



DigitalHouse >
Coding School

DATA SCIENCE

UNIDAD 3
MÓDULO 5

Normalización de Bases
de Datos

Octubre 2017

NORMALIZACIÓN DE BASES DE DATOS



- Las formas normales de una base de datos fueron planteadas por Boyce y Codd a principios de la década del '70

¿Para qué normalizar un base de datos?

- Tres objetivos principales:
 - **Garantizar la integridad** de la información
 - **Evitar redundancia** en los datos
 - **Escalabilidad**: que el modelo soporte modificaciones y extensiones con un bajo impacto

La primera forma normal exige los siguientes puntos:

- Eliminar los grupos repetidos en celdas individuales, cada celda debe contener un atributo “atómico” o indivisible
- Crear tablas separadas para cada conjunto de observaciones relacionadas
- Identificar a cada tabla con una clave primaria

NO CUMPLE

Customer

Customer ID	First Name	Surname	Telephone Number
123	Pooja	Singh	555-861-2025, 192-122-1111
456	San	Zhang	(555) 403-1659 Ext. 53; 182-929-2929
789	John	Doe	555-808-9633

Customer

Customer ID	First Name	Surname	Telephone Number1	Telephone Number2
123	Pooja	Singh	555-861-2025	192-122-1111
456	San	Zhang	(555) 403-1659 Ext. 53	182-929-2929
789	John	Doe	555-808-9633	



CUMPLE

Customer Name

<u>Customer ID</u>	First Name	Surname
123	Pooja	Singh
456	San	Zhang
789	John	Doe

Customer Telephone Number

Customer ID	<u>Telephone Number</u>
123	555-861-2025
123	192-122-1111
456	(555) 403-1659 Ext. 53
456	182-929-2929
789	555-808-9633

NO CUMPLE

DATA	
Curso	Contenido
Programación	Java, C++
Web	HTML, CSS, Php



CUMPLE

CURSO	
id_curso	descripcion
1	Programación
2	Web

CONTENIDO	
id_contenido	descripcion
1	Java
2	C++
3	HTML
4	CSS
5	php

CURSO_CONTENIDO	
id_curso	id_contenido
1	1
1	2
2	3
2	4
2	5

Además de cumplir con la primera forma normal la segunda forma normal exige:

- Que todos los atributos **que no forman parte** de la clave primaria, **dependan de todos** los componentes de la clave primaria.
- Si uno de los atributos depende únicamente de una parte de la clave primaria, entonces no se cumple la segunda forma normal.

Electric Toothbrush Models

<u>Manufacturer</u>	<u>Model</u>	Model Full Name	Manufacturer Country
Forte	X-Prime	Forte X-Prime	Italy
Forte	Ultraclean	Forte Ultraclean	Italy
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZbrush	USA
Kobayashi	ST-60	Kobayashi ST-60	Japan
Hoch	Toothmaster	Hoch Toothmaster	Germany
Hoch	X-Prime	Hoch X-Prime	Germany



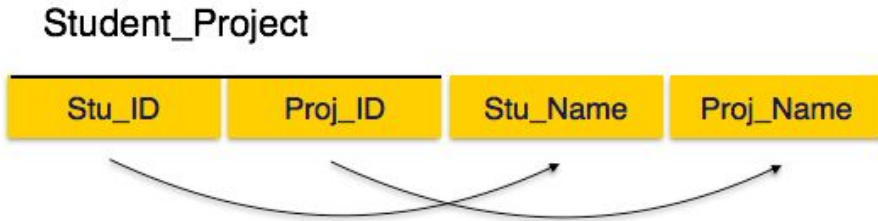
Electric Toothbrush Manufacturers

<u>Manufacturer</u>	Manufacturer Country
Forte	Italy
Dent-o-Fresh	USA
Kobayashi	Japan
Hoch	Germany

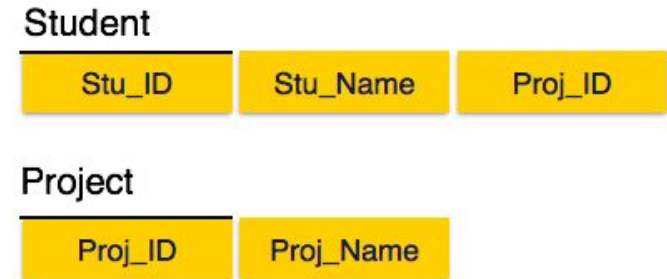
Electric Toothbrush Models

<u>Manufacturer</u>	<u>Model</u>	Model Full Name
Forte	X-Prime	Forte X-Prime
Forte	Ultraclean	Forte Ultraclean
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZbrush
Kobayashi	ST-60	Kobayashi ST-60
Hoch	Toothmaster	Hoch Toothmaster
Hoch	X-Prime	Hoch X-Prime

NO CUMPLE



CUMPLE



Además de cumplir con la segunda forma normal la tercera forma normal exige:

- Que ninguno de los atributos que no forman parte de la clave primaria dependan transitivamente de alguno de los otros atributos

La tercera forma normal se puede parafrasear de la siguiente manera:

"Every non-keyattribute must provide a fact about the key, the whole key, and nothing but the key."

