

Introducción a las bases de datos

Jaime Lorenzo Sánchez

24 de julio de 2022

Índice general

1. Introducción a las bases de datos	1
1.1. Características de las bases de datos	2
1.2. Las diferentes visiones de los datos en la base de datos	5
1.2.1. Independencia del nivel de descripción conceptual	8
1.2.2. Granularidad y ligadura	9
1.3. Bases de datos y Sistemas de Gestión de Bases de Datos	11
1.4. Componentes de los SGBD	12
1.4.1. El lenguaje de definición de datos	12
1.4.2. El lenguaje de definición del almacenamiento de los datos	13
1.4.3. El lenguaje de manipulación de datos	14
1.4.4. El diccionario de datos	15
1.4.5. El gestor de la base de datos	16
1.4.6. El administrador de la base de datos	17
1.4.7. Los usuarios de la base de datos	19

Capítulo 1

Introducción a las bases de datos

El término de bases de datos no apareció hasta mediados de los años 60, época en la cuál la información era representada haciendo uso de un conjunto de ficheros, generalmente planos. Estos ficheros no estaban relacionados entre sí, y los datos almacenados representaban las relaciones existentes en la información que representaban mediante referencias simbólicas y/o físicas. La redundancia era grande y la integridad de la información representada dejaba mucho que desear.

Con la aparición de los dispositivos de almacenamiento que permitían el acceso directo, generalmente aleatorio, las estructuras mediante las cuales se podía representar la información se fueron haciendo más complejas.

Si se desea que cualquier modificación en la cantidad, contenido y estructura de la información que se desea mantener acerca de un determinado problema no afecte a los procedimientos desarrollados previamente para el mantenimiento de la misma, es necesario tener en cuenta la existencia de una independencia de los datos con respecto a los procedimientos.

La independencia de los datos con respecto a los procedimientos supone que la visión conceptual de los datos no tienen por qué ser la misma que la visión física de los mismos.

Esta independencia de la información con respecto a los procedimientos que la maneja debe satisfacerse a dos niveles de abstracción para que sea efectiva:

Independencia lógica de los datos: La modificación de la representación lógica general del dominio del problema no afecta a los programas de aplicación que la manipulan, siempre que esta modificación no elimine ninguno de los ítems de datos que estos programas requieran.

Independencia física de los datos: La distribución de los datos en las unidades de almacenamiento y la estructura física de la información almacenada es independiente de los cambios de la estructura lógica general de la información y, por tanto, de los procedimientos que manejan la misma.

1.1. Características de las bases de datos

La información que forma parte de una base de datos puede organizarse de múltiples formas pero con independencia de la arquitectura de la base de datos, ésta debe cumplir una serie de características para ser considerada como tal:

Versatilidad para la representación de la información: Si bien la información que forma parte del dominio de un problema es única y caracteriza a ese problema o sistema, pueden existir diferentes versiones de esa información. La organización de la información en la base de datos debe permitir que distintos procedimientos pueden construir diferentes registros a partir de la información existente en la base de datos. Estos registros (lógicos) contruidos por los procedimientos deben ser independientes de los registros físicos existentes en la base de datos para almacenar la información.

Desempeño: Las bases de datos deben asegurar un tiempo de respuesta adecuado en la comunicación hombre-máquina, permitiendo el acceso simultáneo al mismo o distinto conjunto de ítems de datos por el mismo o distinto procedimiento.

Mínima redundancia: La existencia de redundancia es nefasta debido a la posibilidad de inconsistencia en la información almacenada en la base de datos. La redundancia implica la existencia de varias copias de un mismo ítem de datos, las cuáles pueden en un momento dado tener distintos valores, lo que implica unas necesidades de almacenamiento innecesarias que siempre implicarán un coste innecesario.

Un objetivo principal de las bases de datos es eliminar la redundancia siempre que ello

no implique una complejidad de la misma y/o una disminución en el desempeño.

Capacidad de acceso: Los usuarios de la base de datos reclaman a ésta continuamente información sobre los datos almacenados. Una base de datos debe ser capaz de responder, en un tiempo aceptable, a cualquier consulta sobre la información que mantiene, sin restricciones graves en cuanto a los ítems, relaciones, etc, solicitados en la misma, y respondiendo al usuario rápidamente.

Esta característica va a depender directamente de la organización física de los datos en la base de datos. Una organización física muy completa garantiza una respuesta rápida a las consultas, aunque requiere un mayor coste computacional en actualizaciones y viceversa.

