



# Bases de datos

UNIDAD 2

INTRODUCCIÓN AL MODELADO DE DATOS



# Bases de datos

## UNIDAD 2

EN ESTA SEGUNDA PARTE VEREMOS EL DISEÑO LÓGICO DE UNA BASE DE DATOS: DER

# Introducción al Modelo Relacional -> DER

---

- Desarrollado Edgar Frank Codd en los años setenta.
- Los datos se estructuran lógicamente en forma de relaciones (tablas).
- Elemento principal: RELACIÓN
  - En MER una relación es una asociación o correspondencia entre entidades (interrelación).
  - En el DER, cada relación se representa mediante una **tabla bidimensional**

“El esquema de una base de datos relacional es un conjunto de relaciones que cumplen unas determinadas propiedades.”

# Diseño lógico

---

Proceso de construcción de un esquema de la información relacionada con el sistema objetivo a partir del esquema conceptual, en base a un modelo lógico de base de datos e **independiente del SGBD concreto** que se vaya a utilizar.

- El diseño conceptual trataba de obtener un esquema conceptual completo y expresivo.
- El diseño lógico perseguirá obtener una representación que estructure lo anterior con **estructuras lógicas** que capturen *eficientemente* los datos y sus restricciones.
- Tanto el diseño conceptual como el diseño lógico son **procesos iterativos**, se van refinando continuamente a partir de la idea inicial.
- Constituyen un proceso de **aprendizaje** en el que el diseñador va comprendiendo las necesidades y el significado de los datos que maneja.
- Si la base de datos no es una **representación fiel** de la necesidad a cubrir, será difícil definir todas las vistas de los usuarios o mantener la *integridad* de la misma.
- **Objetivo** de esta etapa: obtener el **esquema lógico**, que estará formado por el **conjunto de relaciones** de la base de datos a partir de la especificación realizada en la etapa del diseño conceptual.

# Terminología del modelo relacional

---

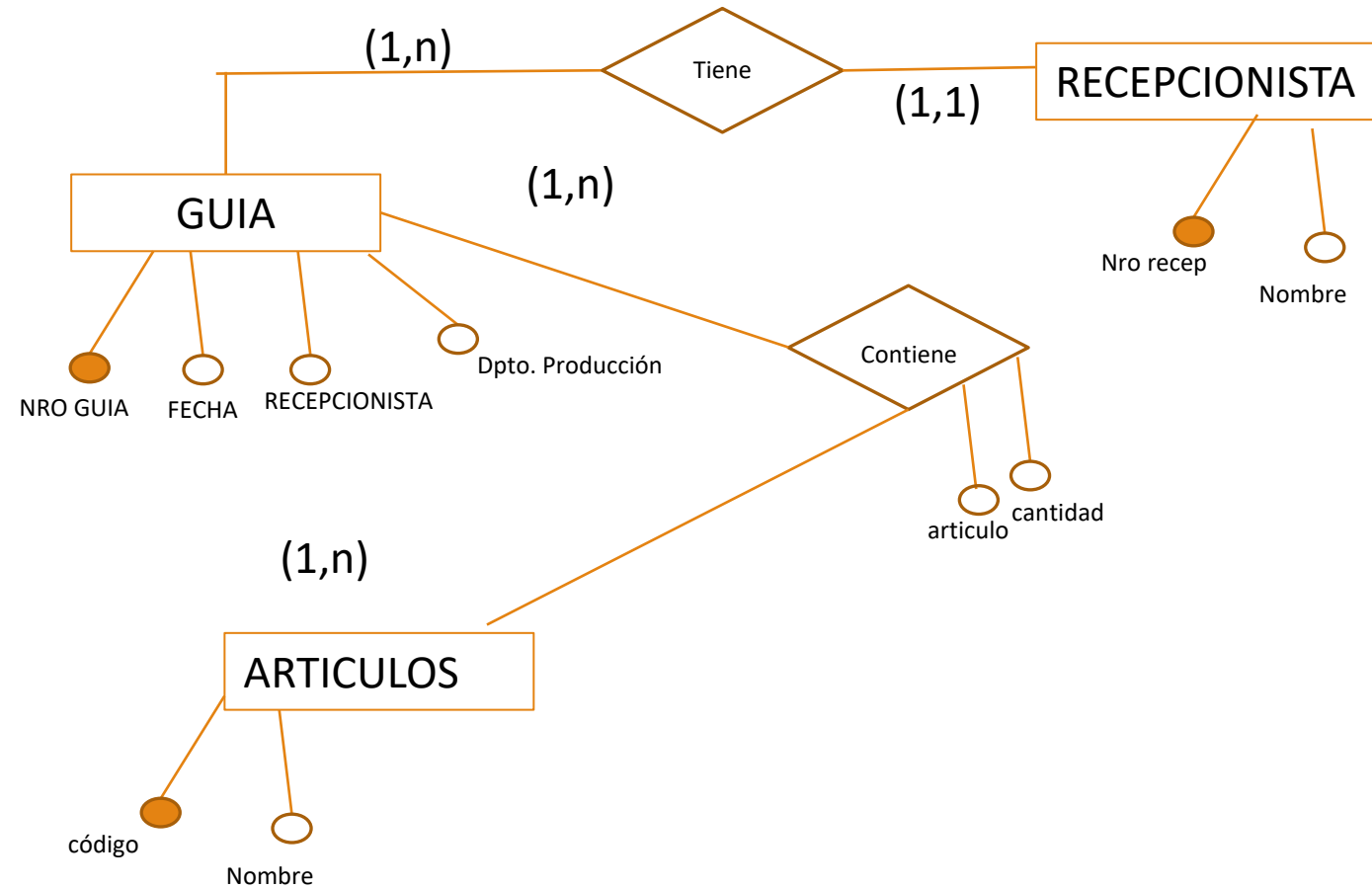
- **Tupla** o registro. Es cada una de las filas de la tabla.
- **Atributo** o campo. Es cada una de las columnas de la tabla. Todos los registros o tuplas tienen igual número de atributos o campos.
- **Grado**. Es el número de atributos de la tabla. (cantidad de columnas)
- **Dominio de un atributo**. Es el conjunto de valores que puede tomar dicho atributo.
- **Clave candidata**. Conjunto mínimo de atributos que identifican de forma unívoca cada tupla de una relación.
- **Clave principal o primaria**. Clave candidata elegida para identificar las tuplas de una relación.
- **Clave foránea o externa**. Conjunto de atributos en una relación que es una clave primaria en otra (o incluso en la misma).

