

2. Modelado de datos

2.1 Modelos de datos

2.1.1 Definición

Un modelo es un conjunto de herramientas conceptuales para describir datos, sus relaciones, su significado y sus restricciones de consistencia.

2.1.2 Características

- Es el proceso de analizar los aspectos de interés para una organización y la relación que tienen unos con otros.
- Resulta en el descubrimiento y documentación de los recursos de datos del negocio.
- El modelado hace la pregunta " Qué ? " en lugar de " Cómo ? ", ésta última orientada al procesamiento de los datos.
- Es una tarea difícil, bastante difícil, pero es una actividad necesaria cuya habilidad solo se adquiere con la experiencia.

2.1.3 Metas y beneficios

- Registrar los requerimientos de datos de un proceso de negocio.
- Dicho proceso puede ser demasiado complejo y se tendrá que crear un "enterprise data model", el cual deberá estar constituido de líneas individuales.
- Permite observar:
 - Patrones de datos
 - Usos potenciales de los datos

2.1.4 Tipos de modelado de datos

Basicamente son 3:

- Conceptual: muy general y abstracto, visión general del negocio/institución.
- Lógico: versión completa que incluye todos los detalles acerca de los datos.
- Físico: esquema que se implementara en un manejador de bases de datos (DBMS).

En las siguientes secciones se analizarán los aspectos relacionados con el modelado conceptual, más adelante y teniendo ya un modelo lógico se procederá a estudiar la representación física del mismo.

2.1.5 Modelado de Datos Conceptual

2.1.5.1 Conceptos básicos

Algunos aspectos a considerar al momento de realizar el modelado/análisis

- No pensar físicamente, pensar conceptualmente
- No pensar en procesos, pensar en estructura
- No pensar en navegación, pensar en términos de relaciones

2.1.5.2 Modelos conceptuales

Existen distintos tipos de modelos conceptuales:

Basados en registros

- Jerárquico: datos en registros, relacionados con apuntadores y organizados como colecciones de árboles
- Redes: datos en registros relacionados por apuntadores y organizados en gráficas arbitrarias
- Relacional: datos en tablas relacionados por el contenido de ciertas columnas

Basados en objetos

- Orientado a objetos: datos como instancias de objetos (incluyendo sus métodos)
- Entidad-relación: datos organizados en conjuntos interrelacionados de objetos (entidades) con atributos asociados



2.2 Modelo Entidad-Relación

2.2.1 Definición

Generalmente todo modelo tiene una representación gráfica, para el caso de datos el modelo más popular es el modelo entidad-relación o digrama E/R.

Se denomina así debido a que precisamente permite representar relaciones entre entidades (objetivo del modelado de datos).

El modelo debe estar compuesto por:

- Entidades
- Atributos
- Relaciones
- Cardinalidad
- Llaves

2.2.2 Conjuntos de entidades y atributos

- Entidades: todo lo que existe y es capaz de ser descrito (sustantivo).
- Atributos: es una característica (adjetivo) de una entidad que puede hacer 1 de tres cosas:
 - Identificar
 - Relacionar
 - Describir

321-12-3123	Jones	Main	Harrison
019-28-3746	Smith	North	Rye
677-89-9011	Hayes	Main	Harrison
555-55-5555	Jackson	Dupont	Woodside
244-66-8800	Curry	North	Rye
963-96-3963	Williams	Nassau	Princeton
335-57-7991	Adams	Spring	Pittsfield

customer

L-17	1000
L-23	2000
L-15	1500
L-14	1500
L-19	500
L-11	900
L-16	1300

loan

Ejemplos de entidades con sus atributos

En el diseño se pueden considerar 3 categorías de atributos

- **Simple o compuestos:** ya sea que el atributo sea un todo o bien este compuesto
 - Color es simple, toma valores rojo, azul, etc
 - Nombre es compuesto, contiene nombre de pila, apellido materno, apellido materno
- **Con valores simples o multivaluados:** en base a si consisten de un solo valor o un conjunto de valores.
 - Telefono o Teléfonos
- **Derivados:** que se pueden calcular en base a otros atributos
 - El promedio de préstamos se puede derivar si tenemos los valores de cada préstamo realizado a un persona

NOTA: en la práctica es mejor considerar "siempre" a todos los atributos como simples y con valores simples

2.2.3 Llaves

- Super llave: conjunto de uno o más atributos que "juntos" identifican de manera única a una entidad
- Llave candidata: es una super llave mínima

- Llave primaria: la seleccionada para identificar a los elementos de un conjunto de entidades.

Ejemplo:

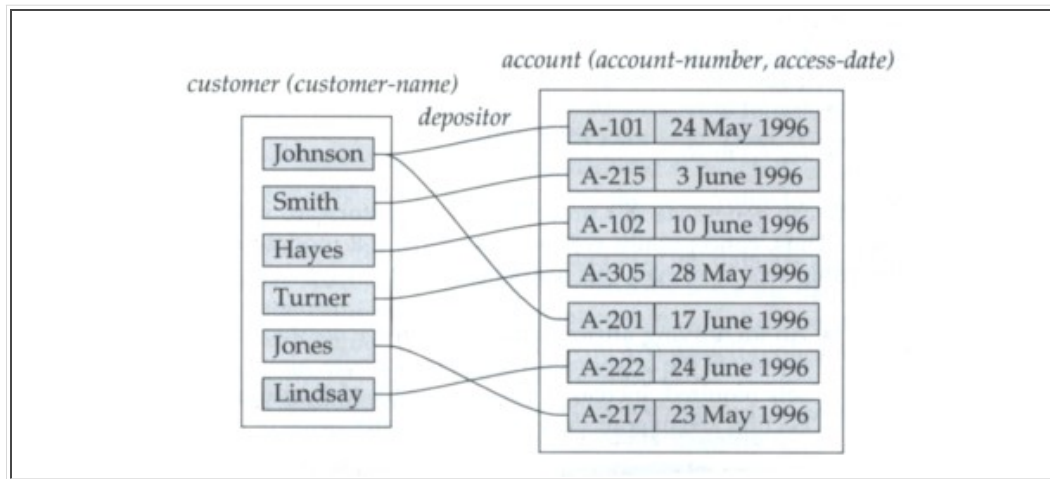
Teniendo los atributos de la entidad "persona"

Nombre	Dirección	Teléfono	CURP
--------	-----------	----------	------

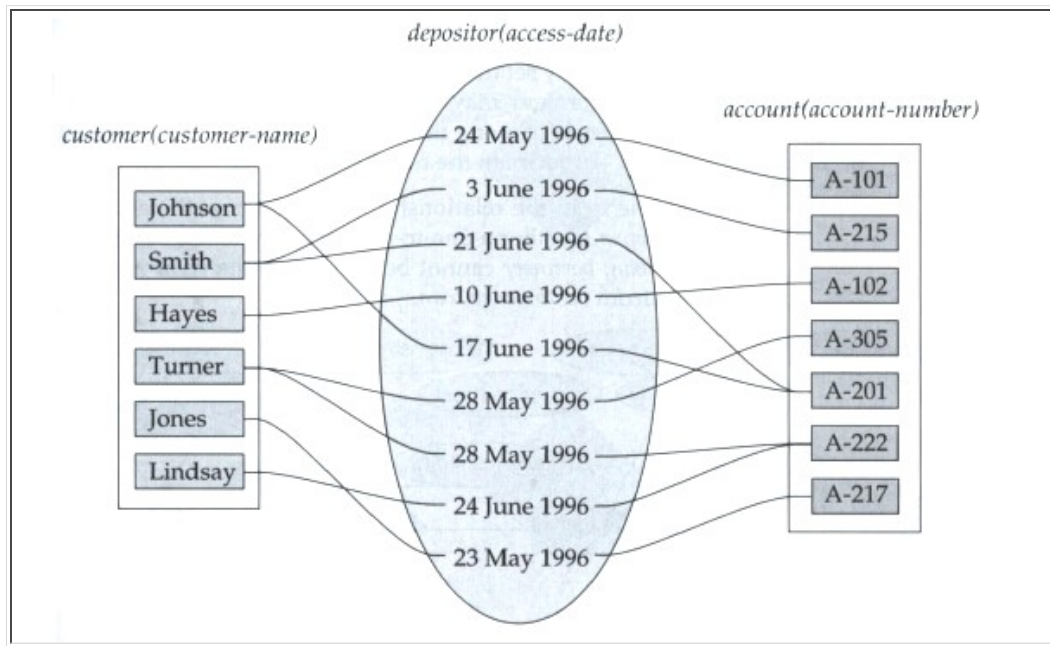
- Las superllaves serían:
 - Nombre y Dirección
 - Nombre y CURP
 - CURP
- Las llaves candidatas serían
 - Nombre y Dirección
 - CURP
- La llave primaria sería
 - CURP

2.2.4 Conjuntos de relaciones

- Relaciones: la conexión que existe entre 2 entidades (verbo).

