

TEMA 2: MODELO **ENTIDAD-RELACIÓN**

David Bataller Signes

ÍNDICE

1. Conceptos.

1.1. Entidades y atributos.

1.2. Interrelaciones

1.3 Entidades débiles

2. Diagramas Entidad-Relación (ER).

2.1. Diseño de bases de datos.

2.1.1 Fases del diseño de BD.

2.1.2. Diseño conceptual.

2.1.3. Captura y abstracción de los requerimientos de datos.

2.1.4. Identificación de entidades.

2.1.5. Designación de interrelaciones.

2.1.6. Establecimiento de claves.

2.1.7. Establecimiento de cardinales.

2.1.8. Restricciones de participación y límites de cardinalidad.

2.1.9. Elaboración de un esquema conceptual.

2.2. Extensiones del modelo ER.

2.2.1. Especialización y generalización.

2.2.2. Agregaciones de entidades.

3. Decisiones de diseño.

3.1. Alternativas de diseño.

3.2. Trampas de diseño.

4. Ejemplos de diseño.

Un modelo de datos consiste en un conjunto de herramientas conceptuales para describir los datos, sus interrelaciones, su significado, y las limitaciones necesarias con el fin de garantizar su coherencia.

En esta unidad estudiaremos el modelo de datos más ampliamente utilizado, el modelo Entidad-Relación (modelo ER). El modelo ER es un modelo de datos de alto nivel. Se basa en una percepción del mundo real que se traduce en una colección de objetos llamados entidades y de relaciones entre estas.

Su éxito se debe, probablemente, a que su notación se basa en una serie de diagramas muy sencillos y comprensibles. Por este motivo, actualmente, la mayoría de las herramientas de diseño de bases de datos (BD) utilizan los conceptos del modelo ER.

A lo largo de toda la unidad, se recomienda analizar con detenimiento los ejemplos que se exponen, ya que permitirán comprender mejor los conceptos teóricos que ilustran.

1. Conceptos del modelo Entidad-Relación. Entidades. Relaciones

Las estructuras básicas del modelo Entidad-Relación (modelo ER) se corresponden, fundamentalmente, con los conceptos propuestos en la formulación original de este modelo que hizo el Dr. Peter Pin Shan Chen en 1976.

La notación de estas construcciones es fundamentalmente diagramática, aunque en algunos casos se puede añadir alguna especificación textual. Estos diagramas son generalmente conocidos como diagramas ER o diagramas Chen.

Los diagramas ER son muy eficaces a la hora de modelizar la realidad para obtener un esquema conceptual comprensible. Debido a ello, muchas de las herramientas de ingeniería del software asistida por ordenador, que también ayudan en el diseño de BD, utilizan los conceptos del modelo ER en sus diagramas.

La utilización de los elementos más simples del modelo ER, entidades, atributos e interrelaciones, y quizás de alguna otra construcción adicional,

como las entidades débiles, pueden ser de gran utilidad en la comunicación entre los diseñadores de BD y los usuarios.

1.1 Entidades y atributos

Una entidad es algo que existe en el mundo real, distinguible del resto de cosas, y de la que nos interesan algunas propiedades.

Así pues, con el término entidad se puede hacer referencia a un objeto específico del mundo real, pero también a un conjunto de objetos similares, de los que nos interesan las mismas características. Por lo tanto, tenemos que distinguir:

- Entidades-instancia, como objetos concretos del mundo real (por ejemplo, el alumno Manuel es una entidad-instancia).
- Entidades-tipo, como conjuntos de entidades-instancia (por ejemplo, la entidad tipo alumno).

Llamamos atributos las características que nos interesan de las entidades.

Habitualmente, sólo nos interesará modelizar una parte de los atributos de una entidad, ya que podrá haber datos que sólo serán de utilidad en ámbitos muy específicos.

Los atributos de cada entidad-instancia adoptan valores concretos. Estos valores han de ser válidos.

Para que un valor de un atributo sea válido, debe pertenecer al conjunto de valores aceptables para el atributo en cuestión. Este conjunto de valores válidos se denomina dominio.

Los atributos de una entidad-instancia pueden no tener ningún valor para alguno atributo concreto. En estos casos, también se dice que el atributo tiene valor nulo.

Se pueden considerar dos tipos diferenciados de atributos: los atributos simples y los compuestos.

Un atributo simple no se puede dividir en partes más pequeñas sin que ello comporte la pérdida de su significado.

Un atributo compuesto es el que está subdividido en partes más pequeñas (que también tienen la consideración de atributos), las cuales tienen un significado propio.

Puede resultar interesante utilizar atributos compuestos si nos consta que los usuarios se referirán, en ocasiones, en el atributo globalmente considerado, ya veces a sus componentes por separado.

Por otra parte, los atributos compuestos agrupan los atributos relacionados, estructurándolos jerárquicamente, por lo que normalmente contribuyen a la comprensibilidad de los modelos.

Otra forma de caracterizar los atributos es en función de si son atributos monovaluados o multivaluados.

Un atributo monovaluado es lo que sólo puede almacenar, como máximo, un solo valor para cada entidad instancia concreta, en un momento determinado.

Un atributo multivaluado puede almacenar, para cada entidad instancia concreta, diferentes valores al mismo tiempo.

Si es necesario, se pueden especificar a continuación del nombre del atributo, entre paréntesis y separados por comas, el límite máximo y el mínimo de valores que se han de almacenar, esto es la cardinalidad de los atributos. Y se pueden presentar las siguientes opciones:

- NombreAtributo (1, 1): atributo univaluado obligatorio (predeterminado, si no especifica nada).
- NombreAtributo (0, 1): atributo univaluado opcional (admite valores nulos).
- NombreAtributo (1, n): atributo multivaluado obligatorio (no admite valores nulos).
- NombreAtributo (0, n): atributo multivaluado opcional (admite valores nulos).

Se dice que un atributo es derivado cuando su valor se puede calcular a partir otros atributos o bien de otras entidades interrelacionadas.

Cuando un atributo sirve para calcular el valor de un atributo derivado, se le considera atributo base de este.

Los atributos derivados constituyen una redundancia, es decir, una repetición normalmente innecesaria de datos. Por este motivo, los datos de los atributos derivados incluidos en los diagramas ER no suelen almacenarse, sino que se calculan cuando son necesarios.

Una entidad instancia concreta debe poder distinguir del resto de objetos del mundo real. Por tanto, cualquier modelización ER debe indicar, para toda entidad tipo, un atributo o un conjunto de atributos que la permita identificar unívocamente.

El atributo o conjunto de atributos que identifican unívocamente las entidades instancia denominan clave primaria de la entidad.

Notación

El modelo ER nos permite representar entidades y atributos mediante una sencilla notación diagramática.

En esta representación respetaremos las características siguientes:

- Como regla general, no usaremos acentos ni caracteres especiales, sólo letras y cifras.
- Representaremos las entidades tipo escribiendo su nombre en mayúsculas y en singular, dentro de un rectángulo.
- Representaremos cada atributo escribiendo su nombre con la primera letra en mayúscula y el resto en minúsculas, dentro de una elipse unida con un guion con el rectángulo que representa la entidad tipo de la que forman parte:
 - Si un atributo tiene un nombre compuesto, cada nombre comenzará con mayúscula para hacerlo más diferenciable. Por ejemplo, TelefonoFijo, TelefonoMovil.
 - Si el nombre de un atributo corresponde a unas siglas, debe ir íntegramente en mayúsculas, como DNI (documento nacional de identidad).