Simplicidad: La base de datos representa el dominio de un problema que se necesita tratar computacionalmente. La naturaleza de este problema puede ser muy variada y, por tanto, existir en el mismo un número de objetos variables que se relacionan de múltiples formas. Por ello, es la naturaleza del problema un factor de complejidad de partida de las bases de datos que es conveniente eliminar para que se garanticen otras de las características que se le requieren.

Las bases de datos deben estar basadas en representaciones lógicas simples que permitan la verificación en la representación del problema que representan y la modificación de requisitos en el mismo, de tal forma que la inclusión y/o modificación de nuevos ítems de datos y relaciones no ocasione una complejidad excesiva.

Integridad: La integridad de una base de datos hace referencia a la veracidad de los datos almacenados con respecto a la información existente en el dominio del problema que trata la misma. Como los datos de la base de datos son manejados por muchos usuarios haciendo uso de muchos procedimientos que tratan los mismos datos de muchas formas, es necesario garantizar que estos datos no sean destruidos ni modificados de forma anómala.

Durante el procedimiento se pueden producir fallos de muy diversa naturaleza. Así, los procedimientos que manejan la información deben asegurar que el sistema pueda garantizar la integridad de la información a pesar de los errores que se puedan producir, temporalmente, a causa de los fallos con independencia de su naturaleza.

Además, esta integridad debe garantizarse con respecto a la veracidad de los ítems de datos y sus relaciones con respecto al dominio del problema. Así, en una base de datos deben establecerse procedimientos que verifiquen que los valores de los datos se ajustan a los requisitos y restricciones extraídas del análisis del problema.

Seguridad y privacidad: La seguridad de una base de datos hace referencia a la capacidad de ésta para proteger los datos contra su pérdida total o parcial por fallos del sistema o por accesos accidentales o intencionados a los mismos. Mientras que la privacidad de una base de datos hace referencia a la reserva de la información de la misma a personas no autorizadas.

Para conseguir estas características, una base de datos debe satisfacer, al menos, los siguientes requisitos:

- Seguridad contra la destrucción de los datos causada por el entorno.
- Seguridad contra la destrucción de los datos causada por fallos del sistema, de forma que los datos puedan reconstruirse.
- Seguridad contra accesos no autorizados a la base de datos.
- Seguridad contra accesos indebidos a los datos.

Por lo que deben existir en la base de datos tanto procedimientos de recuperación de la información perdida total o parcialmente por cualquier causa.

Afinación: Hace referencia a la organización física de la información de la base de datos, la cual determina directamente el tiempo de respuesta de los procedimientos que operan la misma.

La base de datos evoluciona con el tiempo, lo que implica que una buena organización física de los datos en un momento dado puede no ser tan buena en otro.

Por ello, la base de datos debe ser flexible a la modificación de esta organización física, lo que puede suponer más de una migración de los datos según evolucione la base de datos, sin que por ello se vean afectados los procedimientos u otras representaciones de los datos pero, sin embargo, se consiga un desempeño más alto.

Interfaz con el pasado y el futuro: Es aceptado que el dominio de un problema cambia evolucionando con el tiempo. Las necesidades de la organización cambian diariamente y, por lo tanto, cambia la información correspondiente al subsistema o dominio del problema de la misma. Una base de datos debe estar abierta a estos cambios de forma que no afecten, o afecten lo mínimo posible, a los procedimientos existentes para manejar la información que mantiene.

Por otro lado, una base de datos debe estar abierta a reconocer información organizada físicamente por otro software de distinta forma a la que utiliza la base de datos.

1.2. Las diferentes visiones de los datos en la base de datos

Para que una base de datos puedan satisfacer las anteriores características, y otras más, es necesario que los usuarios de la misma tengan una visión abstracta de los datos almacenados. Es decir, el usuario no tiene necesidad de conocer cómo se organizan los datos físicamente en la base de datos.

El usuario conoce el dominio del problema que va a ser tratado con una base de datos. Si es ésta la visión de los datos que la base de datos presenta al usuario, éste podrá utilizar el sistema e integrarse en él adecuadamente.

Se puede hablar de que existen 3 visiones de los datos en una base de datos:

Visión externa: Es la visión de los datos que tienen los usuarios finales de una base de datos. Un usuario trata sólo una visión parcial de la información, sólo aquella que interviene en el dominio de actividad (el subsistema de la organización en el que interviene).

Estas visiones particulares de los usuarios son proporcionadas por los procedimientos o programas de aplicación que sólo manejan una parte de la información de la base de datos.

