

Ángel

CAPÍTULO 1: INTRODUCCIÓN A LAS BASES DE DATOS

Cuando se reconoce que los sistemas evolucionan y que, por tanto, la información y la estructura de la misma no es estática sino que va cambiando con el tiempo, es cuando aparece el concepto de las Bases de Datos. Esta independencia de la información con respecto a los procedimientos que la maneja debe satisfacerse a dos niveles abstracción para que sea efectiva; por tanto, se habla de:

Independencia lógica de los datos, por la que la modificación de la representación lógica general del dominio del problema no afecta a los programas de aplicación que la manipulan.

Independencia física de los datos, por la que la distribución de los datos en la unidades de almacenamiento y 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. CARACTERISTICAS DE LAS BASES DE DATOS

- **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 visiones de esa información. Visiones parciales en la que solo se tiene en cuenta parte del dominio del problema y/o visiones globales que observan el problema desde diferentes puntos de vista.
- **Desempeño:** las bases de datos deben asegurar un tiempo de respuesta adecuado en la comunicación hombre-máquina, permitiendo el acceso simultaneo al mismo o distinto conjunto de ítems de datos por el mismo o distinto procedimiento.
- **Mínima redundancia:** 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 esta continuamente información sobre los datos almacenados. Esta característica va a depender directamente de la organización física de los datos en la base de datos.
- **Simplicidad:** las bases de datos deben estar basadas en representaciones lógicas simples que permitan la verificación en la representación del problema que representan, de tal forma que la inclusión y/o modificación de nuevos ítems de datos y relaciones no ocasionen 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.

- **Seguridad y Privacidad:** la seguridad de una base de datos hace referencia a la capacidad de esta 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.
- **Afinación:** la 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 sobre la misma.
- **Interfaz con el pasado y futuro:** una base de datos debe estar abierta a los cambios que ocasionan las necesidades de la organización de forma que no afecten, o afecten lo menos posible, a los procedimientos existentes para manejar la información que mantiene. Por otro lado debe estar abierta a reconocer información organizada físicamente por otro software.

1.2. LAS DIFERENTES VISIONES DE LOS DATOS EN LAS BASES DE DATOS

Dependiendo de quién acceda o use la base de datos, esta debe presentarle una visión de los datos que sea capaz de reconocer, interpretar y manejar.

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

- 1—**Visión externa:** es la visión de los datos que tienen los usuarios finales de una base de datos. Un usuario trata solo una visión parcial de la información, solo aquella que interviene en el dominio de actividad. Estas visiones son proporcionadas por los procedimientos o programas de aplicación que solo manejan parte de la información del problema.
- 2—**Visión conceptual:** es la visión o representación del problema tal y como este se presenta en el mundo real, es decir, es una representación abstracta del problema e independiente de cómo va a ser tratada esta información, de las aplicaciones externas y de cómo esta información pueda ser almacenada físicamente.
- 3—**Visión física:** es la representación de como la información es almacenada en los dispositivos de almacenamiento.

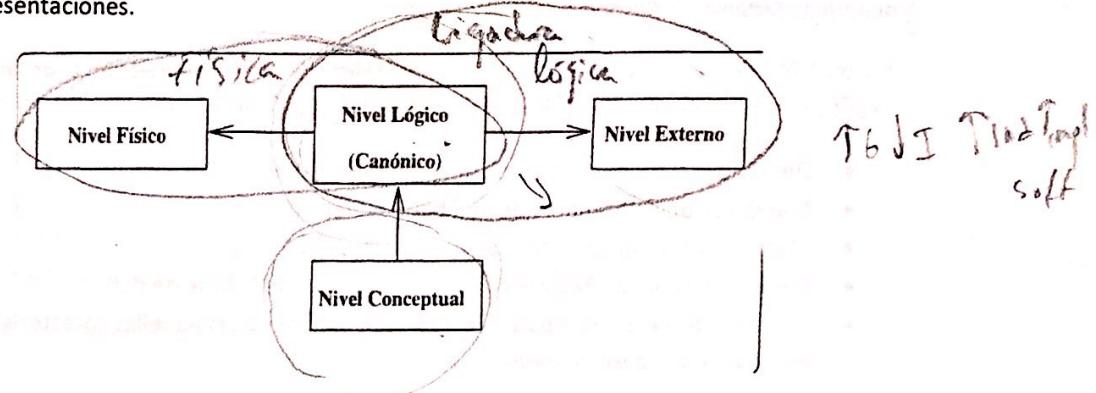
1.3. INDEPENDENCIA DEL NIVEL DE DESCRIPCION CONCEPTUAL

El nivel de descripción conceptual es seguramente el más importante, o por lo menos en el que se apoyan el resto de los niveles, en base a su calidad, se garantiza que la base de datos solucione el problema de la organización.

El fenómeno del mundo real o problema 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 canonica es derivada de la descripción conceptual en base a la aplicación de una serie de reglas y restricciones que tiene en cuenta cómo la información representada puede ser tratada por los procedimientos que van a manejar y definir la información en base a las otras representaciones.