## Reglas de integridad

---

- **De entidad:** PK no valores nulos
- **De clave:** la PK es única para cada tupla,  
Pueden existir PK compuestas
- **Referencial:** la FK debe existir en la tabla  
que corresponde a la PK

# MER



# DER

---

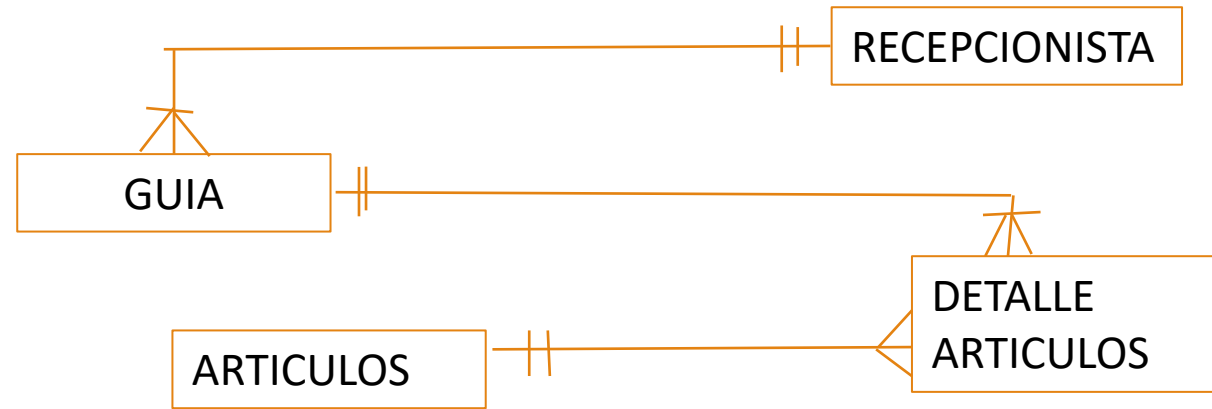
## Entidades: Tablas

Relaciones

Claves Primarias

Claves Foráneas

Cardinalidad





# MER -> DER

---

Entidad	Fuerte	Tabla
	Débil	Tabla (con FK)
Atributo	Simple	Columna - campo
	Compuesto	Columna atributo simple
	Multivaluados	Tabla
	Derivados	No almacenar (inconsistencias)
Clave primaria		Clave primaria
Interrelaciones	N:M	Tabla Intermedia
	1:N	Propagación de claves o Tabla Intermedia
	1:1	Tabla o propagación de clave (pocos casos)
Especialización/Gene- ralización		No hay definición: varias Tablas, varias relaciones