- Las elipses de los atributos en que se puede descomponer un atributo deben ir unidas con un guion con la elipse del atributo compuesto.
- La elipse de un atributo multivaluado estará formada por un trazo doble.
- Los límites de un atributo multivaluado, en caso de existir, se especificarán a continuación del nombre del atributo, entre paréntesis y separados por una coma.
- La elipse de un atributo derivado estará formada por un trazo punteado.
- Los atributos que forman parte de una clave primaria deben ir subrayados.

Si hemos de establecer cualquier otra característica de los datos que no tenga predefinida una notación diagramática concreta, tendremos que añadir al diagrama las especificaciones textuales necesarias.

1.2 Interrelaciones

Una interrelación consiste en una asociación entre dos o más entidades.

Con el término interrelación podemos hacer referencia tanto a una asociación concreta entre diferentes entidades instancia, así como a una asociación de carácter más genérico, entendida como un conjunto de asociaciones de la misma tipología, entre diferentes entidades tipo.

A veces, nos puede interesar reflejar algunas características de determinadas interrelaciones. La manera de hacerlo es añadir los atributos necesarios, como haríamos si trabajáramos con entidades. Estos atributos son los atributos de la interrelación.

Las propiedades de los atributos de las interrelaciones son idénticas a las descritas previamente en relación con los atributos de las entidades.

El grado de una interrelación depende del número de entidades que esta asocia.

Las interrelaciones de grado dos también se llaman binarias. Y las de grado superior a dos se denominan genéricamente n-arias.

La conectividad (también llamada cardinalidad) de una interrelación indica el tipo de correspondencia que hay entre las ocurrencias de las entidades que ella misma permite asociar.

Tratándose de interrelaciones binarias, la cardinalidad expresa el número máximo de instancias de una de las entidades con las que una instancia de la otra entidad puede estar asociada según la interrelación en cuestión.

Las interrelaciones binarias pueden ofrecer tres tipos de conectividad:

- Uno a uno (1: 1)
- Un a unos cuantos (1: N)
- Unos cuantos a unos cuantos (N: M)

Un 1 junto a una entidad indica que, como máximo, sólo una de sus instancias (la cual podrá variar en cada caso) tendrá la posibilidad de estar asociada con cada una de las instancias de la otra entidad.

Una N en un lado de la interrelación también se representa frecuentemente con un asterisco (*)

Si más de un extremo de la interrelación tiene una N, por razón de elegancia se representa con consonantes sucesivas, comenzando por M: M, N, P, Q, etc.

La cardinalidad 1 también se puede representar convirtiendo la línea que une la interrelación con la entidad en una flecha que apunte hacia la entidad.

En cambio, una N (o una M) junto a una entidad indica que será una pluralidad de sus instancias (las cuales también podrán variar en cada caso) la que tendrá la posibilidad de estar asociada con cada una de las instancias de la otra entidad.

La cardinalidad N (o M) también se puede representar con una flecha de doble punta que vaya de la interrelación hacia la entidad.

Es muy importante darse cuenta de que, independientemente del tipo de conectividad, una interrelación sólo permite asociar una sola vez unas entidades instancia determinadas entre ellas.

A veces, una entidad instancia sólo tiene sentido si existe al menos otra entidad instancia asociada con ella mediante una interrelación binaria determinada. En estos casos, se dice que la última entidad es una entidad obligatoria para la interrelación. De lo contrario, se dice que se trata de una entidad opcional para interrelación.

Un círculo en la línea de conexión entre una entidad y una interrelación indica que la entidad es opcional en la interrelación. La obligatoriedad de una entidad en una interrelación indica con un guion perpendicular a la línea que une la entidad con la interrelación. Si no se consigna ni un círculo ni una línea perpendicular, se considera que la dependencia de existencia es desconocida.

Tendremos en cuenta esta característica sólo por lo que haga a las interrelaciones binarias, pero no a las n-arias.

La cardinalidad de las interrelaciones n-arias expresa el número máximo de instancias de una de las entidades con las que una combinación concreta de instancias de las otras entidades puede estar asociada según la interrelación en cuestión.

Las interrelaciones ternarias pueden ofrecer cuatro tipos de conectividad:

- 1: 1: 1
- 1: 1: N
- 1: M: N
- M: N: P

Donde 1 indica que como máximo sólo una de sus instancias (la cual podrá variar en cada caso) tendrá la posibilidad de estar asociada con cada combinación concreta de instancias de las otras entidades. Y en que N, M o P indica que varias instancias pueden estar relacionadas con cada combinación de instancias de las otras entidades.

En general, las interrelaciones n-arias pueden ofrecer $n + 1$ tipo de conectividad.

A veces, puede resultar útil establecer límites mínimos y máximos a las cardinalidades de las interrelaciones. Para ello, basta con añadir una etiqueta del tipo mín..màx, para expresar los límites respectivos, junto a la línea que une cada entidad con la interrelación.

Los valores mín y máx podrán tener los siguientes valores:

- Cero, para indicar la posibilidad de que no exista ninguna asociación entre instancias.
- Cualquier número entero, para indicar un límite mínimo o máximo concreto de posibilidades de asociación entre instancias.
- Un asterisco (*), para indicar la posibilidad de un número ilimitado de asociaciones entre instancias.

Aunque otras interrelaciones asocian instancias de diferentes entidades, esta característica no es aplicable a las interrelaciones recursivas.

Una interrelación recursiva asocia las instancias de una entidad con otras instancias de la misma entidad.

Se dice que una interrelación recursiva es de grado 2 (o binaria) si sólo participa una entidad, la que se relaciona con ella misma.

Si en una interrelación recursiva participan, adicionalmente, más entidades, hablaremos de interrelaciones recursivas de grado 3 (o ternarias), de grado 4 (o cuaternarias), y así sucesivamente.

En el caso de las interrelaciones recursivas, puede tener importancia especificar los diferentes papeles o roles que interpretan las instancias de una misma entidad, si estos roles no coinciden plenamente. Si el rol es exactamente el mismo, no hay que especificarlo.

Notación

Como con las entidades, la notación diagramática para representar las interrelaciones y sus propiedades también es bastante sencilla:

- Toda interrelación se representa con un rombo, que va unido, mediante líneas, a todas las entidades que asocia.
- Los atributos de una interrelación, cuando existen, se representan de la misma por lo que los atributos de una entidad.
- La conectividad de una interrelación se representa añadiendo una etiqueta con un 1 o un N, según sea necesario, a cada una de las líneas que la une con las entidades que participan.
- La opcionalidad se representa superponiendo un círculo a la línea de conexión correspondiente, y la obligatoriedad, superponiendo un pequeño guion perpendicular a la línea de conexión de que se trate.
- Si hay que establecer límites (0, entero, *) en la cardinalidad de una interrelación, se ha de añadir a cada una de sus líneas de conexión una etiqueta con el límite inferior y el superior separados por dos puntos seguidos.
- La recursividad de una interrelación se representa haciendo llegar dos líneas de conexión a la misma entidad. Si participan más entidades de la misma interrelación recursiva, se harán llegar las líneas de conexión correspondientes desde la interrelación.
- Si hay que hacer una diferenciación de los roles de una interrelación recursiva, se ha de añadir una etiqueta, con la especificación textual adecuada, junto cada una de las líneas de conexión.