NO CUMPLE

Tournament Winners

<u>Tournament</u>	<u>Year</u>	<u>Winner</u>	<u>Winner Date of Birth</u>
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975
Indiana Invitational	1999	Chip Masterson	14 March 1977



CUMPLE

Tournament Winners

<u>Tournament</u>	<u>Year</u>	<u>Winner</u>
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

Winner Dates of Birth

<u>Winner</u>	<u>Date of Birth</u>
Chip Masterson	14 March 1977
Al Fredrickson	21 July 1975
Bob Albertson	28 September 1968

NO CUMPLE

Student_Detail

Stu_ID	Stu_Name	City	Zip
--------	----------	------	-----



CUMPLE

Student_Detail

Stu_ID	Stu_Name	Zip
--------	----------	-----

ZipCodes

Zip	City
-----	------



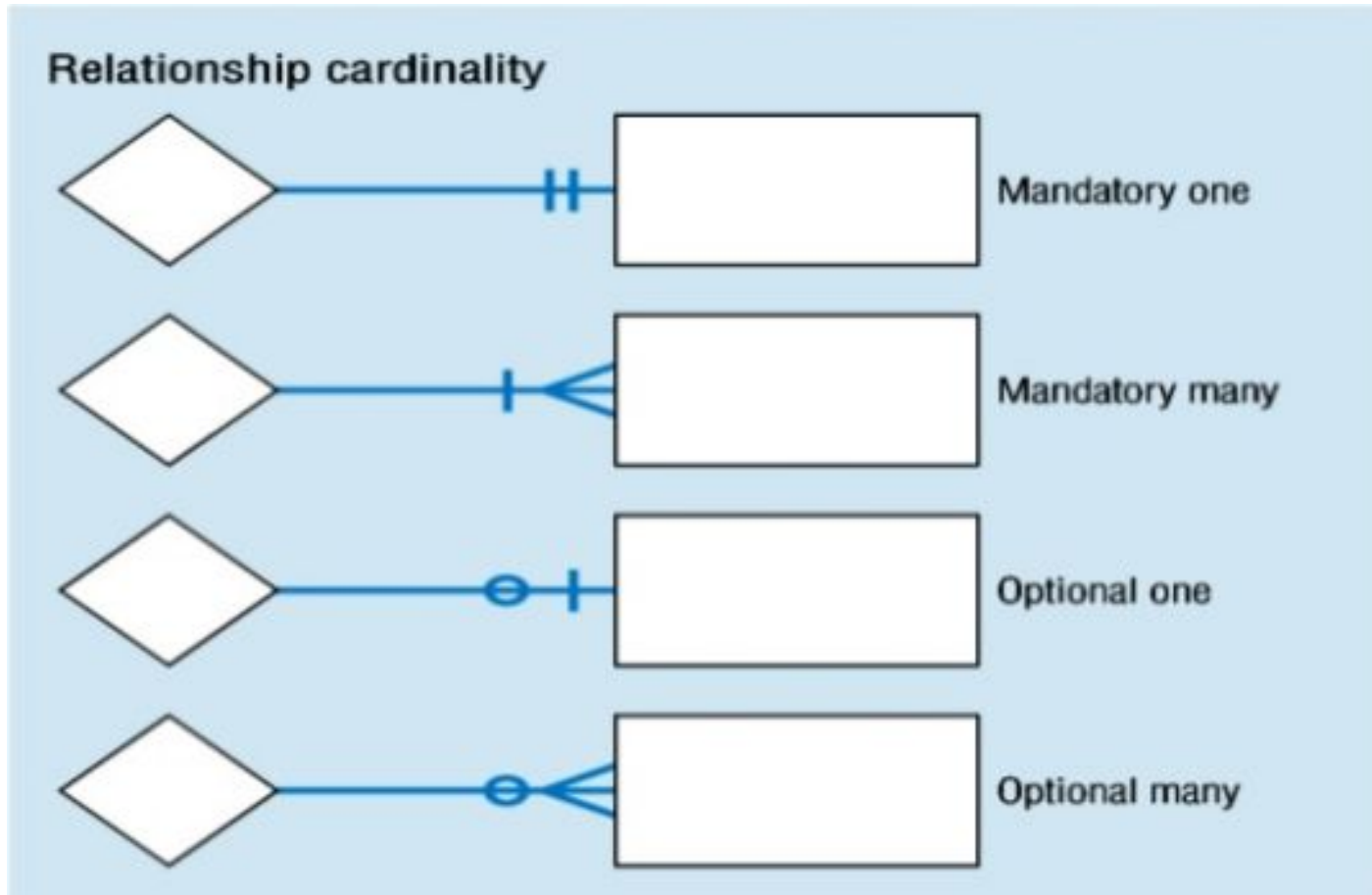
- Toda base de datos se puede modelar como un conjunto de entidades y las relaciones entre esas entidades.
- El DER es un esquema que representa las entidades y sus relaciones. No es lo mismo que el diagrama de tablas. Es una representación gráfica de la lógica de la base.
- El DER modela todas las **decisiones de diseño** que se deben tomar a la hora de construir una base de datos ¿Qué entidades va a haber? ¿Cómo se relacionan? ¿Qué restricciones hay sobre los datos?