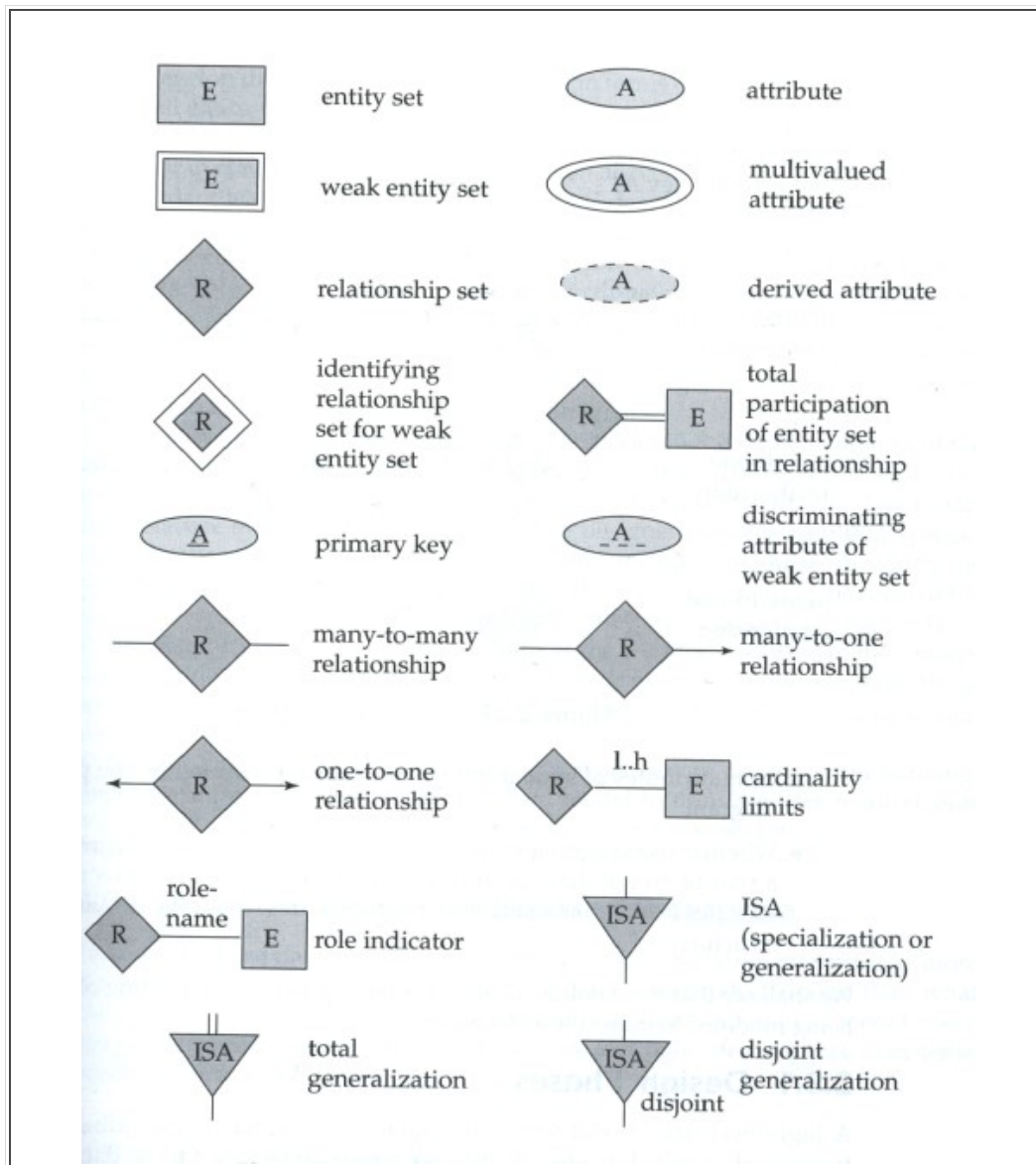


Relación entre 2 entidades



Relación entre 2 entidades incluyendo un atributo en la relación

2.2.5 Diagrama E-R



Notación empleada para elaborar modelos E-R

2.2.5.1 Diagramas E-R de relaciones entre entidades

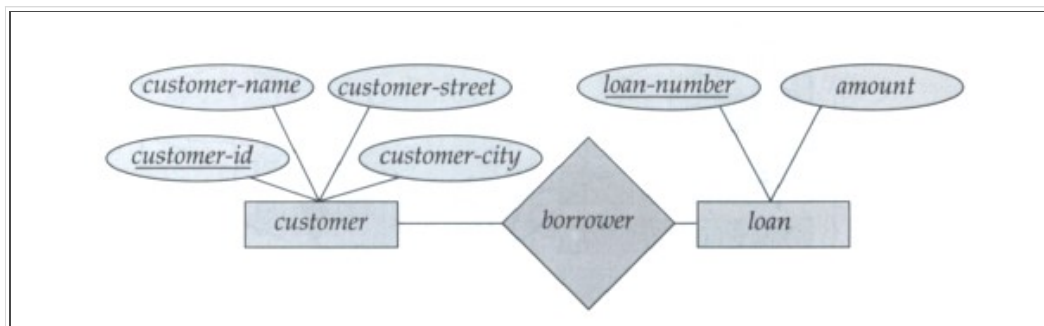


Diagrama E-R mostrando una relación entre 2 entidades

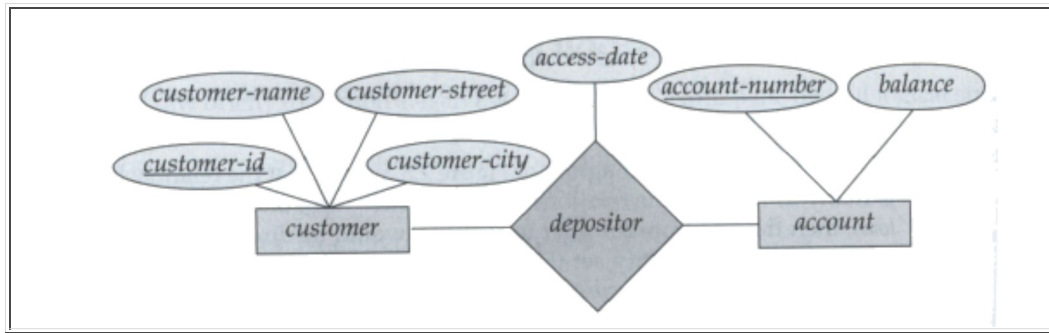


Diagrama E-R mostrando una relación entre 2 entidades, con atributo en la relación

