivos.
- ❖ **Falta de seguridad y compartición limitada de datos**, el compartir datos entre múltiples usuarios dispersos desde el punto de vista geográfico introduce numerosos riesgos de seguridad.
- ❖ **Programación extensa**, cualquier modificación en la estructura de archivos existentes puede ser difícil en un ambiente de sistema de archivos.

6.2.1. Dependencia Estructural y de Datos

Un sistema de archivos presenta:

- ❖ **Dependencia estructural**, significa que cualquier modificación en la estructura del archivo (añadir un campo) supone modificar todos los programas para que funcionen con la nueva estructura. Por el contrario, existe **independencia estructural** cuando es posible hacer cambios en la estructura de los archivos sin afectar la capacidad del programa para tener acceso a los datos.
- ❖ **Dependencia de datos**, significa que cualquier modificación en las características de los datos, requieren modificaciones en todos los programas que tienen acceso al archivo. Por el contrario, existe independencia de datos cuando es posible hacer cambios en las características de almacenamiento de datos sin afectar la capacidad del programa de aplicación para tener acceso a los datos. La importancia práctica de la dependencia de datos es la diferencia entre el **formato lógico de datos** (el modo en que una persona ve los datos) y el **formato físico de datos** (el modo en que la computadora debe trabajar con los datos).

6.2.2. Redundancia de Datos

La estructura del sistema de archivos hace difícil combinar datos desde múltiples fuentes y su falta de seguridad lo hace vulnerable. La estructura de la organización promueve el almacenamiento de los mismos datos básicos en diferentes lugares, es lo que los profesionales denominan **islas de información**. Como es improbable que los datos almacenados en diferentes lugares sean actualizados siempre de manera consistente, es frecuente que las islas de información contengan diferentes versiones de los mismos datos. Por ejemplo los nombres y números de teléfono de los agentes aparecen en los archivos CUSTOMER y AGENT. Sólo se necesita una copia correcta de estos datos. Hacer que aparezcan en más de un lugar produce **redundancia de datos**. Existe **redundancia de datos** cuando los mismos datos se guardan innecesariamente en lugares diferentes.

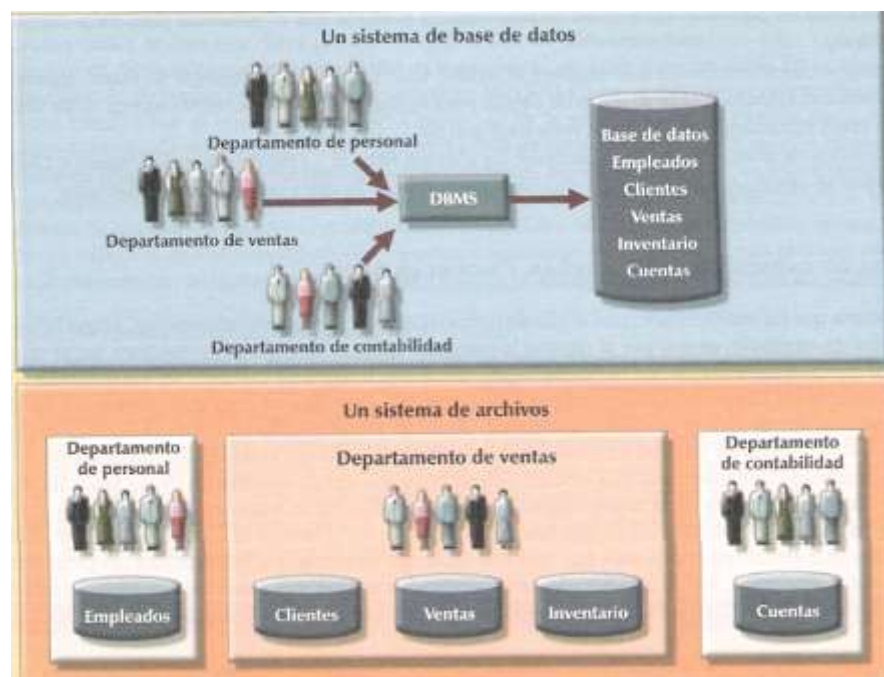
La redundancia de datos no controlada de datos prepara el escenario para:

- ❖ **Deficiencia en la seguridad de los datos**, tener múltiples copias de datos aumenta las probabilidades de que estos sean susceptibles a un acceso no autorizado.
- ❖ **Inconsistencia de datos**, existe cuando diferentes y conflictivas versiones del mismo dato se encuentran en lugares diferentes.
- ❖ **Anomalías de datos**, se crea cuando no todos los cambios requeridos en los datos redundantes se hacen correctamente.