1.3 Entidades débiles

Las entidades que disponen de un atributo o, si no, de un conjunto de atributos capaces de establecer una clave primaria que sirva para distinguir cada instancia de la entidad del resto de ocurrencias se pueden llamar, más específicamente, entidades fuertes.

Las entidades débiles son aquellas que no disponen de suficientes atributos para designar unívocamente sus instancias. Para conseguirlo, han de estar asociadas, mediante una interrelación, con una entidad fuerte que las ayude.

La interrelación entre una entidad débil y su fuerte asociada es siempre de cardinalidad 1: N, y se resta del 1 al lado de la entidad fuerte, y la N junto a la débil.

Cada instancia de una entidad débil está asociada con una única ocurrencia de la entidad fuerte (por eso es en el lado 1 de la interrelación), y así es posible completar su identificación de manera única.

Por otra parte, la entidad del lado 1 debe ser obligatoria en la interrelación que, si no fuera así, alguna instancia de la entidad débil podría no estar asociada con ninguna de las ocurrencias de la entidad fuerte y, entonces, no se podría identificar completamente.

Las entidades débiles, pues, no tienen clave primaria, pero sí un atributo (o un conjunto de atributos) llamado discriminante, que permite distinguir entre ellas todas las instancias de la entidad débil que dependen de una misma instancia de la entidad fuerte.

Adicionalmente a la interrelación que les sirve para identificarse completamente, las entidades débiles pueden participar en otras interrelaciones, como cualquier otra entidad.

Notación

Para incorporar las entidades débiles a los diagramas ER, hay que aplicar unas pocas reglas de notación adicionales:

- Las entidades débiles se representan escribiendo su nombre en mayúsculas y en singular, dentro de un rectángulo dibujado con una línea doble.
- La interrelación que une la entidad débil con su fuerte se representa con un rombo también de línea doble.
- Este rectángulo debe unirse, igualmente con una línea doble, con la interrelación que lo asocia con su entidad fuerte de la que depende.
- El atributo o conjunto de atributos que actúen como discriminantes deben ir subrayados con una línea discontinua.

2. Diagramas Entidad-Relación

Los diagramas Entidad-Relación son un estándar actual en el diseño de bases de datos. De hecho, se trata de una herramienta sin la que, posiblemente, las bases de datos tal como las entendemos actualmente no existirían.

Los diagramas Entidad-Relación son herramientas gráficas clave en el diseño de bases de datos. Su confección debe ser sistemática y rigurosa si se quiere obtener un sistema válido y eficiente, ya que será a partir de estos diagramas con los que se desarrollará toda la implementación de bases de datos en los sistemas gestores de bases de datos concretos que correspondan a cada empresa u organismo.

2.1 Diseño de bases de datos

El diseño de BD estructura, fundamentalmente, en tres grandes etapas:

- Diseño conceptual
- Diseño lógico
- Diseño físico

Aunque nos centramos en el estudio y en la práctica del diseño conceptual, no se han de perder de vista las otras fases del diseño, que también son importantes, a fin de obtener una visión de conjunto de todos estos procesos.

Conocer cómo se estructura el modelo Entidad-Relación es importante. Y también lo son cuestiones que nos deben permitir aprovechar la tecnología proporcionada por las BD y los correspondientes sistemas gestores, como los siguientes:

- ¿Qué entidades debe incluir una BD determinada?
- ¿Qué interrelaciones se deben considerar?
- ¿Qué atributos deben existir y en qué entidades o interrelaciones han de incorporarse?

- ¿Qué claves primarias ya se pueden establecer en la fase de diseño conceptual?

2.1.1 Fases del diseño de BD

Diseñar BD no es una tarea sencilla. Aunque la porción del mundo real que se quiera modelizar en un caso concreto sea relativamente pequeña, las estructuras de datos resultantes pueden llegar a tener un cierto grado de complejidad. Sin embargo, a medida que aumenta la información a considerar, y su complejidad, el modelo de datos necesario para representarla puede convertirse, ciertamente, en una construcción complicada.

Querer resolver de golpe toda la problemática que puede conllevar la modelización de una BD no es, pues, una opción muy realista. Al afrontar una tarea de esta envergadura, es preferible dividirla en subtareas para simplificarla.

Así pues, resulta conveniente descomponer el diseño de BD en diferentes etapas, de modo que en cada una sólo se examinen ciertos aspectos, o tipo de problemas, a fin de minimizar la posibilidad de error. Entonces, a partir del resultado obtenido en cada fase, se puede continuar trabajando en la fase siguiente, hasta llegar al resultado esperado, al final de la última fase.

Es habitual estructurar el diseño de BD en las tres etapas o fases:

1. Diseño conceptual.
2. Diseño lógico.
3. Diseño físico.

Fase de diseño conceptual

Lo primero que hay que hacer, durante la fase de diseño conceptual, es recopilar toda la información necesaria de la parte del mundo real que nos proponemos modelizar con una BD.

Esta recopilación de información se realizará por diferentes vías, como éstas:

- Entrevistas con los futuros usuarios de la BD que se está diseñando.
- Examen de la documentación proporcionada por estos mismos usuarios.
- Observación directa de los procesos a informatizar.

A continuación, se deben estructurar convenientemente los datos necesarios para dar respuesta a todas las necesidades derivadas del conjunto de informaciones compendiadas.

El objetivo del diseño conceptual consiste en la obtención de una especificación sistemática.

Dicha especificación sistemática resultado del diseño conceptual debe cumplir dos tipos de requisitos:

- De datos. El modelo resultante debe tener en cuenta la estructura completa de los datos y su integridad.
- Funcionales. Un buen esquema conceptual también deberá prever las necesidades básicas en materia de manipulación de datos (es decir, las operaciones de inserción, borrado, consulta y modificación, de éstas).

Durante las fases posteriores, puede ser conveniente depurar el diseño con el fin de optimizar las operaciones a realizar sobre los datos.

Finalmente, hay que elegir un modelo de datos de alto nivel y traducir los requisitos anteriores a un esquema conceptual de la futura BD expresado con los conceptos y la notación correspondientes. Uno de los modelos de datos de alto nivel más utilizados es el modelo entidad-interrelación.

Expresado en la terminología del modelo ER, el esquema de datos desarrollado durante la fase de diseño conceptual debe especificar todas las entidades necesarias, y las interrelaciones entre ellas, con las cardinalidades adecuadas, y también los atributos en cada caso.

El modelo resultante debe revisarse para garantizar la satisfacción de todas las necesidades detectadas, por un lado, y para evitar redundancias de los datos (es decir, repeticiones indeseadas de estas), de otra.

El resultado de la fase de diseño conceptual pertenece al llamado mundo de las concepciones, pero aún no en el mundo de las representaciones, ya que no se especifica ninguna representación informática concreta.

