

## 1. PERSPECTIVA DEL ÁREA DE BASE DE DATOS

### Introducción

#### 1.1 Funciones de un DBMS

#### 1.2 Usuarios de un DBMS: DBA

#### 1.3 Desarrolladores y usuarios finales

#### 1.4 Componentes de un DBMS

### Introducción

Antes de empezar con DBMS es muy importante tener muy bien definidos los conceptos básicos para poder hacer un adecuado manejo, diseño y/o especificación del DBMS.

Empecemos por definir DBMS, estos son programas denominados Sistemas Gestores de Base de Datos, abreviado SGBD, **en inglés Data Base Management System (DBMS)** que permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada. Las propiedades de estos SGBD o DBMS, así como su utilización y administración, se estudian dentro del ámbito de la informática.

Es importante aquí establecer que dentro del área de la Informática es convencional utilizar como terminología a los acrónimos como: SGBD o DBMS que para uniformizar emplearemos la terminología en inglés e iremos definiendo conforme avancemos. Lo anterior es considerando que los mayores desarrollos en Informática se establecen por lo general en inglés, en este caso usaremos siempre DBMS, DB, IT, SQL, etc...

### Base de Datos (database DB)

Es un almacén de datos relacionados con diferentes modos de organización. Una base de datos representa algunos aspectos del mundo real, aquellos que le interesan al usuario. Y que almacena datos con un propósito específico. Con la palabra “datos” se hace referencia a hechos conocidos que pueden registrarse, como ser números telefónicos, direcciones, nombres, etc.

### El Concepto de Datos

Datos son los hechos que describen sucesos y entidades. Datos es una palabra en plural que se refiere a más de un hecho. A un hecho simple se le denomina “data-ítem” o elemento de dato. Los datos son

comunicados por varios tipos de símbolos tales como las letras del alfabeto, números, movimientos de labios, puntos y rayas, señales con la mano, dibujos, etc.

Lo importante es considerar que estos símbolos se pueden ordenar y reordenar de forma utilizable y se les denomina **información**.

Los datos son símbolos que describen condiciones, hechos, situaciones o valores. Los datos se caracterizan por no contener ninguna información. Un dato puede significar un número, una letra, un signo ortográfico o cualquier símbolo que represente una cantidad, una medida, una palabra o una descripción.

La importancia de los datos está en su capacidad de asociarse dentro de un contexto para convertirse en información. Por si mismos los datos no tienen capacidad de comunicar un significado y por tanto no pueden afectar el comportamiento de quien los recibe. Para ser útiles, los datos deben convertirse en información para ofrecer un significado, conocimiento, ideas o conclusiones

El concepto de Información:

La información no es un dato conjunto cualquiera de ellos. Es más bien una colección de hechos significativos y pertinentes, para el organismo u organización, que los percibe. La definición de información es la siguiente: Información es un conjunto de datos significativos y pertinentes que describan sucesos o entidades.

**DATOS SIGNIFICATIVOS.** Para ser significativos, los datos deben constar de símbolos reconocibles, estar completos y expresar una idea no ambigua.

Los símbolos de los datos son reconocibles cuando pueden ser correctamente interpretados. Muchos tipos diferentes de símbolos comprensibles se usan para transmitir datos. La integridad significa que todos los datos requeridos para responder a una pregunta específica están disponibles. Por ejemplo, un marcador de béisbol debe incluir el tanteo de ambos equipos. Si se oye el tanteo "New York 6" y no oyes el del oponente, el anuncio será incompleto y sin sentido.

Los datos son inequívocos cuando el contexto es claro. Por ejemplo, el grupo de signos 2-x puede parecer "la cantidad 2 menos la cantidad

desconocida llamada  $x$ ” para un estudiante de álgebra, pero puede significar “2 barra  $x$ ” a un vaquero que marca ganado. Tenemos que conocer el contexto de estos símbolos antes de poder conocer su significado.

Otro ejemplo de la necesidad del contexto es el uso de términos especiales en diferentes campos especializados, tales como la contabilidad. Los contables utilizan muchos términos de forma diferente al público en general, y una parte de un aprendizaje de contabilidad es aprender el lenguaje de contabilidad. Así los términos Debe y Haber pueden significar para un contable no más que “derecha” e “izquierda” en una contabilidad en T, pero pueden sugerir muchos tipos de ideas diferentes a los no contables.

**DATOS PERTINENTES.** Decimos que tenemos datos pertinentes (relevantes) cuando pueden ser utilizados para responder a preguntas propuestas.

Disponemos de un considerable número de hechos en nuestro entorno, solo los hechos relacionados con las necesidades de información son pertinentes. Así la organización selecciona hechos entre sucesos y entidades particulares para satisfacer sus necesidades de información.

Desde épocas prehistóricas el humano ha tenido la necesidad de ir almacenando datos y podemos remitirnos a las pinturas rupestres en donde se fueron registrando símbolos y dibujos (datos) que finalmente representaban una información necesaria para nuestros antepasados.

Conforme fue evolucionando el ser humano el volumen de los datos registrados y cada vez más necesarios fue creciendo y de tal forma que se empezaron a desarrollar sistemas básicos para el almacenamiento, que conjunto con el desarrollo de las lenguas y posteriormente de sistemas de escritura fue necesario el ir “automatizando” el almacenamiento y procesamiento de estos así fueron surgiendo separadamente el ábaco, la imprenta y la primera calculadora (Pascalina) solo por mencionar brevemente algunos. Lo importante es tener muy claro que cada vez es más importante el almacenamiento y procesamiento de los datos (DB y DBMS) y para las DB comprender un mejor los DBMS haremos un breve resumen de su desarrollo más reciente.

## Historia de los sistemas de bases de datos

Los predecesores de los sistemas de bases de datos fueron los sistemas de ficheros, No hay un momento concreto en que los sistemas de ficheros hayan cesado y hayan dado comienzo los sistemas de bases de datos. De hecho, todavía existen sistemas de ficheros en uso.