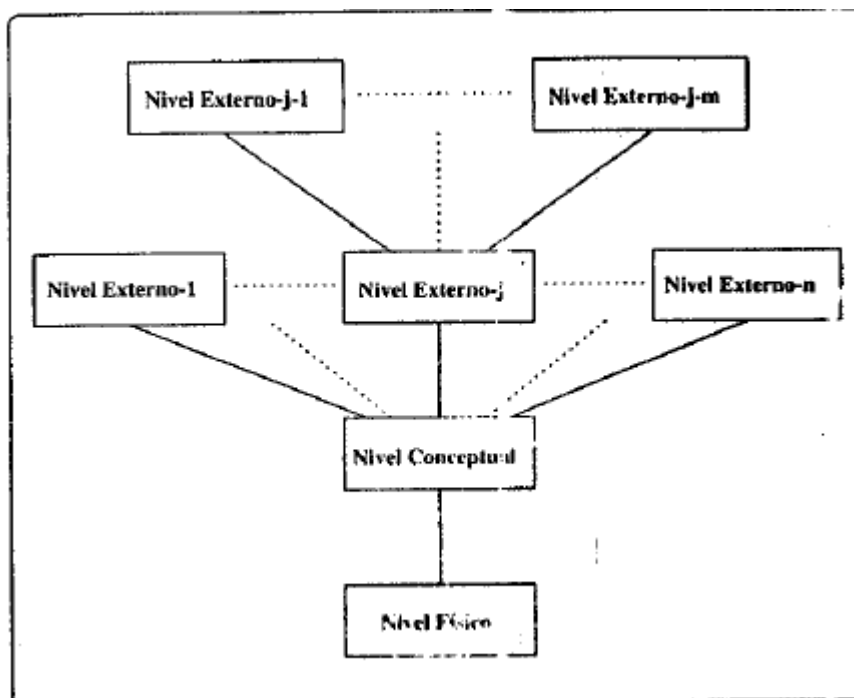
Visión conceptual: Es la visión o representación del problema tal y como éste se presenta en el mundo real. Una base de datos representa la información que es observada en el mundo real con respecto a un determinado problema. En el análisis del problema, se

determinan los objetos o entidades que intervienen en el mismo, las propiedades o características de estas entidades y las relaciones o dependencias que existen entre ellos.

La visión conceptual de una base de datos es una representación abstracta del problema e independiente, en principio, de cómo va a ser tratada ésta información, de qué visiones externas pueden tener y de cómo ésta información pueda ser almacenada físicamente. Así, la visión conceptual de una base de datos no cambia a no ser que cambie la naturaleza del problema.

Visión física: La visión física de una base de datos es la representación de cómo la información es almacenada en los dispositivos de almacenamiento. Esta visión describe las estructuras de mayor o menor complejidad que representan el dominio del problema de una forma entendible por el sistema informático.

Si los usuarios de las bases de datos pueden percibir 3 visiones diferentes de los datos, es debido a que en una base de datos existen, al menos, 3 formas diferentes de descripción de la información que almacena. Pero, como el dominio del problema que representa una base de datos es el mismo, entonces estas 3 formas de descripción son, en realidad, 3 niveles de abstracción diferentes que describen un mismo problema.



En cada nivel se describen aquellos objetos de interés que pueden ser entendidos por los usuarios, de ese nivel, de la base de datos. Así, el usuario final sólo entiende de registros

(visión o nivel de abstracción externo) o formas de comunicación similares a los documentos externos que son manejados en la organización, mientras que el diseñador o analista de sistemas sólo entiende de tipos de entidades o clases de objetos que intervienen en el dominio del problema que la organización desea que se trate mediante una base de datos, de las relaciones existentes entre ellas y de los procedimientos que son llevados a cabo en la organización para la solución del problema que se está tratando (nivel de abstracción conceptual) y, por otro lado, es el administrador de la base de datos el encargado de describir el nivel físico para determinar aquella organización física que pueda garantizar el desempeño óptimo del sistema (nivel de abstracción físico o interno).

La descripción de los datos a éstos 3 niveles de abstracción diferentes garantiza la independencia de los datos, uno de los principales objetivos de las bases de datos. Es decir:

- Que pueda ser modificada la organización física de los datos sin que por ello haya cambiado o deba cambiar la descripción conceptual, y sin que por ello tengan que ser modificados los programas de aplicación (nivel de abstracción externo) que manipula esta información.
- Que pueda ser modificada la representación conceptual del problema que está siendo presentado en la base de datos, debido a la consideración de nuevas entidades, o relaciones, o cambios en las características de las mismas, sin que por ello tenga que ser modificada la estructura física de la información (nivel interno), ni los programas de aplicación (nivel de abstracción externo), naturalmente siempre y cuando no se eliminen de la representación conceptual objetos necesarios o requeridos en estos otros niveles.
- Evidentemente, las visiones externas pueden cambiar conforme nuevos requisitos o necesidades funcionales o de operación son incorporados al dominio del problema por la organización y/o su entorno, y sin que por ello deba ser modificada ninguna de las descripciones de los datos a ninguno de los restantes niveles de abstracción.

