

DEPARTAMENTO DE INGENIERÍA Y
CIENCIA DE LOS COMPUTADORES

F47. FICHEROS Y BASES DE DATOS

< <http://www3.uji.es/~mmarques/f47> >

Segundo curso. I.T.I.G.

Curso 2001/2002
Segundo Cuatrimestre
7,5 Créditos
(4 Teoría + 3,5 Prácticas)

Profesores de teoría: Merche Marqués, Miguel Pérez,
Profesores de prácticas: Juanma Chiva, Salvador García, Merche Marqués,
Raúl Montoliu, Juanma Pérez, Gregorio Quintana

1.- OBJETIVOS

Teoría

En esta asignatura se introducen los sistemas de bases de datos como el modo fundamental de organizar los datos en los sistemas de información. Se centra el estudio en los sistemas de bases de datos relacionales, ya que son los más extendidos hoy en día y poseen una sólida base teórica que los sustenta. Además de aprender el uso y funcionamiento de los sistemas relacionales, se estudia una metodología de diseño que es también la más utilizada, tanto en el ámbito profesional como en el académico.

El objetivo general es, por lo tanto, el de ofrecer una base sólida en los aspectos fundamentales de la tecnología de bases de datos desde el enfoque relacional. Dichos aspectos comprenden las ideas, tanto teóricas como prácticas, que son la base de los sistemas relacionales, y el problema del diseño de bases de datos.

Prácticas

El objetivo es aprender el lenguaje SQL y utilizarlo para interactuar con bases de datos relacionales, tanto en el ámbito interactivo, como embebido en un lenguaje de programación de alto nivel. También se aprende el manejo de herramientas que permiten la generación de aplicaciones.

2.- TEMARIO

Teoría

Tema 1. Introducción

En este tema se presentan los sistemas de bases de datos, haciendo antes un repaso por sus predecesores, los sistemas de ficheros. Aunque los sistemas de ficheros se han quedado obsoletos, hay dos buenas razones para estudiarlos. En primer lugar, el conocer los problemas de este tipo de sistemas nos previene de volver a cometerlos. En segundo lugar, si en algún momento fuera necesario convertir un sistema de ficheros en un sistema de bases de datos, comprender cómo trabaja este sistema puede ser una ayuda esencial.

Este tema consta de los siguientes apartados:

1. Sistemas de ficheros
2. Sistemas de bases de datos
3. Papeles en el entorno de las bases de datos
4. Historia de los sistemas de bases de datos

Resumen

Los predecesores de los sistemas de bases de datos son los sistemas de ficheros. Un sistema de ficheros está formado por un conjunto de programas que dan servicio a los usuarios finales. Cada programa define y gestiona sus propios datos. Aunque los sistemas de ficheros supusieron un gran avance sobre los sistemas manuales, tienen inconvenientes bastante importantes, como la redundancia de datos y la dependencia entre programas y datos.

Los sistemas de bases de datos surgieron con el objetivo de resolver los problemas que planteaban los sistemas de ficheros. Una base de datos es un conjunto de datos relacionados que recogen las necesidades de información de una empresa u organización. Estos datos se comparten por todos los usuarios. El sistema de gestión de bases de datos es un conjunto de programas que permiten a los usuarios definir, crear y mantener la base de datos, además de proporcionar un acceso controlado a la misma.

La base de datos contiene tanto los datos como su definición. Todos los accesos a la base de datos se realizan a través del sistema de gestión de bases de datos. El sistema de gestión de bases de datos proporciona un lenguaje de definición de datos que permite a los usuarios definir la base de datos, y un lenguaje de manejo de datos que permite a los usuarios la inserción, actualización, eliminación y consulta de datos de la base de datos. El sistema de gestión de bases de datos también proporciona un acceso controlado a la base de datos: proporciona seguridad, integridad, concurrencia y controla la recuperación ante fallos. Además, proporciona un mecanismo de vistas que permite mostrar a los usuarios sólo aquellos datos que les interesan.