Como se puede ver, durante la fase de diseño conceptual no hay que tener en cuenta, aún, ni el tipo de BD que se utilizará posteriormente ni, mucho menos, el SGBD o el lenguaje concreto con el que se implementará.

Fase de diseño lógico

En la fase de diseño lógico, se trabaja con el modelo abstracto de datos obtenido al final de la etapa de diseño conceptual, para traducirlo al modelo de datos utilizado por el sistema gestor de bases de datos (SGBD) con el que se quiere implementar y mantener la BD.

Por tanto, a partir de esta fase de diseño, sí que hay que tener en cuenta la tecnología concreta que se debe emplear en la creación de la BD, ya que la BD resultante se puede adecuar a diferentes modelos lógicos, tales como los siguientes:

- Jerárquico
- Relacional
- Distribuido
- Orientado a objetos

A pesar de la diversidad de posibilidades, lo cierto es que lo más frecuente, a la hora de diseñar una BD, aunque consiste en expresar el esquema conceptual en un modelo ER y, a continuación, traducirlo a un modelo relacional.

Cuando el producto de una fase de diseño lógico es una BD relacional, esta consiste en un conjunto de relaciones (de lo contrario, llamadas representaciones tabulares) compuestas por atributos, algunos de los cuales forman parte de claves primarias o de claves foráneas.

Resulta evidente, pues, que el resultado de la fase de diseño lógico ya se sitúa dentro del llamado mundo de las representaciones informáticas.

Fase de diseño físico

El diseño físico consiste en hacer ciertos tipos de modificaciones sobre el esquema lógico obtenido en la fase anterior de diseño lógico, a fin de incrementar la eficiencia.

La eficiencia de un esquema puede conllevar la modificación de algunas operaciones que deban hacerse con los datos, aunque conlleven un cierto grado de redundancia de éstas, como por ejemplo:

- Añadir algún atributo calculable en alguna relación.
- Dividir una relación en otros dos o en más.
- Incluir en la BD una relación que sea el producto de combinar dos o más relaciones.

Pero la fase de diseño físico también se caracteriza por la posibilidad de adoptar otras decisiones, relacionadas con aspectos de implementación física a más bajo nivel, y estrechamente vinculadas con el SGBD con el que se trabaja en cada caso, como los siguientes:

- Definición de índices.
- Asignación del espacio inicial para las tablas, y previsión de su crecimiento ulterior.
- Selección del tamaño de las memorias intermedias.
- Parametrización del SGBD según las opciones que éste ofrezca.

Para tomar acertadamente estos tipos de decisiones, hay que tener en cuenta las características de los procesos que operan con los datos, la frecuencia de ejecución los diferentes tipos de consulta, el grado de volatilidad de los datos, los volúmenes de información a almacenar, etc.

2.1.2 Diseño conceptual de una BD

Podemos considerar cualquier empresa, organización, institución, etc. como un sistema con reglas propias de funcionamiento. Este sistema, susceptible de ser informatizado, está compuesto por tres subsistemas:

- Subsistema de producción. Se encarga de realizar las actividades propias de la organización de que se trate en cada caso (por ejemplo, fabricar coches, o repararlos, o venderlos, etc.).
- Subsistema de decisión. Se encarga de dirigir, coordinar y planificar las actividades realizadas dentro del ámbito del subsistema de producción.
- Subsistema de información. Se encarga de recoger, almacenar, procesar y distribuir, todas las informaciones necesarias para el buen funcionamiento los otros dos subsistemas.

Las BD sirven para almacenar las representaciones de las informaciones utilizadas dentro del ámbito de los sistemas de información.

Los elementos que conforman un sistema de información son de dos tipos:

- Datos: representaciones de las informaciones.
- Procesos: acciones ejercidas sobre los datos (consultas, modificaciones, cálculos, etc.).

En función de las observaciones anteriores, podemos afirmar que hay dos premisas que todo diseñador de BD debería tener bien asumidas antes de empezar a trabajar en cualquier proyecto:

- No es competencia del diseñador de BD, como tal, tomar decisiones sobre la porción del mundo real que quiere modelizar.
- En principio, el diseñador de BD tampoco no debe inventar características de la realidad a modelizar: simplemente las debe reflejar de la manera más fiel posible en el modelo resultante.

Podemos subdividir esta etapa de diseño conceptual en dos fases sucesivas, las cuales conllevan tareas de diferentes tipos: la recogida y abstracción de las necesidades de la organización, por un lado, y la elaboración de un esquema conceptual mediante un modelo de datos concreto, de la otra.

La recogida y abstracción de las necesidades de la organización se lleva a cabo mediante diferentes procedimientos (entrevistas, examen de

documentación, etc.), y en esta fase hay que recoger toda la información necesaria para cubrir todos los requerimientos de datos. Pero, con esta información, se debe seguir un proceso de abstracción que nos permita estructurarla y diferenciar entre las cuestiones esenciales y las accesorias.

La elaboración de un esquema conceptual mediante un modelo de datos concreto conlleva que toda la información recogida y debidamente estructurada se expresará en una notación estandarizada (como los diagramas ER).

2.1.3 Captura y abstracción de los requerimientos de datos

Para realizar la captura y abstracción de los requerimientos, en primer lugar, es necesario averiguar qué necesidades tienen los usuarios de la futura BD. A menudo, estos usuarios potenciales sólo tienen una percepción muy general de lo que necesitan.

El diseñador debe saber seleccionar las cuestiones esenciales, diferenciarla de los aspectos accesorios, y descubrir cuáles son los verdaderos intereses los que han encargado el diseño de la BD.

Ahora bien, el diseñador de BD, como tal, no debe decidir nada. Su tarea consiste más bien en ayudar a los usuarios potenciales a descubrir qué necesitan exactamente. Es un trabajo pesado y complicada, que sirve para descubrir el verdadero flujo de datos que se reflejará en el modelo conceptual resultante.

Un buen diseñador de BD debe llegar, como mínimo, a una solución plausible para cada problema que se le haya planteado. A veces, también podrá tomar en consideración soluciones parciales alternativas. Y, en todo caso, finalmente, debe presentar un análisis completo de los requerimientos en la que concrete las especificaciones de la organización que le ha encargado el proyecto de diseño.

2.1.4 Identificación de entidades

La especificación de los requerimientos de datos sirve como punto de partida para la elaboración del esquema conceptual de la futura BD. A

continuación, lo primero que hay que hacer es identificar las entidades y sus atributos.

La identificación de las entidades y de sus atributos se puede documentar con un listado que siga el siguiente formato:

ENTIDAD1 (ATRIBUTO1, Atributo2, ...)

ENTIDAD2 (ATRIBUTO1, Atributo2, ...)

....

ENTIDADn (ATRIBUTO1, Atributo2, ...)

En el listado, los nombres de las entidades irán con mayúsculas, y los nombres de los atributos comenzarán con mayúsculas.

Debemos ser conscientes de que cualquier especificación que no se desprenda directamente los documentos de trabajo inicial en que hemos sintetizado los diferentes requerimientos de datos detectados debe validar mediante nuevas entrevistas, o el examen de nueva documentación o bien revisión de la antigua, o, si no, observando in situ el proceso en ejecución que pueda conllevar una innovación en la especificación.