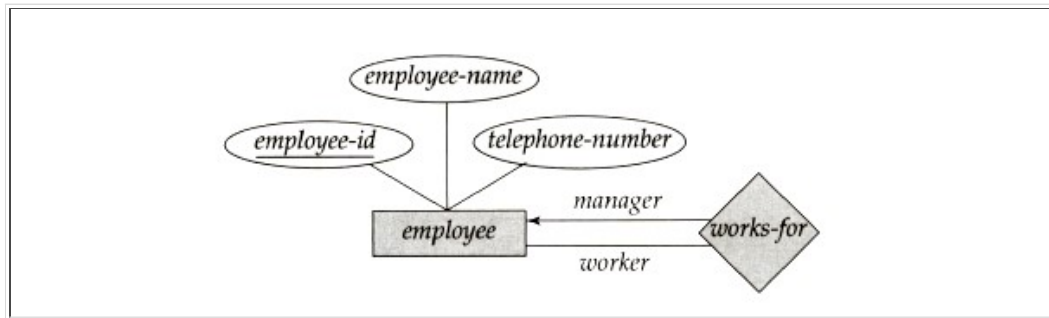
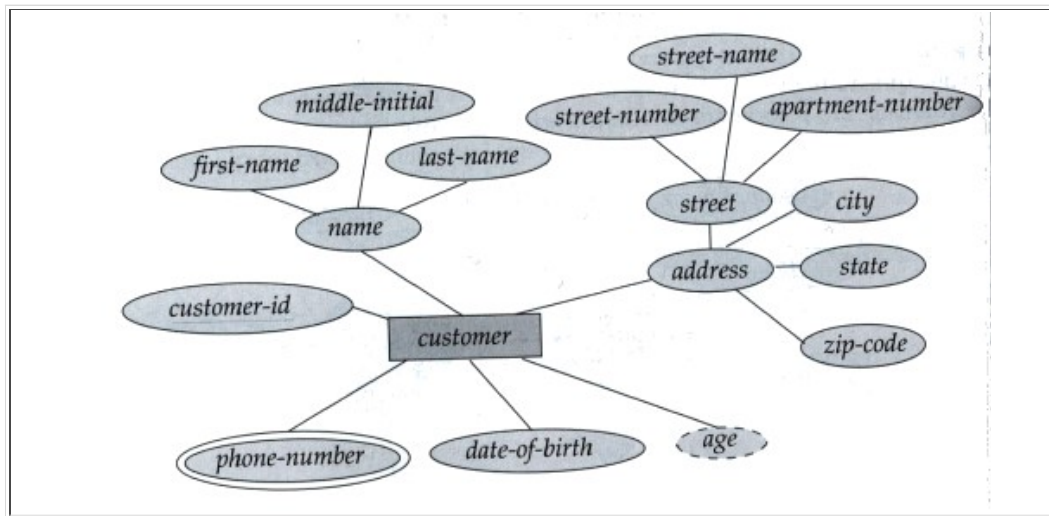


Diagrama E-R mostrando una relación entre una misma entidad (útiles para elaborar jerarquías)

2.2.5.2 Categorías de atributos



Ejemplos de atributos derivados, compuestos y multivaluados

NOTA: como se mencionó anteriormente **NO** es lo mejor el emplear estos atributos

2.2.5.3 Entidades débiles

- Una entidad débil es aquella que no posee una llave primaria
- Para existir dependen de una relación con una entidad fuerte
- Pueden contener algún atributo "discriminante" que podría considerarse como aquel que lo distingue pero no de manera única, de ahí que no se considere como llave

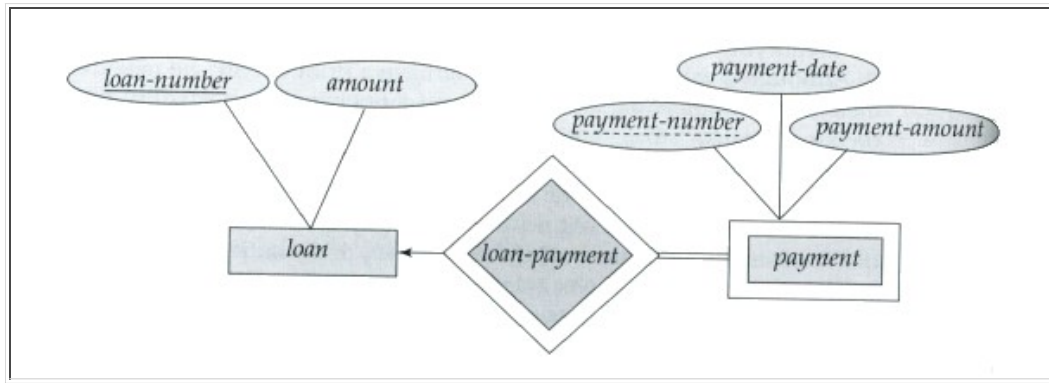


Diagrama E-R mostrando una relación entre 2 entidades, una de ellas fuerte y otra débil