# NORMALIZACIÓN DE BASE DE DATOS

---

El proceso de normalización de **bases de datos** consiste en designar y aplicar una serie de reglas a las relaciones obtenidas tras el paso del **modelo entidad relación** al **diagrama entidad relación** .

Las bases de datos relacionales se normalizan para:

- Evitar la **redundancia** de los datos.
- Disminuir problemas de actualización de los datos en las tablas.
- Proteger la **integridad** de los datos.
- Producir un diseño que sea una buena representación del mundo real

## 4 ETAPAS DE NORMALIZACIÓN

# NORMALIZACIÓN DE BASE DE DATOS

[illegible]

#### 2.2.1.1 Primera forma normal

Nro_GI	fecha	nom_recepcionista	cod_art	cant_art	nom_art
GI/000010	01/06/2010	Marcos Gonzales	AR410	12	pantalón
GI/000011	02/06/2010	José Julón	AR411	36	polos
GI/000012	03/06/2010	Samuel Jiménez	AR412	24	Medias
GI/000013	04/06/2010	Manuel Pérez	AR413	36	polos

#### 2.2.1.2 Segunda forma normal

Guia_remision			Articulo		
Nro_Gl	fecha	nom_recepcionista	cod_art	cant_art	nom_art
GI/000010	01/06/2010	Marcos Gonzales	AR410	12	pantalón
GI/000011	02/06/2010	José Julón	AR411	36	polos
GI/000012	03/06/2010	Samuel Jiménez	AR412	24	Medias
GI/000013	04/06/2010	Manuel Pérez	AR413	36	polos

### 2.2.1.3 Tercera forma normal

Guia_interna		Recepcionista		
Nro_Gl	fecha	cod_recepcionista	nom_recepcionista	cant_art
GI/000010	01/06/2010	R0001	Marcos Gonzales	12
GI/000011	02/06/2010	R0002	José Julón	36
GI/000012	03/06/2010	R0003	Samuel Jiménez	24
GI/000013	04/06/2010	R0004	Manuel Pérez	36

Artículo	
cod_art	nom_art
AR410	pantalón
AR411	polos
AR412	Medias
AR413	polos

# NORMALIZACIÓN: 1 FORMA(1FN)

---

**Una tabla relacional R esta en primera forma normal (1FN) si NO contiene campos multivaluados.**

- Todos los atributos son atómicos. (simples e indivisibles).
- La tabla contiene una clave primaria única. (La clave primaria no contiene atributos nulos.)
- No debe existir variación en el número de columnas.
- Debe Existir una independencia del orden tanto de las filas como de las columnas, es decir, si los datos cambian de orden no deben cambiar sus significados
- Una tabla no puede tener múltiples valores en cada columna.

Esta forma normal elimina los valores repetidos dentro de una Base de Datos.

## *2.2.1.1 Primera forma normal*

Nro_GI	fecha	nom_recepcionista	cod_art	cant_art	nom_art
GI/000010	01/06/2010	Marcos Gonzales	AR410	12	pantalón
GI/000011	02/06/2010	José Julon	AR411	36	polos
GI/000012	03/06/2010	Samuel Jiménez	AR412	24	Medias
GI/000013	04/06/2010	Manuel Pérez	AR413	36	polos

# NORMALIZACIÓN: 2 FORMA(2FN)

Una relación está en 2FN si está en 1FN y si los atributos que no forman parte de ninguna clave dependen de forma completa (dependencia funcional plena) de la clave principal.

Es decir que no existen dependencias parciales. (Todos los atributos que no son clave principal deben depender únicamente de la clave principal). DEPENDENCIA FUNCIONAL

## 2.2.1.2 Segunda forma normal

Guia_remision			Articulo		
Nro_GI	fecha	nom_recepcionista	cod_art	cant_art	nom_art
GI/000010	01/06/2010	Marcos Gonzales	AR410	12	pantalón
GI/000011	02/06/2010	José Julon	AR411	36	polos
GI/000012	03/06/2010	Samuel Jiménez	AR412	24	Medias
GI/000013	04/06/2010	Manuel Pérez	AR413	36	polos

# NORMALIZACIÓN: 3 FORMA(3FN)

La 3NF fue definida originalmente por **E.F.Codd** en 1971.

