**TEMA2. DISEÑO DE BASES DE DATOS RELACIONALES. - DISEÑO CONCEPTUAL -**

1. FASES DEL DISEÑO
2. EL MODELO ENTIDAD RELACIÓN (Modelo E-R)
   1. Entidades
   2. Relaciones.
   3. Grado y Cardinalidad de una Relación.
   4. Atributos

Enfoque lingüístico.

1. EL MODELO ENTIDAD RELACIÓN EXTENDIDO
   1. Generalización y Especialización.
   2. Agregación.

# FASES DEL DISEÑO

La creación de una base de datos es un proceso complejo que parte de la necesidad de almacenar información del mundo real, para que su acceso sea rápido y eficiente. Un buen diseño de nuestra base de datos nos va dar un mayor rendimiento, mejor velocidad de acceso a los datos, eliminar las redundancias, y en general un mejor aprovechamiento de los recursos de los que disponemos.

Parece claro que requerimos de una metodología lo suficientemente potente para realizar esta transformación de una forma metódica y eficiente.

Sin esta metodología podemos llegar a un diseño inapropiado de nuestra base de datos, que puede derivar en alguno de los siguientes problemas:

* **Redundancia:** podemos definir una redundancia como una repetición innecesaria de información. La redundancia en una base de datos no sólo incrementa el espacio ocupado por la misma, sino que en el futuro provocará inconsistencia en los datos que se almacenan.
* **Inconsistencia:** como hemos comentado anteriormente, al existir redundancias pueden producirse inconsistencias. Supongamos una base de datos en la que tenemos elementos duplicados y modificamos sus valores en una sola de las copias; esto produce incoherencia, la base de datos es inconsistente consigo misma, ya que tenemos dos elementos que deberían ser iguales y, por lo tanto, contener los mismos valores, pero sus valores son contradictorios.
* **Pérdida de dependencias funcionales**: las dependencias funcionales las veremos más adelante, simplemente comentar que pueden pasarse por alto si no disponemos de una metodología de diseño.

Estos problemas parecen justificar por sí mismos la adopción de una metodología para el diseño de la base de datos.

Esta metodología la podríamos dividir en las siguientes fases:

* **Diseño conceptual:** este diseño es independiente del modelo de base de datos usado, del ordenador, del sistema gestor de la base de datos. Simplemente estudiamos el problema, y seleccionamos qué elementos del mundo real vamos a modelar.

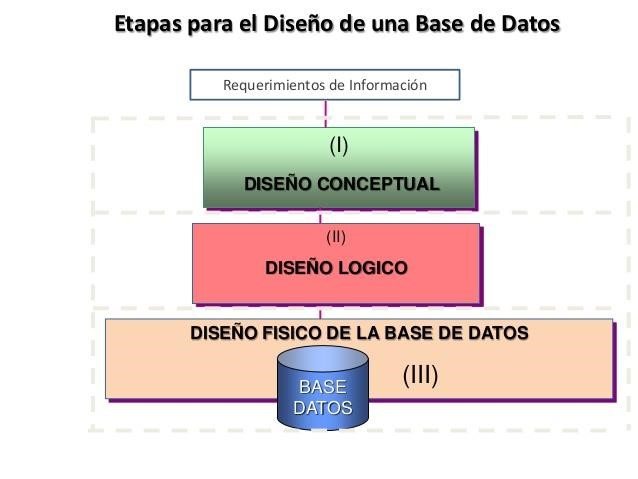
Este diseño, en el caso del modelo conceptual, es el modelo Entidad-Relación.

* **Diseño lógico:** partiendo del diseño conceptual obtenido en la fase anterior, llegamos a un diseño lógico. Transformamos las entidades y relaciones obtenidas del modelo anterior en tablas.
* **Diseño físico**: este diseño si depende del ordenador, sistema gestor de base de datos, etc. En este caso, empleando el gestor de la base de datos implementamos nuestra base en una o varias máquinas.

Cada fase del diseño genera un resultado diferente:

* **Diseño conceptual:** el producto de esta fase es un modelo Entidad-Relación.
* **Diseño lógico**: obtenemos las tablas que van a componer nuestra base de datos.
* **Diseño físico**: el resultado de esta fase es la base de datos almacenada en un SGBD determinado y sobre un soporte informático.