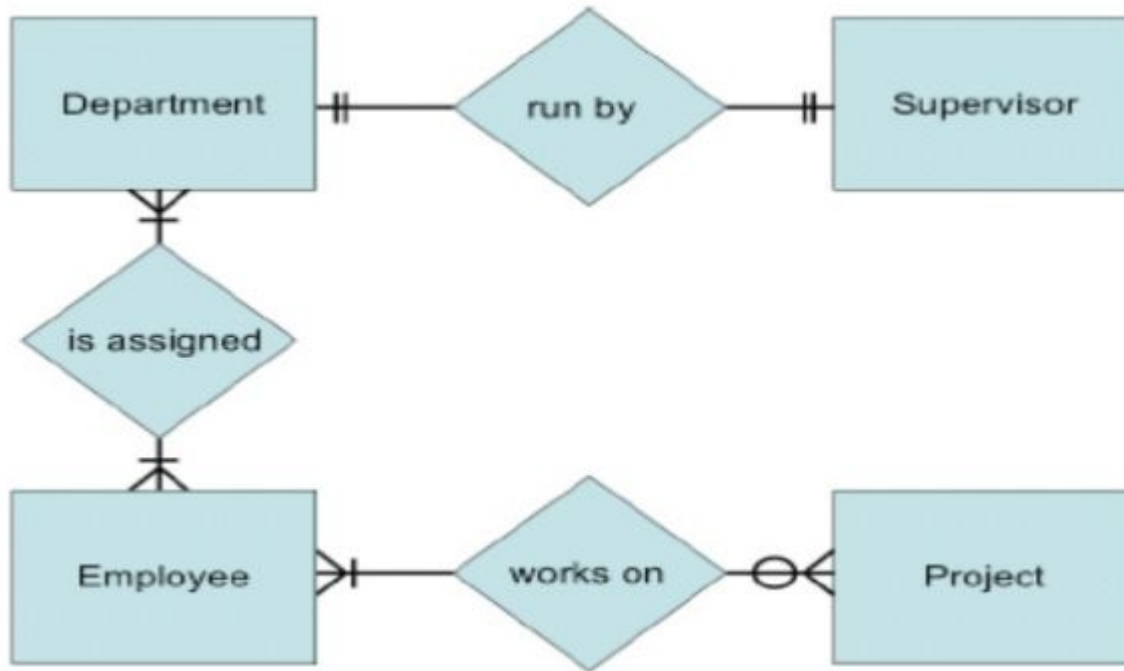
- Relaciones “no obligatorias”, por ejemplo:
 - Un empleado puede estar asignado o no a algún departamento
 - Un paciente puede estar asignado o no a una cama

- Relaciones “obligatorias”, por ejemplo:
 - Cada curso tiene que ser dictado por al menos un docente
 - Todo conductor conduce al menos un auto

Las relaciones binarias entre dos entidades se pueden dividir en las siguientes clases:

- Relaciones uno a uno, por ejemplo:
 - Un gerente dirige un departamento y cada departamento es dirigido por un único gerente.
- Relaciones uno a muchos, por ejemplo:
 - Un empleado trabaja en un departamento pero un departamento tiene muchos empleados.
- Relaciones muchos a muchos, por ejemplo:
 - Un profesor con varios estudiantes, cada estudiante con varios profesores.



