

Introducción al diseño de bases de datos

Dolors Costal Costa

P06/M2109/02150

Introducción al diseño de bases de datos

Dolors Costal Costa

P06/M2109/02150

Índice

Introducción	5
Objetivos	5
1. Introducción al diseño de bases de datos	7
1.1. Etapas del diseño de bases de datos.....	7
2. Diseño conceptual: el modelo ER	10
2.1. Construcciones básicas.....	11
2.1.1. Entidades, atributos e interrelaciones.....	11
2.1.2. Grado de las interrelaciones.....	13
2.1.3. Interrelaciones binarias.....	16
2.1.4. Ejemplo: base de datos de casas de colonias.....	18
2.1.5. Interrelaciones <i>n</i> -arias.....	21
2.1.6. Interrelaciones recursivas.....	23
2.1.7. Entidades débiles.....	25
2.2. Extensiones del modelo ER.....	26
2.2.1. Generalización/especialización.....	26
2.2.2. Entidades asociativas.....	28
2.3. Ejemplo: base de datos del personal de una entidad bancaria.....	30
3. Diseño lógico: la transformación del modelo ER al modelo relacional	35
3.1. Introducción a la transformación de entidades e interrelaciones.....	35
3.2. Transformación de entidades.....	35
3.3. Transformación de interrelaciones binarias.....	36
3.3.1. Conectividad 1:1.....	36
3.3.2. Conectividad 1:N.....	37
3.3.3. Conectividad M:N.....	38
3.3.4. Influencia de la dependencia de existencia en la transformación de las interrelaciones binarias.....	39
3.4. Transformación de interrelaciones ternarias.....	40
3.4.1. Conectividad M:N:P.....	40
3.4.2. Conectividad M:N:1.....	41
3.4.3. Conectividad N:1:1.....	42
3.4.4. Conectividad 1:1:1.....	43
3.5. Transformación de interrelaciones <i>n</i> -arias.....	44
3.6. Transformación de interrelaciones recursivas.....	44
3.7. Transformación de entidades débiles.....	46
3.8. Transformación de la generalización/especialización.....	47

Índice

Introducción.....	5
Objetivos.....	5
1. Introducción al diseño de bases de datos	7
1.1. Etapas del diseño de bases de datos.....	7
2. Diseño conceptual: el modelo ER.....	10
2.1. Construcciones básicas.....	11
2.1.1. Entidades, atributos e interrelaciones.....	11
2.1.2. Grado de las interrelaciones.....	13
2.1.3. Interrelaciones binarias	16
2.1.4. Ejemplo: base de datos de casas de colonias.....	18
2.1.5. Interrelaciones n-arias	21
2.1.6. Interrelaciones recursivas	23
2.1.7. Entidades débiles.....	25
2.2. Extensiones del modelo ER	26
2.2.1. Generalización/especialización	26
2.2.2. Entidades asociativas.....	28
2.3. Ejemplo: base de datos del personal de una entidad bancaria	30
3. Diseño lógico: la transformación	
del modelo ER al modelo relacional	35
3.1. Introducción a la transformación de entidades e interrelaciones.....	35
3.2. Transformación de entidades	35
3.3. Transformación de interrelaciones binarias.....	36
3.3.1. Conectividad 1:1	36
3.3.2. Conectividad 1:N	37
3.3.3. Conectividad M:N	38
3.3.4. Influencia de la dependencia de existencia en la transformación de las interrelaciones binarias	39
3.4. Transformación de interrelaciones ternarias.....	40
3.4.1. Conectividad M:N:P	40
3.4.2. Conectividad M:N:1	41
3.4.3. Conectividad N:1:1	42
3.4.4. Conectividad 1:1:1	43
3.5. Transformación de interrelaciones n-arias	44
3.6. Transformación de interrelaciones recursivas.....	44
3.7. Transformación de entidades débiles.....	46
3.8. Transformación de la generalización/especialización.....	47

3.9. Transformación de entidades asociativas.....	48
3.10. Resumen de la transformación del modelo ER al modelo relacional	49
3.11. Ejemplo: base de datos del personal de una entidad bancaria	49
Resumen	51
Ejercicios de autoevaluación	53
Solucionario	55
Glosario	59
Bibliografía	60