Las personas involucradas en el entorno de una base de datos son: el administrador de la base de datos, los diseñadores de la base de datos, los programadores de aplicaciones y los usuarios finales.

Los sistemas jerárquico y de red representan la primera generación de los sistemas de gestión de bases de datos, surgida hacia los años sesenta. El modelo relacional, propuesto por E. F. Codd en 1970, representa la segunda generación de los sistemas de gestión de bases de datos. Este modelo es el más extendido en la actualidad. La tercera generación de los sistemas de gestión de bases de datos se encuentra representada por el modelo relacional extendido y el modelo orientado a objetos.

Los sistemas de bases de datos presentan una serie de ventajas, tanto por el hecho de compartir los datos, como por la existencia del sistema de gestión de bases de datos. Algunas de estas ventajas son el control de la redundancia, la consistencia de datos, la mejora en los aspectos de seguridad y la integridad. Algunos de sus inconvenientes son su elevada complejidad, su coste y su vulnerabilidad ante fallos.

Bibliografía

En la mayoría de los libros sobre sistemas de bases de datos se presentan estos sistemas haciendo un repaso de sus predecesores, los sistemas de ficheros. Para la elaboración de este tema en el libro de apuntes de la asignatura se ha utilizado el capítulo 1 de CONNOLLY, BEGG Y STRACHAN (1996).

Tema 2. Organizaciones de ficheros y estructuras de acceso

En este tema se presentan los conceptos fundamentales sobre el almacenamiento físico de la base de datos en dispositivos de almacenamiento secundario, como los discos.

Este tema consta de los siguientes apartados:

1. Introducción
2. Conceptos fundamentales de organizaciones de ficheros
3. Dispositivos de almacenamiento secundario
 - 3.1 Discos
 - 3.2 Acceso a los datos
4. Ficheros desordenados
5. Ficheros ordenados

6. Ficheros dispersos
 - 6.1 Dispersión dinámica
 - 6.2 Dispersión extensible
 - 6.3 Dispersión lineal
7. Agrupamiento
8. Índices
 - 8.1 Índices de un solo nivel
 - 8.2 Índices multinivel
 - 8.3 Árboles B y árboles B+
 - 8.4 Ficheros dispersos como índices

Resumen

Los discos magnéticos son los dispositivos de almacenamiento secundario más utilizados para almacenar los ficheros de una base de datos. Los datos se almacenan en el disco en sectores o en bloques. El acceso a un sector o bloque es costoso debido al tiempo de búsqueda, el tiempo de rotación y el tiempo de transferencia del mismo.

Hay tres organizaciones primarias para los ficheros: desordenados, ordenados y dispersos. Los ficheros desordenados requieren una búsqueda lineal para localizar registros, pero la inserción de éstos es muy sencilla. La eliminación plantea problemas en el sentido de que se desperdicia espacio, recuperable mediante una reorganización del fichero. Los ficheros ordenados reducen el tiempo requerido para leer registros en orden, según el campo de ordenación. El tiempo necesario para buscar un registro cualquiera, dado el valor de su campo de ordenación, también se reduce, ya que se puede realizar una búsqueda binaria. Sin embargo, la necesidad de mantener los registros en orden hace muy costosa la inserción. Por esta razón, se puede utilizar un fichero de desborde no ordenado para reducir el coste de la inserción. Los registros de desborde se fusionan con el fichero de datos periódicamente.

La dispersión ofrece acceso muy rápido a un registro cualquiera dado el valor de su campo de dispersión. Las colisiones que causan desborde en los bloques se pueden resolver de varios modos, siendo el más frecuente el encadenamiento. El acceso a un registro a través de un campo que no es el campo de dispersión, es lento, y lo mismo sucede con el acceso secuencial a los registros a través de cualquier campo. Hay varias técnicas que permiten al fichero expandirse y contraerse dinámicamente: la dispersión dinámica, la extensible y la lineal.