- **Anomalías de actualización**, si un agente tiene un nuevo número de teléfono, ese número debe introducirse en cada uno de los registros del archivo CUSTOMER en los que se encuentra el número de teléfono antiguo.
- **Anomalías de inserción**, si solo existiera el archivo CUSTOMER, para agregar un nuevo agente se agregaría también una entrada ficticia de datos de cliente.
- **Anomalías de borrado**, si elimina un cliente también se eliminarán los datos de su agente.

7. Sistemas de Bases de Datos

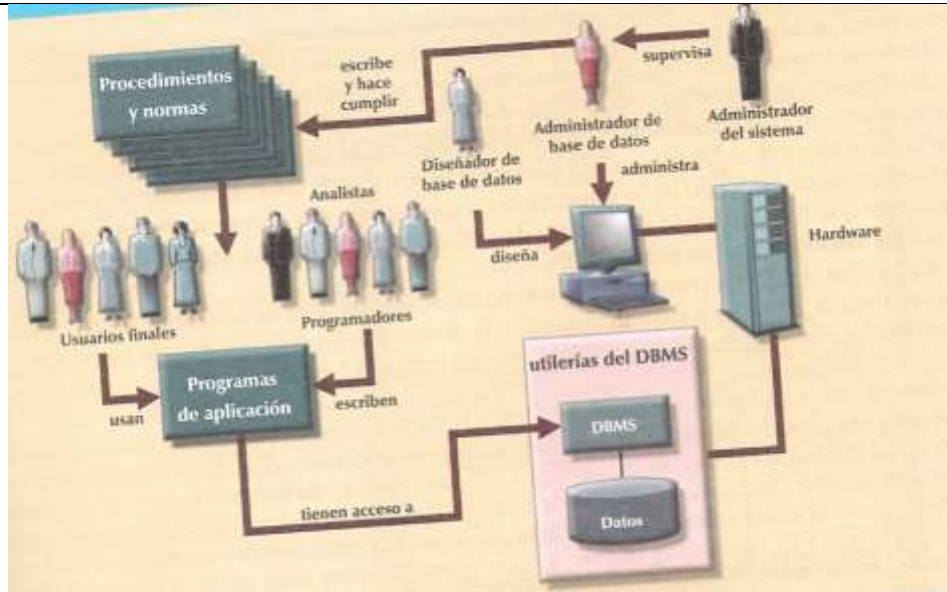
Los problemas propios de los sistemas de archivos hacen bastante deseable un sistema de BD. A diferencia del sistema de archivos, con sus numerosos archivos separados y sin relación entre sí, el sistema de BD está formado por datos relacionados lógicamente en un solo depósito lógico.



Comparación de una BD con un sistema de archivos

7.1. El ambiente de un Sistema de Bases de Datos

El término de base de datos se refiere a una organización de componentes que define y regula captura, almacenamiento, administración y uso de datos dentro de un ambiente de BD. El sistema de BD está compuesto de cinco partes principales: **hardware, software, personas, procedimientos y datos**.