En cambio, otros aspectos que pueden parecer más osados, (como representar lo que era inicialmente el atributo de una entidad como otra entidad independiente) no deben implicar necesariamente unos procesos de validación como los que acabamos de describir, si se tratan puramente de decisiones técnicas de diseño.

Hay que hacer notar que, a poco complicada que sea la porción del mundo real a modelizar, normalmente es posible obtener varios modelos conceptuales, en algunos casos alternativos y equivalentes, y en otros orientados a satisfacer en mayor o menor medida diferentes fines, pero con resultados igualmente correctas.

2.1.5 Designación de interrelaciones

Tras la captura y abstracción de los requerimientos de datos y de la identificación de entidades, el siguiente paso consiste en establecer las interrelaciones necesarias entre las entidades detectadas.

Las interrelaciones se pueden documentar con el siguiente formato:

Interrelación (ATRIBUT1, Atribut2, ...), entre Entidad1 y Entidad2.

En que el nombre de la interrelación comenzará con mayúsculas y, en caso disponer de atributos, estos irán entre paréntesis y con la inicial del nombre también con mayúsculas. Se deberá especificar el nombre de las entidades que interrelaciona.

En algún caso, las interrelaciones pueden incorporar algún atributo, al igual que las entidades.

Cuando se repite un mismo nombre en más de una interrelación, hay que enumerarlas correlativamente, para evitar confusiones al hacer referencia a cada una.

2.1.6 Establecimiento de claves

Recordemos que la clave primaria de una entidad está constituida por un atributo, o por un conjunto de atributos, los valores de los que son capaces de identificar unívocamente las instancias de aquella.

A veces, una entidad dispone de más de un atributo, o conjunto de atributos, que están en condiciones de constituir la clave primaria. En este caso, hablamos de claves candidatas.

Y una vez que el diseñador elige un atributo o conjunto de atributos concretos para identificar unívocamente las instancias, entre los que cumplen las condiciones (claves candidatas), este o estos atributos pasan a ser la clave primaria, y los no seleccionados permanecen como claves alternativas.

Un criterio importante a la hora de decidirse porque uno o más atributos formen la clave primaria de una entidad, es que su valor no cambie nunca o, cuando menos, muy raramente.

El atributo o conjunto de atributos que forman la clave primaria deben aparecer subrayados en la documentación.

Por ejemplo, no sería muy prudente decidirse por el número de teléfono móvil como clave primaria de una entidad que almacena personas, porque, a pesar de ser un número habitualmente de uso personal, puede cambiar a lo largo del tiempo. En cambio, el número de DNI puede ser, en general, una buena opción, ya que, en principio, este número es personal e intransferible.

2.1.7 Establecimiento de cardinalidades

El establecimiento correcto de las cardinalidades adecuadas para cada interrelación depende de las características del mundo real que se quieren modelizar.

Hay que añadir, pues, la información referente a las cardinalidades de las interrelaciones en la documentación de diseño.

2.1.8 Restricciones de participación y límites de cardinalidad

Se dice que la participación de una entidad en una interrelación es total si cada una de sus instancias participa una vez, al menos, en la interrelación y parcial en caso contrario.

Por lo tanto, las interrelaciones se documentarán en el siguiente formato:

Interrelación (Atributo1, Atributo2, ...), entre ENTIDADi (participación) y ENTIDADj (participación): tipo_de_cardinalidad

En el que el nombre de la interrelación comenzará con mayúsculas y, en caso de atributos, estos serán entre paréntesis, separados por comas, y con la inicial del nombre también con mayúsculas. Será necesario especificar el nombre de entidades interrelacionadas y el tipo de cardinalidad, que puede ser 1-1, 1-N, N-1 o M-N. La participación puede ser total o parcial.

2.1.9 Elaboración de un esquema conceptual

Una vez que tengamos los requisitos; y hemos identificado las entidades, atributos e interrelaciones; y gracias al conocimiento alcanzado de las

necesidades que deben satisfacer los datos, el diseñador está en condiciones de elaborar un conceptual de estos y sus interrelaciones.

Para crear este modelo conceptual, el diseñador de BD utilizará modelo de datos que se adapte a las necesidades del proyecto, y permita seguir trabajando en ello durante la fase posterior del diseño lógico.

El modelo de datos más utilizado para desarrollar el esquema conceptual es el modelo ER, y la documentación del diseño conceptual culmina con diagrama ER.

Aunque el modelo ER sigue siendo el modelo de datos más utilizado, con variaciones en los aspectos de notación, el modelo UML es muy útil en proyectos que basan la implementación de la programación en el paradigma de la orientación de los objetos.

Para empezar a trabajar con diagramas de ER, puede ser conveniente establecer el esqueleto modelo, sin incluir todos los detalles (atributos, cardinalidades, etc.).

Al diseñar el diagrama ER, puede ser una buena práctica comenzar con un primer diagrama inacabado e ir incluyendo más detalles progresivamente, por ejemplo, las cardinalidades y restricciones de participación.

Y el último paso que el diseñador tomará para completar el diagrama, para que refleje todos los requisitos detectados anteriormente es especificar todos los atributos y claves.

2.2 Extensiones del modelo Entidad-Relación

Las estructuras básicas del modelo Entidad-Relación (Modelo ER) permiten representar la mayoría de las situaciones del mundo real que normalmente son necesarias para ser incorporadas a la BD. Sin embargo, a veces ciertos aspectos de los datos deben describirse por medio de construcciones más avanzadas del modelo ER, las cuales son una extensión del modelo básico de ER. Estas extensiones del modelo de ER consisten en especialización, generalización y agregación de entidades.

2.2.1 Especialización y generalización

Podemos encontrar el caso de algún tipo de entidad en la que -además de las características generales, comunes a todos sus instancias- nos interesa modelar, además, ciertas características específicas aplicables a sus instancias.

Entonces, podemos considerar que esta entidad de tipo contiene otras entidades de tipo, con sus propias características.

La especialización refleja la existencia de una entidad general, la llamada entidad de superclase, que puede especializarse en diferentes entidades subclase.

La entidad superclase permite representar las características comunes de la entidad desde un punto de vista general. Por otra parte, las entidades de subclase permiten características de las especializaciones de la entidad de superclase.

Las instancias de las subclases deben ser, al mismo tiempo, instancias de la respectiva superclase.

El proceso de nombramiento de subclases de una superclase se llama Especialización.

Por lo tanto, la especialización refleja las diferencias entre las instancias de una misma entidad, estableciendo diferentes entidades de nivel inferior, que agrupan el subconjunto de instancias con características específicas comunes.

Estas características de las subclases pueden consistir tanto en la existencia de atributos como en la participación en interrelaciones, pero bajo ninguna circunstancia pueden ser aplicable a todas las instancias de la superclase considerada como tal.

La generalización, por otro lado, es el resultado de observar condiciones preexistentes que comparten ciertas características comunes (es decir, la identidad de atributos o interrelaciones en las que participan).

Dependiendo de las similitudes detectadas entre diferentes entidades, estas se sintetizan en una sola entidad, a un nivel superior, a través de un proceso de generalización.