El agrupamiento es otro modo de organizar los registros, cuyo resultado es un fichero ordenado de un modo especial. Mediante el agrupamiento se implementan físicamente las relaciones entre los registros de varios ficheros, almacenándolos consecutivos en el mismo fichero.

Los índices son estructuras de acceso adicionales con las que se mejora la eficiencia en la obtención de registros de un fichero de datos. Dichas estructuras de acceso pueden usarse junto con las organizaciones de ficheros vistas anteriormente. Hay tres tipos de índices ordenados de un solo nivel: primarios, secundarios y de agrupamiento. Cada índice se basa en un campo del fichero. Los índices primarios y de agrupamiento se construyen según el campo de ordenamiento físico del fichero, en tanto que los índices secundarios se basan en campos que no son de ordenamiento. El campo de un índice primario debe ser, además, un campo con valores únicos (campo clave), cosa que no sucede con un índice de agrupamiento. Los índices de un solo nivel son ficheros ordenados, y se examinan mediante búsqueda binaria. Se pueden construir índices de varios niveles (multinivel) para mejorar la eficiencia en las búsquedas.

Los árboles B y los árboles B+ son índices de varios niveles cuya estructura permite al índice expandirse y contraerse dinámicamente. Los nodos (bloques) de estas estructuras se mantienen ocupados entre el 50% y el 100% de su capacidad gracias a sus algoritmos de inserción y eliminación. Después de cierto tiempo, los nodos se estabilizan en un grado de ocupación medio del 69%, lo que deja espacio para hacer inserciones sin tener que reorganizar el índice con mucha frecuencia. En general, los árboles B+ pueden contener más entradas en sus nodos internos que los árboles B, por lo que es posible que un árbol B+ tenga menos niveles o incluya más entradas que el árbol B equivalente.

También es posible construir índices mediante estructuras de datos que utilizan la técnica de la dispersión.

Bibliografía

Para elaborar este tema en el libro de apuntes de la asignatura se han utilizado los capítulos 4 y 5 del texto de ELMASRI Y NAVATHE (1997), en donde se describen las principales organizaciones de ficheros para el almacenamiento de datos en disco. En el texto se realiza un estudio de este dispositivo de almacenamiento secundario. Estos autores también realizan un análisis de las distintas estructuras de índices para el acceso a estos ficheros.

FOLK Y ZOELICK (1992) presentan con mayor detalle todos los conceptos básicos sobre estructuras de ficheros en el capítulo 4 y estudian los dispositivos de almacenamiento secundario magnéticos más utilizados: discos y cintas, así como el CD-ROM (capítulo 3 y apéndice A). Además, analizan de modo muy exhaustivo los distintos tipos de índices (capítulo 6), los árboles B y los árboles B+ (capítulos 8 y 9), y la dispersión, haciendo un extenso estudio de la dispersión extensible (capítulos 10 y 11).

Tema 3. Sistemas de bases de datos

En este tema se presenta un análisis de los modelos de datos y se definen los conceptos de esquema y estado de una base de datos. Se trata la arquitectura de los sistemas de bases de datos y la independencia respecto a los datos. También se definen los distintos lenguajes que proporcionan los sistemas de gestión de bases de datos y se estudian varias clasificaciones de estos sistemas. Se comentan las funciones de los sistemas de gestión de bases de datos y los módulos de software que los componen.

Este tema consta de los siguientes apartados:

1. Modelos de datos
2. Arquitectura de los sistemas de bases de datos
3. Lenguajes de los sistemas de gestión de bases de datos
 - 3.1 Lenguaje de definición de datos
 - 3.2 Lenguaje de manejo de datos
 - 3.3 Lenguajes de cuarta generación
4. Clasificación de los sistemas de gestión de bases de datos
5. Funciones de los sistemas de gestión de bases de datos
6. Componentes de un sistema de gestión de bases de datos

Resumen

Un modelo de datos es un conjunto de conceptos que se utilizan para describir el esquema de una base de datos, las operaciones para manejar los datos y el conjunto de reglas de integridad. Hay tres categorías principales de modelos de datos: modelos conceptuales, modelos lógicos y modelos

físicos. Es importante distinguir entre el esquema (descripción de una base de datos) y la base de datos en sí misma. El esquema no cambia a menudo, en tanto que la base de datos cambia cada vez que se insertan, eliminan o modifican datos.