- ❖ **Hardware:** se refiere a todos los equipos físicos del sistema: computadoras, equipos de almacenamiento, periféricos, equipos de redes y otros equipos como cajeros automáticos, lectores de identificación ...
- ❖ **Software:** el software que se identifica con más facilidad es el SGBD, pero para que el sistema de BD funcione plenamente se requieren 3 tipos de software:
 - **Sistema operativo**, maneja todos los componentes del hardware y hace posible que todo el software se ejecute en las computadoras. Microsoft Windows, Linux, Mac OS....
 - **Software del SGBD**, maneja la base de datos dentro del sistema de BD. SQL Server de MS, MySQL, Oracle...
 - **Programas de aplicación**, se usan para tener acceso y manipular datos que se encuentran dentro del SGBD.
- ❖ **Personas**, incluye a todos los usuarios del sistema de BD. A partir de las funciones que desarrollan dentro del sistema, se pueden identificar 5 tipos de usuarios:
 - Los **administradores del sistema**, supervisan las operaciones generales del sistema.
 - Los **administradores de la BD (DBA)**, manejan el SGBD y aseguran que la BD se encuentre funcionando correctamente.
 - Los **diseñadores de BD**, proyectan la estructura de la BD.
 - Los **analistas y programadores**, diseñan y ponen en práctica los programas de aplicación.
 - Los **usuarios finales**, son las personas que usan los programas de aplicación para ejecutar las tareas diarias de la organización.
- ❖ **Procedimientos**, son las instrucciones y reglas que gobiernan la BD.
- ❖ **Datos**, comprende el conjunto de datos almacenados en la BD.

7.2. Funciones del SGBD

Un **SGBD** realiza **funciones** importantes que garantizan la integridad y consistencia de los datos de la BD. Casi todas estas funciones son transparentes al usuario final y la mayor parte se pueden realizar con el SGBD.

- ❖ **Administración de un diccionario de datos**, se denomina **diccionario de datos** y contiene información que describe los datos de la base de datos (metadatos) y debe ser accesible por los usuarios de la base de datos. Normalmente, un diccionario de datos describe entre otras cosas:
 - ✓ Nombre, tipo y tamaño de los datos.
 - ✓ Relaciones entre los datos.
 - ✓ Restricciones de integridad.
 - ✓ Usuarios autorizados a acceder a los objetos de la base de datos.
 - ✓ Estadística de utilización, tales como la frecuencia de las transacciones y el número de accesos realizados a los objetos de la base de datos.
- ❖ **Administración de almacenamiento de datos**, crea y maneja las complejas estructuras requeridas para el almacenamiento de éstos datos.
- ❖ **Transformación y presentación de los datos**, transforma los datos introducidos para adaptarse a las estructuras requeridas.
- ❖ **Administración de seguridad**, crea un sistema de seguridad que hace cumplir la seguridad del usuario y la privacidad de los datos.
- ❖ **Control de acceso a múltiples usuarios**, para dar integridad y consistencia a los datos, utiliza algoritmos complejos que aseguran que el acceso concurrente de múltiples usuarios no compromete la integridad de los datos.
- ❖ **Lenguajes de acceso a datos**, proporciona el acceso a los datos a través de un lenguaje de consulta (SQL).
- ❖ **Interfaz de comunicación de una BD**, los SGBD actuales aceptan peticiones de un usuario hechas a través de múltiples y diferentes ambientes de red.

Aun cuando los sistemas de bases de datos dan considerables ventajas sobre anteriores métodos de administración de datos, también tienen **desventajas** considerables:

- ❖ **Costos más altos**, los sistemas requieren hardware y software complejos y personal altamente cualificado.
- ❖ **Complejidad de administración**, los sistemas interactúan con numerosas tecnologías y tienen un considerable impacto en los recursos y cultura de la compañía.
- ❖ **Mantener una actualización general**, para maximizar la eficiencia de los sistemas es necesario mantener actualizado el sistema.
- ❖ **Dependencia de vendedores**, dada la fuerte inversión en tecnología y capacitación de personal, las empresas son reacias a cambiar de vendedores, en consecuencia es menos probable que los vendedores ofrezcan precios ventajosos a clientes existentes.
- ❖ **Ciclos frecuentes de actualización/reemplazo**.

ÍNDICE

4.	Introducción a las Bases de Datos.....	1
4.1.	Sistemas Gestores de Bases de Datos	1
4.2.	Función y Ventajas de SGBD	1
4.3.	Tipos de Bases de Datos	2
5.	¿Por qué es importante el diseño de BD?	3
6.	Evolución del procesamiento de datos de un Sistema de Archivos	3
6.1.	Sistemas de Archivos Manuales	3
6.2.	Sistemas de Archivos Computarizados.....	3
6.2.1.	Dependencia Estructural y de Datos	5
6.2.2.	Redundancia de Datos.....	5
7.	Sistemas de Bases de Datos	6
7.1.	El ambiente de un Sistemas de Bases de Datos	6
7.2.	Funciones del SGBD.....	8