La generalización sirve para destacar las similitudes entre entidades, por encima de las diferencias, así como para simplificar las representaciones de los datos, evitando la repetición de atributos compartidos por diferentes subclases.

Por lo tanto, el producto resultante de la especialización y la generalización es idéntico.

La diferencia entre los dos radica en el tipo de proceso que conduce a cada uno:

- La especialización deriva de un proceso de diseño descendiente, durante el cual, a partir de una entidad preexistente, considerada como una superclase se detecta la utilidad de establecer ciertas subclases, debido a la existencia de ciertas características (atributos y acciones en interrelaciones) no aplicables a todas las instancias de la superclase.
- La generalización responde a un proceso considerado de diseño ascendente. Durante este tipo de diseño se valora la utilidad de contemplar entidades preexistentes, denominadas subclases, dependientes de una misma superclase común a todas ellas. La superclase presenta características comunes (atributos y acciones en las interrelaciones) a todas las subclases que dependen de ella.

Herencia de propiedades

Tanto en el caso de la generalización como de la especialización, las características de la entidad superclase se extiende a las entidades de subclase. Como ya sabemos, estas características pueden consistir en atributos de la entidad de superclase, o en la participación en diferentes interrelaciones.

Llamamos herencia de propiedades a la transmisión de características (atributos e interrelaciones) de la entidad de superclase a la subclase.

Puede suceder que la misma entidad adopte el papel de subclase en un proceso de generalización o especialización y, al mismo tiempo, asumir el papel de superclase en otro de estos procesos en los que participa.

Cuando se produce una jerarquía de entidades, las entidades en niveles inferiores pueden heredar características no sólo de la superclase respectiva, sino también otras clases de nivel superior.

Llamamos herencia múltiple a la recepción, por una entidad de subclase, de características (atributos e interrelaciones) de su superclase, así como las de otras entidades de nivel superior, dentro de una estructura de entidades jerárquicas con generalizaciones o especializaciones encadenadas.

Restricciones

Para modelar más exactamente la trama del mundo real que nos interesa, pueden establecerse ciertas restricciones a las especializaciones o generalizaciones detectadas.

Un primer tipo de restricciones define si las instancias pueden pertenecer simultáneamente a más de una subclase de una estructura simple (que tiene una sola superclase y un solo nivel de subclases) de generalización o especialización. En estos casos, las entidades de subclase pueden ser dos tipos:

- Disjuntas (exclusivas). La misma entidad de instancia no puede aparecer en dos entidades diferentes subclase. Se representa en el diagrama añadiendo una etiqueta con letra D.
- Superpuestas (solapadas). La misma entidad de instancia puede aparecer en dos (o incluso, en más de dos) entidades de subclase diferentes. Está representado en el diagrama mediante una etiqueta con la letra E.

Un segundo tipo de restricciones especifica si cualquier instancia de la superclase debe simultáneamente pertenecer a una o más de las subclases o no. Aquí las entidades de subclase también pueden ser de dos tipos:

- Totales. Todas las instancias de la entidad de superclase deben pertenecer simultáneamente a al menos una de sus entidades de subclase. Se denota con etiqueta T.
- Parcial. Es posible que algunas instancias de la entidad de superclase no pertenezcan simultáneamente a cualquiera de sus entidades de subclase. Se denota con etiqueta P.

Al combinar estas restricciones, obtenemos cuatro posibilidades aplicables a las subclases de una generalización o especificación. Es necesario separar las letras que se incluyen en la etiqueta con un coma:

- D, T (disjunta y total)
- D, P (disjunta y parcial)
- E, T (superpuestas y totales)
- E, P (superpuesta y parcial)

Notación

Tanto la especialización como la generalización están representadas por un triángulo que incluye, en su interior, la etiqueta ES. Esta etiqueta indica que cualquier instancia de cualquiera de las subclases es, al mismo tiempo, una instancia de la superclase.

Para distinguir claramente la superclase en los casos en que hay un gran número de entidades de subclase implicadas en la estructura, o cuando es difícil, para las características del diagrama, alinee claramente todas las subclases, es conveniente indicar los límites de cardinalidad de la generalización o especialización. Para esto, sólo tiene que añadir una etiqueta del tipo min.. máx., con el fin de expresar los límites respectivos, junto a la línea que conecta cada entidad con el triángulo que representa la generalización o especialización, en la que min y max pueden tener los siguientes valores:

- 1..1 en la línea que une la superclase, ya que cualquier caso de cualquier subclase siempre constituirá, simultáneamente, una única instancia de la superclase.

- 0..1 en la línea que une cada subclase, porque no necesariamente toda instancia de la superclase debe ser simultáneamente una instancia de la subclase en cuestión (puede ser de otra subclase, o puede no serlo de ninguna, si estamos en presencia de una restricción de parcialidad).

Las entidades que forman parte de una estructura de generalización o especialización se representan como las otras entidades: cada una con un rectángulo que incorpora el nombre respectivo, y los atributos respectivos rodeados dentro de elipses vinculadas a la entidad con una línea. Si los atributos forman una clave principal, su nombre debe subrayarse.

En términos de notación esquemática, no hay diferencia entre una generalización y una especialización. Las diferencias entre los dos fenómenos son reducir al proceso que se ha seguido para derivar de cada uno de ellos, pero no en el resultado, que siempre es el mismo: el establecimiento de una superclase y subclases con restricciones específicas, que se representan añadiendo, a la etiqueta ES, las iniciales de las dos restricciones aplicables separadas por una coma:

- D, T (disjunta y total)
- D, P (disjunta y parcial)
- E, T (superpuestas y totales)
- E, P (superpuesta y parcial)

2.2.2 Agregaciones de entidades

Con las reglas básicas del modelo de ER, sólo se pueden modelizar interrelaciones en que participan exclusivamente entidades, pero no es posible expresar la posibilidad de que una interrelación participa directamente en otra interrelación. Pero hay un mecanismo, llamado agregación, que permite superar la limitación descrita anteriormente, considerando una interrelación entre entidades como si se tratara de una entidad, y usarla como tal.

La agregación de entidades es una abstracción, a través de la cual, una interrelación es tratada como una entidad de nivel superior, que reúne las

entidades interrelacionadas gracias a ella. La agregación debe tener el mismo nombre que la interrelación en la que se define.

Por lo tanto, la utilidad de una agregación de entidades consiste en el hecho de que la interrelación en la que se basa puede estar interrelacionada con otras entidades. Una agregación de entidades se representa encuadrando todas las entidades que participan en una interrelación de terminada, con el fin de construir una nueva entidad que pueda establecer sus propias interrelaciones.

Notación

Las agregaciones de entidades se representan incluyendo dentro de un cuadro todas las entidades que participan en una interrelación particular.

A partir de una interrelación, se puede hacer una flecha (punto único o doble, para expresar la cardinalidad 1 o N, respectivamente) hasta el rombo incluido dentro del recuadro que indica la existencia de una agregación (o hasta el mismo recuadro, exactamente como si fuera una entidad simple).

Toda agregación debe tener el mismo nombre que la interrelación que la define.

