

1.-Sistemas de Bases de Datos

1.1.-Evolución: De los Sistemas Orientados a los Procesos a los Sistemas Orientados a los datos

Los sistemas orientados a los datos se caracterizan porque los datos no son de una aplicación sino de una Organización entera que los va a utilizar. Se tiende a integrar las aplicaciones, evitando aplicaciones aisladas. Se diferencian las estructuras lógicas y físicas, de manera que el usuario final solo se vincule con las estructuras lógicas. La descripción de la estructura lógica se separa de los lenguajes de programación. El concepto de relación cobra importancia, de modo que se requiere de herramientas que permitan definir las y almacenarlas.

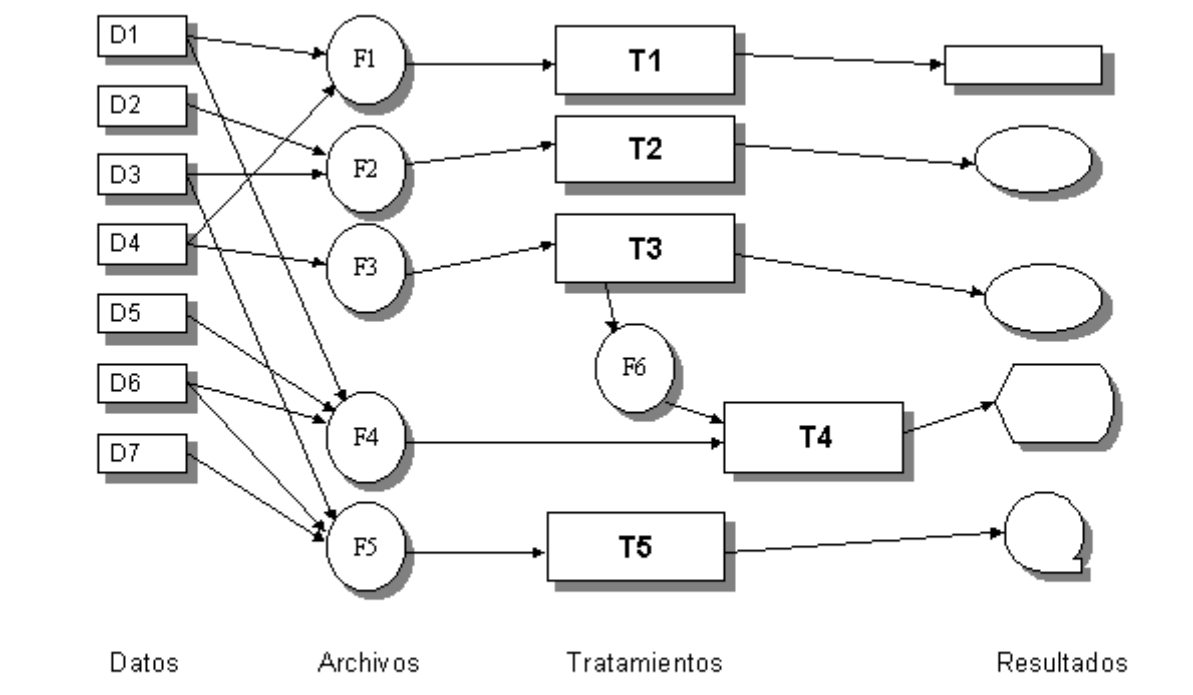


Figura 1.1.- Sistemas Orientados a los Procesos

Originalmente las aplicaciones que se desarrollaban para en las organizaciones estaban orientadas a cubrir necesidades muy específicas de procesamiento, por lo cual tanto los lenguajes de programación como las estructuras de datos se centraban en realizar de manera más eficiente una tarea específica. En la Figura 1.1.- se puede ver claramente como la tarea (tratamientos) es un elemento central en el diseño de las estructuras de datos que utiliza, por ejemplo si el tratamiento T5 no existiese, ¿tendría sentido que exista el archivo F5?. Más aún, si profundizamos el análisis, de la Figura 1.1.- vemos que existen datos (D1, D3, D4 y D6) que residen en distintos archivos –siendo conceptualmente únicos- esta situación resultaba claramente peligrosa por las inconsistencias que acarrea, por ejemplo, si el dato D6 tiene un valor distinto en el archivo F4 al que tiene en el archivo F5. ¿Cuál es el correcto?

Así, las bases de datos buscan resolver principalmente aquel problema, evitar las inconsistencias que se producían por la utilización de los mismos datos lógicos desde distintas plataformas físicas (archivos) a través de procesos independientes.

