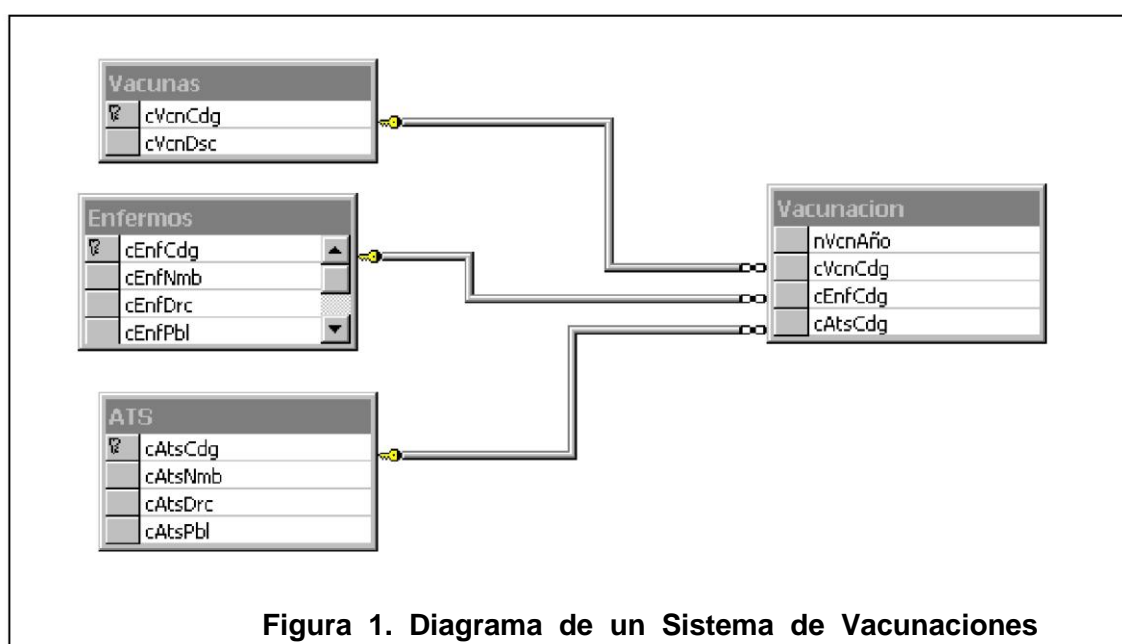


1. Introducción.

Con la aparición de los SGBD, surgen metodologías de **análisis de datos**. La idea central del análisis de datos es la representación de una parcela del mundo real, correspondiente a una “*organización*”, mediante un SGBD dado.

Las metodologías de Análisis de Datos se centran únicamente en los datos que maneja el sistema y en qué relaciones se establecen entre estos datos, dejando de lado las funciones que realiza el sistema y el control del mismo. Se basan en una representación de los objetos del problema y de las relaciones entre los mismos, lo que haremos mediante Diagramas Entidad/Relación (DE/R).



Esta representación deberá contener todos los datos necesarios referentes a la propia *organización*. El primer paso será entonces examinar todos los *objetos* de la *organización* sobre los que queramos mantener información en el sistema, a fin de identificarlos, describirlos y en un momento posterior incluirlos en el Diccionario de Datos del SGBD.

El objetivo final del análisis de datos será la implementación de todos los datos necesarios en el sistema de información en un SGBD comercial.

2. Etapas en el análisis de datos

Bajo la denominación de análisis de datos se encuentran varias actividades que hay que realizar a lo largo del análisis de un sistema:

- **Análisis de la información.** En esta etapa realizaremos un análisis de la realidad existente, con el fin de obtener un modelo conceptual.