El uso de sistemas de bases de datos automatizadas, se desarrollo a partir de la necesidad de almacenar grandes cantidades de datos, para su posterior consulta, producidas por las nuevas industrias que creaban gran cantidad de información.

Herman Hollerit (1860-1929) fue denominado el primer ingeniero estadístico de la historia, ya que invento una computadora llamada "Máquina Automática Perforadora de Tarjetas. Para hacer el censo de Estados Unidos en 1880 se tardaron 7 años para obtener resultados, pero Herman Hollerit en 1884 creo la máquina perforadora, con la cual, en el censo de 1890 dio resultados en 2 años y medio, donde se podía obtener datos importantes como número de nacimientos, población infantil y número de familias. La máquina uso sistemas mecánicos para procesar los datos de las tarjetas y para tabular los resultados.

A diferencia con la maquina de Babbage, que utilizaba unas tarjetas similares, estas se centraban en dar instrucciones a la máquina. En el invento de Herman Hollerit, cada perforación en las tarjetas representaba un número y cada dos perforaciones una letra, cada tarjeta tenia capacidad para 80 variables. La máquina estaba compuesta por una perforadora automática y una lectora, la cual por medio de un sistema eléctrico leía los orificios de las tarjetas, esta tenía unas agujas que buscaban los orificios y al tocar el plano inferior de mercurio enviaba por medio del contacto eléctrico los datos a la unidad.

Este invento disparó el desarrollo de la tecnología, la industria de los computadores, abriendo así nuevas perspectivas y posibilidades hacia el futuro.

### Década de 1950

En este lapso de tiempo se da origen a las cintas magnéticas, las cuales sirvieron para suplir las necesidades de datos de las nuevas industrias. Por medio de este mecanismo se empezó a automatizar los datos de las nóminas, como por ejemplo el aumento de salario.

Consistía en leer una cinta o más y pasar los datos a otra, y también se podían pasar desde las tarjetas perforadas. Simulando un sistema de

Backup, que consiste en hacer una copia de seguridad o copia de respaldo, para guardar en un medio extraíble los datos importantes. La nueva cinta a la que se transfiere los datos pasa a ser una cinta maestra. Estas cintas solo se podían leer secuencial y ordenadamente.

#### Década de 1960

El uso de los discos en ese momento fue un adelanto muy efectivo, ya que por medio de este soporte se podía consultar los datos directamente, esto ayudo a ahorrar tiempo. No era necesario saber exactamente donde estaban los datos en los discos, ya que en milisegundos eran recuperables los datos. A diferencia de las cintas magnéticas, ya no era necesaria la secuencialidad, y este tipo de soporte empieza a ser ambiguo.

Los discos dieron inicio a las Bases de Datos, de red y jerárquicas, pues los programadores con su habilidad de manipulación de estructuras junto con las ventajas de los discos era posible guardar estructuras de datos como listas y árboles.

#### Década de 1970

Edgar Frank Codd (23 de agosto de 1923 – 18 de abril de 2003), en un artículo “Un modelo relacional de datos para grandes bancos de datos compartidos” (“A Relational Model of Data for Large Shared Data Banks”) en 1970, definió el modelo relacional y publicó una serie de reglas para la evaluación de administradores de sistemas de datos relacionales y así nacieron las bases de datos relacionales.

A partir de los aportes de Codd el multimillonario Larry Ellison desarrollo la base de datos Oracle, el cual es un sistema de administración de base de datos, que se destaca por sus transacciones, estabilidad, escalabilidad y multiplataforma.

Inicialmente no se uso el modelo relacional debido a que tenía inconvenientes por el rendimiento, ya que no podían ser competitivas con las bases de datos jerárquicas y de red. Ésta tendencia cambio por un proyecto de IBM el cual desarrolló técnicas para la construcción de un sistema de bases de datos relacionales eficientes, llamado System R.

#### Década de 1980

Las bases de datos relacionales con su sistema de tablas, filas y columnas, pudieron competir con las bases de datos jerárquicas y de red, ya que su nivel de programación era bajo y su uso muy sencillo.

En esta década el modelo relacional ha conseguido posicionarse del mercado de las bases de datos. Y también en este tiempo se iniciaron grandes investigaciones paralelas y distribuidas, como las bases de datos orientadas a objetos.

#### Principios década de los 90

Para la toma de decisiones se crea el lenguaje SQL, que es un lenguaje programado para consultas. El programa de alto nivel SQL es un lenguaje de consulta estructurado que analiza grandes cantidades de datos, el cual permite especificar diversos tipos de operaciones frente a los mismos datos a diferencia de las bases de datos de los 80 que eran diseñadas para las aplicaciones de procesamiento de transacciones. Los grandes distribuidores de bases de datos incursionaron con la venta de bases de datos orientada a objetos.

#### Finales de la década de los 90

El boom de esta década fue la aparición de la WWW “Word Wide Web” ya que por éste medio se facilitaba la consulta de las bases de datos. Actualmente tienen una amplia capacidad de almacenamiento de información, también una de las ventajas es el servicio de siete días a la semana las veinticuatro horas del día, sin interrupciones a menos que haya planificaciones de mantenimiento de las plataformas o el software.

#### Siglo XXI

En la actualidad existe gran cantidad de alternativas en línea que permiten hacer búsquedas orientadas a necesidades específicas de los usuarios, una de las tendencias más amplias son las bases de datos que cumplan con el protocolo Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) los cuales permiten el almacenamiento de gran cantidad de artículos que permiten una mayor visibilidad y acceso en el ámbito científico y general.

Como respuesta a la creciente complejidad de las aplicaciones que requieren bases de datos, han surgido dos nuevos modelos: el modelo de datos orientado a objetos y el modelo relacional extendido. Sin embargo, a diferencia de los modelos que los preceden, la composición de estos modelos no está clara. Esta evolución representa la tercera generación de los DBMS

### Diferencia entre Datos e información

Los Datos a diferencia de la información son utilizados como diversos métodos para comprimir la información a fin de permitir una transmisión o almacenamiento más eficaces.