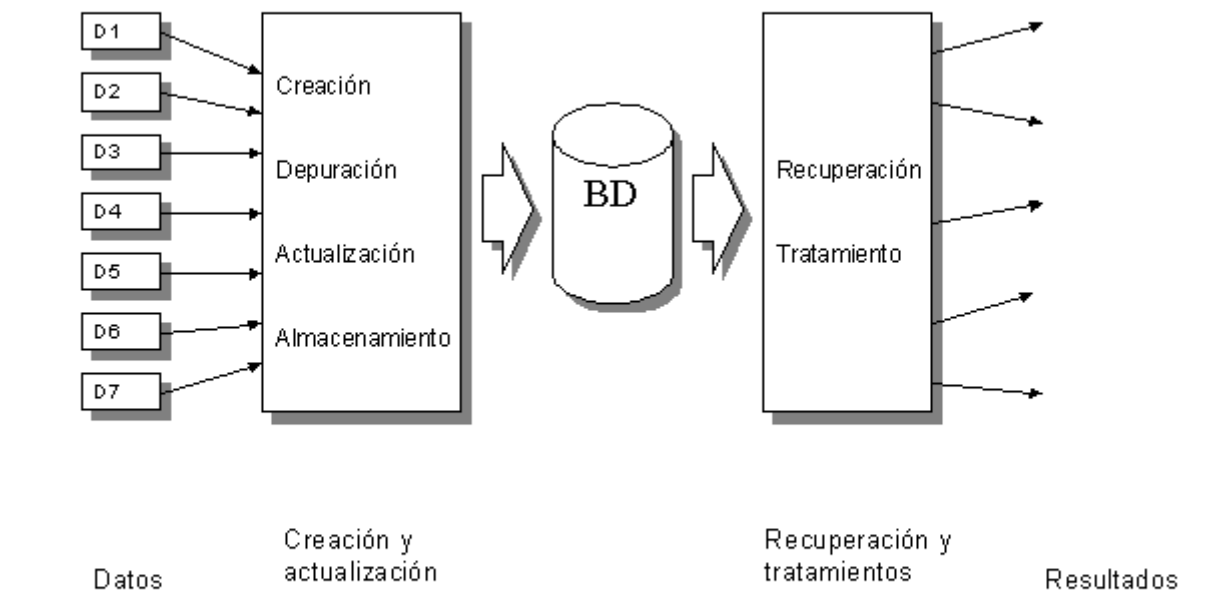


Figura 1.2.- Sistema orientado a las Bases de Datos

El análisis entonces comienza por formular la lógica de los datos organizacionales como un todo, para después vincular aquellos con los procesos que los utilizan. Es en este análisis en que las Bases de Datos como una unidad tanto teórica como conceptual y física cobran importancia. Y el mecanismo sobre el cual esto se articula es el de disminuir la redundancia a través del establecimiento de relaciones entre los datos de una organización.

1.2.- Concepto de Base de Datos

La idea de base de datos surge como una necesidad de mantener datos relacionados. Veamos un ejemplo.

Supongamos 3 archivos: artículos, clientes y pedidos. Un cliente puede tener varios pedidos, pueden existir varios pedidos para un mismo artículo, diferentes pedidos en diferentes fechas, etc.

Si los requerimientos son:

- Los pedidos del cliente X: una solución es generar una lista por cada cliente, generando una relación entre los archivos clientes y pedidos.
- Todos los pedidos de un artículo: se tendrá una lista por cada artículo, generando una relación entre pedidos y artículo.

Las relaciones anteriores son complejas, por lo que se hace necesario contar con alguna herramienta que facilite estos requerimientos. Se debe considerar que los requerimientos hacen uso de los mismos datos.

Las definiciones de base de datos son numerosas. Todas coinciden en que es un conjunto de datos almacenados en un soporte de acceso directo. Los datos están interrelacionados y estructurados de acuerdo a un modelo que sea capaz de conseguir el máximo contenido semántico.

Definición 1:

"Colección de datos interrelacionados almacenados en conjunto sin redundancias perjudiciales o innecesarias; su finalidad es servir a una o más aplicaciones de la mejor forma posible; los datos se almacenan de modo que resulten independientes de los programas que los usan; se emplean métodos bien determinados para incluir nuevos datos y para modificar o extraer los datos almacenados". J.Martin,

Definición 2:

"Colección integrada y generalizada de datos, estructurada atendiendo a las relaciones naturales de modo que suministre todos los caminos de acceso necesarios a cada unidad de datos con objeto de poder atender todas las necesidades de los diferentes usuarios". Deen.

Definición 3:

"Colección de datos integrados, con redundancia controlada y con una estructura que refleje las interrelaciones y restricciones existentes en el mundo real; los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción, únicas para cada tipo de datos, han de estar almacenadas junto con los mismos. Los procedimientos de actualización y recuperación, comunes y bien determinados, habrán de ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de los datos". A.de Miguel.

Definición 4:

"Una base de datos consiste en alguna colección de datos persistentes e independientes usados por una organización determinada." J.Date.

1.2.1.- Características del dato:

- No efímero, en el sentido que permanece en el tiempo
- Estructurado, para que facilite el compartirlos por aquellos que lo necesiten
- Operacional Transaccional (OLTP), manipularlos aplicando operadores para obtener resultados
- Sentido semántico
- Integro, en el sentido que refleja una realidad existente

De hecho los datos que contiene una base de datos tiene una característica especial, se les reconoce como datos de operación.

Datos de Operación: los datos de una Base de Datos se consideran DATOS de OPERACION, distinguiéndose de los datos de entrada y de salida. Una Base de Datos, es un conjunto de datos de operación almacenados y utilizados por los sistemas de aplicación de una organización específica. Cualquier organización necesita disponer de una gran cantidad de datos acerca de su funcionamiento. Estos constituyen sus datos de operación. Los datos de Operación no incluyen datos de entrada o de salida, colas de espera de trabajo o cualquier otro dato de índole transitoria. Así los datos de entrada, por ejemplo, se refiere a los datos que entran al sistema desde el exterior, tales datos pueden ocasionar un cambio en los datos de operación, pero no constituyen parte de la Base de Datos. El conjunto de los datos de operación tiene directa relación con aquellos elementos de datos que modelan el sentido de la organización como un sistema dotado de identidad propia, así, si comparamos el conjunto de los datos de operación estructurados para organizaciones distintas, la semántica de aquellos necesariamente será diferente, por mucho que sus sintaxis sean idénticas. Los datos de operación reflejan contenidos semánticos, tanto así que el diseño lógico de una base de datos también se conoce como modelado semántico.

