**Capítulo 1: Introducción**

**1.3 - Visión de los Datos**

Un sistema de bases de datos es una colección de archivos interrelacionados y un conjunto de programas que permitan a los usuarios acceder y modificar estos archivos. Uno de los propósitos principales es proporcionar a los usuarios una visión abstracta de los datos.

Abstracción de datos

Para que el sistema sea útil debe recuperar los datos eficientemente. Los desarrolladores esconden la complejidad a los usuarios a través de varios niveles de abstracción para simplificar la interacción de los usuarios con el sistema:

* Nivel físico: el nivel más bajo de abstracción describe cómo se almacenan realmente los datos. Se describen en detalle las estructuras de datos complejas de bajo nivel.
* Nivel lógico: describe qué datos se almacenan en la base de datos y qué relaciones existen entre esos datos. La base de datos completa se describe así en términos de un número pequeño de estructuras relativamente simples.
* Nivel de vistas: nivel más alto de abstracción, describe sólo parte de la base de datos completa. En el nivel lógico, queda algo de complejidad, debido a la variedad de información almacenada en una gran base de datos. Muchos usuarios no necesitan toda esta información, sino que necesitan acceder sólo a una parte de la base de datos. Para que su interacción con el sistema se simplifique, se define la abstracción del nivel de vistas. El sistema puede proporcionar muchas vistas para la misma base de datos.

Además de esconder detalles del nivel lógico de la base de datos, las vistas también proporcionan un mecanismo de seguridad para evitar que los usuarios accedan a ciertas partes de la base de datos.

Ejemplares y esquemas

Las bases de datos van cambiando a lo largo del tiempo conforme la información se inserta y borra. La colección de información almacenada en la base de datos en un momento particular se denomina un *ejemplar* de la base de datos. El diseño completo de la base de datos se llama el *esquema* de la base de datos. Los esquemas son raramente modificados, si es que lo son alguna vez.

**1.4 – Modelos de los datos**

Bajo la estructura de la base de datos se encuentra el modelo de datos: una colección de herramientas conceptuales para describir los datos, las relaciones, la semántica y las restricciones de consistencia. Para ilustrar el concepto de un modelo de datos, describimos dos modelos de datos en este apartado: el modelo entidad-relación y el modelo relacional.

Modelo entidad-relación

El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real que consta de una colección de objetos básicos, llamados entidades, y de relaciones entre estos objetos. Una entidad es una «cosa» u «objeto» en el mundo real que es distinguible de otros objetos. Las entidades se describen en una base de datos mediante un conjunto de atributos.

Una relación es una asociación entre varias entidades. El conjunto de todas las entidades del mismo tipo, y el conjunto de todas las relaciones del mismo tipo, se denominan respectivamente **conjunto de entidades** y **conjunto de relaciones**.

La estructura lógica general de una base de datos se puede expresar gráficamente mediante un diagrama ER,

que consta de los siguientes componentes:

* **Rectángulos**, que representan conjuntos de entidades.
* **Elipses**, que representan atributos.
* **Rombos**, que representan relaciones entre conjuntos de entidades.
* **Líneas**, que unen los atributos con los conjuntos de entidades y los conjuntos de entidades con las relaciones.

El modelo E-R representa ciertas restricciones que los contenidos de la base de datos deben cumplir. Una restricción importante es la correspondencia de cardinalidades, que expresa el número de entidades con las que otra entidad se puede asociar a través de un conjunto de relaciones.

Modelo relacional

En el modelo relacional se utiliza un grupo de tablas para representar los datos y las relaciones entre ellos. Cada tabla está compuesta por varias columnas, y cada columna tiene un nombre único.

El modelo relacional es un ejemplo de un modelo basado en registros. Los modelos basados en registros se denominan así porque la base de datos se estructura en registros de formato fijo de varios tipos. Cada tabla contiene registros de un tipo particular. Cada tipo de registro define un número fijo de campos, o atributos. Las columnas de la tabla corresponden a los atributos del tipo de registro.

El modelo de datos relacional es el modelo de datos más ampliamente usado, y una amplia mayoría de sistemas de bases de datos actuales se basan en el modelo relacional.