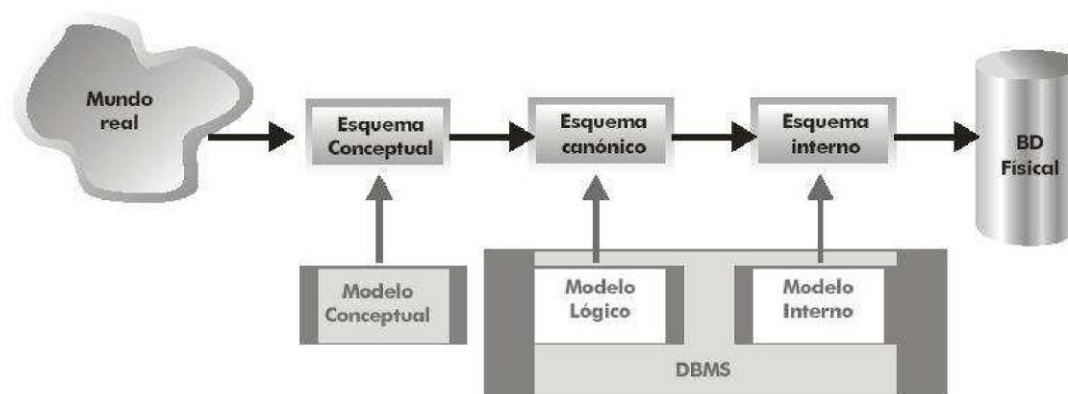
- **Diseño de la base de datos.**

Esta etapa podría dividirse en dos fases:

- **Diseño del modelo conceptual.** El modelo conceptual representa la forma en que percibimos la información existente en el mundo real y que tratamos de plasmar en nuestro sistema de información.
- **Diseño del modelo lógico.** Es una representación de la información orientada a la implementación en el SGBD elegido (FIGURA 1).

- **Implementación** del diseño obtenido sobre el sistema de gestión de base de datos comercial que utilice la empresa.

Sea cual sea el método empleado, se tratará de pasar del mundo real (*modelo externo*) existente en la organización, que puede estar o no informatizado, a un modelo que podríamos llamar interno, soportado por un SGBD. Gráficamente las etapas a seguir se pueden ver en la siguiente figura:



Los modelos de datos lo que intentan reproducir es una información real que deseamos almacenar en un sistema informático.

Se denomina esquema a una descripción específica en términos de un modelo de datos. El conjunto de datos representados por el esquema forma la base de datos.

Son muchos los objetivos que se le piden a un buen diseño de bases de datos, entre ellos:

- Flexibilidad en cuanto a modificaciones y ampliaciones en caso de crecimiento o cambios en el sistema de información.
- Accesibilidad, que facilite el acceso a los datos.
- Comprensibilidad, que el diseño sea fácil de entender.
- Almacenamiento óptimo.
- Rendimiento, máximo número de transacciones por unidad de tiempo y mínimo tiempo de proceso.
- Seguridad de datos.

Como actividades fundamentales del diseño de bases de datos nos centraremos en los modelos conceptual y lógico, entre otros motivos por ser independientes de cualquier sistema de bases de datos comercial concreto.

3. El modelo conceptual

La primera fase del análisis de datos será realizar un estudio de la *realidad* o *modelo externo* existente en la organización, con el fin de obtener un primer diseño llamado modelo conceptual de datos.

El método más utilizado en el modelado conceptual es el modelo **entidad-relación** (E-R, *entity-relationship*, quizá una traducción más acertada hubiera sido entidad-interrelación, para no confundirlo con el término “relación” del modelo lógico relacional, pero desgraciadamente la primera traducción está muy difundida y es perfectamente aceptada en el contexto que nos ocupa). Nosotros estudiaremos el modelo **E-R** propuesto por **Chen** en 1976.

3.1 Definiciones previas

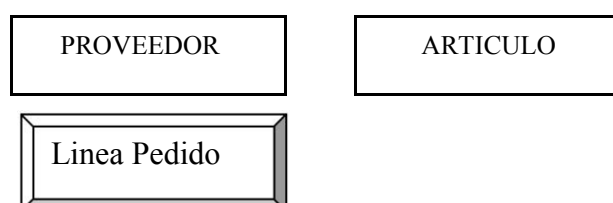
Antes de profundizar en modelado conceptual, veremos algunas definiciones de diversos conceptos que posteriormente utilizaremos.