Aunque para el procesador de la computadora hace una distinción vital entre la información entre los programas y los datos, la memoria y muchas otras partes de la computadora no lo hace. Ambos son registradas temporalmente según la instrucción que se le de. Es como un pedazo de papel no sabe ni le importa lo que se le escriba: un poema de amor, las cuentas del banco o instrucciones para un amigo. Es lo mismo que la memoria de la computadora.

Sólo el procesador reconoce la diferencia entre datos e información de cualquier programa Para la memoria de la computadora y también para los dispositivos de entrada y salida (I / O) y almacenamiento en disco, un programa es solamente más datos, más datos que deben ser almacenados, movidos o manipulados.

La cantidad de información de un mensaje puede ser entendida como el número de símbolos posibles que representan el mensaje.”Los símbolos que representan el mensaje no son más que datos significativos.

En su concepto más elemental, la información es un mensaje con un contenido determinado emitido por una persona hacia otra y, como tal, representa un papel primordial en el proceso de la comunicación, a la vez que posee una evidente función social. A diferencia de los datos, la información tiene significado para quien la recibe, por eso, los seres humanos siempre han tenido la necesidad de cambiar entre sí información que luego transforman en acciones. “La información es, entonces, conocimientos basados en los datos a los cuales, mediante un procesamiento, se les ha dado significado, propósito y utilidad.

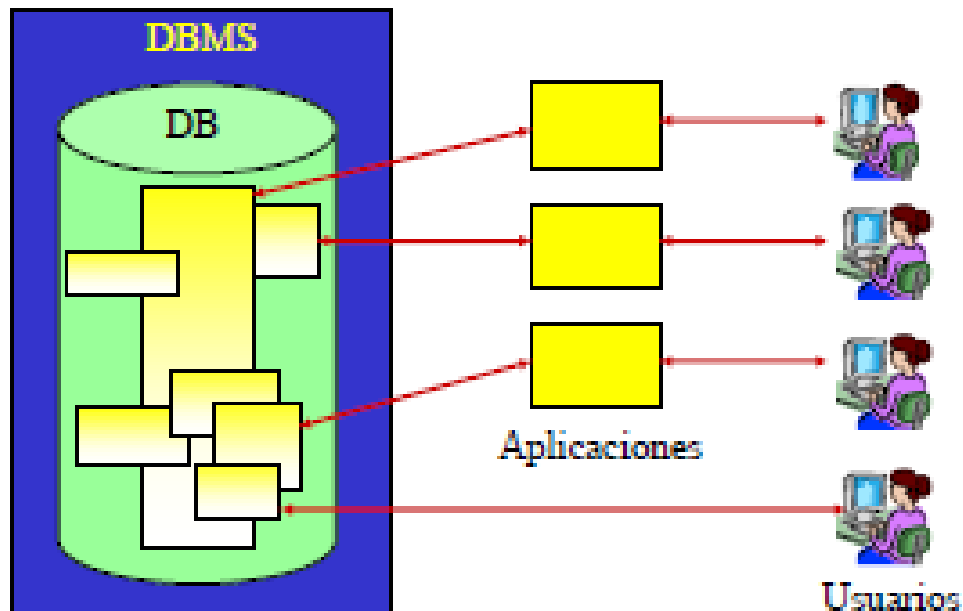
### **DATO(s)**

Técnicamente, los datos son hechos y cifras en bruto, tales que pueden registrarse, como ser números telefónicos, direcciones, nombres, órdenes y pagos, los cuales se procesan para obtener información, por ejemplo el saldo deudor y el monto disponible.

Nota: Los datos no contienen información

Datos de una DB se refiere a archivos, bases de datos, documentos de texto, imágenes y, voz y video codificados en forma digital

## Sistema de base de datos



En el diagrama anterior se percibe claramente como la DB y DBMS son entes separados aunque interrelacionados.

### BASE DE DATOS (DB)

Una base de datos es una colección de datos almacenados y organizados de forma que un programa del ordenador pueda seleccionarlos rápidamente y capaces de ser: recobrados, actualizados, insertados y borrados. En un DBMS una base de datos es un sistema de archivos electrónico.

### INFORMACIÓN

La información no es un dato conjunto cualquiera de ellos. Es más bien una colección de hechos significativos y pertinentes, para el organismo u organización que los percibe. La definición de información es la siguiente: Información es un conjunto de datos significativos y pertinentes que describen sucesos o entidades.

A diferencia de los datos, la información tiene significado para quien la recibe siendo esta el resultado del procesamiento de los datos extraídos de una DB.



Es común que los términos datos e información se tomen incorrectamente como sinónimos. Recordemos que en una DB puede ser lo que para algunos usuarios es información sin embargo, esta deberá considerarse siempre como datos hasta que estos hayan salido del sistema a través de un **DBMS**.

Las Bases de Datos y el sistema administrador resultan ser la columna vertebral de cualquier Empresa, siendo esta una unidad económico-social, integrada por elementos humanos, materiales y técnicos, que tiene por objeto obtener un resultado a través de su participación en la sociedad, con o sin afán de lucro., como pueden ser:

- Industrias manufactureras,
- Hospitales,
- Bancos,
- Escuelas,
- Instituciones Gubernamentales,
- Etc.

Donde para operar se deben tener una gran cantidad de datos como:

- Datos de producción,
- Información de pacientes ,
- Cuentas contables,
- Datos de alumnos y profesores,
- Censos de población y de recursos,
- Etc.

### **Tipos de bases de datos**

Las bases de datos pueden clasificarse de varias maneras, de acuerdo al criterio elegido para su clasificación:

#### **Según la variabilidad de los datos almacenados**

##### **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 los datos almacenados se modifican con el tiempo, permitiendo operaciones como actualización 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, etc.