La definición de Codd indica que una tabla está en 3NF **si y solo si** las tres condiciones siguientes se cumplen:

- La tabla está en la (2NF).
- Ningún atributo no-primario de la tabla es dependiente transitivamente de una clave primaria.
- Es una relación que no incluye ningún atributo clave.

Guia_interna		Recepcionista		
Nro_GI	fecha	cod_recepcionista	nom_recepcionista	cant_art
GI/000010	01/06/2010	R0001	Marcos Gonzales	12
GI/000011	02/06/2010	R0002	José Julon	36
GI/000012	03/06/2010	R0003	Samuel Jiménez	24
GI/000013	04/06/2010	R0004	Manuel Pérez	36

Articulo	
cod_art	nom_art
AR410	pantalón
AR411	polos
AR412	Medias

\* Dependencia transitiva: Z depende de X a través de Y.

# NORMALIZACIÓN: Forma Boyce/Codd

---

La Forma Normal de Boyce-Codd (BCNF) es una forma normal en la que todas las dependencias funcionales no triviales en una tabla se basan en una clave candidata. Esto significa que, en una tabla que cumple con la BCNF, cada atributo no clave debe ser dependiente funcionalmente de todas las claves candidatas de la tabla y no de otros atributos no clave.

La BCNF se aplica después de la Tercera Forma Normal (3NF), y se utiliza para evitar problemas de redundancia y anomalías de actualización en las tablas de la base de datos.



# NORMALIZACIÓN: Forma Boyce/Codd

Estudiante	Curso	Asesor
Gomez	Matemática I	Arias
Gomez	Fisica	Flores
Perez	Matemática I	Arias
Perez	Algebra	Sanchez
Ramos	Física	Flores
Ramos	Matemática I	García

Cada Curso, cada Estudiante tiene un solo Asesor **(E,C)** -> A

Cada Curso tiene varios Asesores, pero cada professor asesora en un solo Curso **(E,A)** -> C

## EstudiantexAsesor

Estudiante	Asesor
Gomez	Arias
Gomez	Flores
Perez	Arias
Perez	Sanchez
Ramos	Flores
Ramos	Garcia

## AsesorXCurso

Asesor	Curso
Arias	Matemática I
Flores	Física
Sanchez	Algebra



# NORMALIZACIÓN: 4 FORMA(4FN)

Una relación está en 4FN si está en BCFN y no contiene dependencias multivalor

**CURSO\_PROFESOR\_TEXTO**

<u>Curso</u>	<u>Profesor</u>	<u>Texto</u>
Química	Moreno	Físico Química
Química	Moreno	Química Orgánica
Química	Mora	Físico Química
Química	Mora	Química Orgánica
Matemáticas	Merino	Análisis Vectorial
Matemáticas	Merino	Álgebra
Matemáticas	Merino	Trigonometría

*Esta relación especifica que: el curso impartido puede ser dictado por varios profesores utilizando varios textos". Existen dos DMV:*

Curso  $\twoheadrightarrow$  Profesor      y      Curso  $\twoheadrightarrow$  Texto

# NORMALIZACIÓN: 4 FORMA(4FN)

La redundancia de datos causada por la DMV, se puede eliminar siguiendo **uno** de los siguientes métodos:

- ✓ Crear una nueva relación para cada atributo DMV.

Curso  $\rightarrow$  Profesor      Curso  $\rightarrow$  Texto

R1 = (Curso, Profesor)    R2 = (Curso, Texto)

Curso  $\rightarrow$  Profesor

<u>Curso</u>	<u>Profesor</u>
Química	Moreno
Química	Mora
Matemáticas	Merino

Curso  $\rightarrow$  Texto

<u>Curso</u>	<u>Texto</u>
Química	Físico Química
Química	Química Orgánica
Matemática	Análisis Vectorial
Matemática	Algebra
Matemática	Trigonometría

- ✓ Reemplazar un atributo DMV con atributos funcionalmente dependientes DF.