1.2.2.- *Análisis del concepto de base de datos*

El concepto de Base de Datos determina algunas características que le son propias, por ejemplo:

- El mundo real considera interrelaciones entre datos y restricciones semánticas que deben estar presentes en una base de datos. Una base de datos no solo debe almacenar entidades y atributos (recordar los sistemas tradicionales de archivos), sino que también debe almacenar interrelaciones entre datos. Por otro lado, actualmente se le está dando mucha importancia a las restricciones semánticas, de manera que éstas se almacenan junto con los datos.
- La redundancia de datos debe ser controlada, de forma que no existan duplicidades perjudiciales ni innecesarias. Las redundancias físicas, convenientes muchas veces a fin de responder a objetivos de eficiencia, sean tratadas por el mismo sistema, de modo que no puedan producirse incoherencias. Esto significa que en las bases de datos NO está permitida la redundancia lógica, pero si se admite cierta redundancia física por motivos de eficiencia.
- Las bases de datos pretenden servir a toda la organización, es decir a múltiples usuarios y a diferentes aplicaciones (recordar los sistemas tradicionales de archivos).
- La independencia, tanto lógica como física, de los tratamientos sobre los datos y estos mismos, ha tenido una enorme influencia en la arquitectura de los SGBD (recordar los sistemas tradicionales de archivos).
- La definición y descripción del conjunto de datos contenido en la base debe ser única e integrada con los mismos datos. (recordar los sistemas tradicionales de archivos). En las bases de datos, la descripción, y en algunos casos, también una definición y documentación completas (metadatos) se almacenan junto con los datos, de modo que éstos están documentados, y cualquier cambio que se produzca en la documentación debe quedar incorporado en el sistema.
- La actualización y recuperación de las bases de datos debe realizarse mediante procesos bien determinados, incluidos en el SGBD; procedimientos que han de estar diseñados de modo que se mantenga la integridad, seguridad y confidencialidad de la base.

1.2.3.- *Las características elementales de una Base de Datos*

El objetivo de disminuir la redundancia de un conjunto de datos determina dos características fundamentales que poseerá cualquier sistema de Bases de Datos:

Integrada: se entiende que una base de datos puede considerarse como una unificación de varios archivos de datos independientes, donde se elimina parcial o totalmente cualquier **redundancia** entre los mismos. Por ej. una base de datos específica puede contener registros de EMPLEADO, que incluyen el nombre, dirección, departamento, salario, etc. y, existir registros de INSCRIPCION que representan inscripciones de empleados en cursos de capacitación. Supongamos que para llevar a cabo el proceso de administración de los cursos se necesita conocer el departamento de cada estudiante inscrito. Desde luego, no hay necesidad de incluir este dato (redundante) en los registro de INSCRIPCION, siempre se puede obtener recurriendo a los registros de EMPLEADO correspondiente.

Compartida: Se entiende que partes individuales de la Base de Datos pueden compartirse entre varios usuarios distintos, en el sentido que cada uno de ellos puede tener acceso a la misma parte de la Base de Datos y utilizarla con propósitos diferentes. Tal comportamiento es en verdad consecuencia del hecho de que la Base de Datos es integrada. En el caso del ejemplo anterior se tiene que los datos de los registros de EMPLEADOS es compartido por usuarios del departamento de personal y capacitación. Consecuencia del mismo hecho, que la Base de Datos es integrada, se advierte en que cualquier usuario tendrá acceso sólo a algún subconjunto de la Base completa, además, los subconjuntos de diferentes usuarios se procesarán de muy diversas maneras. En otras palabras, diferentes usuarios percibirán de modos muy distintos una base de datos.

1.2.4.- La independencia Dato-Proceso

Una de las principales ventajas que provee una base de Datos (ver siguiente punto) es la independencia entre los datos y los tratamientos que se hacen de ellos. A diferencia de los sistemas orientados al procesos, en los cuales los datos eran sumamente dependientes de los programas al extremo que lenguajes como COBOL definían en su código la estructura de los archivos –esto lo podemos ver actualmente en C y Pascal. Con ello la estructura de los datos de las tablas que los contienen son altamente dependiente de los procesos que los utilizan, de hecho, para que un proceso pueda utilizar un determinado dato que se encontraba almacenado en un archivo debía hacer la declaración completa de la estructura de este archivo, esta declaración era sólo modificable en tiempo de edición quedando fija ya en tiempo de compilación. Lo anterior era válido para cualquier acceso a los datos de un conjunto de archivos.