Esta idea la podemos resumir en el siguiente esquema:



El diseño o modelo conceptual es la primera fase de diseño, y puede subdividirse en dos etapas muy diferenciadas: la etapa de **análisis de requisitos** y la de **conceptualización.**

* En la **etapa de** **análisis de requisitos** se ha de responder a la pregunta: “¿Qué representar?”. A través de entrevistas a los usuarios de los niveles de la organización, y de las reglas de la misma, se elabora un esquema descriptivo de la realidad. El esquema queda expresado en lenguaje natural, porque, a pesar de la ambigüedad, el usuario establece los problemas a resolver, facilitando la comunicación entre usuarios y analistas. Los posibles problemas de esta especificación se solucionarán con el resto de etapas del diseño.
* En la segunda etapa, la **etapa de** **conceptualización**, se responde a la pregunta: “¿Cómo representar?”. Se busca una representación normalizada, apoyándose en un modelo de datos (como el E/R o el EE/R) que cumpla varias propiedades: coherencia, plenitud, no redundancia, simplicidad, fidelidad, etc., para llegar al esquema conceptual, que presenta la información con su contenido semántico.

# EL MODELO ENTIDAD RELACIÓN (Modelo E-R)

El modelo conceptual en el que se apoya el modelo relacional es el que denominaremos modelo E-R. Este modelo fue desarrollado por Peter Chen en 1976 en su obra “The Entity-Relationship Model”

En 1988, el ANSI lo seleccionó como modelo estándar para los sistemas de diccionarios de recursos de información.

Es el modelo conceptual más extendido en todo el mundo. Su utilidad se basa en llegar a un nivel de abstracción, que permita definir los elementos que componen nuestra base de datos, de forma que sean independientes de la implementación de la misma.

Este modelo, como su nombre indica, se compone de dos elementos: por un lado, tenemos las **entidades (E),** y por el otro las **relaciones (R).**

Tanto las entidades como las relaciones tienen **atributos.** Los atributos almacenan información.

2.1. Entidades

Las entidades se representan con sustantivo dentro de un rectángulo, dentro del mismo situamos el nombre de la entidad.

Una entidad es un objeto sobre el que almacenamos información. Dicho de una manera más formal, podríamos decir que una entidad es “una persona, lugar, cosa, concepto, o suceso real o abstracto de interés para la empresa " ANSI (1977).

*También podemos definir una entidad como cualquier elemento diferenciado sobre el que nos interesa almacenar información.*

Ejemplos de entidad:



Dentro de las entidades distinguimos entre dos tipos:

* **Entidades fuertes**: su existencia no depende de ninguna otra entidad. Se representa con el nombre de la entidad encerrado dentro de un recuadro.
* **Entidades débiles:** su existencia está condicionada a la aparición de otra entidad. Se representa mediante un doble recuadro, encerrando el nombre de la entidad.

Por ejemplo: Para que exista una cuenta bancaria, es necesario que exista otra entidad, en la que pondríamos titular, que es el beneficiario de la misma. Si no hay ningún titular no existe la cuenta. Dicho de otra forma, si eliminásemos el titular de la cuenta de nuestra base de datos, nos veríamos obligados a borrar todas sus cuentas. Esto demuestra que cuenta es una entidad débil que depende de titular.

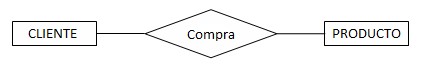


* 1. **Relaciones.**

Las relaciones se representan con un verbo, dentro de un rombo. Las relaciones representan asociaciones entre entidades.

**compra**

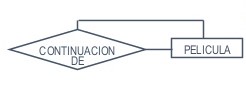
Podemos poner un pequeño ejemplo representado en este modelo: Supongamos que tenemos un almacén, y lo queremos informatizar. Entonces tenemos dos entidades, por un lado, la entidad **CLIENTE**, y por otro lado tenemos la entidad **PRODUCTO.** Entre los clientes y los productos del almacén existe una relación, que hemos denominado **compra.**



* 1. **Grado y Cardinalidad de una Relación.**

El **grado** de una relación nos indica el número de entidades que asocia. Una interrelación puede ser:

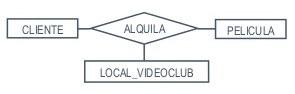
* **Unaria**: asocia una entidad consigo misma (de forma reflexiva).