El modelo relacional se encuentra a un nivel de abstracción inferior al modelo de datos E-R. Los diseños de bases de datos a menudo se realizan en el modelo E-R, y después se traducen al modelo relacional.

Otros modelos de datos

El modelo de datos orientado a objetos es otro modelo de datos que está recibiendo una atención creciente. Se puede observar como una extensión del modelo E-R con las nociones de encapsulación, métodos (funciones) e identidad de objeto.

El modelo de datos relacional orientado a objetos combina las características del modelo de datos orientado a objetos y el modelo de datos relacional.

**1.5 – Lenguajes de Bases de Datos**

Un sistema de bases de datos proporciona un **lenguaje de definición de datos (DDL)** para especificar el esquema de la base de datos y un **lenguaje de manipulación de datos (DML)** para expresar las consultas a la base de datos y las modificaciones. Forman partes de un único lenguaje de bases de datos, tal como SQL.

Lenguaje de definición de datos

Un esquema de base de datos se especifica mediante un conjunto de definiciones expresadas mediante un lenguaje especial llamado lenguaje de definición de datos.

Lenguaje de manipulación de datos

La manipulación de datos es:

* La recuperación de información almacenada en la base de datos.
* La inserción de información nueva en la base de datos.
* El borrado de información de la base de datos.
* La modificación de información almacenada en la base de datos.

Permite a los usuarios acceder o manipular los datos organizados mediante el modelo de datos apropiado. Dos tipos:

* **DMLs procedimentales**. Requieren que el usuario especifique qué datos se necesitan y cómo obtener esos datos.
* **DMLs declarativos** (DMLs no procedimentales). Requieren que el usuario especifique qué datos se necesitan sin especificar cómo obtener esos datos.

Una consulta es una instrucción de solicitud para recuperar información.

**1.6 – Usuarios y administradores de la Base de Datos**

Usuarios de bases de datos e interfaces de usuario

Cuatro tipos diferentes de usuarios de un sistema de base de datos:

* **Usuarios normales**. Son usuarios no sofisticados que interactúan con el sistema mediante la invocación de alguno de los programas de aplicación permanentes que se ha escrito previamente. La interfaz de usuario normal para los usuarios normales es una interfaz de formularios, donde el usuario puede rellenar los campos apropiados del formulario. Los usuarios normales pueden también simplemente leer informes generados de la base de datos.
* **Programadores de aplicaciones**. Son profesionales informáticos que escriben programas de aplicación.
* **Usuarios sofisticados**. Interactúan con el sistema sin programas escritos. En su lugar, ellos forman sus consultas en un lenguaje de consulta de bases de datos. Los analistas que envían las consultas para explorar los datos en la base de datos entran en esta categoría.
* **Usuarios especializados**. Son usuarios sofisticados que escriben aplicaciones de bases de datos especializadas que no son adecuadas en el marco de procesamiento de datos tradicional.

Administrador de la base de datos

Una de las principales razones de usar SGBDs es tener un control centralizado tanto de los datos como de los programas que acceden a esos datos. La persona que tiene este control central sobre el sistema se llama administrador de la base de datos (ABD). Las funciones del ABD incluyen las siguientes:

* **Definición del esquema**. El ABD crea el esquema original de la base de datos escribiendo un conjunto de instrucciones de definición de datos en el LDD.
* **Definición de la estructura y del método de acceso**.
* **Modificación del esquema y de la organización física**. Los ABD realizan cambios en el esquema y en la organización física para reflejar las necesidades cambiantes de la organización, o para alterar la organización física para mejorar el rendimiento.
* **Concesión de autorización para el acceso a los datos**. La concesión de diferentes tipos de autorización permite al administrador de la base de datos determinar a qué partes de la base de datos puede acceder cada usuario.
* **Mantenimiento rutinario**.

**1.8 – Estructura de un Sistema de Bases de Datos**

Un sistema de bases de datos se divide en módulos que se encargan de cada una de las responsabilidades del sistema completo. Los componentes funcionales de un sistema de bases de datos se pueden dividir a grandes rasgos en los componentes *gestor de almacenamiento* y *procesador de consultas*.