Ocurre que históricamente la tasa de variación de los procesos es mayor que la de los datos que aquellos procesos manejan, más aún cualquier actualización de los datos que maneja un proceso determina que éste necesariamente sea actualizado, lo que no siempre ocurre en el sentido contrario. No siempre que un proceso cambia esto determina cambios en los datos que ese proceso utiliza. Así, es claro que la situación de datos dependientes de los procesos era inadecuada, ya que claramente la dependencia es en el sentido contrario. Lo anterior es asimilable a los cambios que sufren las organizaciones, generalmente aquellos son de forma, un banco que cambia su imagen corporativa, que se agreguen o eliminen funciones de atención a público o de producción, reducción de personal, etc. Los cambios de fondo no son habituales en las organizaciones, ellos tiene que ver con cambios en la misión de la organización, en los objetivos. Si un banco cambia de imagen corporativa –logo, forma de atención, etc.- aquellos son cambios de forma y tienen básicamente que ver con variaciones en los procesos y eventualmente con algunos datos; ahora si el banco quiere transformarse en una AFP, entonces la situación es más compleja y profunda, la misión organizacional cambia, el objetivo cambia y, por lo mismo, el cambio es radical afectando todos los elementos que componen la organización, todos aquellos elementos donde se refleja el sentido organizacional, es decir afecta directamente a los datos y sus relaciones cambio que indirectamente afectaría a los procesos que manejan estos datos.

El concepto de Base de Datos rescata aquella dependencia que tienen los procesos de los datos y la radicaliza priorizando la independencia de estos últimos, determinando mecanismos de definición y de descripción que no requieren de procesos. Con lo anterior se crea un sustrato de datos modelando la organización, sobre el cual se establecen los procesos que los utilizan. Donde la definición y mantención de este sustrato es teóricamente independiente de los procedimientos de la organización y los procesos que apoyan su realización. Por otra parte esta independencia también apunta en el sentido contrario, es decir, si hay claridad en el problema de la influencia de los datos sobre los procesos, es perfectamente definible mecanismos que minimicen este influjo y que posibiliten una, también, independencia de los procesos.

1.3.-Ventajas de las bases de datos

Una sistematización de las ventajas de las Bases de Datos bosquejadas e los puntos anteriores los podemos resumir en el siguiente cuadro:

Cuadro resumen de las ventajas de las bases de datos	
Referidas a	Ventajas
Los datos	<ul style="list-style-type: none"> • Independencia de estos respecto de los tratamientos y viceversa • Mejor disponibilidad de los mismos • Mayor eficiencia en la recuperación, codificación y entrada
Los resultados	<ul style="list-style-type: none"> • Mayor coherencia • Mayor valor informativo • Mejor y más normalizada documentación de la información
Los usuarios	<ul style="list-style-type: none"> • Acceso más rápido y sencillo de los usuarios finales • Más facilidades para compartir los datos por el conjunto de los usuarios • Mayor flexibilidad para atender a demandas cambiantes.

Análisis cuadro anterior

- Independencia de los datos respecto a los tratamientos y viceversa: esto supone que un cambio en los tratamientos no imponga un nuevo diseño lógico y/o físico de la base de datos. Por otro lado, cambios en la incorporación, desaparición de datos, cambios en la estructura física o caminos de acceso no deben obligar a alterar los programas. Así se evita la reprogramación de las aplicaciones. Es el punto de partida para la adaptación de los sistemas de información a la evolución de las organizaciones.
- Coherencia de los resultados: debido a que la información de la base de datos se recupera y se almacena una sola vez, en todos los tratamientos se utilizan los mismos datos, por lo que los resultados de estos son coherentes y comparables. Así, se eliminan las divergencias en los resultados.
- Mejor disponibilidad de los datos para el conjunto de los usuarios: en una base de datos ningún usuario es propietario de los datos, pues éstos se comparten entre las aplicaciones, existiendo una mayor disponibilidad y transparencia.
- Mayor valor informativo: esto se refiere al concepto de sinergia, en donde el valor informativo del conjunto de datos es superior a la suma del valor informativo de los elementos individuales.
- Mejor y más normalizada documentación: la mayoría de los SGBD proporcionan herramientas para reflejar el contenido semántico de los datos, es decir, incluyen una descripción de los datos dentro del sistema.
- Mayor eficiencia en la captura, validación e ingreso de datos al sistema: al no existir redundancias, los datos se capturan y validan una sola vez aumentando el rendimiento del proceso previo al almacenamiento
- Reducción del espacio de almacenamiento: por un lado, la disminución de redundancias y las técnicas de compactación hacen que disminuya el espacio en disco. Sin embargo, los diccionarios, referencias, punteros, listas invertidas también ocupan espacio.

1.4.-Desventajas de las bases de datos

Las desventajas de una base de datos no las podemos dejar de mencionar. El siguiente cuadro resume algunas de aquellas.

Cuadro resumen de las desventajas de las bases de datos	
Relativas a	Desventajas
La implantación	<ul style="list-style-type: none"> • Costosa en equipos (lógico y físico) • Ausencia de estándares • Larga y difícil puesta en marcha • Rentabilidad a mediano plazo
Los usuarios	<ul style="list-style-type: none"> • Personal especializado • Desfase entre teoría y práctica

Análisis del cuadro anterior

- Instalación costosa: equipos: nuevas instalaciones o ampliaciones, sistemas operativos, compiladores, SGBD comerciales, computadores más poderosos, etc.
- Personal especializado: es clave la administración de la base de datos, se requiere de conocimientos específicos.
- Desfase entre teoría y práctica: muchos ejecutivos asumen que ciertas funcionalidades son ya un hecho, cuando en realidad son estudios teóricos.