3.9. Transformación de entidades asociativas.....	48	3.10. Resumen de la transformación del modelo ER	
al modelo relacional	49	3.11. Ejemplo: base de datos del personal de una entidad bancaria	49
Resumen.....	51		
Ejercicios de autoevaluación.....	53		
Solucionario.....	55		
Glosario.....	59		
Bibliografía.....	60		

Introducción al diseño de bases de datos

Introducción

En otras unidades didácticas se estudian las bases de datos relacionales y un lenguaje relacional, SQL, que nos proporciona mecanismos para crear, actualizar y consultar estas bases de datos.

Es necesario complementar estos conocimientos con un aspecto que es fundamental para poder utilizar adecuadamente la tecnología de las bases de datos relacionales: el diseño. Éste será el objeto de estudio de esta unidad, que tratará el **diseño de bases de datos** para el caso específico del modelo relacional. 

Concretamente, en esta unidad explicaremos en qué consiste el diseño de una base de datos, analizaremos las etapas en las que se puede descomponer y describiremos con detalle las etapas del diseño conceptual y lógico de una base de datos relacional.

Objetivos

En los materiales didácticos de esta unidad encontraréis las herramientas indispensables para alcanzar los siguientes objetivos:

1. Conocer las etapas que integran el proceso del diseño de una base de datos.
2. Conocer las estructuras del modelo ER.
3. Saber hacer el diseño conceptual de los datos de un sistema de información mediante el modelo ER.
4. Saber hacer el diseño lógico de una base de datos relacional partiendo de un diseño conceptual expresado con el modelo ER.

Introducción al diseño de bases de datos

Introducción

En otras unidades didácticas se estudian las bases de datos relacionales y un lenguaje relacional, SQL, que nos proporciona mecanismos para crear, actualizar y consultar estas bases de datos.

Es necesario complementar estos conocimientos con un aspecto que es fundamental para poder utilizar adecuadamente la tecnología de las bases de datos relacionales: el diseño. Éste será el objeto de estudio de esta unidad, que tratará el diseño de bases de datos para el caso específico del modelo relacional.

Concretamente, en esta unidad explicaremos en qué consiste el diseño de una base de datos, analizaremos las etapas en las que se puede descomponer y describiremos con detalle las etapas del diseño conceptual y lógico de una base de datos relacional.

Objetivos

En los materiales didácticos de esta unidad encontraréis las herramientas indispensables para alcanzar los siguientes objetivos:

1. Conocer las etapas que integran el proceso del diseño de una base de datos.
2. Conocer las estructuras del modelo ER.
3. Saber hacer el diseño conceptual de los datos de un sistema de información mediante el modelo ER.
4. Saber hacer el diseño lógico de una base de datos relacional partiendo de un diseño conceptual expresado con el modelo ER.

1. Introducción al diseño de bases de datos

En otras unidades hemos aprendido cómo es una base de datos relacional y hemos estudiado un lenguaje, el SQL, que nos proporciona mecanismos para crear estas bases de datos, así como para actualizarlas y consultarlas.

Recordad que las bases de datos relacionales y los lenguajes SQL se han estudiado en las unidades "El modelo relacional y el álgebra relacional" y "El lenguaje SQL", respectivamente.

Sin embargo, todavía debemos resolver algunas cuestiones fundamentales para poder emplear la tecnología de las bases de datos relacionales; por ejemplo, cómo se puede decidir qué relaciones debe tener una base de datos determinada o qué atributos deben presentar las relaciones, qué claves primarias y qué claves foráneas se deben declarar, etc. La tarea de tomar este conjunto de decisiones recibe el nombre de *diseñar la base de datos*.

Una base de datos sirve para almacenar la información que se utiliza en un sistema de información determinado. Las necesidades y los requisitos de los futuros usuarios del sistema de información se deben tener en cuenta para poder tomar adecuadamente las decisiones anteriores.

En resumen, el **diseño de una base de datos** consiste en definir la estructura de los datos que debe tener la base de datos de un sistema de información determinado. En el caso relacional, esta estructura será un conjunto de esquemas de relación con sus atributos, dominios de atributos, claves primarias, claves foráneas, etc.