1.4. GRANULARIDAD Y LIGADURA

Al nivel de detalle en que pueden ser descritas las representaciones externas derivadas de la representación lógica se le denomina *Granularidad*. A mayor granularidad de una representación externa (menor información a considerar) mayor será la independencia, pero al mismo tiempo una mayor complejidad en el software utilizado al igual que en la organización física de la información; y viceversa.

Para garantizar la integridad de la base de datos es, por tanto, necesario que en algún instante esos procedimientos que manipulan las representaciones de un determinado nivel de abstracción tenga en cuenta como se representa la información en los otros niveles. 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 dos 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 física.

1.5. 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.

Para que la información pueda ser almacenada y el acceso a la misma satisfaga las características exigidas a una base de datos, es necesario que exista una serie de procedimientos que sea capaz de llevar a cabo tal labor. A este sistema software es lo se le denomina *Sistema de Gestión de Bases de Datos*.

Así, 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 de una base de datos.
- Mantenimiento de la integridad de la base de datos.
- Control de la privacidad y seguridad de los datos de la base de datos.
- Y los medios necesarios para el establecimiento de todas aquellas características exigibles a una base de datos.

1.6. 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.

1.6.1. El lenguaje de definición de datos

El lenguaje de definición de los datos - *Data Definition Language (DDL)* - es un lenguaje artificial basado en un determinado modelo de datos que permite la representación lógica de los datos. Son lenguajes simples basados en una gramática sencilla que cuenta con un conjunto muy reducido de morfemas.

La representación obtenida en este proceso de compilación es almacenada en otro componente del SGBD llamado *Diccionario de Datos*.

1.6.2. 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 - *Data Storage Definition Language (DSDL)* -.

Además del DSDL, el DDL cuenta con un sublenguaje encargado del control y seguridad de los datos, el cual se denomina lenguaje de control de datos - *Data Control Language (DCL)* - y permite el control del acceso a la información almacenada en el diccionario de datos.

1.6.3. El lenguaje de manipulación de datos

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

1. La definición del nivel externo o de usuario de los datos.

- La manipulación de los datos; es decir, la inserción, borrado, modificación y recuperación de los datos almacenados en la base de datos.

Dependiendo del modelo de datos en el cual se soportan y, por supuesto, del SGBD, existen dos 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.

1.6.4. 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; es decir, una base de datos (intencional) que contiene información sobre otra base de datos (extensional).

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 tres niveles de abstracción.

CSIE

1.6.5. 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.

1.6.6. El administrador de la base de datos

Otro de los componentes de los SGBD es el administrador de la base de datos - Data Base Administrator (DBA) -. Se trata de un componente humano de suma importancia en el resultado que el uso de las bases de datos van a tener en la resolución de un determinado problema.

Entre las tareas asignadas al DBA se encuentran:

- La definición del esquema canónico o lógico de la base de datos.
- La definición del esquema físico de la base de datos.
- La definición de los subesquemas o visiones externas o de usuarios de la base de datos.
- El control de la privacidad de los datos, mediante la concesión de privilegios usuarios o grupos de estos para el acceso de la información almacenada en la base de datos.

- Mantenimiento de los esquemas.
- La especificación de los procedimientos necesarios para el mantenimiento de la seguridad de los datos almacenados en la base de datos.

1.4.7. Los usuarios de la base de datos

- **Usuarios terminales:** aquellos usuarios que, a través de programas de aplicación, interactúan con la base de datos. Son usuarios no especializados que tiene la visión del problema que les proporcionan las visiones eternas o subesquemas que utilizan los programas de aplicación a los cuales tiene privilegio 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.
- **Usuarios especializados:** aquellos que utilizan el SGBD como una herramienta en el desarrollo de otros sistemas más o menos complejos.
- **Usuarios críticos:** estos usuarios pueden tener desde mucho, hasta ningún conocimiento técnico de la tecnología de la base de datos, y/o del SGBD en el cual se soporta la base de datos con la cual interactúan pero, independientemente de ello, requieren de la base de datos información en un formato, detalle y bajo unos requisitos que generalmente no está previsto de antemano y en un tiempo mínimo.



BASES DE DATOS. Curso 2013-2014
Convocatoria de Febrero

3. Describa cada uno de los etapas y tareas del análisis y diseño de una base de datos relacionales, indicando en cada ítem su objetivo y actividades que se deben llevar a cabo (2.5 puntos)