- **Entidad:** Cualquier objeto relativo a la organización acerca del cual queremos guardar información; denotan personas, lugares, cosas o eventos de interés. Por ejemplo: Proveedor, pedido, artículo.

Hay que tener en cuenta que cada entidad debe tener un nombre lo más significativo posible. Están formadas por ocurrencias o registros. Existen dos tipos de entidades :

- a) **Entidades fuertes o regulares** en las que cada ocurrencia es identificable por sí misma, como por ejemplo “pedido”, identificable por su clave “número de pedido” (clave primaria o principal).
- b) **Entidades débiles** en las que sus ocurrencias son identificables solamente por el hecho de estar asociadas a otras entidades de las que dependen. La mayoría de las veces no se puede establecer la clave con sus propios atributos, necesitando el atributo de clave de la entidad regular de la que depende. Por ejemplo: la entidad “línea de pedido” quedará identificada por el “número de pedido” más el atributo propio “número de línea”.

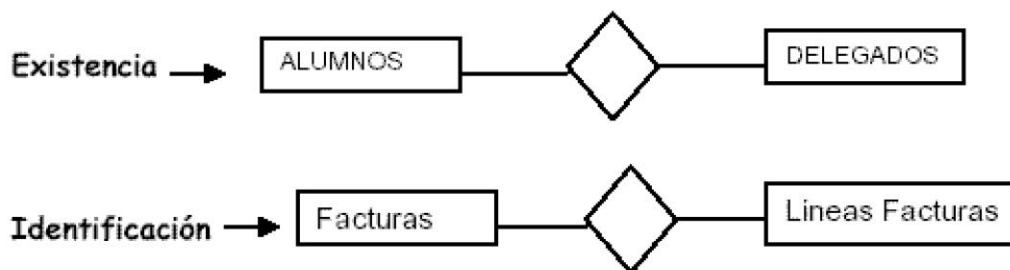
La representación gráfica de una **entidad fuerte** se hace por medio de un rectángulo con el nombre de la entidad en su interior. Si se trata de una **entidad débil** se la representa por un doble rectángulo.



- **Relación** : Describe la correspondencia existente entre dos o más entidades o incluso entre la misma entidad. Se podría utilizar el término asociación o interrelación (*Relationship*) para no confundirlo con relación (*Relation*) del modelo relacional. Por ejemplo: La asociación “ *suministra* ” asociará a la entidad proveedor con la entidad artículo. Se representa gráficamente mediante un rombo donde se escribe el nombre de la asociación y una línea que conecta las entidades relacionadas.



La relación entre una entidad fuerte y una débil puede ser de dos tipos : de **existencia** y de **identificación**. Es de existencia cuando si desaparece la entidad fuerte desaparece la entidad débil (Profesores, Tutores) ;y es de identificación cuando, además de cumplirse la condición anterior, la entidad débil necesita de atributos de la entidad regular para identificarse (Clientes, Facturas Clientes).



Facturas

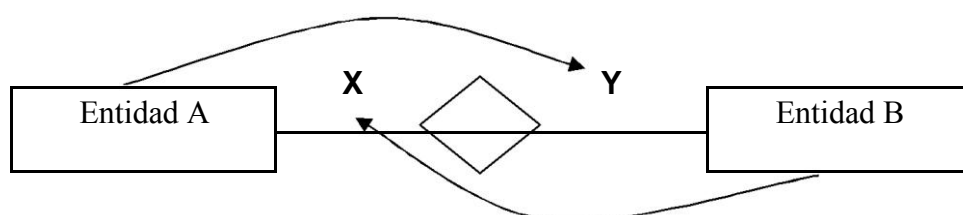
NºFAC	FECHA
1	1-1-98
2	3-1-98
3	8-1-98

Lineas de Facturas

REFER	NOMBRE	PVP	CANT	Nº FACT
001	AGUA	100	04	1
002	CAFÉ	300	02	5
003	JUDIA	200	06	2

Sin el atributo de la entidad fuerte, no podemos establecer la Relación entre las entidades.