Si recordáis los tres mundos presentados –el real, el conceptual y el de las representaciones–, observaréis que el diseño de una base de datos consiste en la obtención de una representación informática concreta a partir del estudio del mundo real de interés. 

1.1. Etapas del diseño de bases de datos

El diseño de una base de datos no es un proceso sencillo. Habitualmente, la complejidad de la información y la cantidad de requisitos de los sistemas de información hacen que sea complicado. Por este motivo, cuando se diseñan bases de datos es interesante aplicar la vieja estrategia de dividir para vencer.

Por lo tanto, conviene descomponer el proceso del diseño en varias etapas; en cada una se obtiene un resultado intermedio que sirve de punto de partida de la etapa siguiente, y en la última etapa se obtiene el resultado deseado. De este modo no hace falta resolver de golpe toda la problemática que plantea el diseño, sino que en cada etapa se afronta un solo tipo de subproblema. Así se divide el problema y, al mismo tiempo, se simplifica el proceso.

1. Introducción al diseño de bases de datos

En otras unidades hemos aprendido cómo es una base de datos relacional y hemos estudiado un lenguaje, el SQL, que nos proporciona mecanismos para crear estas bases de datos, así como para actualizarlas y consultarlas.

Sin embargo, todavía debemos resolver algunas cuestiones fundamentales para poder emplear la tecnología de las bases de datos relacionales; por ejemplo, cómo se puede decidir qué relaciones debe tener una base de datos determinada o qué atributos deben presentar las relaciones, qué claves primarias y qué claves foráneas se deben declarar, etc. La tarea de tomar este conjunto de decisiones recibe el nombre de diseñar la base de datos.

Una base de datos sirve para almacenar la información que se utiliza en un sistema de información determinado. Las necesidades y los requisitos de los usuarios del sistema de información se deben tener en cuenta para poder tomar adecuadamente las decisiones anteriores.

Si recordáis los tres mundos presentados –el real, el conceptual y el de las representaciones–, observaréis que el diseño de una base de datos consiste en la obtención de una representación informática concreta a partir del estudio del mundo real de interés.

1.1. Etapas del diseño de bases de datos

El diseño de una base de datos no es un proceso sencillo. Habitualmente, la complejidad de la información y la cantidad de requisitos de los sistemas de información hacen que sea complicado. Por este motivo, cuando se diseñan bases de datos es interesante aplicar la vieja estrategia de dividir para vencer.

Por lo tanto, conviene descomponer el proceso del diseño en varias etapas; en cada una se obtiene un resultado intermedio que sirve de punto de partida de la etapa siguiente, y en la última etapa se obtiene el resultado deseado. De este modo no hace falta resolver de golpe toda la problemática que plantea el diseño, sino que en cada etapa se afronta un solo tipo de subproblema. Así se divide el problema y, al mismo tiempo, se simplifica el proceso.

Recordad que las bases de datos relacionales y los lenguajes SQL se han estudiado en las unidades “El modelo relacional y el álgebra relacional” y “El lenguaje SQL”, respectivamente.

En resumen, el diseño de una base de datos consiste en definir la estructura de los datos que debe tener la base de datos de un sistema de información determinado. En el caso relacional, esta estructura será un conjunto de esquemas de relación con sus atributos, dominios de atributos, claves primarias, claves foráneas, etc.

Descompondremos el diseño de bases de datos en tres etapas: 

1) **Etapa del diseño conceptual:** en esta etapa se obtiene una estructura de la información de la futura BD independiente de la tecnología que hay que emplear. No se tiene en cuenta todavía qué tipo de base de datos se utilizará –relacional, orientada a objetos, jerárquica, etc.–; en consecuencia, tampoco se tiene en cuenta con qué SGBD ni con qué lenguaje concreto se implementará la base de datos. Así pues, la etapa del diseño conceptual nos permite concentrarnos únicamente en la problemática de la estructuración de la información, sin tener que preocuparnos al mismo tiempo de resolver cuestiones tecnológicas.

El resultado de la etapa del diseño conceptual se expresa mediante algún modelo de datos de alto nivel. Uno de los más empleados es el **modelo entidad-interrelación** (*entity-relationship*), que abreviaremos con la sigla ER.