Curso  $\rightarrow$  (Texto1, Texto2, Texto3)

R3 = (Curso, texto1, texto2, texto3)

Curso	Texto1	Texto2	Texto3
Química	Físico química	Química orgánica	
Matemática	Análisis Vectorial	Algebra	Trigonometría

# Reglas de Codd

---

## Regla No. 1 - La Regla de la información

“

Toda la información en un RDBMS está explícitamente representada de una sola manera por valores en una tabla".

## Regla No. 2 - La regla del acceso garantizado

"Cada ítem de datos debe ser lógicamente accesible al ejecutar una búsqueda que combine el nombre de la tabla, su clave primaria, y el nombre de la columna".

# Reglas de Codd

---

## **Regla No. 3 - Tratamiento sistemático de los valores nulos**

"La información inaplicable o faltante puede ser representada a través de valores nulos".

## **Regla No. 4 - La regla de la descripción de la base de datos**

"La descripción de la base de datos es almacenada de la misma manera que los datos ordinarios, esto es, en tablas y columnas, y debe ser accesible a los usuarios autorizados".

# Reglas de Codd

---

## **Regla No. 5 - La regla del sub-lenguaje Integral**

"Debe haber al menos un lenguaje que sea integral para soportar la definición de datos, manipulación de datos, definición de vistas, restricciones de integridad, y control de autorizaciones y transacciones".

## **Regla No. 6 - La regla de la actualización de vistas**

"Todas las vistas que son teóricamente actualizables, deben ser actualizables por el sistema mismo".

# Reglas de Codd

---

## **Regla No. 7 - La regla de insertar y actualizar**

"La capacidad de manejar una base de datos con operandos simples aplica no solo para la recuperación o consulta de datos, sino también para la inserción, actualización y borrado de datos".

## **Regla No. 8 - La regla de independencia física**

"El acceso de usuarios a la base de datos a través de terminales o programas de aplicación, debe permanecer consistente lógicamente cuando quiera que haya cambios en los datos almacenados, o sean cambiados los métodos de acceso a los datos".

# Reglas de Codd

---

## **Regla No. 9 - La regla de independencia lógica**

"Los programas de aplicación y las actividades de acceso por terminal deben permanecer lógicamente inalteradas cuando quiera que se hagan cambios (según los permisos asignados) en las tablas de la base de datos".

## **Regla No. 10 - La regla de la independencia de la integridad**

"Todas las restricciones de integridad deben ser definibles en los datos, y almacenables en el catalogo, no en el programa de aplicación".

# Reglas de Codd

---

## **Regla No. 11 - La regla de la distribución**

"El sistema debe poseer un lenguaje de datos que pueda soportar que la base de datos esté distribuida físicamente en distintos lugares sin que esto afecte o altere a los programas de aplicación".