3. Decisiones de diseño

El diseño del BD consiste en definir una estructuración de los datos para satisfacer las necesidades de los futuros usuarios del sistema de información que se quiere construir.

Con el fin de satisfacer los requisitos funcionales de los usuarios, el diseñador de BD considerará los diferentes tipos de operaciones que se llevarán a cabo sobre los datos.

El diseñador tendrá que tomar ciertas decisiones en la modelización de los datos, que incluso pueden conducir a la revisión del esquema inicial, como en las siguientes áreas:

- Uso de entidades o atributos
- Uso de entidades o interrelaciones

- Uso de una interrelación n-aria o diferentes interrelaciones binarias
- Ubicación de los atributos de interrelación
- Uso de la entidad DATA

El diseñador también tendrá que detectar y evitar las trampas de diseño que se producen mediante conceptualizaciones erróneas del mundo real, tales como:

- Encadenamiento erróneo de las interrelaciones binarias 1-N
- Uso incorrecto de las interrelaciones binarias M-N
- Falsas interrelaciones ternarias

A continuación, todas estas decisiones deben reflejarse en una documentación que permita continuar trabajando en las fases de diseño posteriores.

3.1 Alternativas de diseño

Una de las características fundamentales del modelo DE ER es que es muy flexible. Tanto es así, que la misma realidad se puede modelar de diferentes maneras por el diseñador, que a veces tiene alternativas a la hora de definir las entidades y sus interrelaciones.

3.1.1 Uso alternativo de entidades o atributos

A veces, el mismo objeto del mundo real puede ser representado por medio de un atributo o entidad.

Tratar un concepto del mundo real como una entidad en lugar de como un atributo implica ciertas ventajas:

- Evita la redundancia de datos, ya que el mismo valor sólo se introducirá una, que permite lo siguiente:
 - Ahorrar espacio en la BD.
 - Minimizar la posibilidad de error del usuario y facilitar la corrección.

– Optimizar las consultas sobre la BD y mejorar su consistencia de resultados.

- Asigna una cardinalidad (1 o N), y unos límites sobre esta, sin recurrir al uso de atributos multivalor, para que podamos asignar 0, 1 o más valores en cada caso, de acuerdo con la realidad que queramos modelizar.
- Incluye información adicional mediante la adición de nuevos atributos a la entidad creada o, si no, relacionándolos con otras entidades.

En resumen, tratar un concepto como una entidad es una opción más general que tratarlo como un atributo, lo que permite almacenar información adicional, agregar nuevos atributos o establecer nuevas interrelaciones.

3.1.2 Uso alternativo de entidades o de interrelaciones

A veces, es mejor representar un objeto del mundo real mediante una entidad y, otras veces, como una interrelación.

Como regla general, podemos hacer las siguientes afirmaciones:

- Las entidades consisten en objetos del mundo real, independientemente del hecho que existan físicamente como, por ejemplo un coche, o que tengan un carácter más bien abstracto, como una póliza de seguro. Habitualmente, nos referimos con sustantivos.
- En cambio, las interrelaciones, deberían servir para representar acciones o procesos que tienen lugar entre entidades. Es frecuente referirse a ella utilizando verbos (aunque sea con participios).

3.2 Trampas de diseño

Llamamos trampas de diseño las conceptualizaciones erróneas del mundo real, producidas durante la fase de diseño conceptual, que tienen repercusiones negativas tanto en la modelización inicial como en la implementación a posterior de la BD.

Estas trampas pueden conllevar la imposibilidad de representar los datos tal como son, o bien la imposibilidad de realizar determinadas consultas sobre estas.

3.2.1 Encadenamiento erróneo de interrelaciones binarias 1-N

El encadenamiento erróneo de interrelaciones se puede producir siempre que nos encontramos con dos (o más) interrelaciones de cardinalidad 1-N mal encadenadas. Consideramos una entidad (A) que está asociada con otra (B), que al mismo tiempo lo está con una tercera entidad C. Entonces, si en la aplicación errónea de la transitividad asocia directamente la entidad A con la C, se puede producir un error conceptual que provoque pérdida de información.

3.2.2 Uso incorrecto de interrelaciones binarias M-N

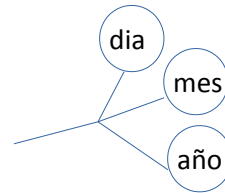
El uso de dos interrelaciones binarias, encadenadas y de cardinalidad M-N será erróneo siempre que en el mundo real exista algún tipo de asociación entre las instancias de las entidades de ambos extremos, ya que ésta no quedará reflejada en el modelo. La solución consistirá en sustituir las dos interrelaciones binarias por una de ternaria, con cardinalidad M-N-P.

3.2.3 Falsas interrelaciones ternarias

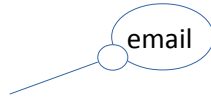
Cuando alguna interrelación ternaria (o n-aria de orden superior) tiene asociada alguna entidad con cardinalidad 1, se debe estudiar detenidamente, ya que es posible que esta entidad esté directamente relacionada sólo con una sola de las otras entidades y que, por tanto, no tenga que participar en la interrelación examinada, sino en una de binaria con la entidad con la que mantiene realmente una asociación.

4. Ejemplos de diseño.

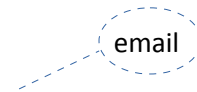
- Atributos. Compuesto(fecha).



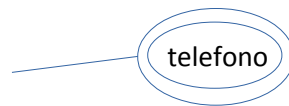
Opcional.



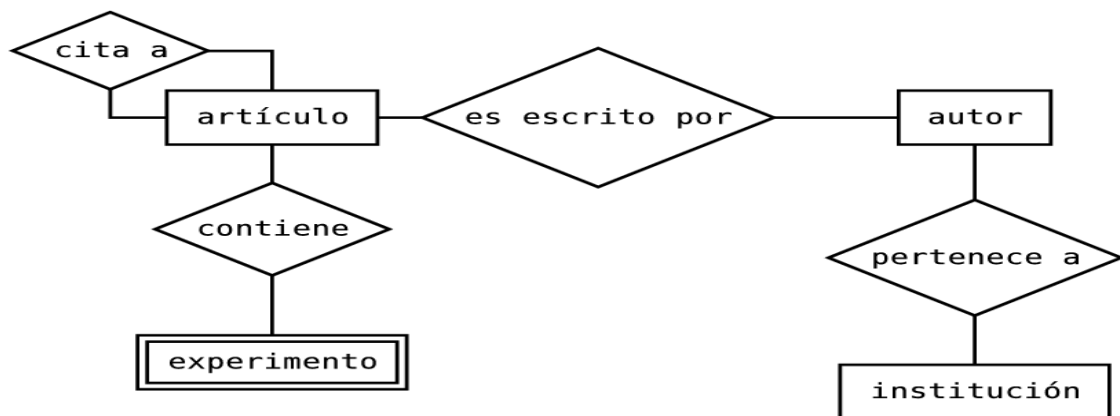
Derivado.