La arquitectura de los sistemas de bases de datos establecida por el comité ANSI-SPARC utiliza tres niveles de abstracción: externo, conceptual e interno. En el nivel externo, el esquema consta de las distintas visiones que tienen los usuarios de la base de datos. En el nivel conceptual, el esquema es la visión común de la base de datos; especifica el contenido de información de la base de datos independientemente de las consideraciones de almacenamiento. En el nivel interno, el esquema es la visión que el ordenador tiene de la base de datos; especifica cómo se representan los datos, en qué orden se almacenan los registros, qué índices y punteros se han creado y qué esquema de dispersión se ha utilizado, si es el caso.

Todo sistema de gestión de bases de datos que separe los tres niveles deberá tener correspondencias entre los esquemas para transformar las peticiones de los usuarios y los resultados, de un nivel al siguiente. La mayoría de los sistemas de gestión de bases de datos no separan los tres niveles por completo. La independencia de datos hace que cada nivel de la arquitectura sea inmune a los cambios en los niveles de debajo. La independencia de datos lógica se refiere a la inmunidad de los esquemas externos frente a los cambios en el esquema conceptual. La independencia de datos física se refiere a la inmunidad del esquema conceptual frente a los cambios en el esquema interno.

Un lenguaje de base de datos consta de dos partes: un lenguaje de definición de datos y un lenguaje de manejo de datos. El lenguaje de definición de datos se utiliza para especificar el esquema de la base de datos, las vistas de los usuarios y las estructuras de almacenamiento, mientras que el lenguaje de manejo de datos se utiliza para leer y actualizar los datos de la base de datos.

Los sistemas de gestión de bases de datos se pueden clasificar según varios criterios: el modelo lógico que soportan, el número de usuarios, el número de puestos, el coste y la generalidad. La clasificación más importante es la que se basa en el modelo lógico, siendo los principales modelos que se utilizan en el mercado el modelo relacional, el de red, el jerárquico y el orientado a objetos.

Los sistemas de gestión de bases de datos son sistemas informáticos muy complejos formados por una serie de componentes, cada uno de ellos con una función específica. Además de estos módulos, los sistemas de gestión de bases de datos cuentan con una serie de herramientas que ayudan al administrador de la base de datos a manejar el sistema.

Bibliografía

La mayoría de los libros de texto sobre bases de datos analizan los diversos conceptos que se han presentado en este tema. En el libro de apuntes de la asignatura se ha elaborado este tema a partir de los capítulos 1 y 2 del texto de ELMASRI Y NAVATHE (1997). Los mismos contenidos se encuentran en el capítulo 2 del texto de CONNOLLY, BEGG Y STRACHAN (1996) y también en los capítulos 1 y 2 del texto de DATE (1993).

Tema 4. El modelo relacional

En este tema se presenta el modelo relacional, que es el modelo lógico en el que se basan la mayoría de los sistemas de gestión de bases de datos comerciales en uso hoy en día. En primer lugar, se trata la descripción de los principios básicos del modelo relacional: la estructura de datos relacional y las reglas de integridad. A continuación, se presenta un tratamiento detallado del álgebra relacional, que es un conjunto de operaciones para manipular la estructura de datos

relacional y especificar consultas de datos. El álgebra relacional es un lenguaje procedural, mientras que el cálculo relacional, que también se estudia en este tema, es un lenguaje equivalente no procedural.