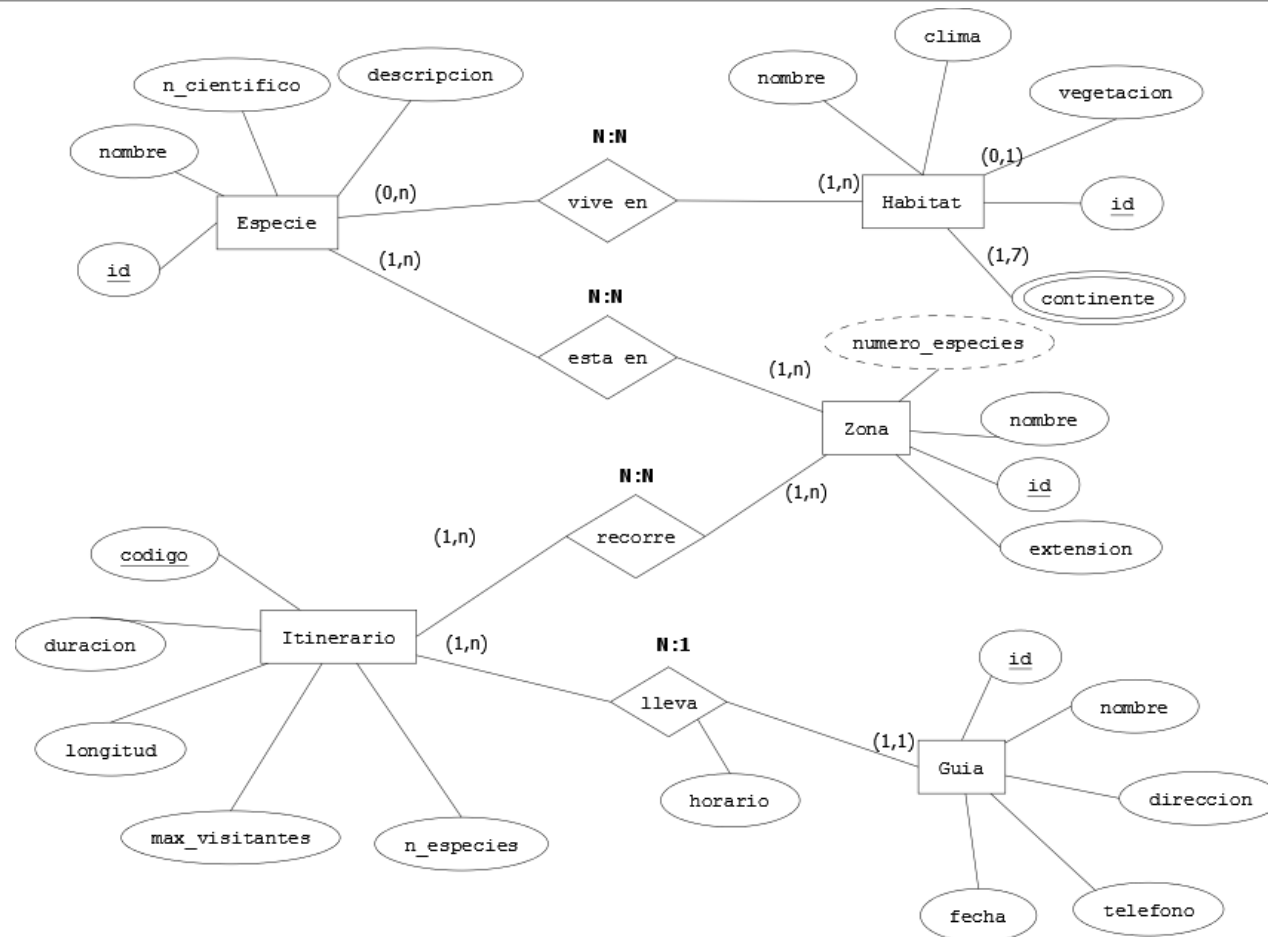
## **Regla No. 12 - Regla de la no-subversión**

"Si el sistema tiene lenguajes de bajo nivel, estos lenguajes de ninguna manera pueden ser usados para violar la integridad de las reglas y restricciones expresadas en un lenguaje de alto nivel (como SQL)".