## **Según el contenido**

### **Bases de datos bibliográficas**

Solo contienen un surrogante (representante) de la fuente primaria, que permite localizarla. Un registro típico de una base de datos bibliográfica contiene datos sobre el autor, fecha de publicación, editorial, título, edición, de una determinada publicación, etc. Puede contener un resumen o extracto de la publicación original, pero nunca el texto completo, porque sino estaríamos en presencia de una base de datos a texto completo (o de fuentes primarias—ver más abajo). Como su nombre lo indica, el contenido son cifras o números. Por ejemplo, una colección de resultados de análisis de laboratorio, entre otras.

### **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.

### **Directorios**

Un ejemplo son las guías telefónicas en formato electrónico.

### **Bases de datos fuente:**

Serían aquellas bases de datos que ofrecen el documento completo, no una representación del mismo.

- **Numéricas:**

Contienen datos de tipo numérico, como un censo o indicadores cuantitativos.

- **Mixtas:**

Combinan ambos tipos de datos, como por ejemplo informes económicos de empresas, datos geo-económicos, etc.

### **Bases de datos o "bibliotecas" de datos sobre Biológica**

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

- Aquellas 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)

Además de la clasificación por la función de las bases de datos, éstas también se pueden clasificar de acuerdo a su modelo de administración de datos.

Un modelo de datos es básicamente una "descripción" de algo conocido como *contenedor de datos* (algo en donde se guarda los datos), así como de los métodos para almacenar y recuperar datos de esos contenedores. Los modelos de datos no son cosas físicas: son abstracciones que permiten la implementación de un sistema eficiente de *base de datos*; por lo general se refieren a algoritmos, y conceptos matemáticos.

Algunos modelos con frecuencia utilizados en las bases de datos:

### **Bases de datos jerárquicas**

#### **Base de datos de red**

#### **Base de datos relacional**

#### **Bases de datos multidimensionales**

#### **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 (estado y comportamiento).

Una 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 los datos 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.

En bases de datos orientadas a objetos, los usuarios pueden definir operaciones sobre los datos como parte de la definición de la base de datos. Una operación (llamada función) se especifica en dos partes. La interfaz (o *signatura*) de una operación incluye el nombre de la operación y los tipos de datos de sus argumentos (o *parámetros*). La implementación (o *método*) de la operación se especifica separadamente

y puede modificarse sin afectar la interfaz. Los programas de aplicación de los usuarios pueden operar sobre los datos invocando a dichas operaciones a través de sus nombres y argumentos, sea cual sea la forma en la que se han implementado. Esto podría denominarse independencia entre programas y operaciones.

Se está trabajando en **SQL3**, que es el estándar de SQL92 ampliado, que soportará los nuevos conceptos orientados a objetos y mantendría compatibilidad con SQL92.

### **Bases de datos documentales**

Permiten la indexación a texto completo, y en líneas generales realizar búsquedas más potentes. Tesauro es un sistema de índices optimizado para este tipo de bases de datos.

### **Base de datos deductivos**

Un sistema de **base de datos deductivos**, 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. También las bases de datos deductivas son llamadas **base de datos lógica**, a raíz de que se basan en lógica matemática.

### **Gestión de bases de datos distribuida**

La base de datos está almacenada en varias computadoras conectadas en red. Surgen debido a la existencia física de organismos descentralizados. Esto les da la capacidad de unir las bases de datos de cada localidad y acceder así a distintas universidades, sucursales de tiendas, etc.

### **Tipo de Procesos:**

#### **Multitarea**

La ejecución de más de un programa a la vez en una computadora. Los sistemas operativos multitarea (Windows, OS/2, Unix y otros) permiten que múltiples operaciones se ejecuten simultáneamente. Sus mayores beneficios se obtienen cuando se realizan tareas largas y complejas (como consultar una base de datos). La cantidad de programas que pueden ejecutarse depende de la memoria disponible, de la velocidad de la CPU y de la eficiencia del sistema operativo.

#### **Multiusuario**

Un SO multiusuario permite a mas de un solo usuario acceder una computadora. Claro que, para llevarse esto a cabo, el SO también debe

ser capaz de efectuar multitareas. Unix es el Sistema Operativo Multiusuario más utilizado. Debido a que Unix fue originalmente diseñado para correr en una minicomputadora, era multiusuario y multitarea desde su concepción. Actualmente se producen versiones de Unix para PC tales como The Santa Cruz Corporation Microport, Esix, IBM, y Sunsoft. Apple también produce una versión de Unix para la Machintosh llamada: A/UX. Unix, proporciona tres maneras de permitir a múltiples personas utilizar la misma PC al mismo tiempo.

1. Mediante Módems

2. Mediante conexión de terminales a través de puertos seriales

3. Mediante Redes

## **Multiproceso**

Las computadoras que tienen más de un CPU son llamadas multiproceso. Un sistema operativo multiproceso coordina las operaciones de las computadoras multiprocesadoras. Ya que cada CPU en una computadora de multiproceso puede estar ejecutando una instrucción, el otro procesador queda liberado para procesar otras instrucciones simultáneamente. Al usar una computadora con capacidades de multiproceso incrementamos su velocidad de respuesta y procesos. Casi todas las computadoras que tienen capacidad de multiproceso ofrecen una gran ventaja. Los primeros Sistemas Operativos Multiproceso realizaban lo que se conoce como: Multiproceso asimétrico: Una CPU principal retiene el control global de la computadora, así como el de los otros procesadores. Esto fue un primer paso hacia el multiproceso pero no fue la dirección ideal a seguir ya que la CPU principal podía convertirse en un cuello de botella. Multiproceso simétrico: En un sistema multiproceso simétrico, no existe una CPU controladora única. La barrera a vencer al implementar el multiproceso simétrico es que los SO tienen que ser rediseñados o diseñados desde el principio para trabajar en un ambiente multiproceso. Las extensiones de Unix, que soportan multiproceso asimétrico ya están disponibles y las extensiones simétricas se están haciendo disponibles. Windows NT de Microsoft soporta multiproceso simétrico. Lenguajes que tienen estas características: C, SQL, El UNIX y sus clones permiten múltiples tareas y múltiples usuarios. Su sistema de archivos proporciona un método sencillo de organizar archivos y permite la protección de archivos. Sin embargo, las instrucciones del UNIX no son intuitivas.