Este tema consta de los siguientes apartados:

1. Introducción
2. El modelo relacional
3. Estructura de datos relacional
 - 3.1 Relaciones
 - 3.2 Propiedades de las relaciones
 - 3.3 Tipos de relaciones
 - 3.4 Claves
 - 3.5 Esquema de una base de datos relacional
4. Reglas de integridad
 - 4.1 Nulos
 - 4.2 Regla de integridad de entidades
 - 4.3 Regla de integridad referencial
 - 4.4 Reglas de negocio
5. Lenguajes relacionales
 - 5.1 Álgebra relacional
 - 5.2 Cálculo relacional
 - 5.3 Otros lenguajes
6. Vistas

Resumen

La relación es la estructura de datos del modelo relacional. Las relaciones se representan gráficamente como tablas, donde las filas corresponden a las tuplas y las columnas corresponden a los atributos. Los atributos se definen sobre dominios. Las relaciones de una base de datos tienen una serie de propiedades: en la intersección de cada fila con cada columna hay un solo valor (valor atómico), los nombres de los atributos de una relación son todos distintos entre sí, los atributos no están ordenados, las tuplas no están ordenadas y no hay tuplas repetidas. El grado de una relación es el número de atributos y la cardinalidad es el número de tuplas.

Una superclave es un conjunto de atributos que identifica las tuplas de una relación de modo único. Una clave candidata es una superclave irreducible. La clave primaria es la clave candidata que se escoge para identificar las tuplas de una relación. Toda relación tiene siempre clave primaria. Una clave ajena es un atributo o un conjunto de atributos que hacen referencia a la clave primaria de otra relación.

Cuando un atributo no tiene valor para una determinada tupla, bien porque se desconoce, o bien porque no tiene sentido para dicha tupla, se dice que es nulo.

La regla de integridad de entidades es una restricción que dice que ninguno de los atributos que forman la clave primaria puede ser nulo. La regla de integridad referencial dice que los valores de las claves ajenas deben coincidir con alguno de los valores de la clave primaria a la que hacen referencia, o bien ser completamente nulos.

Los lenguajes relacionales de manejo de datos se pueden clasificar como procedurales, no procedurales, orientados a transformaciones, gráficos, de cuarta generación o de quinta generación. El álgebra relacional es un lenguaje procedural formal. Sus operaciones son: restricción, proyección, producto cartesiano, unión, intersección, diferencia, división y varios tipos de concatenación. El cálculo relacional es un lenguaje no procedural formal que utiliza predicados. El álgebra relacional y el cálculo relacional son lenguajes equivalentes.

Una vista es una relación virtual. Las vistas proporcionan seguridad y permiten que el diseñador haga esquemas a medida de cada usuario. Las vistas se generan dinámicamente y no todas son actualizables.

Bibliografía

Este tema ha sido elaborado en el libro de apuntes de la asignatura, fundamentalmente, a partir del capítulo 3 del texto de CONNOLLY, BEGG Y STRACHAN (1996). La sintaxis del álgebra y del cálculo relacional se ha tomado de los capítulos 13 y 14 del texto de DATE (1993), ya que se ha considerado que es sencilla y fácil de recordar.

Tema 5. Planificación, diseño y administración de bases de datos

En este tema se comentan las fases principales del ciclo de vida de un sistema de información y se ve cómo se relaciona esto con el desarrollo de aplicaciones de bases de datos. Después, se describen las tareas que se deben realizar en cada etapa del ciclo de vida de una aplicación de bases de datos para que se produzca un sistema que funcione correctamente. Por último, se presenta todo el personal responsable de la planificación, diseño y administración de una base de datos.

Este tema consta de los siguientes apartados:

1. Introducción
2. Ciclo de vida de los sistemas de información
3. Ciclo de vida de las aplicaciones de bases de datos
4. Diseño de bases de datos
 - 4.1 Diseño conceptual
 - 4.2 Diseño lógico
 - 4.3 Diseño físico
5. Diseño de aplicaciones
 - 5.1 Diseño de transacciones
 - 5.2 Diseño de interfaces de usuario
6. Herramientas CASE
7. Administración de datos y de la base de datos

Resumen

Un sistema de información es el conjunto de recursos que se utilizan para recoger, gestionar, controlar y divulgar la información dentro de una empresa u organización. Desde los años setenta, los sistemas de bases de datos han ido reemplazando a los sistemas de ficheros en los sistemas de información de las empresas, de modo que éstos constan de los siguientes componentes: la base de datos, el sistema de gestión de bases de datos, los programas de aplicación, los equipos informáticos, el personal que utiliza el sistema y el personal que lo desarrolla.