# NORMALIZACION – DER

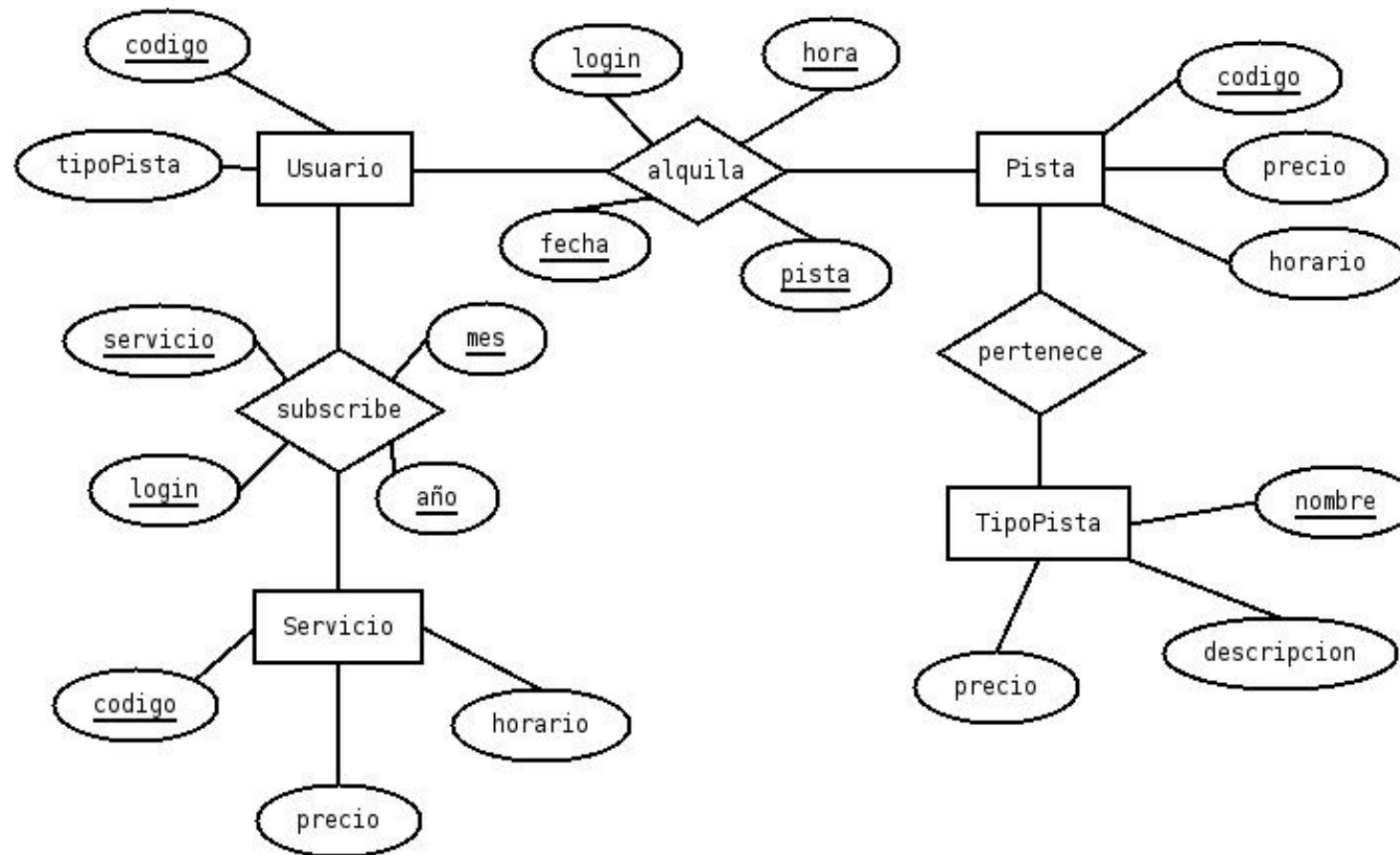
## Modelado Entidad-Relación -> DER



# NORMALIZACION – DER

## Modelado Entidad-Relación -> DER

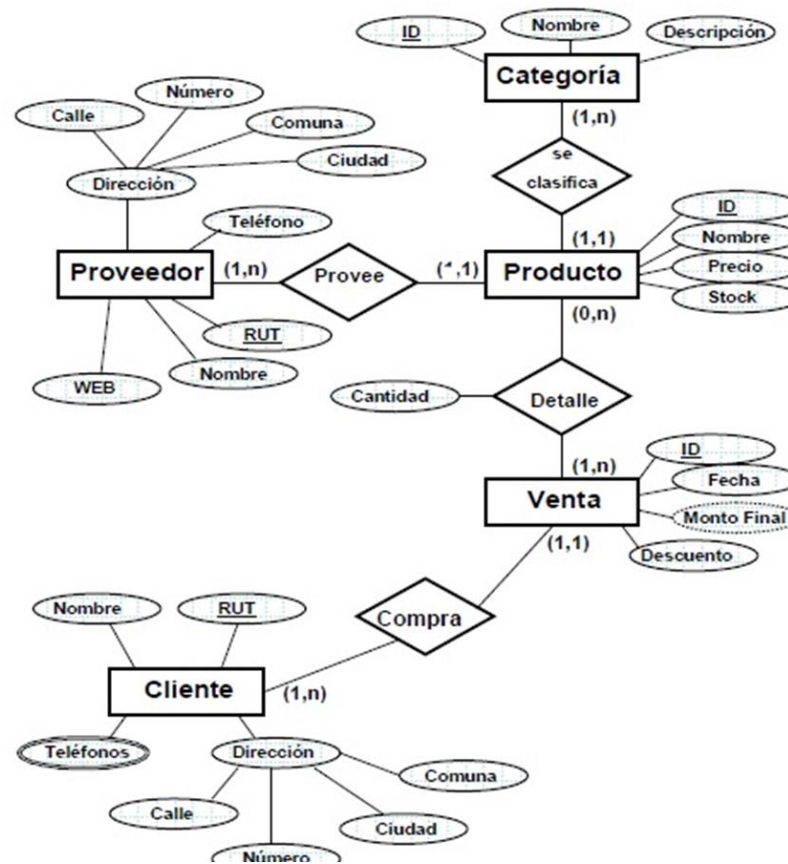
---



# NORMALIZACION – DER

## Modelado Entidad-Relación -> DER

---



# NORMALIZACIÓN DE BASE DE DATOS

---

1- **PRÉSTAMOS DE LIBROS**: La biblioteca del Colegio, necesita generar una base de datos que contemple la siguiente información. Aplicar las reglas de normalización para obtener las tablas que soporten la información requerida. Generar el Diagrama de Relación resultante.

# Normalización de Tablas: Tabla Sin Normalizar

---

CodLibro	Titulo	Autor	Editorial	NombreLector	FechaDev
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez Gómez, Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos Terán, Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca, René	16/04/2005
1006	Oracle University	Nancy Greenberg y Priya Nathan	Oracle Corp.	García Roque, Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez Gómez, Juan	18/04/2005

# Normalización de Tablas: Normalización F1

NO contiene campos multivaluados.

---

CodLibro	Título	Autor	Editorial	Paterno	Materno	Nombres	FechaDev
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez	Gómez	Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos	Terán	Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca		René	16/04/2005
1006	Oracle University	Nancy Greenberg	Oracle Corp.	García	Roque	Luis	20/04/2005
1006	Oracle University	Priya Nathan	Oracle Corp.	García	Roque	Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez	Gómez	Juan	18/04/2005

# Normalización de Tablas: Normalización F2

Los atributos que no forman parte de ninguna clave dependen de forma completa de la clave principal.

CodLibro	Titulo	Autor	Editorial
1001	Variable compleja	Murray Spiegel	McGraw Hill
1004	Visual Basic 5	E. Petroustsos	Anaya