Para que exista esta independencia tan deseada de los datos los 3 niveles de abstracción mediante los cuales son descritos deben ser completamente independientes, lo cual no es del todo cierto en la mayoría de las bases de datos. Sin embargo, puede conseguirse una buena independencia de los datos si:

- La representación interna de los datos no es una traducción dependiente de la representación conceptual. Una misma representación conceptual puede representarse de varias formas físicamente. Serán los requisitos funcionales y de desempeño los que determinen esta representación.
- Si bien las representaciones externas son dependientes de la representación conceptual el hecho de que los registros (representaciones externas) deben estar formados por ítems de datos existentes y, por tanto, representados en el nivel conceptual (igual entre el nivel físico y conceptual), la estructura de estos registros (número de ítems y disposición de los mismos) debe ser independiente de cómo estos ítems han sido representados en el nivel conceptual y de las relaciones que mantienen en el mismo.

1.2.1. Independencia del nivel de descripción conceptual

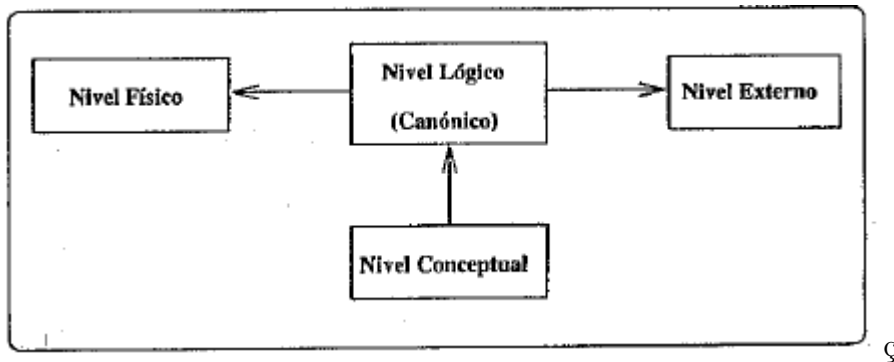
El nivel de descripción conceptual es seguramente el más importante, o por lo menos aquel en el que se apoyan en menor o mayor grado los otros niveles y, con seguridad, en el que, en base a su calidad, se garantiza que la base de datos solucione el problema de la organización.

El nivel conceptual describe el sistema de la organización, o aquella parte del sistema, el dominio del problema, que se desea tratar. Una descripción conceptual de calidad describirá todas y cada una de las entidades o clases de objetos que intervienen en el problema, sus propiedades y atributos, así como las características de las relaciones existentes entre las mismas. En este nivel se describe cada uno de los ítems de datos o elementos de información que intervienen en el comportamiento del sistema y cuya información es necesario considerar.

De hecho, existen muchas formas de representar de forma abstracta un fenómeno observado del mundo real. Una representación abstracta (conceptual) de un problema supone la aplicación de una serie de reglas que restringen y dirigen la forma en que ese problema es representado. Pero, por otro lado, el fenómeno del mundo real que se está representando debe ser, y de hecho lo es, independiente de la forma en que el ser humano sea capaz de representarlo.

Por ello, se puede hablar de un cuarto nivel de abstracción en la representación de la

información en una base de datos, el nivel lógico o canónico.



La descripción conceptual es independiente de las descripciones externas e interna, puesto que sólo depende del problema del mundo real objeto de la representación. Si el problema no cambia, no cambia la representación conceptual, aunque cambien los mecanismos por los cuales el problema será tratado y, por tanto, alguna de las otras representaciones.

La representación canónica es dependiente de la forma, mecanismos o procedimientos por los cuales la información correspondiente al problema va a ser manipulada. La descripción canónica es derivada de la descripción conceptual en base a la aplicación de una serie de reglas y restricciones que tienen en cuenta cómo la información representada puede ser trata por los procedimientos que van a manejar y definir la información en base a las otras representaciones. Así, pueden existir muchas representaciones lógicas de una misma representación conceptual, al igual que de una representación lógica pueden ser derivadas muchas representaciones externas.