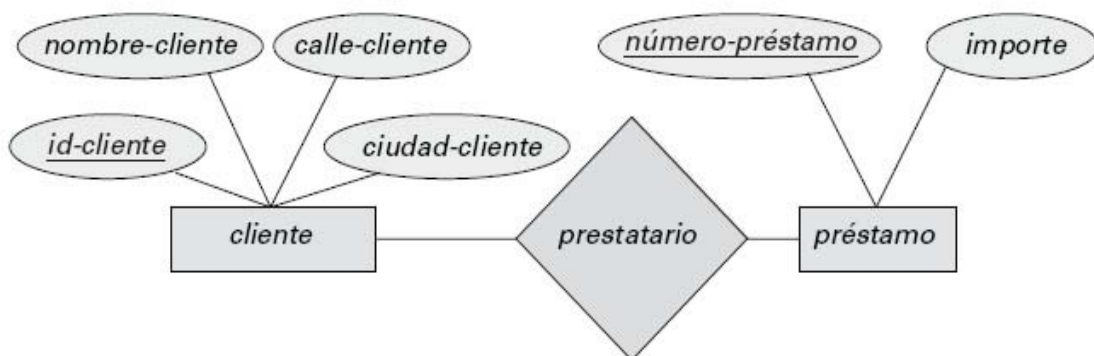
Multivaluado.



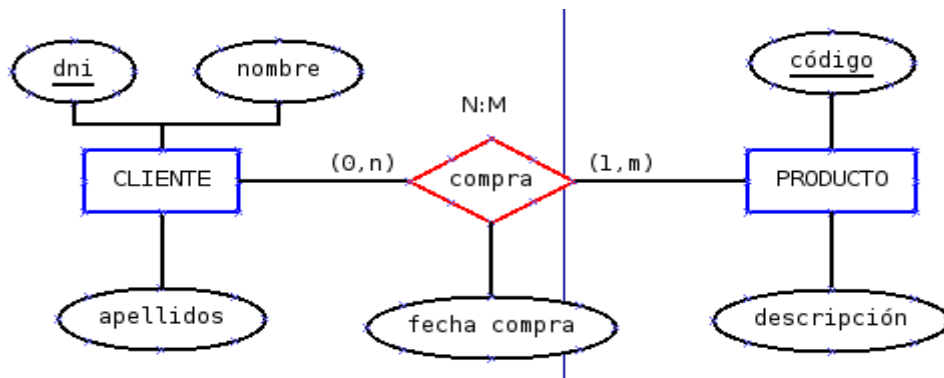
- En el siguiente gráfico podemos ver varias entidades fuertes, representadas por rectángulos (artículo, autor, institución). También podemos ver una entidad débil, representada por un rectángulo doble (experimento).



- En la imagen siguiente, la clave primaria de la entidad CLIENTE es id-cliente. La clave primaria de la entidad PRÉSTAMOS es número-préstamo. Tened en cuenta que la clave primaria puede estar formada por más de un atributo.



- Relación binaria, de nombre COMPRA (entre las entidades CLIENTE y PRODUCTO):



Aquí también podemos ver la cardinalidad de la relación COMPRA:

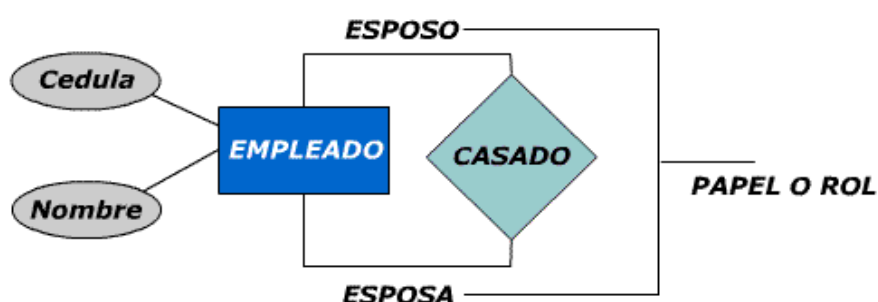
A un cliente le corresponden 1 o n productos. Es decir, un cliente compra uno o muchos productos (en principio no tiene sentido que un cliente compre 0 productos).

Al mismo tiempo, un producto puede ser comprado por 0 a n clientes. Es decir, un mismo producto puede ser comprado por ningún cliente o por muchos clientes.

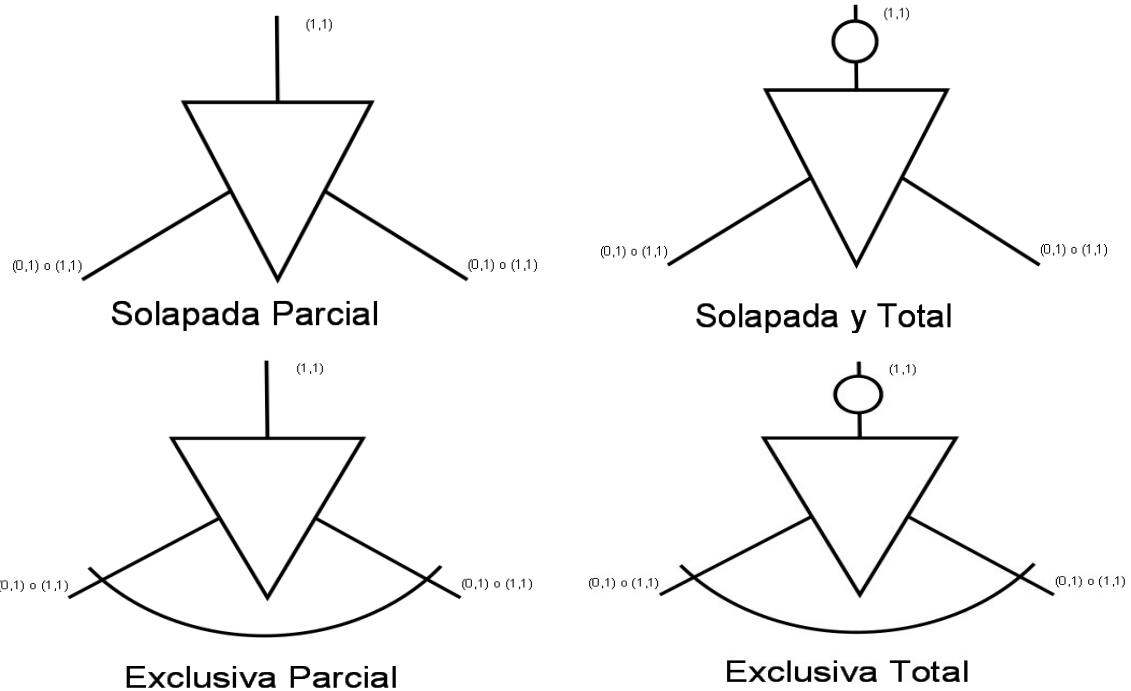
- Relación ternaria de nombre TRABAJA-EN (entidades EMPLEADO, TRABAJO y SUCURSAL):



- En la siguiente imagen se muestra una relación unaria, de nombre CASADO (en la que participa una única entidad de nombre EMPLEADO).



- Tipos de generalización/especialización



La generalización es total si no hay ocurrencia en el supertipo que no pertenezca a ninguno de los subtipos. En otro caso la generalización sería parcial.

La generalización será exclusiva si una ocurrencia no puede aparecer en varios subtipos a la vez.

Ejemplo (solapada y parcial):