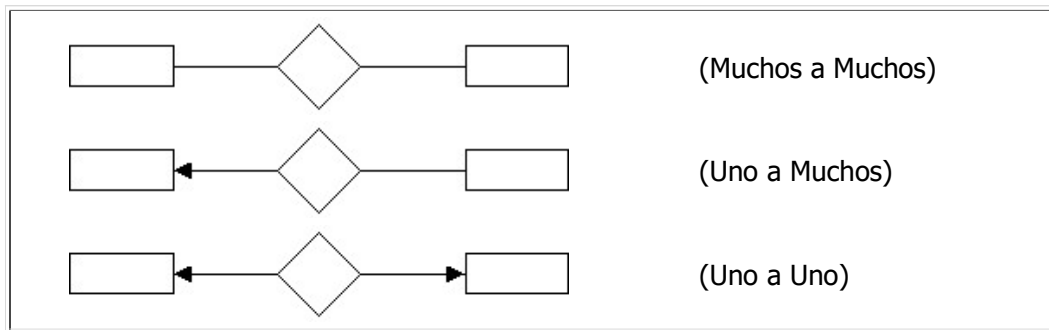
2.2.5.4 Guías de nombramiento

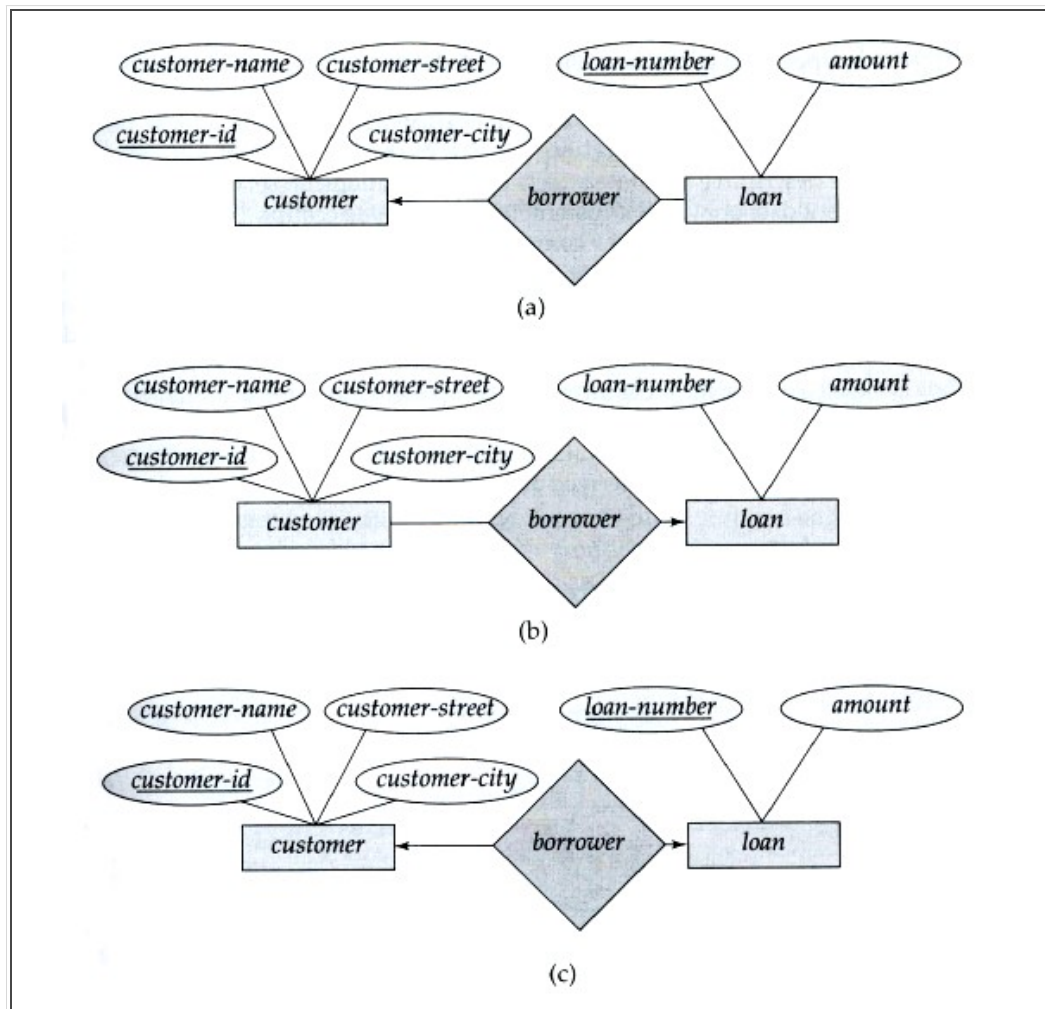
Es importante mantener guías o reglas para poder tener una documentación uniforme y consistente de todos los datos.

- Entidades: una sola palabra (en singular) y con mayúsculas
- Atributos:
 - FirstName
 - first_name
 - de relacion: VendorID, ProductName
- Valores: definir que valores son válidos (NULL no es un valor)

2.2.5.5 Cardinalidades

En base al número de instancias involucradas en cada relación, éstas presentan un cardinalidad, que puede ser:

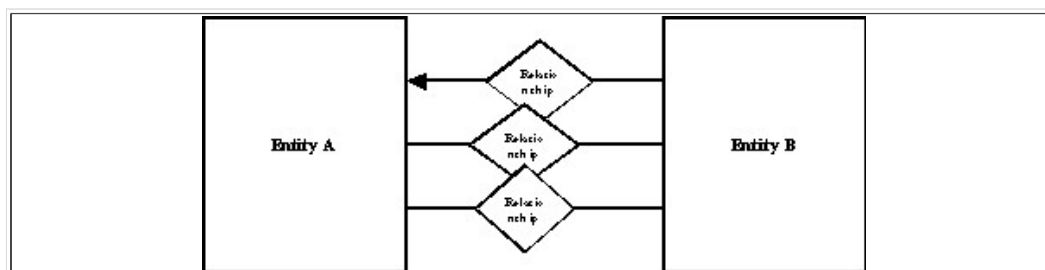




Relaciones (a)uno-muchos, (b)muchos-uno,(c) uno-uno

2.2.5.6 Múltiples relaciones entre 2 entidades

Es posible mantener muchas relaciones entre las mismas entidades, inclusive con distintas cardinalidades siempre y cuando cada una represente algo totalmente independiente de las otras. No se puede asumir que las relaciones se complementan o ni mucho menos que compartan atributos.