2) **Etapa del diseño lógico:** en esta etapa se parte del resultado del diseño conceptual, que se transforma de forma que se adapte a la tecnología que se debe emplear. Más concretamente, es preciso que se ajuste al modelo del SGBD con el que se desea implementar la base de datos. Por ejemplo, si se trata de un SGBD relacional, esta etapa obtendrá un conjunto de relaciones con sus atributos, claves primarias y claves foráneas.

Esta etapa parte del hecho de que ya se ha resuelto la problemática de la estructuración de la información en un ámbito conceptual, y permite concentrarnos en las cuestiones tecnológicas relacionadas con el modelo de base de datos.

Más adelante explicaremos cómo se hace el diseño lógico de una base de datos relacional, tomando como punto de partida un diseño conceptual expresado con el modelo ER; es decir, veremos cómo se puede transformar un modelo ER en un modelo relacional.

3) **Etapa del diseño físico:** en esta etapa se transforma la estructura obtenida en la etapa del diseño lógico, con el objetivo de conseguir una mayor eficiencia; además, se completa con aspectos de implementación física que dependerán del SGBD.

Por ejemplo, si se trata de una base de datos relacional, la transformación de la estructura puede consistir en lo siguiente: tener almacenada alguna relación que sea la combinación de varias relaciones que se han obtenido en la etapa del diseño lógico, partir una relación en varias, añadir algún atributo calculable a una relación, etc. Los aspectos de implementación física que hay que completar consisten normalmente en la elección de estructuras físicas de implementación de las relaciones, la selección del tamaño de las memorias intermedias (*buffers*) o de las páginas, etc.

El resultado del diseño conceptual

Si retomamos la idea de los tres mundos, podemos afirmar que la etapa del diseño conceptual obtiene un resultado que se sitúa en el mundo de las concepciones, y no en el mundo de las representaciones.

La forma de elaborar un diseño conceptual expresado con el modelo ER se explica en el apartado 2 de esta unidad. 

El resultado del diseño lógico

El resultado del diseño lógico se sitúa ya en el mundo de las representaciones.

El diseño lógico de una base de datos relacional se explica en el apartado 3 de esta unidad didáctica. 

El resultado del diseño físico

El resultado de la etapa del diseño físico se sitúa en el mundo de las representaciones, al igual que el resultado de la etapa del diseño lógico. La diferencia con respecto a la etapa anterior es que ahora se tienen en cuenta aspectos de carácter más físico del mundo de las representaciones.

Descompondremos el diseño de bases de datos en tres etapas:

1) Etapa del diseño conceptual: en esta etapa se obtiene una estructura de la información de la futura BD independiente de la tecnología que hay que emplear. No se tiene en cuenta todavía qué tipo de base de datos se utilizará –relacional, orientada a objetos, jerárquica, etc.–; en consecuencia, tampoco se tiene en cuenta con qué SGBD ni con qué lenguaje concreto se implementará la base de datos. Así pues, la etapa del diseño conceptual nos permite concentrarnos únicamente en la problemática de la estructuración de la información, sin tener que preocuparnos al mismo tiempo de resolver cuestiones tecnológicas.

El resultado de la etapa del diseño conceptual se expresa mediante algún modelo de datos de alto nivel. Uno de los más empleados es el modelo entidad-interrelación (entity-relationship), que abreviaremos con la sigla ER.

2) Etapa del diseño lógico: en esta etapa se parte del resultado del diseño conceptual, que se transforma de forma que se adapte a la tecnología que se debe emplear. Más concretamente, es preciso que se ajuste al modelo del SGBD con el que se desea implementar la base de datos. Por ejemplo, si se trata de un SGBD relacional, esta etapa obtendrá un conjunto de relaciones con sus atributos, claves primarias y claves foráneas.

Esta etapa parte del hecho de que ya se ha resuelto la problemática de la estructuración de la información en un ámbito conceptual, y permite concentrarnos en las cuestiones tecnológicas relacionadas con el modelo de base de datos.

Más adelante explicaremos cómo se hace el diseño lógico de una base de datos relacional, tomando como punto de partida un diseño conceptual expresado con el modelo ER; es decir, veremos cómo se puede transformar un modelo ER en un modelo relacional.