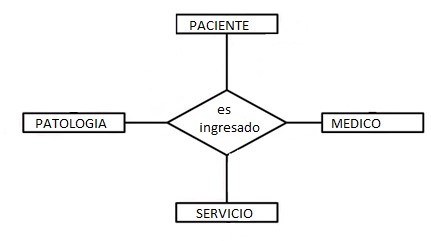
* **Binaria:** Asocia dos entidades.



* **Ternaria**: Asocia 3 entidades.



* **N\_aria**: Asocia n entidades, más de tres.



La **Cardinalidad** define la forma en que se relación las entidades e identifica el número de ocurrencias máximas y mínimas de una entidad en una relación.

Distinguiremos entonces diferentes tipos de cardinalidad:

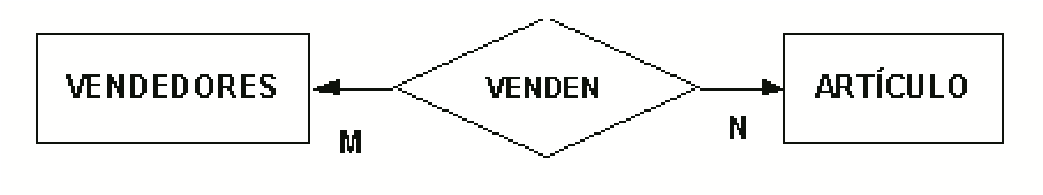
* **1:1.** **uno a uno.** Por cada ocurrencia de una entidad sólo puede aparecer uno de la entidad asociada.



* **1:N. uno a muchos.** Por una ocurrencia de una entidad pueden aparecer muchas ocurrencias de la entidad asociada.



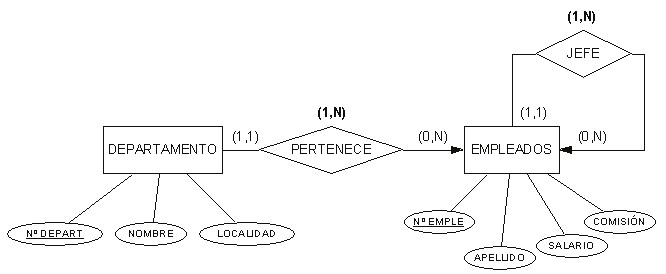
* **N:M. muchos a muchos.** La cantidad de asociaciones de una entidad con otra es múltiple.



La forma de representarlas es situar, encima del rombo que representa la asociación, su correspondiente numeración.

Para representar correctamente la **cardinalidad** escribimos el número mínimo y el número máximo de ocurrencias de la entidad en la relación, separados por 2 puntos, y rodeados por paréntesis, (mínimo: máximo).

En el ejemplo siguiente podemos ver como se representa la cardinalidad.

****

* 1. **Atributos**

Los atributos almacenan las **propiedades** (las informaciones) que nos interesan de las entidades.

Los atributos aparecen encerrados dentro de una elipse, y conectados con la entidad o relación a la que pertenecen mediante una línea.

Los atributos tienen un **dominio**, que es el rango de valores que puede tomar cada atributo, en función del tipo de dato que almacena.

Dentro de los atributos tenemos varias clases. Por un lado, están los atributos “normales” que simplemente dan información, y después tenemos otros atributos que sirven además para identificar las tuplas. Estos atributos son denominados **claves.**

Por ejemplo, la dirección de una persona nos da información sobre ella, pero su DNI la identifica. Existen distintos tipos de claves:

* **Superclave**: atributos que identifican de forma unívoca a una entidad dentro del conjunto de entidades. Por ejemplo: Número de socio es una superclave, ya que es un atributo que sólo poseen los socios, lo que permite diferenciarlos de las demás entidades.
* **Clave candidata:** atributos que identifican de forma inequívoca a una tupla u ocurrencia de la entidad. Por ejemplo: El DNI de un socio identifica a un único socio.
* **Clave primaria:** es la clave candidata por la que ha optado el diseñador de la base de datos. Una entidad puede tener varias claves candidatas, pero sólo puede tener una clave primaria que es la que elegimos como tal.

Para elegir la clave primaria existen varios criterios:

Se prefiere la clave numérica ante la alfanumérica.

La corta ante la larga.

La formada por un campo a la formada por varios campos.