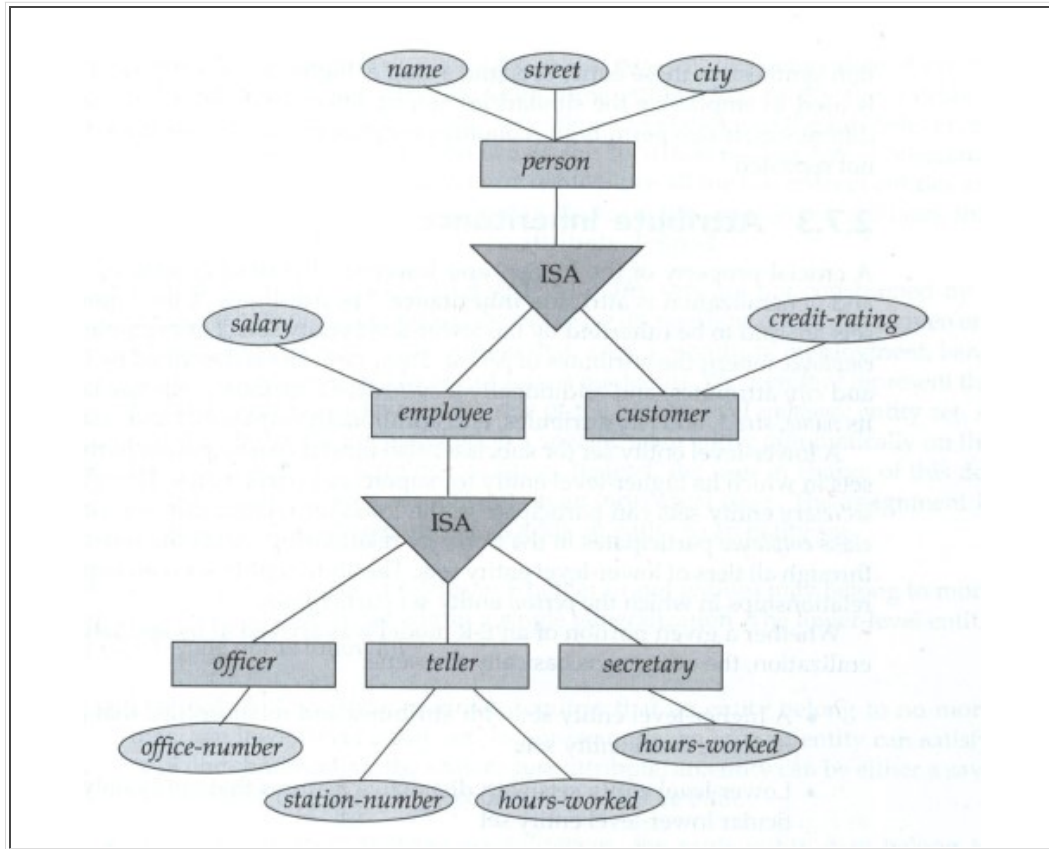


2.2.5.7 Especialización y generalización

Es el principio de "herencia"

Las entidades de bajo nivel heredan todos los atributos de las entidades de mayor nivel

- Si se considera de arriba hacia abajo se considera como especialización
- Si se considera de abajo hacia arriba se considera como generalización



Especialización y generalización

Nota: es importante mencionar que las entidades de menor nivel no poseen una llave primaria, únicamente la entidad de nivel superior es la que tiene entre sus atributos dicha llave y en consecuencia la "hereda" a las entidades especializadas.

Restricciones en las generalizaciones

De pertenencia al nivel más bajo

- **Definido por condición:** alguna condición (inclusive atributo) en el nivel alto define si una entidad puede o no pertenecer al nivel más bajo.
- **Definido por usuario:** dadas ciertas condiciones basadas en el juicio de la experiencia se decide si se puede o no pertenecer a dicho nivel.

De pertenencia entre entidades en el nivel bajo

- **Disjuntas (disjoint):** una entidad no puede pertenecer a 2 conjuntos de entidades de dicho nivel
- **Traslape (overlapping):** una entidad si puede pertenecer a 2 conjuntos de entidades

2.2.5.8 Principios de diseño

Fidelidad: se debe crear siempre un modelo que satisfaga las necesidades del problema, no sirve un modelo correcto si no cumple con la realidad que se pretende representar.