3) Etapa del diseño físico: en esta etapa se transforma la estructura obtenida en la etapa del diseño lógico, con el objetivo de conseguir una mayor eficiencia; además, se completa con aspectos de implementación física que dependen del SGBD.

Por ejemplo, si se trata de una base de datos relacional, la transformación de la estructura puede consistir en lo siguiente: tener almacenada alguna relación que sea la combinación de varias relaciones que se han obtenido en la etapa del diseño lógico, partir una relación en varias, añadir algún atributo calculable a una relación, etc. Los aspectos de implementación física que hay que completar consisten normalmente en la elección de estructuras físicas de implementación de las relaciones, la selección del tamaño de las memorias intermedias (buffers) o de las páginas, etc.

El resultado del diseño conceptual

Si retomamos la idea de los tres mundos, podemos afirmar que la etapa del diseño conceptual obtiene un resultado que se sitúa en el mundo de las concepciones, y no en el mundo de las representaciones.

La forma de elaborar un diseño conceptual expresado con el modelo ER se explica en el apartado 2 de esta unidad.

El resultado del diseño lógico

El resultado del diseño lógico se sitúa ya en el mundo de las representaciones.

El diseño lógico de una base de datos relacional se explica en el apartado 3 de esta unidad didáctica.

El resultado del diseño físico

El resultado de la etapa del diseño físico se sitúa en el mundo de las representaciones, al igual que el resultado de la etapa del diseño lógico. La diferencia con respecto a la etapa anterior es que ahora se tienen en cuenta aspectos de carácter más físico del mundo de las representaciones.

En la etapa del diseño físico –con el objetivo de conseguir un buen rendimiento de la base de datos–, se deben tener en cuenta las características de los procesos que consultan y actualizan la base de datos, como por ejemplo los caminos de acceso que utilizan y las frecuencias de ejecución. También es necesario considerar los volúmenes que se espera tener de los diferentes datos que se quieren almacenar.

En la etapa del diseño físico –con el objetivo de conseguir un buen rendimiento de la base de datos–, se deben tener en cuenta las características de los procesos que consultan y actualizan la base de datos, como por ejemplo los caminos de acceso que utilizan y las frecuencias de ejecución. También es necesario considerar los volúmenes que se espera tener de los diferentes datos que se quieren almacenar.

2. Diseño conceptual: el modelo ER

En este apartado trataremos el diseño conceptual de una base de datos mediante el modelo ER. Lo que explicaremos es aplicable al diseño de cualquier tipo de bases de datos –relacional, jerárquica, etc.–, porque, como ya hemos dicho, en la etapa del diseño conceptual todavía no se tiene en cuenta la tecnología concreta que se utilizará para implementar la base de datos. 

El **modelo ER** es uno de los enfoques de modelización de datos que más se utiliza actualmente por su simplicidad y legibilidad. Su legibilidad se ve favorecida porque proporciona una notación diagramática muy comprensiva. Es una herramienta útil tanto para ayudar al diseñador a reflejar en un modelo conceptual los requisitos del mundo real de interés como para comunicarse con el usuario final sobre el modelo conceptual obtenido y, de este modo, poder verificar si satisface sus requisitos.

El modelo ER resulta fácil de aprender y de utilizar en la mayoría de las aplicaciones. Además, existen herramientas informáticas de ayuda al diseño (herramientas CASE*) que utilizan alguna variante del modelo ER para hacer el diseño de los datos.

El nombre completo del modelo ER es *entity-relationship*, y proviene del hecho de que los principales elementos que incluye son las entidades y las interrelaciones (*entities y relationships*). Traduciremos este nombre por ‘entidad-interrelación’.

El origen del modelo ER se encuentra en trabajos efectuados por Peter Chen en 1976. Posteriormente, muchos otros autores han descrito variantes y/o extensiones de este modelo. Así pues, en la literatura se encuentran muchas formas diferentes del modelo ER que pueden variar simplemente en la notación diagramática o en algunos de los conceptos en que se basan para modelizar los datos.