Existe también una resistencia al cambio, sobre todo que este involucra a toda la organización. En el éxito de esto el papel mediador de los profesionales de informática es fundamental, sobre todo en organizaciones grandes donde una base de datos se puede ver como la centralización del poder en manos de unos pocos, generalmente los encargados de su administración.

1.5.-Componentes de los Sistemas de bases de datos

Un sistema de bases de datos contempla los siguientes componentes:

- La base de datos
- El Sistema de Gestión de Bases de Datos (SGBD, DBMS) o motor, tal como Oracle, Sybase, etc.
- Programas de aplicación
- Un conjunto de usuarios (finales, DBA, programadores de aplicaciones, etc.)
- Máquinas
- Programas utilitarios (generadores de informes, de interfaces, herramientas de desarrollo, de administración, etc.)

En la Figura 1.3.- se puede observar un esquema general de la arquitectura de una base de datos, en la cual se detallan los principales componentes de ella además de las relaciones entre ellos y la base de datos lógica. Veamos una descripción simple de aquellos elementos:

Una Vista Externa es una visión particular de un usuario o un grupo de usuarios de la Base de Datos. El Esquema Externo representa una forma de definición o formalización de esta vista externa.

La Vista Conceptual pretende ser la representación total y abstracta de los datos que componen la Base; la formalización de esta se logra mediante el Esquema Conceptual.

Por último, la Vista Interna es de un nivel muy bajo y corresponde al almacenamiento físico de los datos de la Base, sobre un Esquema Interno que es la formalización de esto, estructuras de datos, tipos de registros almacenados, índices, etc. Las correspondencias se pueden definir como una asociación de distintas representaciones para un mismo dato.

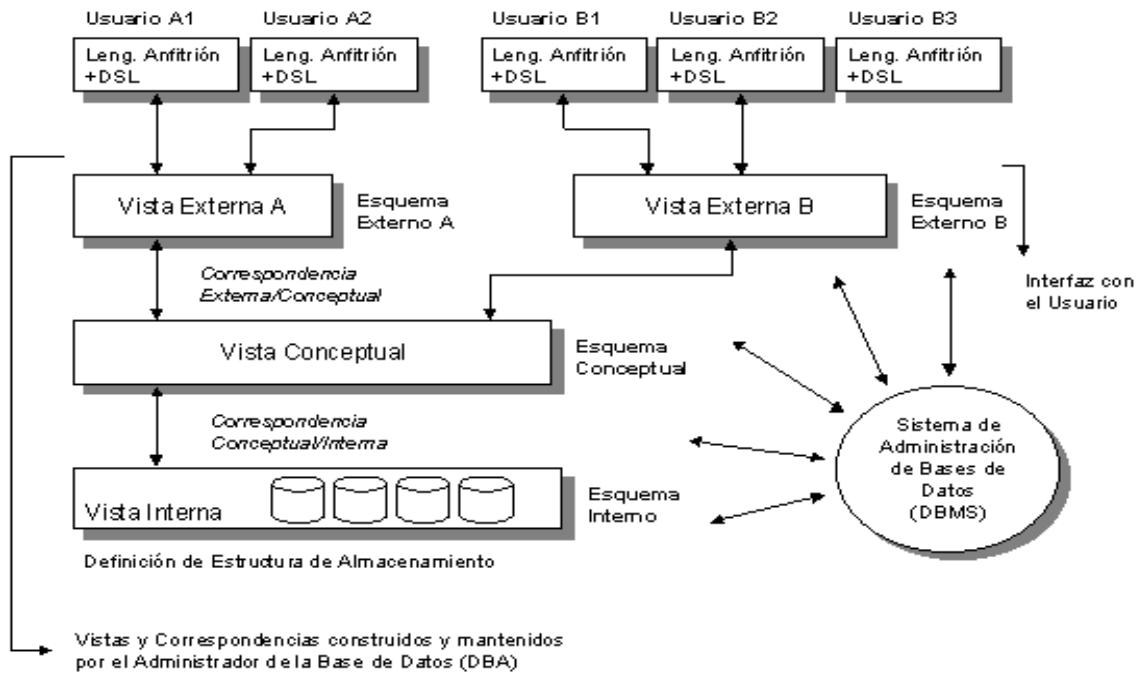


Figura 1.3.- Arquitectura de una Base de Datos

Un **DSL** es un sublenguaje de datos, es una combinación de dos lenguajes: un lenguaje de definición de datos (**DDL**) y un lenguaje de manipulación de Datos (**DML**). Este lenguaje representa un nexo entre el Sistema de Base de Datos y algún lenguaje anfitrión (p.e. COBOL, C, Pascal, etc.); e.d., el DSL provee herramientas a los lenguajes tradicionales para que se integren al Sistema de Base de Datos. Puede haber distintos tipos de DSL para un mismo sistema.

DBMS es la sigla en inglés de Sistema de Administración (Gestión) de Bases de Datos, que corresponde al Software que maneja todos los accesos a la Base de Datos, e.d. cada solicitud de acceso de un usuario al SABD es interpretada e inspeccionadas las correspondencias, generando, a continuación, una respuesta coherente a las necesidades de la pregunta.