* **Clave foránea o ajena:** sirve para establecer relaciones, es parte de la clave primaria de una relación, y simultáneamente es clave primaria de otra entidad. Por ejemplo: si relacionamos un socio con las películas que ha alquilado en la entidad Alquilan, debe aparecer la clave del socio, que puede ser su número.
* **Clave alternativa:** son las claves candidatas por las que no ha optado el diseñador como claves primarias.

Atendiendo a la cantidad de atributos que componen una clave las podemos clasificar en:

* **Claves simples:** la clave está formada por un solo atributo. Por ejemplo, en la entidad socio, el número del socio es una clave simple.
* **Claves compuestas:** la clave está formada por más de un atributo. Por ejemplo: En Alquilan tenemos como clave la calve de la entidad socio, más la clave de la entidad películas, más la fecha de alquiler.

**Caso práctico paso a paso:**

Se desea realizar el diagrama de estructuras de datos en el modelo E-R correspondiente al siguiente enunciado:

*“supongamos que en un centro escolar se imparten muchos cursos. Cada curso está formado por un grupo de alumnos, de los cuales uno de ellos es el delegado del grupo. Los alumnos cursan asignaturas, y una asignatura puede o no ser cursada por los alumnos”.*

**Para su resolución primero identificaremos las entidades, luego las relaciones y las cardinalidades y por último los atributos de las entidades y de las interrelaciones si las hubiera**.

1. **Identificación de entidades**: una entidad es un objeto del mundo real, algo que tiene interés para la empresa, se hace un análisis del enunciado, de donde sacaremos los candidatos a entidades: CENTROS, CURSOS, ALUMNOS, ASIGNATURAS, DELEGADOS, si analizamos esta última veremos que los delegados son alumnos, por lo tanto, los tenemos recogidos en ALUMNOS. Esta posible entidad la eliminaremos. También eliminaremos la posible entidad CENTROS pues se trata de un único centro, si se tratara de una gestión de diferentes centros tendría más sentido incluirla.
2. **Identificar las relaciones**: para ello es muy importante estudiar los requisitos del sistema de información para poder identificar las relaciones que existen entre las entidades y el tipo de relación que existe. Un método que nos puede ayudar mucho es construimos una matriz de entidades en la que las filas y las columnas son los nombres de entidades y cada celda puede contener o no la relación, las relaciones aparecen explícitamente en el enunciado. En este ejemplo las relaciones no tienen atributos. Del enunciado sacamos lo siguiente:

* Un curso está formado por muchos alumnos, la relación entre estas dos entidades la llamamos PERTENECE pues a un curso pertenecen muchos alumnos, relación 1:M. Consideramos que es obligatorio que existan alumnos en un curso. Para calcular los máximos y mínimos hacemos la pregunta: **a un CURSO** ***cuántos* ALUMNOS pertenecen, como mínimo y como máximo** y se ponen los valores en la entidad ALUMNOS, en este caso (1,M). **Para el sentido contrario hacemos lo mismo, un ALUMNO, a cuántos CURSOS va a pertenecer**, como mínimo a 1, y como máximo a 1, en este caso pondremos (1,1) en la entidad CURSOS.
* De los alumnos que pertenecen a un grupo uno de ellos es DELEGADO, hay una relación de grado 1 entre la entidad ALUMNO que la podemos llamar ES DELEGADO. La relación es 1:M, un alumno es delegado de muchos alumnos. Para calcular los valores máximos y mínimos preguntamos: un ALUMNO de cuantos alumnos ES DELEGADO, como mínimo es 0, pues puede que no sea delegado, y como máximo es M, pues si es delegado lo será de muchos, pondremos en el extremo (0,M). Y en el otro extremo pondremos (1,1), pues obligatoriamente el delegado es un alumno.
* Entre ALUMNOS y ASIGNATURAS surge una relación N:M, pues un alumno cursa muchas asignaturas y una asignatura es cursada por muchos alumnos. La relación se llamará CURSA, consideramos que puede haber asignaturas sin alumnos. Las cardinalidades serán (1:M) entre ALUMNO-ASIGNATURA, pues un alumno como mínimo cursa una asignatura, y como máximo muchas, y la cardinalidad entre ASIGNATURA-ALUMNO será (0,N), pues una ASIGNATURA puede ser cursada por 0 alumnos o por muchos.