El gestor de almacenamiento es importante porque las bases de datos requieren normalmente una gran cantidad de espacio de almacenamiento.

El procesador de consultas es importante porque ayuda al sistema de bases de datos a simplificar y facilitar el acceso a los datos.

Gestor de almacenamiento

Un gestor de almacenamiento es un módulo de programa que proporciona la interfaz entre los datos de bajo nivel en la base de datos y los programas de aplicación y consultas emitidas al sistema. El gestor de almacenamiento es responsable del almacenamiento, recuperación y actualización de los datos en la base de datos.

Componentes del gestor de almacenamiento:

* **Gestor de autorización e integridad**
* **Gestor de transacciones**
* **Gestor de archivos**
* **Gestor de memoria compartida**

El gestor de almacenamiento implementa varias estructuras de datos como parte de la implementación física del sistema:

* **Archivos de datos**.
* **Diccionario de datos**.
* **Índices**.

Procesador de consultas

Componentes del procesador de consultas:

* **Intérprete del LDD**
* **Compilador del LMD**
* **Motor de evaluación de consultas**

**Capítulo 2: Modelo Entidad-Relación**

El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real consistente en objetos básicos llamados entidades y de relaciones entre estos objetos.

**2.1-Conceptos Básicos**

Hay tres nociones básicas que emplea el modelo de datos E-R: conjuntos de entidades, conjuntos de relaciones y atributos.

Conjuntos de entidades

Una entidad es una «cosa» u «objeto» en el mundo real que es distinguible de todos los demás objetos.

Un conjunto de entidades es un conjunto de entidades del mismo tipo que comparten las mismas propiedades, o atributos.

Una entidad se representa mediante un conjunto de atributos. Los atributos describen propiedades que posee cada miembro de un conjunto de entidades. Para cada atributo hay un conjunto de valores permitidos, llamados el dominio, o el conjunto de valores, de ese atributo.

Conjuntos de relaciones

Una relación es una asociación entre diferentes entidades. Un conjunto de relaciones es un conjunto de relaciones del mismo tipo.

La asociación entre conjuntos de entidades se conoce como participación.

**2.2 – Restricciones**

Un esquema de desarrollo E-R puede definir ciertas restricciones a las que los contenidos de la BD se deben adaptar.

Correspondencia de cardinalidades

Expresa el número de entidades a las que otra entidad puede estar asociada vía un conjunto de relaciones.

Para un conjunto de relaciones binarias R entre los conjuntos de entidades A y B, la correspondencia de cardinalidades debe ser una de las siguientes:

* **Uno a uno.** Una entidad en A se asocia con a lo sumo una entidad en B, y una entidad en B se asocia con a lo sumo una entidad en A.
* **Uno a varios.** Una entidad en A se asocia con cualquier número de entidades en B (ninguna o varias). Una entidad en B, sin embargo, se puede asociar con a lo sumo una entidad en A.
* **Varios a uno.** Una entidad en A se asocia con a lo sumo una entidad en B. Una entidad en B, sin embargo, se puede asociar con cualquier número de entidades en A.
* **Varios a varios.** Una entidad en A se asocia con cualquier número de entidades en B y una entidad en B se asocia con cualquier número de entidades en A.

Restricciones de participación

La participación de un conjunto de entidades E en un conjunto de relaciones R se dice que es **total** si cada entidad en E participa al menos en una relación en R. Si sólo algunas entidades en E participan en relaciones en R, la participación del conjunto de entidades E en la relación R se llama **parcial**.

**2.3 – Claves**

Una clave permite identificar un conjunto de atributos suficiente para distinguir las entidades entre sí. Las claves también ayudan a identificar unívocamente a las relaciones y así a distinguir las relaciones entre sí.

Conjuntos de entidades

Una **superclave** es un conjunto de uno o más atributos que, tomados colectivamente, permiten identificar de forma única una entidad en el conjunto de entidades. A menudo interesan las superclaves tales que los subconjuntos propios de ellas no son superclave. Tales superclaves mínimas se llaman **claves candidatas**.

Se usa el término **clave primaria** para denotar una clave candidata que es elegida por el diseñador de la base de datos como elemento principal para identificar las entidades dentro de un conjunto de entidades.