Otros sistemas operativos multiusuario y multitarea son OS/2, desarrollado inicialmente por Microsoft Corporation e International Business Machines (IBM) y Windows, desarrollado por Microsoft. El sistema operativo multitarea de las computadoras Apple se denomina Macintosh OS.

## 1.1 Funciones de un DBMS

Siendo entonces la Base de datos una colección de archivos interrelacionados creados con un DBMS. El contenido de una base de datos esta almacenada de tal manera que los datos estén disponibles para los usuarios, una finalidad de la base de datos es eliminar la redundancia o al menos minimizarla.

Como ya se ha establecido la DB no es, en si un cúmulo de información, sino solo un “almacén” de datos, lo que ha hecho indispensable el desarrollo de sistemas que los administren y procesen, siendo estos los DBMS

El propósito general de los DBMS es el de manejar de manera clara, sencilla y ordenada a los datos de una Base de Datos (DB) que posteriormente se convertirán en información relevante, para un buen manejo de los datos.

Existen distintas funciones que deben cumplir los DBMS:

- **Abstracción de los datos.** Los DBMS ahorran a los usuarios detalles acerca del almacenamiento físico de los datos. Da lo mismo si una base de datos ocupa uno o cientos de archivos, este hecho se hace transparente al usuario. Así, se definen varios *niveles de abstracción*.
- **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.
- **Consistencia.** En aquellos casos en los que no se ha logrado eliminar la redundancia, será necesario vigilar que aquellos datos que aparecen repetidos se actualice de forma coherente, es decir, que todos los datos repetidos se actualicen de forma simultánea.

Por otra parte, la base de datos representa una realidad determinada que tiene determinadas condiciones, por ejemplo que los menores de edad no pueden tener licencia de conducir. El sistema no debería aceptar datos de un conductor menor de edad.

En los DBMS existen herramientas que facilitan la programación de este tipo de condiciones.

- **Seguridad.** Los datos almacenados en una base de datos puede llegar a tener un gran valor. Los DBMS deben garantizar que estos datos se encuentren seguros frente a usuarios malintencionados, que intenten leer datos privilegiados; frente a ataques que deseen manipular o destruir los datos; o simplemente ante las torpezas de algún usuario autorizado pero despistado. Normalmente, los DBMS disponen de un complejo sistema de permisos a usuarios y grupos

de usuarios, que permiten otorgar diversas categorías de permisos.

- **Integridad.** Se trata de adoptar las medidas necesarias para garantizar la validez de los datos almacenados. Es decir, se trata de proteger los datos ante fallos de hardware, datos introducidos por usuarios descuidados, o cualquier otra circunstancia capaz de corromper los datos almacenados. Los DBMS proveen mecanismos para garantizar la recuperación de la base de datos hasta un estado consistente (ver Consistencia, más arriba) conocido en forma automática.
- **Respaldo.** Los DBMS deben proporcionar una forma eficiente de realizar copias de respaldo de los datos almacenados y de restaurar a partir de estas copias los datos que se hayan podido perder.
- **Control de la concurrencia.** En la mayoría de entornos (excepto quizás el doméstico), lo más habitual es que sean muchas las personas que acceden a una base de datos, bien para recuperar datos, bien para almacenarla. Y es también frecuente que dichos accesos se realicen de forma simultánea. Así pues, un DBMS debe controlar este acceso concurrente a la DB, que podría derivar en inconsistencias.
- **Manejo de Transacciones.** Una Transacción es un programa que se ejecuta como una sola operación. Esto quiere decir que el estado 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 DBMS 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 DBMS tarda en darnos la información solicitada y en almacenar los cambios realizados.

### **Ventajas de cumplir lo anterior**

Proveen facilidades para la manipulación de grandes volúmenes de datos. Entre éstas:

- Simplifican la programación de chequeos de consistencia.
- Manejando las políticas de respaldo adecuadas garantizan que los cambios de la base serán siempre consistentes sin importar si hay errores en el disco, o hay muchos usuarios accediendo simultáneamente a los

mismos datos, o se ejecutaron programas que no terminaron su trabajo correctamente, etc.

- Permiten realizar modificaciones en la organización de los datos con un impacto mínimo en el código de los programas.
- Permiten implementar un manejo centralizado de la seguridad de los datos (acceso a usuarios autorizados), protección de la DB, de modificaciones, inclusiones, consulta.
- Las facilidades anteriores bajan drásticamente los tiempos de desarrollo y aumentan la calidad del sistema desarrollado si son bien explotados por los desarrolladores.
- Usualmente, proveen interfases y lenguajes de consulta que simplifican la recuperación de los datos.

### **Inconvenientes**

- 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 de los sistemas operativos. Esto puede llegar a incrementar los costos de operación en una empresa. Sin embargo hay que balancear este aspecto con la calidad y confiabilidad del sistema que se obtiene.
- 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 plantilla de cálculo.
- Complejidad: los DBMS son 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 DBMS por lo general son relativamente altos, por lo que estos equipos pueden llegar a costar gran cantidad de dinero.

Para poder evaluar correctamente un DBMS es necesario considerar su



nivel de abstracción,

### **Abstracción de los datos.**

Una base de datos es en esencia una colección de archivos relacionados entre sí, de la cual los usuarios pueden extraer información/datos sin considerar las fronteras de los archivos.

Un objetivo importante de un sistema de base de datos es proporcionar a los usuarios una visión *abstracta* de los datos, es decir, el sistema esconde ciertos detalles de cómo se almacenan y mantienen los datos.