En la tabla se muestra la matriz de entidades y relaciones entre ellas:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CURSOS** | **ALUMNOS** | **ASIGNATURAS** |
| **CURSOS** | *----------------* | *PERTENECE(1:M)* | *---------------* |
| **ALUMNOS** | *x* | *ES DELEGADO(1:M)* | *CURSA(N:M)* |
| **ASIGNATURAS** | *---------------* | *x* | *-------------* |

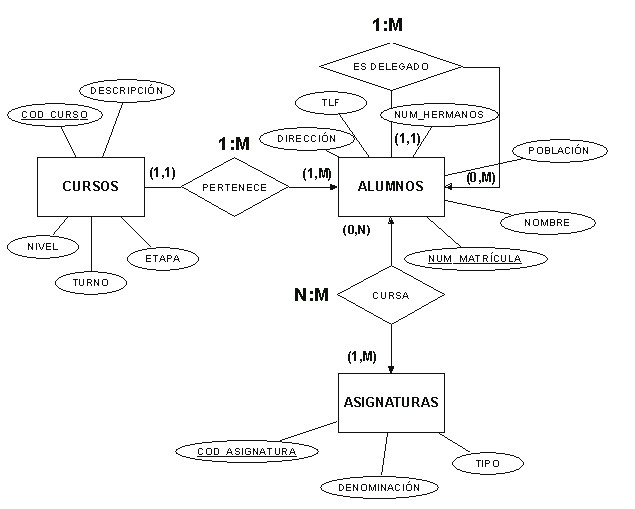
Las celdas que aparecen con una x indican que las relaciones están ya identificadas, las que aparecen con guiones indican que no existe relación. En la siguiente figura se muestra el diagrama de las relaciones y las cardinalidades:

1. **Identificar los atributos**: como el enunciado no explicita ningún tipo de característica de las entidades nos imaginamos los atributos que pueden ser los siguientes:

***CURSOS***: COD\_CURSO (clave primaria), DESCRIPCIÓN, NIVEL, TURNO y ETAPA

***ALUMNOS***: NUM-MATRÍCULA (clave primaria), NOMBRE, DIRECCIÓN, POBLACIÓN, TLF y NUM\_HERMANOS

***ASIGNATURAS***: COD-ASIGNATURA (clave primaria), DENOMINACIÓN y TIPO



* 1. **Enfoque lingüístico.**

Podemos abordar los MER mediante un enfoque lingüístico que nos puede ayudar en la creación de dichos modelos:

* Un sustantivo que actúa como sujeto o complemento directo en una frase es, generalmente, una entidad, aunque también puede ser un atributo.
* Un nombre propio suele indicar ocurrencias de una entidad.
* Un verbo transitivo (los que tienen complemento directo) o una frase verbal es una relación.
* Una preposición o frase preposicional entre dos nombres suele ser una relación, o establecer la asociación entre una entidad y sus atributos.

Otro acercamiento es el de la categorización de los objetos:

* Una entidad es un objeto de datos que tiene más propiedades que su nombre, o que se utiliza como operando en una sentencia de selección, borrado o inserción.
* Un atributo es un objeto de datos al que se le asigna un valor o se utiliza como operando en una operación aritmética, booleana o de cadena de caracteres.
* Una relación es un objeto de datos que hace posible la selección de una entidad por medio de una referencia a un atributo de otra entidad.

Del estudio de las frases que definen los requisitos se puede ir confeccionando el esquema, pero hay que tener una consideración especial con dos verbos muy comunes: ser (que puede indicar una jerarquía de entidades) y tener (que indica normalmente los atributos de una entidad).

**Realizar Ejercicios1 MER y Ejercicios2 MER.**

# EL MODELO ENTIDAD RELACIÓN EXTENDIDO

**3.1. Generalización y Especialización.**

La generalización es el proceso de abstracción inverso a la especialización. Se quitan las diferencias entre varios tipos de entidades y generalizamos sus características comunes para formar una entidad superclase. Dependiendo de si las subclases pueden aparecer en más de una subclase podemos observar dos tipos:

* Subclases disjuntas
* Subclases solapadas

La jerarquía es el proceso de subdividir una entidad en varias subentidades relacionándolas con la entidad a la que se refieren. Puede haber dos tipos:

* Total: que significa que no hay otro subtipo.
* Parcial: significa que puede haber otros subtipos.

Y los dos tipos de subentidades que puede haber, se dividen en dos también:

* Exclusiva: que significa que una subentidad no puede ser otra.
* Solapada: significa que una subentidad también puede ser otra.