La interfaz con el Usuario es el límite de acceso que tiene un Usuario común a la Base, todo lo que está bajo este límite es transparente (desconocido) para él.

Por último, el Administrador de Bases de Datos (DBA) corresponde a la persona o grupo de personas encargada del control general del sistema. Sus responsabilidades o funciones incluyen:

- Decidir el contenido de la Base de Datos: comprende la identificación de entidades de interés para la organización y los datos a registrar de éstas entidades. Luego se define el contenido de la Base de Datos generando un Modelo Conceptual.
- Decidir la estructura de almacenamiento y la estrategia de acceso: esto es decidir como deben representarse los datos en forma interna y hacer la correspondencia entre estos y el modelo conceptual ya definido.
- Vincularse con los usuarios: comprende toda una labor de prestación de servicios que busca garantizar la existencia, en la Base, de los datos necesarios y la formalización de los distintos esquemas externos.
- Definir los controles de autorización y procedimientos de validación: involucra la definición de restricciones de seguridad y protección para la conservación de la integridad de los Datos.
- Definir una estrategia de respaldo y recuperación: esto corresponde a un esquema de seguridad más amplio que lo anterior y, básicamente, su objetivo es la operación exitosa del sistema.
- Controles de desempeño y responder a los cambios de requerimiento: la idea aquí es lograr un desempeño aceptable, según expectativas, del Sistema mediante mecanismos de control.

1.6.- El sistema de Gestión de Bases de Datos (SGBD o DBMS)

Un sistema de gestión de bases de datos consiste de una colección de datos interrelacionados y un conjunto de programas para acceder a esos datos. La colección de datos es la base de datos, y es la que contiene información por ejemplo acerca de una empresa determinada.

El objetivo principal de un SGBD es proporcionar un entorno que sea a la vez conveniente y eficiente para ser utilizado al extraer y almacenar información en la base de datos.

Toda organización puede verse en tres niveles de gestión: operacional, táctico y estratégico. Muchas veces se produce una desconexión de los sistemas que caracterizan a estos niveles, pues constituyen sistemas aislados, sin relación entre ellos. Esto produce un aumento del costo global de creación y mantenimiento del sistema de información, produce redundancias e incoherencias. Esto impide una gestión racional de los datos.

La base de datos es un depósito único de datos para toda la organización, por lo que debe ser capaz de integrar los distintos sistemas y aplicaciones, atendiendo a las necesidades de los usuarios en los tres niveles.

El objetivo del SGBD es suministrar la interfaz entre el conjunto de los datos y dichos usuarios. El SGBD también debe proporcionar a los otros usuarios (analistas, programadores, administradores) las correspondientes herramientas que les permitan un adecuado desarrollo de sus funciones.

Definición del SGBD

El SGBD es un conjunto coordinado de programas, procedimientos y lenguajes que suministra, tanto a usuarios no informáticos como a los analistas, programadores o al administrador (DBA), los medios necesarios para describir, recuperar y manipular los datos almacenados en la base, manteniendo su integridad, confidencialidad y seguridad.

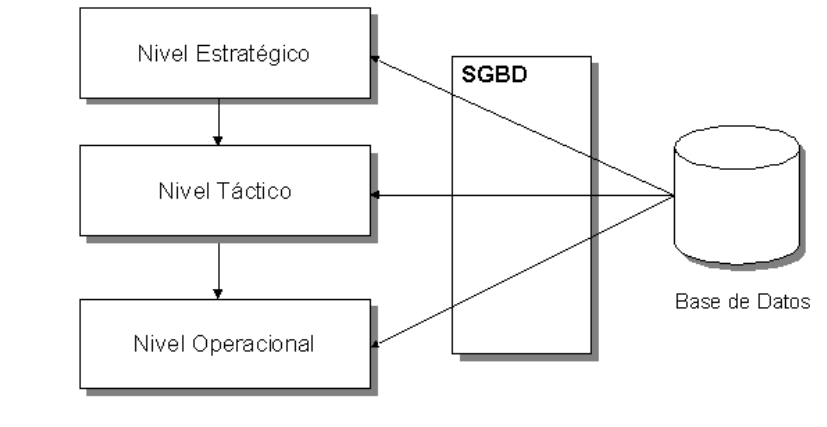


Figura 1.4.- La gestión y las Bases de Datos

Funciones del SGBD*De Descripción o Definición*

Debe permitir al administrador de la base especificar los datos que la integran, su estructura y las relaciones que existen entre ellos, las reglas de integridad semántica, los controles a efectuar antes de autorizar el acceso a la base, etc., así como las características de tipo físico y las vistas lógicas de los usuarios.

Esta función la realiza el lenguaje de definición de datos (LDD), propio del SGBD, y debe ser capaz de definir las estructuras de datos a los tres niveles (nivel externo, nivel lógico global o conceptual y nivel interno).

A nivel interno se define:

- Espacio reservado para la base (volúmenes, cilindros y pistas)
- Longitud de los campos
- Modo de representación de los datos (binario, decimal, alfanumérico, etc.)
- Caminos de acceso como punteros e índices.

A nivel externo y conceptual, la función de descripción proporciona los instrumentos para la definición de entidades, su identificación, atributos, interrelaciones entre ellas, autorizaciones de acceso, restricciones de integridad, etc.

El SGBD, además de describir, debe permitir la correspondencia o mapping entre estos niveles.