**2.5 – Diagrama Entidad-Relación**

Componentes principales:

* **Rectángulos**, que representan conjuntos de entidades.
* **Elipses**, que representan atributos.
* **Rombos**, que representan relaciones.
* **Líneas**, que unen atributos a conjuntos de entidades y conjuntos de entidades a conjuntos de relaciones.
* **Elipses dobles**, que representan atributos multivalorados (atributos que pueden tener más de un valor).
* **Elipses discontinuas**, que denotan atributos derivados.
* **Líneas dobles**, que indican participación total de una entidad en un conjunto de relaciones.
* **Rectángulos dobles**, que representan conjuntos de entidades débiles.

**2.7 – Características del Modelo E-R Extendido**

Especialización

Un conjunto de entidades puede incluir subgrupos de entidades que se diferencian de alguna forma de las otras entidades del conjunto. Por ejemplo, un subconjunto de entidades en un conjunto de entidades puede tener atributos que no son compartidos por todas las entidades del conjunto de entidades.

El proceso de designación de subgrupos dentro de un conjunto de entidades se denomina **especialización**.

Generalización

Es una relación contenedora que existe entre el conjunto de entidades de nivel más alto y uno o más conjuntos de entidades de nivel más bajo. Los conjuntos de entidades de nivel más alto y nivel más bajo también se pueden llamar superclase y subclase, respectivamente.

Herencia de Atributos

Una propiedad crucial de las entidades de nivel más alto y más bajo creadas mediante especialización y generalización es la herencia de atributos. Los atributos de los conjuntos de entidades de nivel más alto se dice que son **heredados** por los conjuntos de entidades de nivel más bajo.

Agregación

Es una abstracción a través de la cual las relaciones se tratan como entidades de nivel más alto.

**2.9 – Reducción de un Esquema E-R a Tablas**

Una base de datos que se ajusta a un esquema de bases de datos E-R se puede representar por una colección de tablas.

Representación tabular de los conjuntos de entidades fuertes

Sea E un conjunto de entidades fuertes con los atributos descriptivos . Esta entidad se representa mediante una tabla llamada E con columnas distintas, cada una de las cuales corresponde a uno de los atributos de E. Cada fila de la tabla corresponde a una entidad del conjunto de entidades E.

Representación tabular de los conjuntos de entidades débiles

Sea A un conjunto de entidades débiles con los atributos . Sea B el conjunto de entidades fuertes del que A depende. Sea la clave primaria de B el conjunto de atributos . Se representa el conjunto de entidades A mediante una tabla llamada A con una columna por cada uno de los atributos del conjunto: .

Representación tabular de los conjuntos de relaciones

Sea R un conjunto de relaciones, sean el conjunto de atributos formados por la unión de las claves primarias de cada uno de los conjuntos de entidades que participan en R, y sean los atributos descriptivos de R (si los hay). El conjunto de relaciones se representa mediante una tabla llamada R con una columna por cada uno de los atributos del conjunto: .

Combinación de tablas

Considérese un conjunto AB de relaciones varios a uno del conjunto de entidades A al conjunto de entidades B. Usando el esquema de construcción de tablas descrito previamente se consiguen tres tablas: A, B y AB. Supóngase además que la participación de A en la relación es total; es decir, cada entidad a en el conjunto de entidades A debe participar en la relación AB. Entonces se pueden combinar las tablas A y AB para formar una única tabla consistente en la unión de las columnas de ambas tablas.

En el caso de relaciones uno a uno, la tabla del conjunto de relaciones se puede combinar con las tablas de cualquiera de los conjuntos de entidades. Las tablas se pueden combinar incluso si la participación es parcial usando valores nulos.

Atributos compuestos

Los atributos compuestos se manejan creando un atributo separado para cada uno de los atributos componentes; no se crea una columna separada para el propio atributo compuesto.

Atributos multivalorados

Para un atributo multivalorado M se crea una tabla T con una columna C que corresponde a la clave primaria del conjunto de entidades o conjunto de relaciones del que M es atributo.

{Falta capítulo 18, puntos 18.1, 18.2, 18.3, 18,4}