Por ejemplo, en una empresa la entidad EMPLEADO con atributos NOMBRE, DNI, DIRECCIÓN, TELÉFONO, FECHA NACIMIENTO, SALARIO y PUESTO se divide en:

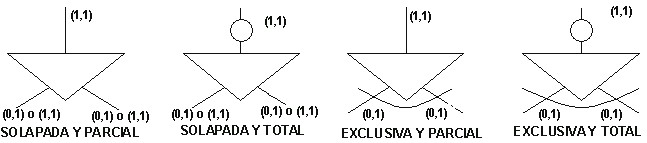
* Arquitectos con atributos COMISIONES Y NUMERO DE PROYECTOS.
* Administrativos con atributos PULSACIONES Y NIVEL
* Ingenieros: con atributos ESPECIALIDAD Y AÑOS DE EXPERIENCIA En el esquema Entidad-Relación quedaría de la siguiente forma:



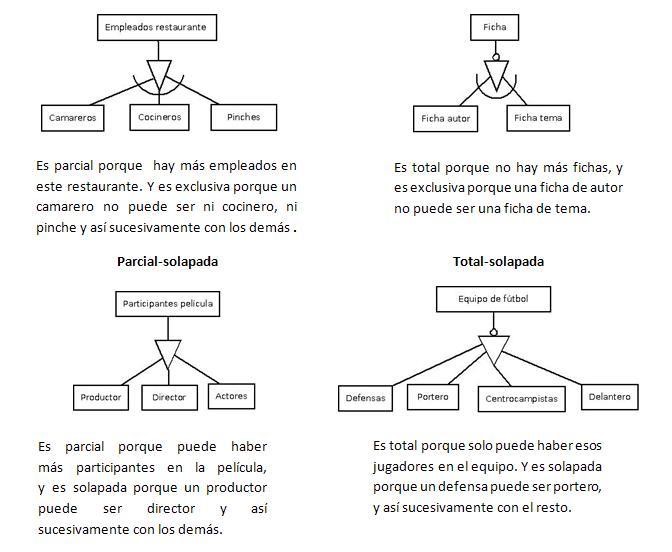
Tipos

* **Total:** todos los elementos de un tipo pertenecen a un subtipo, es decir, que no hay otro subtipo.
* **Parcial**: significa todo lo contrario, que sí hay otros subtipos, muchas veces no aparecen en la jerarquía, pero lo tienes que suponer.
* **Exclusiva**: significa que un subtipo no puede ser otro, simplemente puede ser el mismo sin tener otra segunda opción.
* **Solapada**: un subtipo puede tener la opción de ser otro subtipo, es decir, que no es único.

Por tanto: generalizaciones totales y exclusivas, totales y solapadas, parciales y exclusivas, parciales y solapadas pueden ser las opciones que podemos tener a la hora de hacer una jerarquía.



Por ejemplo



Representar las siguientes jerarquías:

* Un concesionario de coches vende coches nuevos y usados. Los atributos específicos de los nuevos son las: unidades y el descuento, de los usados son: los kilómetros y el año de fabricación.
* Consideramos el conjunto de personas de una ciudad, distinguimos a los trabajadores, estudiantes y parados. De los trabajadores nos interesa el número de la seguridad social, la empresa de trabajo y el salario. De los estudiantes el número de matrícula y el centro educativo, y de los parados la fecha del paro.
* En un campo de fútbol los puestos de los futbolistas pueden ser: portero, defensa, medio y delantero.

**Ejercicio 1**

Disponemos de la información de una serie de fabricantes de automóviles.

Un Fabricante puede fabricar muchos vehículos. Del vehículo sabemos el número de bastidor y el precio. El vehículo será fabricado por un único fabricante.

El vehículo puede ser camión, turismo o motocicleta. Del camión nos interesa el número de ejes y el tonelaje. Del turismo número de puertas y número de plazas. De la motocicleta la cilindrada. La motocicleta puede llevar o no un sidecar. Del sidecar nos interesa su identificación, el precio y el número de plazas.

Los camiones tienen que superar una serie de controles, de los que nos interesa almacenar los datos de los controles, así como la fecha en la que se realizó cada control.

**Ejercicio 2**

Se desea gestionar la información sobre las personas que viven en cada vivienda, de los propietarios de las viviendas.