La inclusión de este nuevo nivel de representación del dominio del problema, un nivel dependiente del software encargado de manipular la información, va a garantizar la independencia de la información en una base de datos.

1.2.2. Granularidad y ligadura

Si bien es posible la descripción de los datos a niveles de abstracción diferentes para que se garantice la independencia de la información con respecto a los procedimientos que la manipulan a cualquiera de los niveles, es obvio pensar que esta independencia tendrá unas restricciones que limitan su ámbito.

Si la descripción externa debe ser derivada de la descripción lógica, el nivel de detalle

de la descripción lógica limitará, por tanto, la independencia por la cual pueden ser las descripciones externas derivadas.

Si el nivel de detalle con que se pueda representar la información en el nivel lógico es al nivel de entidad, entonces las descripciones externas deberán considerar para la descripción de cada registro a entidades descritas en el nivel lógico.

Mientras que si en el nivel lógico la información puede describirse al nivel de agregado de dato, o ítem de dato, los registros descritos en el nivel externo podrán considerar sólo aquellos ítems de datos de las entidades descritas en el nivel lógico que sean de interés para el programa de aplicación.

Al nivel de detalle en que pueden ser descritas las representaciones externas derivadas de la representación lógica se denomina Granularidad. A mayor granularidad de una representación externa (menor información a considerar) mayor será la independencia y viceversa. Una mayor granularidad proporciona una mayor independencia, pero al mismo tiempo una mayor complejidad en el software utilizado para realizar estas representaciones y, por añadidura, en la organización de la información para que estas representaciones puedan ser almacenadas físicamente.

Para garantizar la integridad de las bases de datos es, por tanto, necesario que en algún instante esos procedimientos que manipulan las representaciones a un determinado nivel de abstracción tengan en cuenta cómo se presenta la información en los otros niveles. En ese instante las diferentes representaciones de los datos se vinculan entre sí y, por tanto, la independencia entre ellas se pierde. A este proceso de vinculación de las diferentes representaciones de la información en la base de datos se le denomina Ligadura, y ésta es de 2 tipos:

Ligadura lógica: Correspondiente al proceso de vinculación que se produce entre las representaciones externas y la lógica.

Ligadura física: Correspondiente al proceso de vinculación entre la representación lógica y la física.

Si cuando se produce la ligadura se vinculan las representaciones de los datos en los diferentes niveles de abstracción, ello implica que la independencia de los datos se pierde

y, por tanto, sería conveniente retrasar lo más posible el proceso de vinculación. El proceso de vinculación entre las distintas representaciones puede realizarse en cualquiera de las siguientes fases: Compilación, Enganche, Ejecución y Acceso a la base de datos.

La independencia de los datos será mayor cuanto más tardía se realice la ligadura. Así, aquellos sistemas que realicen la ligadura en la fase de acceso a los datos mantendrán esta independencia hasta el último momento y sólo vincularán aquellos datos a los que se pueda acceder y se vean involucrados en cada uno de los accesos. Mientras que los sistemas que realizan la ligadura en la fase de compilación pierden la independencia de los datos en el momento en que los programas de aplicación se traducen a código objeto.

1.3. Bases de datos y Sistemas de Gestión de Bases de Datos

Una **Base de Datos** es una colección de archivos relacionados que almacenan tanto una representación abstracta del dominio de un problema del mundo real cuyo manejo resulta de interés para una organización, como los datos correspondientes a la información acerca del mismo. Tanto la representación como los datos están sujetos a una serie de restricciones, las cuales forman parte del dominio del problema y cuya descripción está también almacenada en esos ficheros.

De esta definición podemos extraer las siguientes consideraciones:

- Se trata de una colección de archivos relacionados. La base de datos puede ser vista como un único depósito en el cual se almacena toda la información correspondiente al dominio de un problema.
- En estos archivos se encuentra almacenada tanto la representación abstracta del dominio del problema como los datos conocidos acerca del mismo en un momento dado.
- El que tanto la representación como los datos estén sujetos a una serie de restricciones implica:
 - Que las restricciones innatas al problema están representadas. Restricciones

acerca de las propiedades de las entidades, datos y relaciones existentes en el dominio del problema.

- Que el acceso a la información almacenada está sujeto a una serie de restricciones que garantizan la integridad de la misma. Estas restricciones impiden que algún procedimiento viole las reglas que vinculan los datos entre los diferentes niveles de abstracción.