La base de datos es uno de los componentes principales de un sistema de información, por lo que el ciclo de vida de un sistema de información está inherentemente ligado al ciclo de vida de la

base de datos sobre la que se apoya. Las etapas de este ciclo de vida son: planificación de la base de datos, definición del sistema, recolección y análisis de los requisitos, diseño de la base de datos, selección del sistema de gestión de bases de datos, diseño de aplicaciones, elaboración de prototipos, implementación, conversión y carga de datos, prueba y mantenimiento.

En el diseño de una base de datos se debe realizar un modelo de datos que ayude a entender el significado de los datos y que facilite la comunicación en cuanto a los requisitos de información. La primera etapa es el diseño conceptual, en donde se construye un esquema de la información que maneja la empresa, independientemente de todas las consideraciones físicas. Después viene el diseño lógico, en el que el esquema anterior se transforma según el modelo de base de datos que se vaya a utilizar para implementar el sistema. Por último, en la etapa del diseño físico, se produce una descripción de la implementación de la base de datos en memoria secundaria.

El diseño de las aplicaciones, una fase que se debe llevar a cabo en paralelo con el diseño de la base de datos, está compuesta por dos actividades: el diseño de las transacciones y el diseño de las interfaces de usuario de informes y formularios.

Las herramientas CASE permiten que el desarrollo de los sistemas de información se realice de modo eficiente y efectivo.

La administración de datos consiste en la gestión de los datos como recurso, mientras que la administración de la base de datos es la gestión de la base de datos física.

Bibliografía

Siendo éste un tema introductorio, se encuentra reflejado en la mayoría de los textos sobre bases de datos. En el libro de apuntes de la asignatura, este tema se ha elaborado a partir del capítulo 4 del texto de CONNOLLY, BEGGY STRACHAN (1996).

Tema 6. Diseño conceptual. Modelo entidad-relación

En este tema se presenta una metodología para el diseño conceptual de bases de datos que se basa en el modelo de datos más popular en la actualidad, el modelo entidad-relación.

Este tema consta de los siguientes apartados:

1. Introducción
2. Metodología de diseño de bases de datos
3. Modelos de datos
4. El modelo entidad-relación
5. Metodología de diseño conceptual

Resumen

El diseño de bases de datos se descompone en tres etapas: diseño conceptual, diseño lógico y diseño físico. El diseño conceptual es el proceso por el cual se construye un modelo de la información que se utiliza en una empresa u organización, independientemente del sistema de gestión de bases de datos que se vaya a utilizar para implementar el sistema y de los equipos informáticos o cualquier otra consideración física.

Un modelo conceptual es un conjunto de conceptos que permiten describir la realidad mediante representaciones lingüísticas y gráficas. Los modelos conceptuales deben poseer una serie de propiedades: expresividad, simplicidad, minimalidad y formalidad. El modelo conceptual más

utilizado es el modelo entidad-relación, que posee los siguientes conceptos: entidades, relaciones, atributos, dominios de atributos, identificadores y jerarquías de generalización.

En la metodología del diseño conceptual se construye un esquema conceptual local para cada vista de cada usuario o grupo de usuarios. En el diseño lógico se obtiene un esquema lógico local para cada esquema conceptual local. Estos esquemas lógicos se integran después para formar un esquema lógico global que represente todas las vistas de los distintos usuarios de la empresa. Por último, en el diseño físico, se construye la implementación de la base de datos sobre un sistema de gestión de bases de datos determinado. Ya que este diseño debe adaptarse al sistema de gestión de bases de datos, es posible que haya que introducir cambios en el esquema lógico para mejorar las prestaciones a nivel físico.