Los datos de las personas son nombre, DNI y teléfono. De cada vivienda se necesita saber la calle, el Número y los metros cuadrados.

Existen dos tipos de viviendas, Casas y Bloques de Pisos.

De las Casas nos interesa almacenar el número de plantas que tiene, así como el número de metros de fachada. En una casa vive una sola familia.

Los Bloques de Pisos, contiene un número variable de pisos, de los que nos interesa guardar escalera, planta y letra. Los pisos pagan una serie de impuestos como, comunidad, basura, IBI, ….

Toda persona habita en una sola vivienda, considerada como vivienda habitual.

Cada vivienda tiene un único propietario. Las viviendas están en un único barrio de los que no interesa saber el nombre y extensión.

Una persona actúa como jefe de familia.

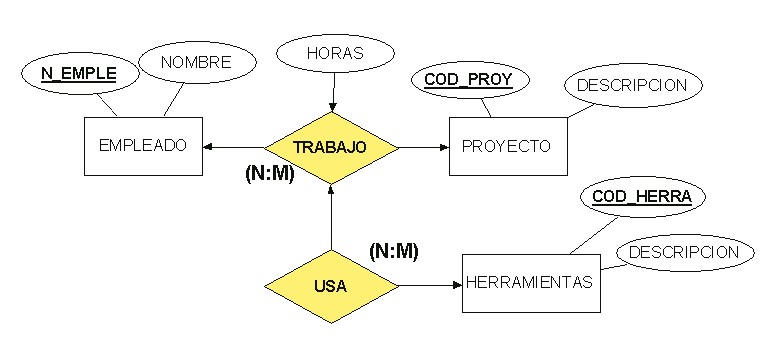
**Realizar Ejercicios3 MER.**

**3.2. Agregación.**

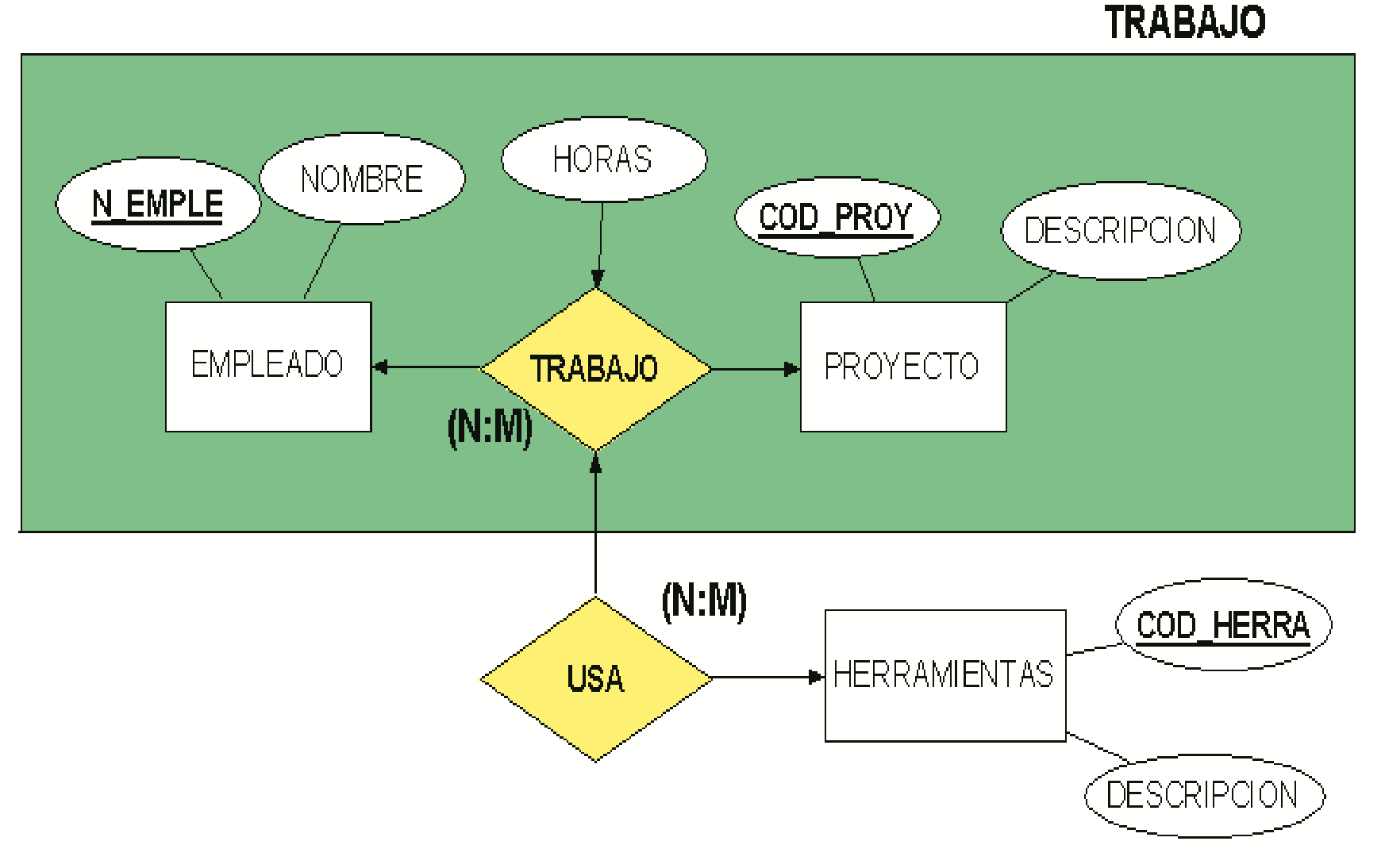
Una limitación del modelo E-R es que **no es posible expresar relaciones entre relaciones**, en estos casos se realiza una agregación, que es una abstracción a través de la cual las relaciones se tratan como entidades de nivel más alto.