Un SGBD es una colección de programas de aplicación que proporcionan al usuario de la base de datos los medios necesarios para realizar las siguientes tareas:

- Definición de los datos a los distintos niveles de abstracción.
- Manipulación de los datos en la base de datos.
- Mantenimiento de la integridad de la base de datos. Integridad en cuanto a los datos en sí, sus valores y las relaciones entre ellos.
- Control de la privacidad y seguridad de los datos en la base de datos.
- Y, en definitiva, los medios necesarios para el establecimiento de todas aquellas características exigibles a una base de datos.

1.4. Componentes de los SGBD

Un SGBD cuenta tanto con herramientas software como con personal humano especializado en la realización de las tareas y acciones necesarias para la gestión adecuada de la información.

1.4.1. El lenguaje de definición de datos

Si para garantizar la independencia de los datos es necesaria la definición de éstos a niveles de abstracción es, por tanto, necesario que el SGBD cuente con un componente que realice esta tarea. El **lenguaje de definición de los datos (DDL)**, es un lenguaje artificial basado en un determinado modelo de datos que permite la representación lógica de los datos.

La representación de los datos obtenida en este proceso de compilación es almacenada en otro componente del SGBD denominado **Diccionario de datos**.

1.4.2. El lenguaje de definición del almacenamiento de los datos

En la mayoría de los SGBD el mismo lenguaje DDL permite la definición de los datos en el nivel de representación físico, si bien en otros es un subcomponente de éste denominado **lenguaje de definición de almacenamiento de los datos (DSDL)**. En cualquier caso, haciendo uso del DDL o del DSDL, se definen los datos correspondientes al dominio de un problema a dos niveles de abstracción, y a esta definición de los datos se le denomina **Esquema de la Base de Datos**.

El esquema de la base de datos es una representación de los datos correspondientes al dominio de un problema mediante un lenguaje de definición de datos, el cual está basado en un modelo de datos. En el esquema están definidas:

- Las características del problema a un nivel de descripción lógico o intencional. Esta definición no variará a no ser que cambie el problema.
 - Cada una de las clases de objetos, y sus propiedades, que formen parte del dominio del problema o sistema que se desea tratar con el SGBD.
 - Cada una de las relaciones, y sus propiedades, existentes entre estas clases de objetos.
 - Todas aquellas restricciones concernientes tanto a las clases de objetos y sus propiedades como a las relaciones entre ellos.
- Las características del problema desde un punto de vista físico u operacional. Esta descripción puede variar con el tiempo en base a las necesidades o requisitos operacionales, y contemplará, por ejemplo:
 - Las unidades físicas en los cuales los datos van a ser almacenados.
 - Los volúmenes y archivos utilizados.

- Las características físicas y lógicas de los medios de almacenamiento y métodos de acceso a la información.

Además del DSDL, el DDL cuenta con un sublenguaje encargado del control y seguridad de los datos, denominado **lenguaje de control de datos (DCL)**, y permite el control del acceso a la información almacenada en el diccionario de datos (definición de privilegios y tipos de acceso), así como el control de la seguridad de los datos.

1.4.3. El lenguaje de manipulación de datos

Otro componente esencial de los SGBD es el lenguaje de manipulación de los datos (DML). El DML es un lenguaje artificial mediante el cual se realizan 2 funciones bien diferentes en la gestión de los datos:

- La definición del nivel externo o de usuario de los datos.
- La manipulación de los datos.

Al igual que el DDL, el DML está basado en un modelo de datos y, por tanto, los SGBD basados en distintos modelos de datos tienen diferente DML. Se trata también de un lenguaje basado en una gramática completa, sencilla y, generalmente, fácil de entender por usuarios no expertos.

Dependiendo del modelo de datos en el cual se soportan y, por supuesto, del SGBD, existen 2 tipos de DML:

- **Procedimentales:** Los cuales requieren que en las sentencias del lenguaje se especifique qué datos se van a manipular, qué se desea obtener y qué acciones/operaciones deben realizarse para ello.
- **No procedimentales:** Los cuales sólo requieren que en las sentencias del lenguaje se especifique qué datos se van a manipular y qué se desea obtener, siendo el propio DML el encargado de determinar los procedimientos más efectivos para ello.