Cada vista de usuario comprende los datos que un usuario maneja para llevar a cabo una determinada tarea. Normalmente, estas vistas corresponden a las distintas áreas funcionales de la empresa, y se pueden identificar examinando los diagramas de flujo de datos o entrevistando a los usuarios, examinando los procedimientos, informes y formularios, y observando el funcionamiento de la empresa.

Cada esquema conceptual local está formado por entidades, relaciones, atributos, dominios de atributos, identificadores y puede haber también jerarquías de generalización. Además, estos esquemas se completan documentándolos en el diccionario de datos.

Bibliografía

En el libro de apuntes de la asignatura, este tema se ha elaborado a partir de los capítulos 2 y 3 del texto de BATINI, CERI Y NAVATHE (1994), en donde se tratan muy bien los aspectos sobre el modelo entidad-relación. También se ha utilizado el capítulo 5 del texto de CONNOLLY, BEGG Y STRACHAN (1996), en el que se presenta la metodología del diseño conceptual proporcionando todo tipo de “trucos” para identificar cada uno de los componentes de un esquema conceptual.

Tema 7. Diseño lógico de bases de datos

En este tema se describen los pasos para llevar a cabo el diseño lógico. Ya que aquí se trata el diseño de bases de datos relacionales, en esta etapa se obtiene un conjunto de relaciones (tablas) que representen los datos de interés. Este conjunto de relaciones se valida mediante la normalización, técnica que se estudia al final del tema.

Este tema consta de los siguientes apartados:

1. Introducción
2. Metodología de diseño lógico en el modelo relacional
3. Normalización

Resumen

El diseño de bases de datos consta de tres etapas: diseño conceptual, lógico y físico. El diseño lógico es el proceso mediante el que se construye un esquema que representa la información que maneja una empresa, basándose en un modelo lógico determinado, pero independientemente del sistema de gestión de bases de datos concreto que se vaya a utilizar para implementar la base de datos e independientemente de cualquier otra consideración física.

Las dos fases de que consta el diseño lógico son la construcción y validación de los esquemas lógicos locales para cada vista de usuario, y la construcción y validación de un esquema lógico global. Cada una de estas fases consta de una serie de pasos. Uno de los pasos más importantes es la

conversión del esquema conceptual a un esquema lógico adecuado al modelo relacional. Para ello, se deben hacer algunas transformaciones: eliminar las relaciones de muchos a muchos, eliminar las relaciones complejas, eliminar las relaciones recursivas, eliminar las relaciones con atributos, eliminar los atributos multievaluados, reconsiderar las relaciones de uno a uno y eliminar las relaciones redundantes.

Los esquemas lógicos se pueden validar mediante la normalización y frente a las transacciones de los usuarios. La normalización se utiliza para mejorar el esquema, de modo que éste satisfaga ciertas restricciones que eviten la duplicidad de datos. La normalización garantiza que el esquema resultante está más próximo al modelo de la empresa, es consistente, tiene la mínima redundancia y la máxima estabilidad.

Las restricciones de integridad son las restricciones que se imponen para que la base de datos nunca llegue a un estado inconsistente. Hay cinco tipos de restricciones de integridad: datos requeridos, restricciones de dominio, integridad de entidades, integridad referencial y reglas de negocio. Para garantizar la integridad referencial se debe especificar el comportamiento de las claves ajenas: si aceptan nulos y qué hacer cuando se borra la tupla a la que se hace referencia, o cuando se modifica el valor de su clave primaria.

Bibliografía

El diseño de bases de datos relacionales es un tema de consenso, coincidiendo la mayoría de autores en las tres etapas de diseño conceptual, diseño lógico y diseño físico. Sin embargo, los pasos de que consta cada una de estas etapas y la terminología utilizada no es muy uniforme. En el libro de apuntes de la asignatura, este tema se ha elaborado siguiendo los pasos que se especifican en el capítulo 7 del texto de CONNOLLY, BEGG Y STRACHAN (1996), combinado con la terminología del capítulo 12 del texto de BATINI, CERI Y NAVATHE (1994). Para el apartado de normalización se ha utilizado el capítulo 6 del texto de CONNOLLY, BEGG Y STRACHAN (1996).