De Manipulación

Permite a los usuarios de la base (todos) buscar, eliminar o modificar los datos de la base, de acuerdo a las especificaciones y normas de seguridad dadas por el administrador.

Esto se realiza mediante el lenguaje de manipulación de datos (LMD), mediante un conjunto de instrucciones (lenguaje huésped) que son admitidas por un lenguaje de programación (lenguaje anfitrión), o bien, mediante un lenguaje autocontenido, que posee todas las instrucciones necesarias para llevar a cabo estas tareas.

De Utilización

Reúne todas las interfaces que necesitan los diferentes tipos de usuarios para comunicarse con la base y proporciona un conjunto de procedimientos para el administrador.

Algunas de estas funciones de servicio son:

- cambiar capacidades de los archivos
- obtener estadísticas de utilización
- respaldos
- cargar y descarga de la base
- seguridad, etc.

Lenguajes de los SGBD

Las distintas funciones que cumple un SGBD, hace necesario contar con diferentes lenguajes y procedimientos que permitan la comunicación con la base de datos.

Por tipo de función, tendremos lenguajes de definición y lenguajes de manipulación.

Por tipo de usuarios tendremos lenguajes para informáticos y lenguajes para no informáticos o usuarios finales. Estos últimos, pueden tener aplicaciones formalizables tal como la gestión de personal o no formalizables como cualquier proceso de toma de decisiones.

Cuando se trata de procesos formalizables, usualmente los programadores de aplicaciones escriben los procedimientos en programas.

Si el proceso no es formalizable, escribir un programa no es aconsejable. Es conveniente que el mismo usuario final resuelva directamente sus requerimientos mediante los instrumentos que el SGBD pone a su alcance.

Por otro lado, los usuarios informáticos, como el DBA, analistas y programadores requerirán medios poderosos por los cuales podrán definir, extraer y manipular los datos en algún lenguaje de programación. A este lenguaje se le llama lenguaje anfitrión (por ejemplo, C).

Casi la totalidad de los SGBD disponen de lenguajes de 4ta generación, que se caracterizan por ser poco procedimentales y el acceso a la base de datos se realiza mediante sentencias embebidas en el lenguaje de 4ta generación y escritas en SQL (SGBD relacionales).

Los lenguajes que por si mismos pueden actuar con la base de datos, sin necesidad de apoyarse en otro lenguaje se llaman autocontenidos.

Lenguajes de manipulación de datos

Para cumplir los objetivos asignados a la función de manipulación, se ha de contar con lenguajes que den a los usuarios la posibilidad de referirse a determinados conjuntos de datos que cumplan ciertas condiciones (criterio de selección).

El SQL como lenguaje de manipulación de datos tiene la propiedad dual, es decir, puede actuar como huésped o autocontenido.

Los LMD pueden ser *procedimentales* o no procedimentales, es decir, si necesitamos especificar con detalle el acceso a la base tendremos un lenguaje procedimental.

Los lenguajes orientados al usuario final deben ser lo menos *procedurales* posible. Aquí basta con decir qué se quiere, sin explicar cómo obtenerlo.

Por otro lado, los LMD pueden ser *navegacionales*, que recuperan o actualizan datos registro a registro. Otros lenguajes actúan sobre un conjunto de registros, de forma que una única sentencia puede dar lugar a la recuperación o actualización del conjunto de registros que cumpla el criterio de selección especificado, tal como el SQL.

1.7.- Distintos niveles de abstracción en una base de datos y arquitectura ANSI/X3/SPARC

Uno de los principales objetivos de las bases de datos es conseguir la independencia entre las estructuras lógica y física de los datos, que tiene como consecuencia la independencia entre datos y aplicaciones. Así, los cambios en la estructura de los datos tengan una repercusión mínima en los programas de aplicación y viceversa.

Este concepto de independencia implica la separación entre el almacenamiento y la organización lógica de los datos, con lo que se consigue:

- Los datos se presentarán de formas distintas, según las necesidades de los usuarios.
- El almacenamiento de los datos, su estructura lógica y los programas de aplicación serán independientes unos de otros.

Este objetivo, ha tenido gran influencia en la arquitectura de los SGBD.

En los SI existen 2 estructuras: la lógica (vista del usuario) y la física (forma en que se encuentran los datos en el almacenamiento). En las bases de datos aparece un nuevo nivel de abstracción llamado nivel conceptual, estructura lógica global, esquema. Esta estructura intermedia es una representación global de los datos independiente de equipos y los usuarios (visión de la empresa).

Los otros dos niveles corresponden al nivel de máquina o interno y al nivel de usuario o externo.

La estructura lógica de usuario o esquema externo (nivel usuario) es la visión que tiene de la base de datos cada usuario en particular. La estructura lógica global o esquema responde al enfoque del conjunto de la empresa (visión del administrador). La estructura física o esquema interno es la forma como se organizan los datos en el medio de almacenamiento físico.