Sin embargo para que el sistema sea manejable, los datos se deben extraer eficientemente.

Existen diferentes niveles de abstracción para simplificar la interacción de los usuarios con el sistema; Interno, conceptual y externo, específicamente el de almacenamiento físico, el del usuario y el del programador.

#### **Nivel físico.**

Es la representación del nivel más bajo de abstracción, en éste se describe en detalle la forma en como se almacenan los datos en los dispositivos de almacenamiento (por ejemplo, mediante señalamientos o índices para el acceso aleatorio a los datos).

#### **Nivel conceptual.**

El siguiente nivel más alto de abstracción, describe que datos son almacenados realmente en la base de datos y las relaciones que existen entre los mismos, describe la base de datos completa en términos de su estructura de diseño. El nivel conceptual de abstracción lo usan los administradores de bases de datos, quienes deben decidir qué datos se van a guardar en la base de datos.

Consta de las siguientes definiciones:

#### **Definición de los datos:**

Se describen el tipo de datos y la longitud de campo todos los elementos direccionables en la base. Los elementos por definir incluyen artículos elementales (atributos), totales de datos y registros conceptuales (entidades).

**Relaciones entre datos:** Se definen las relaciones entre datos para enlazar tipos de registros relacionados para el procesamiento de archivos múltiples.

En el nivel conceptual la base de datos aparece como una colección de registros lógicos, sin descriptores de almacenamiento. En realidad los

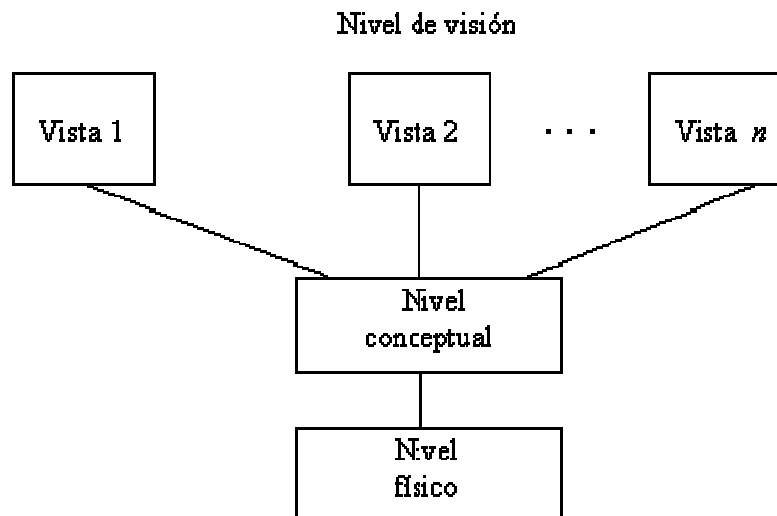
archivos conceptuales no existen físicamente. La transformación de registros conceptuales a registros físicos para el almacenamiento se lleva a cabo por el sistema y es transparente al usuario.

**Nivel de visión.**

Nivel más alto de abstracción, es lo que el usuario final puede visualizar del sistema terminado, describe sólo una parte de la base de datos al usuario acreditado para verla.

El sistema puede proporcionar muchas visiones para la misma base de datos.

La interrelación entre estos tres niveles de abstracción se ilustra en la siguiente figura.

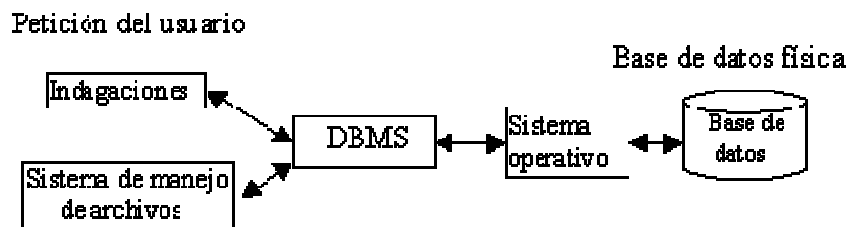


El sistema manejador de bases de datos DBMS es la porción más importante del software de un sistema de base de datos. Un DBMS es una colección de numerosas rutinas de software interrelacionadas, cada una de las cuales es responsable de alguna tarea específica.

**Para cumplir los Objetivos mencionados se tiene que las funciones principales de un DBMS son:**

- Crear y organizar la Base de datos.
- Establecer y mantener las trayectorias de acceso a la base de datos de tal forma que los datos puedan ser accedados rápidamente.
- Manejar los datos de acuerdo a las peticiones de los usuarios.

- Registrar el uso de las bases de datos.
- Interacción con el manejador de archivos. Esto a través de las sentencias en DML al comando del sistema de archivos. Así el Manejador de base de datos es el responsable del verdadero almacenamiento de los datos.
- Respaldo y recuperación.  
Consiste en contar con mecanismos implantados que permitan la recuperación fácilmente de los datos en caso de ocurrir fallas en el sistema de base de datos.
- Control de concurrencia.  
Consiste en controlar la interacción entre los usuarios concurrentes para no afectar la inconsistencia de los datos.
- Seguridad e integridad.  
Consiste en contar con mecanismos que permitan el control de la consistencia de los datos evitando que estos se vean perjudicados por cambios no autorizados o previstos.



La figura muestra el DBMS como interfase entre la base de datos física y las peticiones del usuario. El DBMS interpreta las peticiones de entrada/salida del usuario y las manda al sistema operativo para la transferencia de datos entre la unidad de memoria secundaria y la memoria principal.

En sí, un sistema manejador de base de datos es el corazón de la base de datos ya que se encarga del control total de los posibles aspectos que la puedan afectar.

## 1.2 Usuarios de un DBMS: DBA

Podemos definir a los usuarios como toda persona que tenga todo tipo de contacto con el sistema de base de datos desde que este se diseña, elabora, termina y se usa.