El DML tiene también la función de describir la visión externa de los datos. Mediante el DML se definen las vistas o visiones parciales que los usuarios tienen del esquema de la base de datos definido mediante el DDL. Estas vistas de los datos son denominadas

subesquemas y pueden realizarse de varias formas:

- Haciendo uso única y exclusivamente del DML. Así, con sentencias propias de este lenguaje se definen distintas visiones parciales (orientadas al usuario final) del esquema de la base de datos. Estas vistas son, generalmente, almacenadas en el diccionario de datos.
- Haciendo uso de un lenguaje huésped (host), mediante el cual se realizan los programas de aplicación que permiten al usuario manipular los datos de la base de datos. En el código fuente de estos programas están presentes sentencias del DML, que son las encargadas de estos procesos, mientras que en las sentencias realizadas en el lenguaje huésped tienen como objetivo el control del flujo de la información y la interfaz de usuario.

En este caso, los programas fuentes deben estar precompilados, antes de ser compilados, con el compilador correspondiente, para generar el código máquina, y convertir las sentencias DML inmersas en el código fuente a un código entendible por el compilador del lenguaje huésped.

1.4.4. El diccionario de datos

El diccionario de datos es uno o un conjunto de archivos que contienen información acerca de los datos que pueden ser almacenados en la base de datos. Se trata de una metabase de datos (una base de datos (intencional) que contiene información sobre otra base de datos (extensional)).

En el diccionario de datos se almacenan todas las definiciones realizadas por el DDL sobre el problema que va a ser tratado por el SGBD y, algunas (las que se deseen) de las realizadas por el DML. Así, en el diccionario de datos se encuentra almacenado:

- El esquema lógico de la base de datos.
- El esquema físico de la base de datos.
- Los subesquemas de la base de datos.

Es decir, la representación de los datos a los 3 niveles de abstracción. Además, en el

diccionario de datos se encuentra información correspondiente con:

- Las restricciones de privacidad y acceso a los datos almacenados en la base de datos. Estas restricciones han sido definidas haciendo uso del DDL y/o su sublenguaje, el DCL.
- Las reglas, normas o restricciones referentes a la seguridad de los datos.
- Otra serie de información que permite garantizar la integridad de los datos almacenados en la base de datos.

Si los datos se definen a 3 niveles de abstracción, es necesario que en los procedimientos de acceso a estos datos se haga referencia en algún momento a las distintas representaciones de un mismo dato (el proceso de ligadura o vinculación). Por tanto, en el diccionario de datos, además de almacenarse la representación de los datos al nivel externo, lógico y físico, se almacena un conjunto de reglas que permite vincular los mismos datos desde un nivel de abstracción y representación a otro. A este conjunto de reglas se denomina **Mapa de reglas** y consiste en un conjunto de parámetros que definen qué procedimientos deben realizarse para vincular o transformar un mismo dato desde un nivel de representación a otro. A este conjunto de procedimientos o funciones que realizan la transformación de los datos entre los diferentes niveles de representación se le denomina **Mapa de datos**.

Como existen 3 niveles de representación de los datos, existirán:

- Una serie de reglas para definir la correspondencia entre las representaciones física y canónica de los datos.
- Una serie de reglas para definir la correspondencia entre la representación canónica de los datos y cada una de las representaciones externas de los mismos.

1.4.5. El gestor de la base de datos

El gestor de la base de datos, a veces denominado monitor, es un componente software encargado de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos. Este componente es el encargado de proporcionar una interfaz entre los datos almacenados y los programas de aplicación que los manejan.

Puede verse al gestor de la base de datos como un intérprete entre el usuario (de cualquier tipo) y los datos. Toda operación que se quiera realizar contra la base de datos debe ser previamente permitida por el gestor de la misma, el cual, una vez interpretada y validada, o bien realiza la operación devolviendo el resultado de la misma al programa/procedimiento que la solicitó, o bien la rechaza. Así, el gestor de la base de datos es el responsable de:

- Garantizar la privacidad de los datos, permitiendo sólo el acceso a los mismos a los usuarios autorizados.
- Garantizar la seguridad de los datos, realizando los procedimientos necesarios para que los datos puedan ser recuperados tras un fallo que ocasione una pérdida o deterioro temporal de los mismos.
- Garantizar la integridad de los datos, gestionando que los datos que se almacenan en la base de datos satisfagan las restricciones definidas en el esquema de la misma.
- Garantizar el acceso concurrente a la base de datos de forma que varios usuarios puedan acceder al mismo o distinto dato sin que ello ocasione una pérdida de la integridad de la base de datos.
- Interaccionar con el sistema operativo y, en particular, con el gestor de archivos del mismo, de forma que los procedimientos DML puedan ser entendidos por el sistema operativo para el correcto almacenamiento y recuperación de la información. Para ello, el gestor de la base de datos cuenta con un subcomponente denominado **procesador de consultas**.