Por ejemplo, consideramos una relación entre EMPLEADOS y PROYECTOS, un empleado trabaja en varios proyectos durante unas horas determinadas, y en ese trabajo utiliza unas herramientas determinadas.

La representación del diagrama de estructuras se muestra en la figura siguiente:

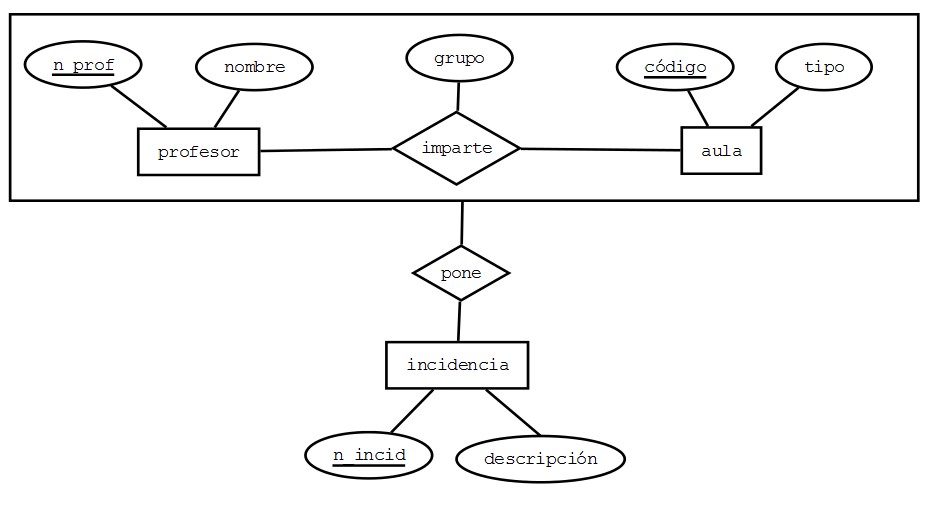


Si consideramos la ***agregación***, tenemos que la relación TRABAJO con las entidades EMPLEADO y PROYECTO se pueden representar como un conjunto de entidades llamadas TRABAJO, que se relaciona con la entidad HERRAMIENTAS mediante la relación USA.



Otro ejemplo de agregación: en el que se representa la situación en la que un profesor, cuando está impartiendo una clase, puede poner una incidencia ocurrida a lo largo de ésta (se fue la luz, falta la configuración de un determinado software, etc.).

# Relación entre impartir clase y poner incidencia



**Ejercicio 1:**

Un instituto de secundaria desea gestionar los partes de los alumnos durante un curso académico.

Los partes son puestos por los profesores a los alumnos por cometer alguna de las incidencias tipificadas con una penalización. Cada incidencia tiene una penalización diferente.

Los alumnos por tener un número de partes, pueden tener una expulsión en la que constarán todos los datos del parte además de la fecha de inicio y de finalización de la expulsión.