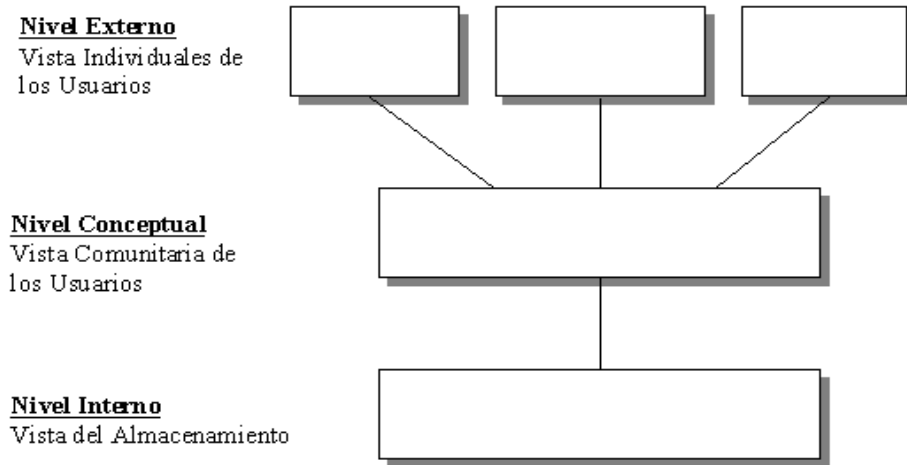


Figura 1.5.- Arquitectura de tres niveles de una Base de Datos

1.7.1.- Esquema Externo / Nivel Externo

En el deberán encontrarse reflejados solo aquellos datos e interrelaciones que necesite un usuario en particular. También deben especificarse las restricciones de uso, como por ejemplo, derecho a borrar o insertar determinados datos.

Habrán tantos esquemas externos como exijan las diferentes aplicaciones y un mismo esquema externo puede ser utilizado por varias aplicaciones.

En algunos SGBD en esta estructura aparecen los caminos de acceso a los datos, lo que no es tan apropiado ya que indicaría una dependencia lógica/física. Esta situación depende del modelo de datos en el cual se apoya el SGBD. En particular, en el modelo relacional los caminos de acceso solo se encuentran en el nivel interno, no siendo nunca visibles por el usuario.

Este nivel se ocupa de las vistas individuales de los usuarios. Los usuarios pueden ser programadores de aplicaciones o usuarios finales. Estos usuarios disponen de un lenguaje, que para los usuarios finales será un lenguaje de consulta o algún lenguaje de aplicación especial, manejado por ejemplo por menús o forms. Los usuarios programadores de aplicaciones tendrán lenguajes de programación convencionales o algún lenguaje propio de 4ta generación.

El SQL (Structured Query Language) es usado en casi todos los sistemas relacionales actuales. En casi todos los sistemas el SQL puede utilizarse como lenguaje interactivo o de consulta o bien embebido en otros lenguajes, tal como Visual Basic por ejemplo. Recordar lo visto anteriormente en los lenguajes de un SGBD.

Esta distinción entre lenguaje anfitrión y lenguaje embebido es transparente para el usuario. Si ambos lenguajes son difícilmente separables, se dice que son fuertemente acoplados, si por el contrario se pueden separar con facilidad son débilmente acoplados. Actualmente la mayoría de los sistemas son débilmente acoplados.

1.7.2.- Esquema o Estructura Lógico Global / Nivel conceptual

Tiene por objetivo describir en términos abstractos pero con absoluta fidelidad una cierta realidad de una organización y de su proceso de gestión.

Por ser la visión general de los datos, deberá incluir la descripción de todos los datos e interrelaciones entre éstos, restricciones de integridad y confidencialidad.

Este nivel se define mediante un esquema conceptual. Para escribirlo se utiliza un DDL conceptual. Es importante señalar que para que exista independencia de los datos, las definiciones en DDL conceptual no deberán implicar consideraciones de estructura de almacenamiento, deben ser definiciones de contenidos de información. Por lo tanto, en el esquema conceptual no debe haber representaciones de campos almacenados, secuencia de registros, indexación, etc.

El paso del mundo real al esquema conceptual corresponde a un proceso de modelización. En este punto es donde se utilizan los modelos conceptuales.

1.7.3.- Esquema interno/ Nivel Interno

Este esquema es dependiente del SGBD. Sin embargo, existen elementos comunes que son:

- Estrategia de almacenamiento: Asignar espacios de almacenamiento para los datos, así como relaciones existentes entre los distintos espacios de almacenamiento. También se especifica la estrategia de emplazamiento de los datos que ha sido utilizada para optimizar tiempo y espacio de memoria secundaria. También debe considerarse como se tratan los desbordes, etc.
- Camino de acceso: Se incluye la especificación de claves primarias, secundarias, índices, claves de ordenación.
- Técnicas de compresión de datos
- Técnicas de criptografía
- Correspondencia conceptual/interna: especifica como se representan los registros y campos conceptuales en el nivel interno. Si se altera la definición de la estructura de almacenamiento, la correspondencia conceptual/interna deberá modificarse también para que no varíe el esquema conceptual. Es tarea del DBA controlar tales modificaciones.
- Técnicas de Tuning y optimización
- Dispositivos de memoria: tamaño de la página, nro. de páginas asignadas a cada área de almacenamiento, tamaño de los buffers de E/S.
- Organizaciones físicas: para mejorar la recuperación y los tiempos de acceso, el sistema debe dar facilidades al DBA para definir hashing, agrupamientos, etc. Dependiendo del SGBD podrá definir punteros entre registros, privilegiando ciertos caminos de acceso.
- Control de acceso: reglas para proteger la confidencialidad y seguridad de la base de datos.