Cuando se quiere utilizar el modelo ER para comunicarse con el usuario, es recomendable emplear una variante del modelo que incluya sólo sus elementos más simples –entidades, atributos e interrelaciones– y, tal vez, algunas construcciones adicionales, como por ejemplo entidades débiles y dependencias de existencia. Éstos eran los elementos incluidos en el modelo original propuesto por Chen. En cambio, para llevar a cabo la tarea de modelizar propiamente dicha, suele ser útil usar un modelo ER más completo que incluya construcciones más avanzadas que extienden el modelo original.

Según la noción de *modelo de datos* que hemos utilizado en los otros módulos, un modelo de datos tiene en cuenta tres aspectos de los datos: la estructura, la manipulación y la integridad. Sin embargo, el modelo ER habitualmente se

* La sigla CASE corresponde al término inglés *Computer Aided Software Engineering*.

El modelo entidad-interrelación

Algunos autores denominan *entidad-relación* al modelo ER, pero en nuestro caso hemos preferido traducir *relationship* por ‘interrelación’ y no por ‘relación’, con el objetivo de evitar confusiones entre este concepto y el de *relación* que se utiliza en el modelo relacional.

Recordad el modelo relacional, que se ha estudiado en la unidad “El modelo relacional y el álgebra relacional”.

2. Diseño conceptual: el modelo ER

En este apartado trataremos el diseño conceptual de una base de datos mediante el modelo ER. Lo que explicaremos es aplicable al diseño de cualquier tipo de bases de datos –relacional, jerárquica, etc.–, porque, como ya hemos dicho, en la etapa del diseño conceptual todavía no se tiene en cuenta la tecnología concreta que se utilizará para implementar la base de datos.

El modelo ER es uno de los enfoques de modelización de datos que más se utiliza actualmente por su simplicidad y legibilidad. Su legibilidad se ve favorecida porque proporciona una notación diagramática muy comprensiva. Es una herramienta útil tanto para ayudar al diseñador a reflejar en un modelo conceptual los requisitos del mundo real de interés como para comunicarse con el usuario final sobre el modelo conceptual obtenido y, de este modo, poder verificar si satisface sus requisitos.

El modelo ER resulta fácil de aprender y de utilizar en la mayoría de las aplicaciones. Además, existen herramientas informáticas de ayuda al diseño (herramientas CASE*) que utilizan alguna variante del modelo ER para hacer el diseño de los datos.

El nombre completo del modelo ER es entity-relationship, y proviene del hecho de que los principales elementos que incluye son las entidades y las interrelaciones (entities y relationships). Traduciremos este nombre por ‘entidad-interrelación’.

El origen del modelo ER se encuentra en trabajos efectuados por Peter Chen en 1976. Posteriormente, muchos otros autores han descrito variantes y/o extensiones de este modelo. Así pues, en la literatura se encuentran muchas formas diferentes del modelo ER que pueden variar simplemente en la notación diagramática o en algunos de los conceptos en que se basan para modelizar los datos.

Cuando se quiere utilizar el modelo ER para comunicarse con el usuario, es recomendable emplear una variante del modelo que incluya sólo sus elementos más simples –entidades, atributos e interrelaciones– y, tal vez, algunas construcciones adicionales, como por ejemplo entidades débiles y dependencias de existencia. Éstos eran los elementos incluidos en el modelo original propuesto por Chen. En cambio, para llevar a cabo la tarea de modelizar propiamente dicha, suele ser útil usar un modelo ER más completo que incluya construcciones más avanzadas que extienden el modelo original.

Según la noción de modelo de datos que hemos utilizado en los otros módulos, un modelo de datos tiene en cuenta tres aspectos de los datos: la estructura, la

manipulación y la integridad. Sin embargo, el modelo ER habitualmente se

*** La sigla CASE corresponde al término inglés Computer Aided Software Engineering.**

El modelo entidad- interrelación

Algunos autores denominan entidad-relación al modelo ER, pero en nuestro caso hemos preferido traducir relationship por 'interrelación' y no por 'relación', con el objetivo de evitar confusiones entre este concepto y el de relación que se utiliza en el modelo relacional.

Recordad el modelo relacional, que se ha estudiado en la unidad "El modelo relacional y el álgebra relacional".