Los usuarios que acceden una base de datos pueden clasificarse como:

### ◆ Programadores de aplicaciones.

Los profesionales en computación que interactúan con el sistema por medio de llamadas en DML (Lenguaje de Manipulación de Datos), las

cuales están incorporadas en un programa escrito en un lenguaje de programación (Por ejemplo, COBOL, PL/I, Pascal, C, etc.)

◆ **Usuarios sofisticados.**

Los usuarios sofisticados interactúan con el sistema sin escribir programas. En cambio escriben sus preguntas en un lenguaje de consultas de base de datos.

◆ **Usuarios especializados.**

Algunos usuarios sofisticados escriben aplicaciones de base de datos especializadas que no encajan en el marco tradicional de procesamiento de datos.

◆ **Usuarios ingenuos.**

Los usuarios no sofisticados interactúan con el sistema invocando a uno de los programas de aplicación permanentes que se han escrito anteriormente en el sistema de base de datos, podemos mencionar al usuario ingenuo como el usuario final que utiliza el sistema de base de datos sin saber nada del diseño interno del mismo por ejemplo: un cajero.

**Administrador de base de datos (DBA):**

Es la persona o equipo de personas profesionales responsables del control y manejo del sistema de base de datos, generalmente tiene(n) experiencia en DBMS, diseño de bases de datos, Sistemas operativos, comunicación de datos, hardware y programación.

Los sistemas de base de datos se diseñan para manejar grandes cantidades de datos, la manipulación de los datos involucra tanto la definición de estructuras para el almacenamiento de la DB como la provisión de mecanismos para la manipulación de la DB, además un sistema de base de datos debe de tener implementados mecanismos de seguridad que garanticen la integridad de la DB, a pesar de caídas del sistema o intentos de accesos no autorizados.

Un objetivo principal de un sistema de base de datos es proporcionar a los usuarios finales una visión abstracta de los datos, esto se logra escondiendo ciertos detalles de como se almacenan y mantienen los datos.

Persona que toma las decisiones estratégicas y de política con respecto a la DB de la empresa, y el DBA es quién proporciona el apoyo técnico necesario para poner en práctica esas decisiones. Por tanto el DBA esta encargado del control general del sistema en el nivel técnico

Para el DBA sus funciones principales son:

<b>Definir el esquema conceptual</b>	Debe decidir cuales son los datos que debe mantenerse en la BD, es decir, identificar las entidades que interesan a la empresa y los datos que deben registrarse acerca de esas entidades. Este proceso se denomina <b>diseño lógico de BD</b> . El DBMS utilizará la versión objeto (compilada) de ese esquema para responder a las solicitudes de acceso. La versión fuente (sin compilar) servirá como documento de referencia para los usuarios del sistema.
<b>Definir el esquema interno</b>	Debe decidir cómo se representará los datos en la BD almacenada. A éste proceso se lo denomina diseño físico de la BD. El DBA se vale del DDL interno para crear la definición de estructura de almacenamiento y la correspondencia pertinente entre los esquemas interno y conceptual (tanto en la versión fuente como objeto).
<b>Vincularse con los usuarios</b>	El DBA debe encargarse de la comunicación con los usuarios, garantizar la disponibilidad de los datos que requieren y escribir los esquemas necesarios.  Las consultas sobre diseño de aplicaciones, la impartición técnica, la ayuda en la localización y resolución de problemas, y otros servicios profesionales similares relacionados con el sistema.
<b>Definir las verificaciones de seguridad e integridad</b>	Las verificaciones de seguridad e integridad pueden considerarse parte del esquema conceptual.
<b>Definir procedimientos de respaldo y recuperación</b>	Cuando una empresa se decide a utilizar un sistema de BD, se vuelve dependiente en grado sumo del funcionamiento correcto de ese sistema. En caso de que sufra daño cualquier porción de la BD resulta esencial poder reparar los datos implicados con un mínimo de retraso y afectando lo menos posible al resto del sistema.  El DBA debe definir y poner en práctica un plan de recuperación adecuado que incluya, por ejemplo, una descarga o “vaciado” periódico de la BD en un medio de almacenamiento de respaldo, y procedimientos para cargar otra vez la BD a partir del vaciado más reciente cuando sea necesario.
<b>Supervisar el desempeño y responder a cambios en los requerimientos</b>	Es responsabilidad del DBA organizar el sistema de modo que se obtenga el desempeño que sea mejor para la empresa, y realizar los ajustes apropiados cuando cambien los requerimientos.

### 1.3 Desarrolladores y usuarios finales

Siendo entonces una DB una colección de datos almacenados en una computadora (discos, tambores u otro medio secundario, en donde hay

una serie de programas que corren con la DB en donde habrá las operaciones de:

- Recuperación,
- Actualización,
- Inserción y
- Eliminación

De los datos en donde habrá usuarios, no solo uno, que accederán a los datos de muchos usuarios, que a veces es el mismo dato que se traslapa y de ahí la importancia que la DB sea integrada conociéndose como BASE DE DATOS ÚNICA.

Es importante entonces identificar el tipo de usuarios que acceden a una DB, que generalmente se clasificarán en dos tipos: desarrolladores y usuarios finales.

Los Desarrolladores o Diseñadores están operando dentro de un DBMS en los Niveles de Diseño: Físico y Conceptual.

**Nivel Físico.-** Es aquel en el que se determinan las características de almacenamiento en el medio secundario. Los diseñadores de este nivel poseen un amplio dominio de cuestiones técnicas y de manejo de hardware (Uno o más computadoras, medios de almacenamiento secundario como: discos duros, CDR, etc. Ver Anexo)...

**Nivel Conceptual.-** Es aquel en el que se definen las estructuras lógicas de almacenamiento y las relaciones que se darán entre ellas. Ejemplos comunes de este nivel son el diseño de los registros y las ligas que permitirán la conexión entre registros de un mismo archivo, de archivos distintos incluso, de ligas hacia archivos.

El Usuario final es el que finalmente accede a la BD a través del DBMS, conocido como Nivel de Edición.