1005	Estadística	Murray Spiegel	McGraw Hill
1006	Oracle University	Nancy Greenberg	Oracle Corp.
1006	Oracle University	Priya Nathan	Oracle Corp.
1007	Clipper 5.01	Ramalho	McGraw Hill

CodLibro	CodLector	FechaDev
1001	501	15/04/2005
1004	502	17/04/2005
1005	503	16/04/2005
1006	504	20/04/2005
1007	501	18/04/2005

CodLector	Paterno	Materno	Nombres
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis

# Normalización de Tablas: Normalización F3

Ningún atributo no-primario de la tabla es dependiente transitivamente de una clave primaria.

CodLibro	Titulo	CodAutor	CodEditorial
1001	Variable compleja	801	901
1004	Visual Basic 5	802	902
1005	Estadística	803	903
1006	Oracle University	804	901
1006	Oracle University	803	902
1007	Clipper 5.01	806	903

CodLibro	CodLector	FechaDev
1001	501	15/04/2005
1004	502	17/04/2005
1005	503	16/04/2005
1006	504	20/04/2005
1007	501	18/04/2005

CodLibro	CodAutor
1001	801
1004	802
1005	803
1006	804
1006	803
1007	806

CodAutor	Autor
801	Murray Spiegel
802	E. Petroustsos
803	Nancy Greenberg
804	Priya Nathan
806	Ramalho

CodEditorial	Editorial
901	McGraw Hill
902	Anaya
903	Oracle Corp.

CodLector	Paterno	Materno	Nombres
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis



# Normalización de Tablas: Normalización BoyceCodd

CodLibro	Titulo	CodEditorial
1001	Variable compleja	901
1004	Visual Basic 5	902
1005	Estadística	903
1006	Oracle University	901
1007	Clipper 5.01	903

CodLibro	CodLector	FechaDev
1001	501	15/04/2005
1004	502	17/04/2005
1005	503	16/04/2005
1006	504	20/04/2005
1007	501	18/04/2005

CodLibro	CodAutor
1001	801
1004	802
1005	803
1006	804
1006	803
1007	806

CodAutor	Autor
801	Murray Spiegel
802	E. Petroustsos
803	Nancy Greenberg
804	Priya Nathan
806	Ramalho

CodEditorial	Editorial
901	McGraw Hill
902	Anaya
903	Oracle Corp.

CodLector	Paterno	Materno	Nombres
501	Pérez	Gómez	Juan
502	Ríos	Terán	Ana
503	Roca		René
504	García	Roque	Luis