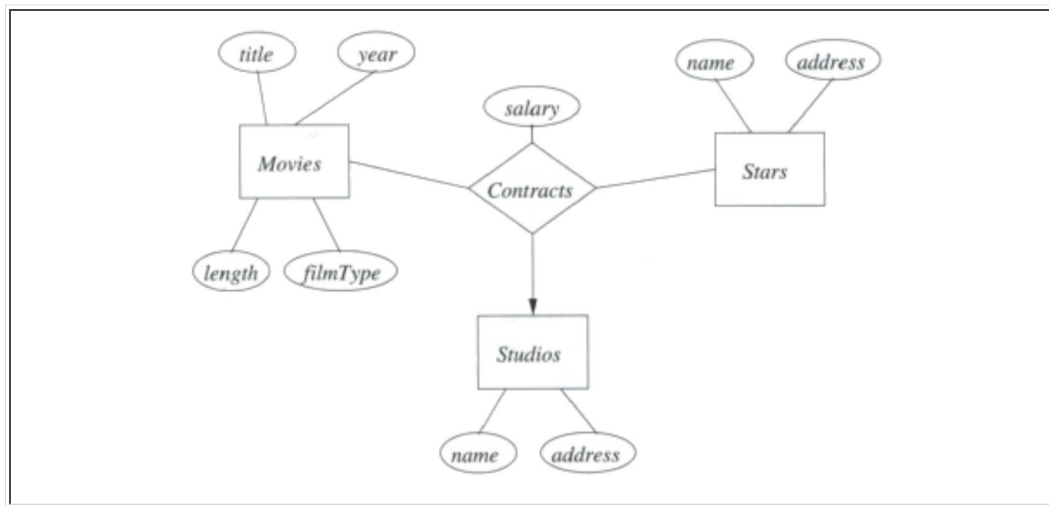
Evitar redundancia: una de las ventajas del diagrama e-r es que nos permite distinguir de una manera fácil y

visual todos los entes y sus relaciones, de manera que es muy fácil identificar si un atributo se esta repitiendo en varias entidades o si una relación es innecesaria.

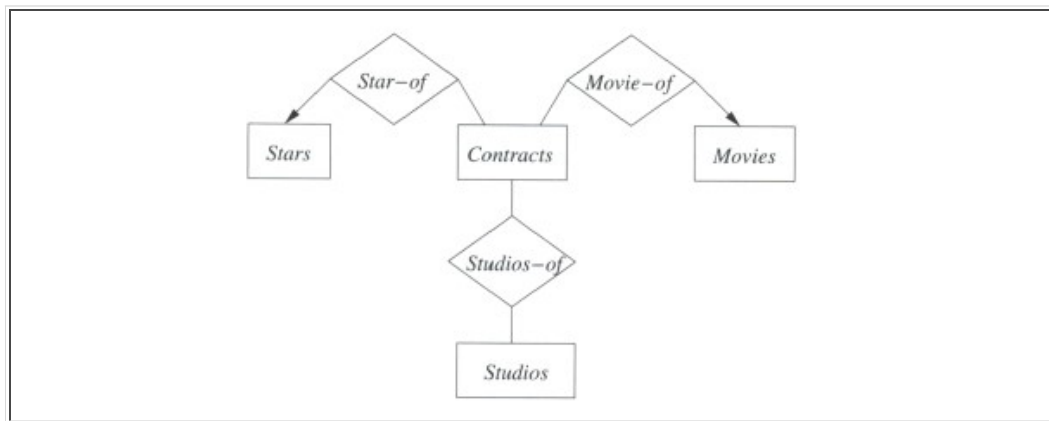
Simplicidad: siempre hay que procurar hacer el modelo tan simple como sea posible (sin olvidar la fidelidad) de manera que sea fácil de entender, fácil de extender y fácil de implementar.

Escoger los elementos correctos: es ocasiones es difícil identificar si una relación, elemento o atributo es correcto, para ello hay que analizar en perspectiva el diagrama y, por ejemplo si se observa una entidad con solo un atributo y que únicamente presenta relaciones de 1, entonces probablemente estamos hablando de un atributo y no de una entidad.

Relaciones n-arias: Aún cuando se pueden presentar casos en los que una relación terciaria o n-aria parezca más conveniente, es mejor siempre pensar en términos de relaciones binarias únicamente. En el peor de los casos de que exista una relación n-aria forzosa, lo que se debe hacer es convertir esa relacion R en entidad E y corregir todas las relaciones que tenía R de manera que ahora esa nueva entidad se relacione con todas las entidades que anteriormente esta.



Relación Ternaria

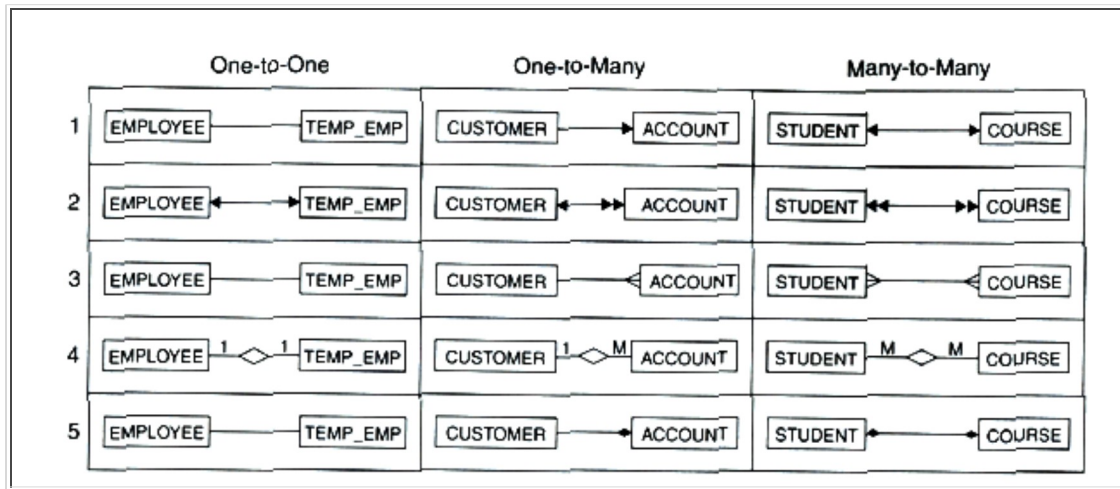


Resultado de la conversión de relación de relación 3-aria a combinación de 2-arias