- **Grado de asociación o de relación** : Definido por un par de números enteros (min : max) que representan el número mínimo y máximo que una **isma** ocurrencia de una entidad puede estar relacionada con ocurrencias de otra entidad. En la mayoría de los casos este número mínimo y máximo coinciden; en estos casos ,la forma de representarlo e interpretarlo es la siguiente :

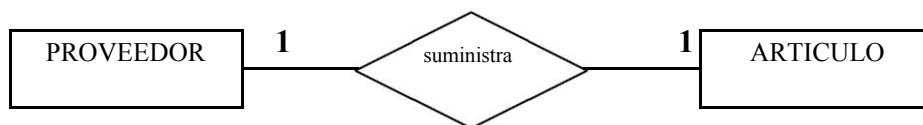


Interpretación : **Cada ocurrencia de la Entidad A se relaciona o asocia con Y ocurrencias de la Entidad B y, a su vez, cada ocurrencia de la Entidad B se relaciona o asocia con X ocurrencias de la Entidad A.**

Genéricamente puede ser de tres tipos:

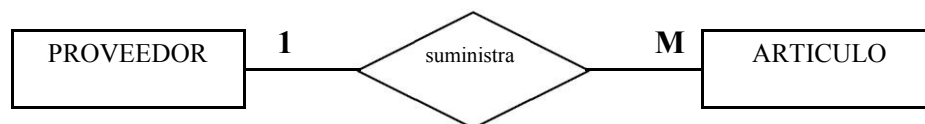
- **De uno a uno (1:1)**, dadas dos entidades A y B, cada ocurrencia de la entidad A está asociada con una ocurrencia de la entidad B y viceversa.

Por ejemplo, para el caso hipotético en el que cada proveedor sólo suministra un artículo y cada artículo sólo es suministrado por un proveedor. Su representación gráfica consistiría en la siguiente:



- **De uno a muchos (1:M)**, cada ocurrencia de la entidad A está asociada con una o muchas ocurrencias de la entidad B y cada ocurrencia de la entidad B está asociada con sólo una ocurrencia de la entidad A.

Por ejemplo, para el caso en el que cada proveedor puede suministrar varios artículos y cada artículo sólo es suministrado por un proveedor. Su representación gráfica consistiría en la siguiente:

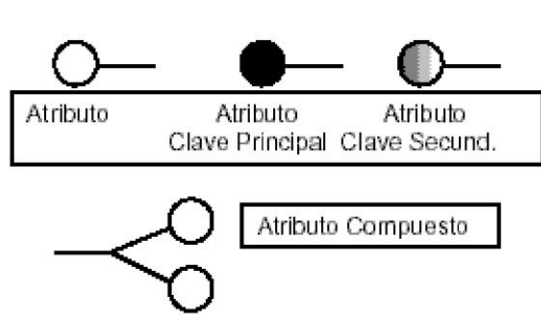


- **De muchos a muchos (M:N)**, cada ocurrencia de la entidad A está asociada con una o muchas ocurrencias de la entidad B y viceversa.

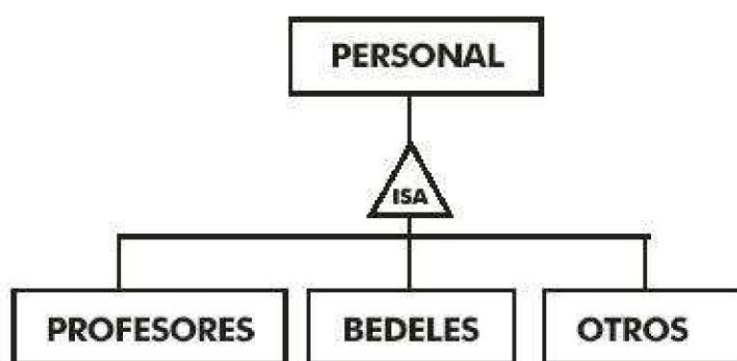
Por ejemplo, para el caso en el que cada proveedor puede suministrar varios artículos y cada artículo puede ser suministrado por varios proveedores. Su representación gráfica consistiría en la siguiente:



- **Atributo de entidad o asociación:** Son aquellos detalles o características que describen, califican, clasifican, cuantifican o expresan el estado de una entidad o de una relación. Una entidad está caracterizada por tener la misma istal de atributos en todas sus ocurrencias. Cada atributo toma valores de un dominio definido y se representan por un círculo y conectado a la entidad o asociación. (El resto de las características se pueden ver en la página 2 del Tema 3).



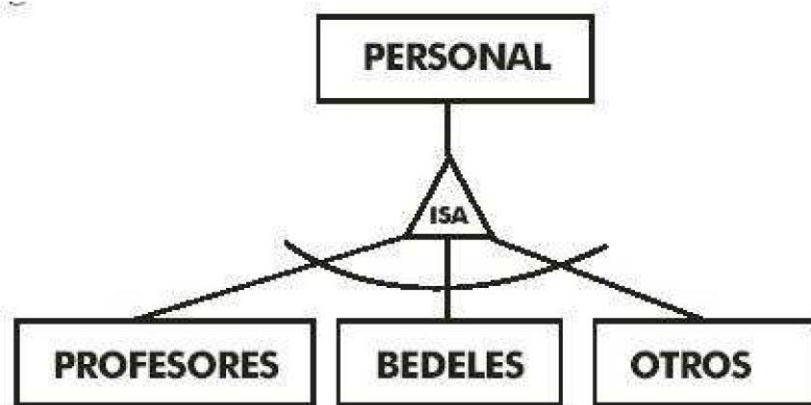
- **Relaciones IS A (Es un):** Son relaciones de tipo **is a** (es un) aquellas en las que una entidad se descompone en entidades especializadas. Hay dos tipos:
 - **Especializaciones:** consisten en que una entidad se divide en entidades más concretas. La entidad general comparte con las especializadas sus atributos. Se observa una especialización cuando hay ejemplares para los que no tienen sentido algunos de los atributos, mientras que para otros sí.
 - **Generalización:** si se agrupan varias entidades en una o más entidades generales. Se observa una generalización si en varias entidades se observan atributos iguales, lo que significa que hay una entidad superior que posee esos atributos. En cualquier caso la representación en el modelo es la misma, se representan con un triángulo con el texto **IS A**.



En estas relaciones se habla también de herencia, ya que tanto los profesores como los bedeles como los otros, heredan atributos de la entidad personal (se habla de la superentidad personal y de la subentidad profesores).

Se puede colocar un círculo (como el del número cero) en lado de la superentidad para indicar que es opcional la especialización, de otro modo se tomará como obligatoria (el personal tiene que ser alguna de esas tres cosas).

Se puede indicar también exclusividad. Esto ocurre cuando entre varias líneas hacia una relación, las entidades sólo pueden tomar una. Se representa con un ángulo en el diagrama:



En el diagrama el ángulo indica que el personal sólo puede ser o profesor o bedel u otros. No puede ser dos cosas a la vez

3.2 Obtención del modelo conceptual.

Aunque el modelo E-R es único se pueden seguir diferentes pasos y enfoques al llevarlo a la práctica. A continuación se proponen los pasos a seguir:

- a. Identificación de entidades
- b. Identificación de claves principales
- c. Identificación de asociaciones indicando grado y clave principal
- d. Representación del modelo E-R
- e. Identificación de atributos de entidades y asociaciones
- f. Verificaciones

a. *Identificación de entidades.*

Para identificar las entidades, lo primero será conocer el funcionamiento del área de la organización o empresa en cuestión, pensaremos en:

- Objetos reales (artículos, edificios, máquinas...).
- Documentos (factura, albarán...)
- Personas (proveedores, clientes, empleados, jefes...).
- Actividades (expedientes, pedidos...).
- Objetos abstractos (tiempos, cursos...).