Respuesta	
0	Descripción y Análisis La actividad se fundamenta en la descripción detallada del problema, y el análisis de requisitos que sirven de base al modelado de la información
1	Elicitación de los tipos de entidad Descripción de los objetos del mundo real que representan. Nominalos
2	Elicitación de los atributos Descripción de las propiedades de los objetos del mundo real que representan. Nominalos. Definir el dominio y restricciones
3	Elicitación de los tipos de Interrelaciones Descripción de las relaciones entre los tipos de entidades. Nominalas. Definir el significado y cardinalidades mínimas y máximas. Definir si existen debilidades. Definir los atributos que las caracterizan, nominalos, establecer su dominio y restricciones.
4	Elicitar toda la semántica Describir la existencia de relaciones jerárquicas, exclusividades, reflexividades o cualquier otro tipo de semántica avanzada. Caracterizar atributos y las características de las relaciones
5	Construcción del modelo EE/R Construir el modelo EE/R y validar con la descripción de los elementos
6	Validación de la información Validar el modelo EE/R con arreglo a los requisitos de información
7	Validación funcional Navegar sobre el modelo validando los requisitos funcionales
8	Simplificación Refinar el modelo, simplificándolo para su más fácil, eficiente y completa traducción al modelo lógico. Eliminar atributos múltiples, compuestos y jerarquía.
9	Traducción al Modelo relacional Aplicar las reglas de transformación del modelo EE/R al modelo Relacional
10	Normalizar Normalizar el modelo
11	Validar con requisitos Incluir todas aquellas restricciones que satisfagan los requisitos. Asertos, triggers, Checks, etc.
12	Refinar y afinar Refinar el modelo, si fuera necesario, para una más eficiente implementación de la base de datos. Definir índices, tablespaces, etc.
13	Generar la BD Definir el script de la BD, depurar y validar.

Los relaciones R_1 y R_2 se dice que son compatibles si ambas relaciones tienen el mismo grado y el atributo inverso de R_1 está definido en el mismo dominio que el atributo inverso de R_2 , si bien los atributos pueden tener nombres diferentes.

Tipo de registro ficticio en las bases de datos jerárquicas

Dado que el modelo jerárquico no permite representar relaciones N:N, en ocasiones los tipos de registros deben ser duplicados. Un tipo de registro ficticio es la duplicación de un tipo de registro existente en el modelo cuyos atributos son punteros a los registros existentes en el tipo de registro original.

Problemas que evita normalizar

¿Qué problemas evita normalizar en FN2 en las operaciones de inserción?

- Evita redundancia de valores de atributos dependientes, ya que se repite su valor para un mismo valor del atributo clave.
- Evita consistencia de valores de atributos dependientes, ya que pueden tener distinto valor para un mismo valor del atributo clave.
- No se puede almacenar la dependencia funcional completa hasta que no se haya insertado la tupla.

¿Por qué la FNBC no requiere que una relación esté en FN2 o FN3?

Porque la FNBC se basa en el concepto de determinante funcional que incorpora el concepto de dependencia funcional completa. Dado que los determinantes funcionales sólo pueden ser claves candidatas de la relación, en su definición está implícita que las dependencias funcionales de los atributos de la clave debe ser completa (FN2) y que no puede haber dependencias funcionales entre atributos que no sean clave (FN3).

Checkpoint

Los checkpoint son puntos de control que se realizan en las copias de seguridad incrementales de las bases de datos. Cada vez que se produce un checkpoint se pasa a definitivas y almacenamiento estable las transacciones realizadas desde el último checkpoint. Esto se realiza sobre los ficheros logs/redo tanto si el procedimiento es con almacenamiento por adelanto o por diferido de las transacciones de las bases de datos.

Transacción

Una transacción es un conjunto de operaciones constituidas por un conjunto de acciones que se ejecutan contra una base de datos y que debe satisfacer las reglas ACID.

Una acción es cualquier operación que se ejecute contra la base de datos (leer, escribir).

Una operación es un conjunto de acciones que actúa sobre el mismo gránulo de la base de datos.

En base de datos se denomina ACID a las características de los parámetros que permiten clasificar las transacciones de los SGBD. Estas características son:

Atomicity, Consistency, Isolation and Durability

(Atomicidad, Consistencia, Aislamiento y Durabilidad)

- Atomicidad: si una operación consiste en una serie de pasos, todos ellos ocurren o ninguno, es decir, las transacciones son completas.
- Consistencia: Integridad. Se ejecutan aquellas operaciones que no van a romper las reglas o directrices de integridad de la base de datos.
- Aislamiento: es la propiedad que asegura que una operación no puede afectar a otra.
- Durabilidad: persistencia. Es la propiedad que asegura que una vez realizada la operación, esta persistirá y no se podrá deshacer aunque falle el sistema y que de esta forma los datos sobrevivan de alguna manera.

Autonomía en las bases de datos distribuidas

La autonomía en las bases de datos distribuidas es la capacidad que tienen los DBA locales de actuar sobre el esquema de la base de datos distribuida, actualizando o modificando dicho esquema definiendo nuevos objetos o modificando los existentes.

Tipo de registro ficticio (dummy) en las bases de datos en red

En el modelo en red las relaciones N:N son mantenidas a través de la definición de un tipo de registro dummy, que puede tener atributos (o no) que mantiene dos relaciones 1:N con los dos tipos de registro que en el problema mantienen una relación N:N.

Conjunto DBTG en las bases de datos en red

Un conjunto DBTG está formado por un tipo de registro owner y un tipo de registro member y un tipo de set que mantiene entre ambos tipos de registros una relación 1:1 o 1:N.

Los tipos de registros representan a conjuntos de objetos del dominio del problema representados por un conjunto de atributos.

El tipo de set representa el conjunto de relaciones existentes en el domino del problema entre objetos.