2.2.5.9 Otras notaciones

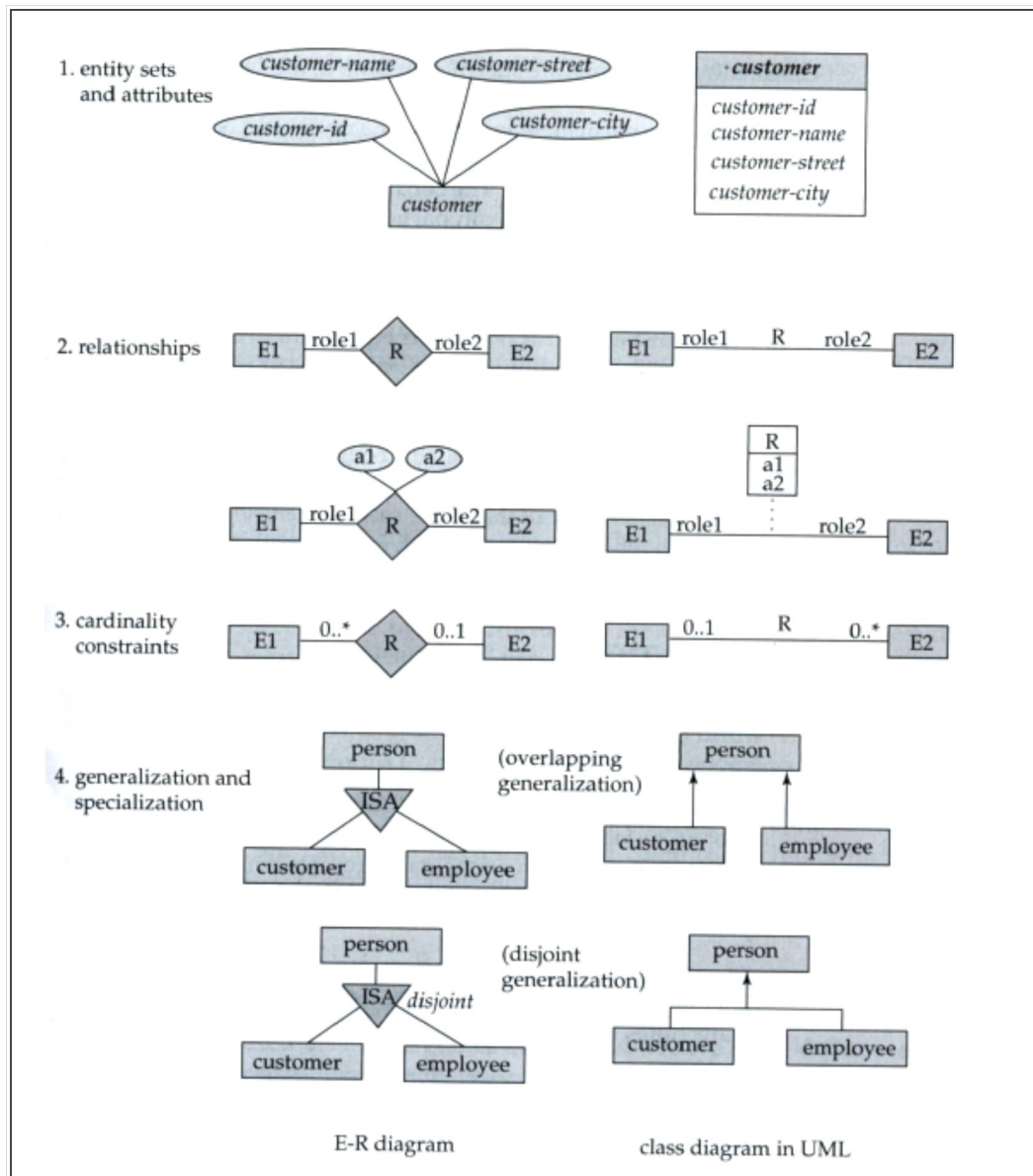
La notación mostrada en las secciones anteriores es solo una de las existentes, aún cuando todas en esencia representen el mismo concepto existen una gran variedad de simbologías y depende de cada persona el escoger

aquella que más le convenga.



Notación E/R (1) Ross, (2) Bachmann, (3) Martin, (4) Chen, (5) Rumbaugh

Por otro lado, Booch con su propuesta de un lenguaje de modelado unificado "UML" (Unified Modeling Language) abarca los aspectos de "relaciones" aplicables no solo al contexto de bases de datos sino al de programación y muchos otros más.



Notación UML para modelos E-R

2.3 Conclusiones:

- El modelado es la actividad más delicada e importante en la realización de una aplicación con base de datos
- Al igual que en el desarrollo de un sistema, toda modificación al esquema de base de datos debe realizarse primero en el modelo conceptual, no en el lógico ni en el físico.
- La habilidad de crear buenos modelos es una cualidad que se adquiere con la experiencia.