Se emplean aquí fundamentalmente las técnicas de entrevistas, encuestas, reuniones con los usuarios, observaciones directas, estudio de documentos, etc.

Es muy importante elegir un nombre adecuado para la entidad, lo ideal sería extraerlo de la propia organización y que indique nítidamente lo que la entidad representa.

b. Identificación de las claves principales

Determinadas las entidades, trataremos de encontrar algún atributo o grupo de ellos, que pueda identificar unívocamente todas las ocurrencias de esa entidad. Si para una entidad concreta existen varios, elegiremos uno de ellos.

c. Identificación de las asociaciones entre entidades

Tratamos ahora de relacionar todas las entidades obtenidas, una a una, con todas las demás, viendo si cada asociación tiene sentido o no en la organización. Como en el caso de las entidades, es muy importante elegir un nombre para la asociación.

Para obtener el grado de una asociación entre dos entidades nos preguntaremos:

- c.1) Se asocia cada ocurrencia de la primera entidad con una y sólo una de la segunda.
- c.2) Se asocia cada ocurrencia de la segunda entidad con una y sólo una de la primera.

Si la respuesta a las dos preguntas es afirmativa, el grado de la asociación es 1-1. Si una de las respuestas fuese afirmativa y la otra no, el grado de la asociación es 1-M. Por último, si las dos respuestas son negativas el grado será M-N.

Hay que tener en cuenta la evolución en el tiempo de la asociación, es decir, además de ver cómo se asocian las entidades actualmente, habrá que pensar cómo se asociarán en un futuro.

Por ejemplo, en una asociación como “suministra”, el grado actual puede ser 1-M, un proveedor suministra varios artículos y un artículo sólo puede ser suministrado por un proveedor, pero en el futuro un mismo artículo podría ser suministrado por varios proveedores, siendo entonces el grado de la asociación “suministra” M-N.

d. Representación del modelo E-R.

Representaremos todas las asociaciones y sus entidades gráficamente, indicando los grados y las claves principales de las mismas. Utilizando la simbología vista anteriormente.

e. Identificación de los atributos de cada entidad y asociación

Realizaremos una lista de todos los atributos de cada entidad o asociación que sean significativos para el sistema.

Cada atributo debe tener un nombre y significado único, además debemos tener claro el rango de valores posibles.

f. Verificaciones

Aunque las verificaciones deben hacerse en todos y cada uno de los pasos. Al final revisaremos de nuevo los resultados obtenidos a fin de identificar si existieran

asociaciones redundantes y comprobar si el modelo conceptual obtenido satisface todas las necesidades de información del sistema en cuestión.

3.3 Ejemplo práctico de modelo conceptual.

Vamos a realizar un pequeño ejemplo práctico donde se puedan observar los pasos anteriormente expuestos. Supondremos un sistema básico de solicitud de pedidos. Tras conversaciones con el supuesto “cliente” obtenemos las siguientes conclusiones: “*Se trata de un comercio de material de papelería. Los artículos son suministrados por varios proveedores, de los cuales, se necesita conocer sus datos fiscales. Además, desea tener un registro de todos los artículos que normalmente vende. Se sabe que un artículo puede ser suministrado por varios proveedores y que un proveedor suministra varios artículos. Por ejemplo, un paquete de 500 folios A4, puede ser suministrado por distintos proveedores a distinto precio. Lógicamente, a un proveedor se le podrán realizar varios pedidos y un pedido podrá constar de uno o varios ítems (artículos distintos dentro del mismo pedido). El precio de cada ítem coincide con el del artículo correspondiente el momento de la realización del pedido, pero los precios de los artículos varían con relativa frecuencia. Los distintos proveedores envían sus tarifas cada vez que cambian sus precios y a nuestro cliente le interesaría saber: ¿Qué proveedores suministran un artículo determinado y a qué precio?. También necesita saber: ¿Qué pedidos se le han realizado a un proveedor?*”.