Prácticas

Parte 1. Consulta de datos en SQL: la sentencia SELECT

En las sesiones de prácticas de esta primera parte se presenta el lenguaje SQL (Structured Query Language), que permite tanto la realización de consultas sobre una base de datos relacional, como la inserción, borrado y modificación de datos de la misma. Además permite crear, definir y mantener los objetos de la base de datos.

Mediante el sistema de gestión de bases de datos Oracle 8, el uso de SQL se realiza tanto de modo interactivo, como embebido en un lenguaje de programación. Para el uso interactivo se utiliza la herramienta de Oracle SQL*Plus.

Las sentencias SQL para crear y cargar la base de datos de prácticas están disponibles en la página web <<http://www4.uji.es/~dba>>. También se pueden encontrar aquí los resultados que se obtienen al ejecutar las sentencias de los ejercicios de SQL propuestos en el libro de prácticas (M. MARQUÉS, J. I. ALIAGA, S. GARCÍA, G. QUINTANA (2001)).

Parte 2. Generador de formularios de Developer 2000

En esta parte se presenta la herramienta Forms incluida en Developer 2000 de Oracle. Esta herramienta permite al usuario desarrollar aplicaciones con interfaces de diseño gráfico, que realizan consultas y transacciones sobre la base de datos.

Se presenta también PL/SQL, un lenguaje procedural proporcionado por Oracle que extiende SQL permitiendo construcciones procedurales y las estructuras básicas de los lenguajes de programación. Los programas de aplicación que utilizan SQL embebido pueden incluir bloques de PL/SQL, lo que permite conseguir mejores prestaciones.

Parte 3. Microsoft Access

Esta última parte es una introducción al SGBD Access 97.

3.- BIBLIOGRAFÍA

Apuntes de Teoría

M. MARQUÉS (2000)

Apuntes de Ficheros y Bases de Datos

Materials. Num. 20

Publicacions de la Universitat Jaume I

Bibliografía de Teoría

C. BATINI, S. CERI, S. B. NAVATHE (1994)

Diseño Conceptual de Bases de Datos. Un Enfoque de Entidades - Interrelaciones

Addison-Wesley / Díaz de Santos

T. CONNOLLY, C. BEGG, A. STRACHAN (1996)

Database Systems. A Practical Approach to Design, Implementation and Management

Addison-Wesley

Segunda edición en 1998

C. J. DATE (1993)

Introducción a los Sistemas de Bases de Datos. Volumen I, Quinta edición

Addison-Wesley Iberoamericana

Sexta edición por Addison-Wesley en 1995 (en inglés)

ELMASRI, S. B. NAVATHE (1997)

Sistemas de Bases de Datos. Conceptos fundamentales. Segunda edición

Addison-Wesley Iberoamericana

Tercera edición por Addison-Wesley en 1999 (en inglés)

M. J. FOLK, B. ZOELLICK (1992)

File Structures. Segunda edición

Addison-Wesley

En el servicio de reprografía estarán disponibles las transparencias que se utilizarán en clase y los boletines de ejercicios.

Prácticas

M. MARQUÉS, J. I. ALIAGA, S. GARCÍA, G. QUINTANA (2001)

SQL y Desarrollo de aplicaciones en Oracle 8

Col·lecció <<Treballs d'Informàtica i Tecnologia>> Num. 9

Universitat Jaume I.

ORACLE

Manuales: <<http://www4.uji.es/~oracle/manuales>>

4.- EVALUACIÓN

Se realiza un examen final que engloba teoría y prácticas, y que combina preguntas de test con preguntas de respuesta corta (cuestiones). Los exámenes de cursos anteriores están disponibles a través de la web de la asignatura.