Es uno de los componentes más complejos del SGBD, aunque su complejidad dependerá del propio SGBD y de las características del sistema operativo y hardware en el cual se implante.

1.4.6. El administrador de la base de datos

Otro de los componentes de los SGBD es el administrador de la base de datos (DBA). Se trata de un componente humano de suma importancia en el resultado que el uso de las bases de datos va a tener en la resolución de un determinado problema. El DBA tiene una

serie de responsabilidades en cuanto a la definición, administración, seguridad, privacidad e integridad de la información que es tratada, así como en el desempeño del SGBD en el procesamiento de la misma.

Entre las tareas asignadas al DBA se encuentran:

- La definición del esquema lógico o canónico de la base de datos. Es decir, la codificación mediante sentencias del DDL del conjunto de definiciones que representan las características del problema que va a ser tratado haciendo uso del SGBD. En esta definición se incluyen aquellas especificaciones necesarias para que el SGBD pueda mantener la integridad de los datos almacenados en la base de datos.
- La definición del esquema físico de la base de datos. Es decir, la especificación de las estructuras de almacenamiento y los métodos de acceso a la información almacenada en los dispositivos físicos de almacenamiento. Esta definición se realiza haciendo uso del DSDL (o el propio DDL) mediante un conjunto de sentencias que son compiladas y traducidas a una especificación entendible por la máquina y que es almacenada en el diccionario de datos junto con el esquema canónico.
- La definición de los subesquemas o visiones externas o de usuarios de la base de datos. Aquellas vistas parciales de la base de datos que son almacenadas en el diccionario de datos son definidas por el DBA, el cual es el único que tiene acceso y, por tanto, privilegios suficientes para la gestión de este componente.
- El control de la privacidad de los datos, mediante la concesión de privilegios a usuarios o grupos de éstos para el acceso a la información almacenada en la base de datos. Esta tarea se realiza en base al esquema de la base de datos y en base a las operaciones básicas que pueden realizarse con los datos, concediéndose privilegios para una o varias de estas acciones a grupos de datos definidos en el esquema de la base de datos.
- Mantenimiento de los esquemas. Así, el DBA es el responsable de:
 - Introducir las modificaciones necesarias en el esquema lógico; modificaciones producidas por un cambio en el problema tratado por el SGBD o una ampliación del mismo.

- Introducir las modificaciones necesarias en la representación física de los datos, de forma que esta representación evolucione paralelamente a la extensión de la base de datos y a la introducción de nuevos requisitos funcionales y/o desempeño.
- Introducir las modificaciones y nuevas definiciones de los subesquemas o visiones externas o de usuario, aportando una explotación efectiva de la base de datos.
- La especificación de los procedimientos necesarios para el mantenimiento de la seguridad de los datos almacenados en la base de datos. Es decir, cuándo, cómo y de qué forma se deben realizar y definir los procesos que garanticen que los datos puedan ser recuperados aun después de un fallo que dé lugar a una pérdida temporal de los mismos.

1.4.7. Los usuarios de la base de datos

Tradicionalmente, se ha considerado a los usuarios de la base de datos como un componente más del SGBD, posiblemente debido a que estos sistemas fueron de los primeros en considerar a éstos como una parte importante del correcto, adecuado y necesario funcionamiento de los mismos.

Naturalmente, en un sistema de esta complejidad existen muchos tipos de usuarios que acceden e interaccionan con el mismo, así se pueden considerar:

Usuarios terminales: Aquellos usuarios que, a través de programas de aplicación, interaccionan con la base de datos. Son usuarios no especializados que tienen la visión del problema que les proporcionan las visiones externas o subesquemas que utilizan los programas de aplicación a los cuales tienen privilegios de ejecución.

Usuarios técnicos: Aquellos que desarrollan los programas de aplicación que van a ser utilizados por los usuarios terminales de la base de datos. Son profesionales informáticos que, haciendo uso de lenguajes de programación, preparan procedimientos que son invocados desde una interfaz orientada al usuario para realizar las operaciones necesarias en la gestión del problema.

Usuarios especializados: Aquellos que utilizan el SGBD como una herramienta en el desarrollo de otros sistemas más o menos complejos. Estos usuarios necesitan una buena gestión de la información que es procesada por otro sistema comercial o desarrollado por ellos y, por tanto, utilizan al SGBD como un submódulo de sus sistemas particulares, interaccionando con él en la medida que le es necesario.

Usuarios críticos: Se trata de aquellos usuarios gerenciales o pertenecientes al staff de las empresas en las cuales se ha instalado la base de datos.