1. Identificación de entidades.

- PROVEEDOR: Entidad regular que representa a las empresas que suministran artículos. Aúna todos los datos necesarios para poder mantener relaciones comerciales con la misma.
- ARTICULO: Entidad regular que representa los distintos artículos que puede requerir la empresa.
- PEDIDO: Entidad regular que representa la solicitud de artículos a un proveedor.
- ITEM: Entidad débil, dependiente de la entidad PEDIDO. Representa los distintos artículos solicitados en un pedido. Si el pedido desaparece, los ítems de dicho pedido deben desaparecer.

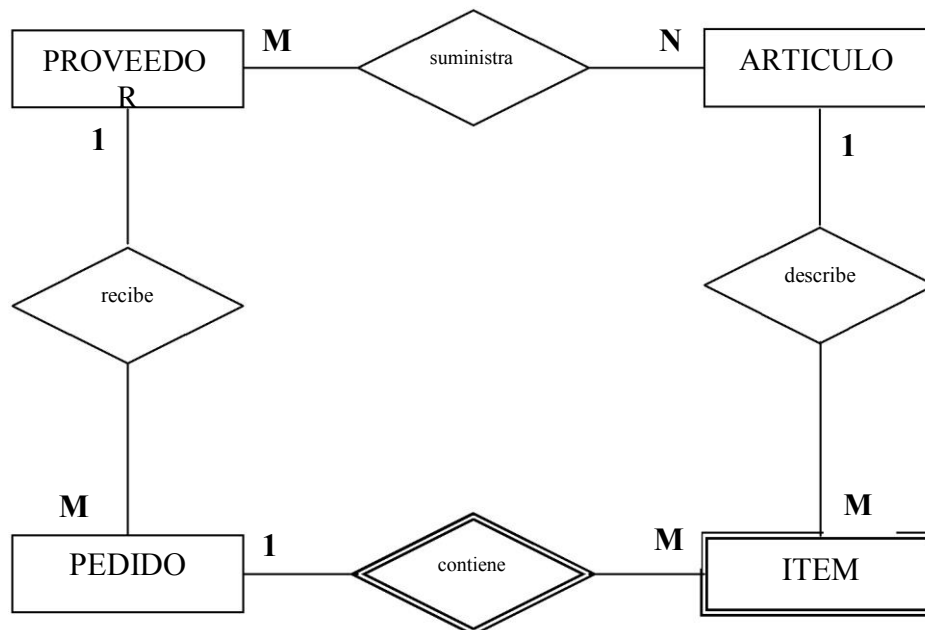
2. Identificación de claves principales.

- De PROVEEDOR la clave es CODPROV
- De ARTICULO la clave es CODART
- De PEDIDO la clave es NUMPED
- De ITEM la clave está formada NUMPED+ NUMITEM ya que existe una relación por Identificación.

3. Identificación de asociaciones indicando grado.

- Un proveedor “**suministra**” artículos. De grado M:N.
- Un proveedor “**recibe**” pedidos. De grado 1:M.
- Un pedido “**posee**” ítems. De grado 1:M.
- Un artículo “**describe**” ítems. De grado 1:M.

4. Representación del modelo E-R.



5. Identificación de atributos de entidades y asociaciones. Además de los atributos que forman la clave primaria o principal de cada entidad hay que indicar el resto de atributos de cada entidad, así como los atributos de las relaciones (en el caso de que existan).

- Entidad PROVEEDOR: Codprov, nombre, nif, dirección, tlf, fax,...
- Entidad ARTICULO: Codart, descripción.
- Entidad PEDIDO: Numped, codprov, fecha.
- Entidad ITEM: Numped, Numitem, codart, cantidad, precioitem.
- Asociación “ **suministra**”: precio. Ya que el precio depende del artículo y del proveedor que lo suministra, es decir, en realidad depende de la asociación.

7. Verificaciones. Al ser un ejemplo pequeño las verificaciones son mínimas.