**Nivel de Edición.-** Es aquel en el que se presenta al usuario final y que puede tener combinaciones o relaciones entre los datos que conforman a la base de datos global. Puede definirse como la forma en el que el usuario aprecia la DB y sus relaciones.

## 1.4 Componentes de un DBMS

Un sistema de base de datos se encuentra dividido en módulos cada uno de los cuales controla una parte de la responsabilidad total de sistema.

En la mayoría de los casos, el sistema operativo proporciona únicamente los servicios más básicos y el sistema de la base de datos debe partir de esa base y controlar además el manejo correcto de los datos. Así el diseño de un sistema de base de datos debe incluir la interfaz entre el sistema de base de datos y el sistema operativo.

Los componentes funcionales de un sistema de base de datos, son:

◆ **Gestor de archivos.**

Gestiona la asignación de espacio en la memoria del disco y de las estructuras de datos usadas para representar datos.

◆ **Manejador de base de datos.**

Sirve de interfaz entre los datos y los programas de aplicación.

◆ **Procesador de consultas.**

Traduce las proposiciones en lenguajes de consulta a instrucciones de bajo nivel. Además convierte la solicitud del usuario en una forma más eficiente.

◆ **Compilador de DDL.**

Convierte las proposiciones DDL en un conjunto de tablas que contienen metadatos, estas se almacenan en el diccionario de datos.

◆ **Archivo de datos.**

En él se encuentran almacenados físicamente los datos de una organización.

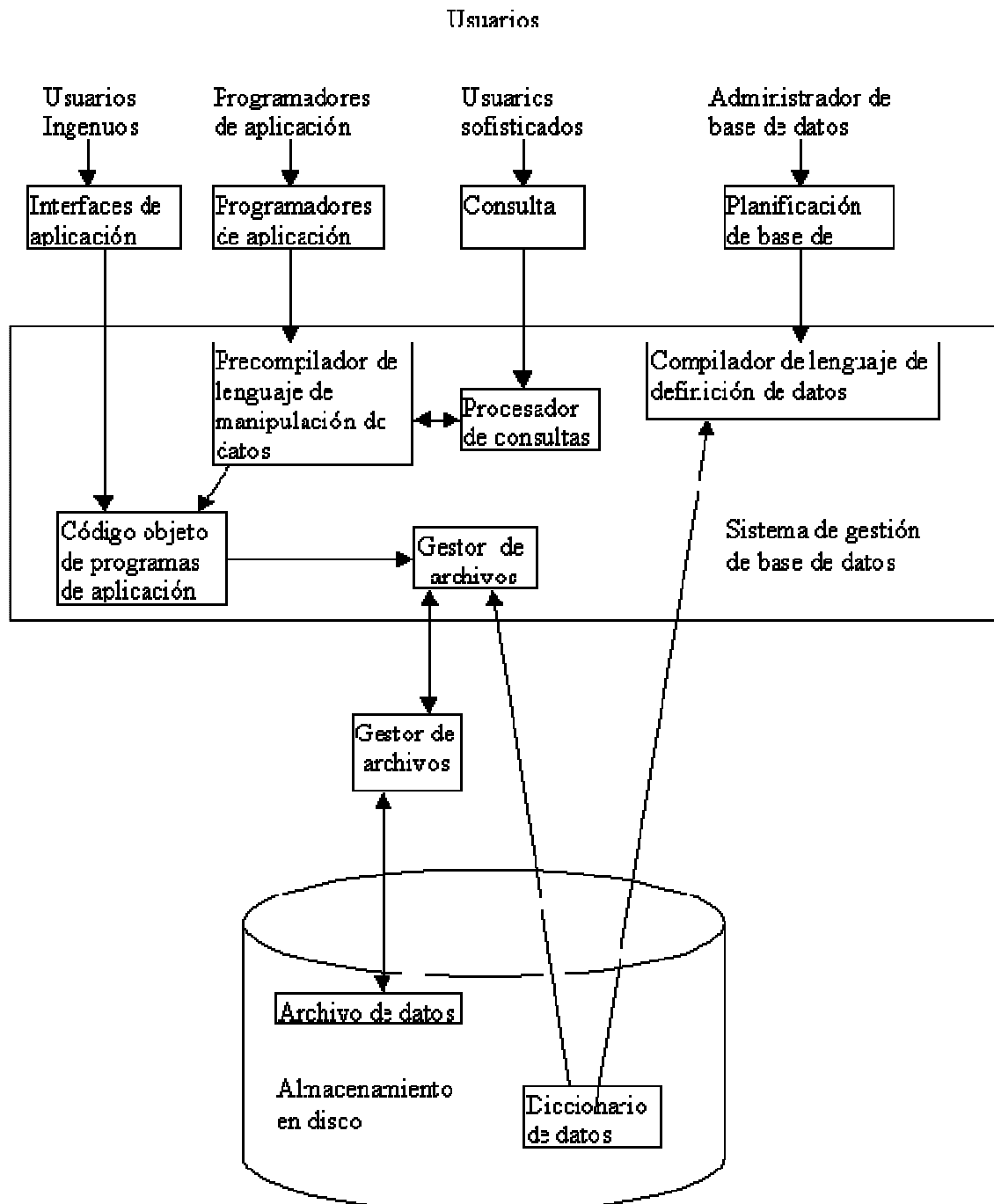
◆ **Diccionario de datos.**

Contiene la información referente a la estructura de la base de datos.

◆ **Índices.**

Permiten un rápido acceso a registros que contienen valores específicos.

Una forma gráfica de representar los componentes antes mencionados y la relación que existe entre ellos sería la siguiente.



El DBMS es entonces un conjunto de programas (software) que se encargan de manejar la creación y todos los accesos a las DB y esta compuesto por:

DDL: Data Definition Language ( Lenguaje de Definición de Datos)

DML Data Manipulation Language (Lenguaje de Manipulación de Datos)

SQL Structured Query Language (Lenguaje Estructurado de Consulta)