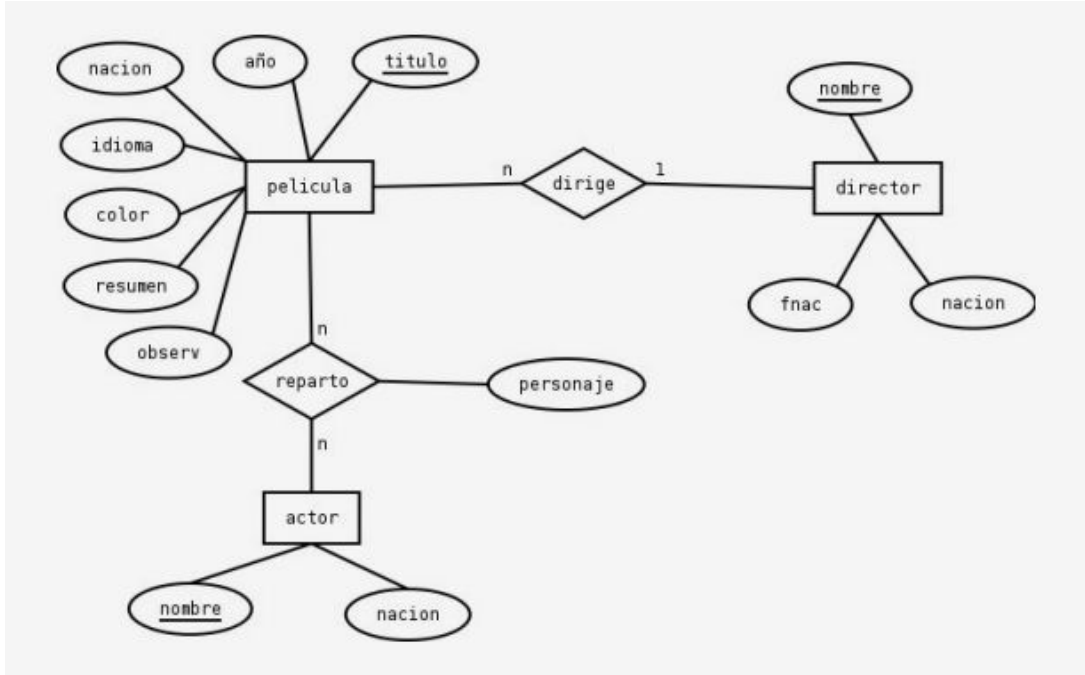


- ¿Cuántos supervisores hay por departamento?

- ¿Puede haber temporalmente un departamento sin supervisor?

- ¿Puede haber un empleado no asignado a ningún proyecto? ¿Y a más de uno?

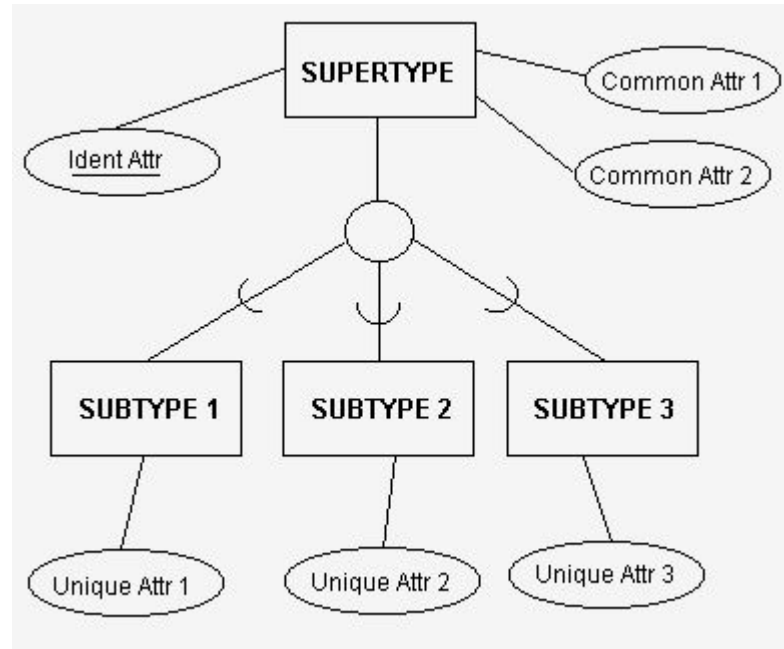
- ¿Cuántos supervisores puede tener un empleado? ¿Y un proyecto?

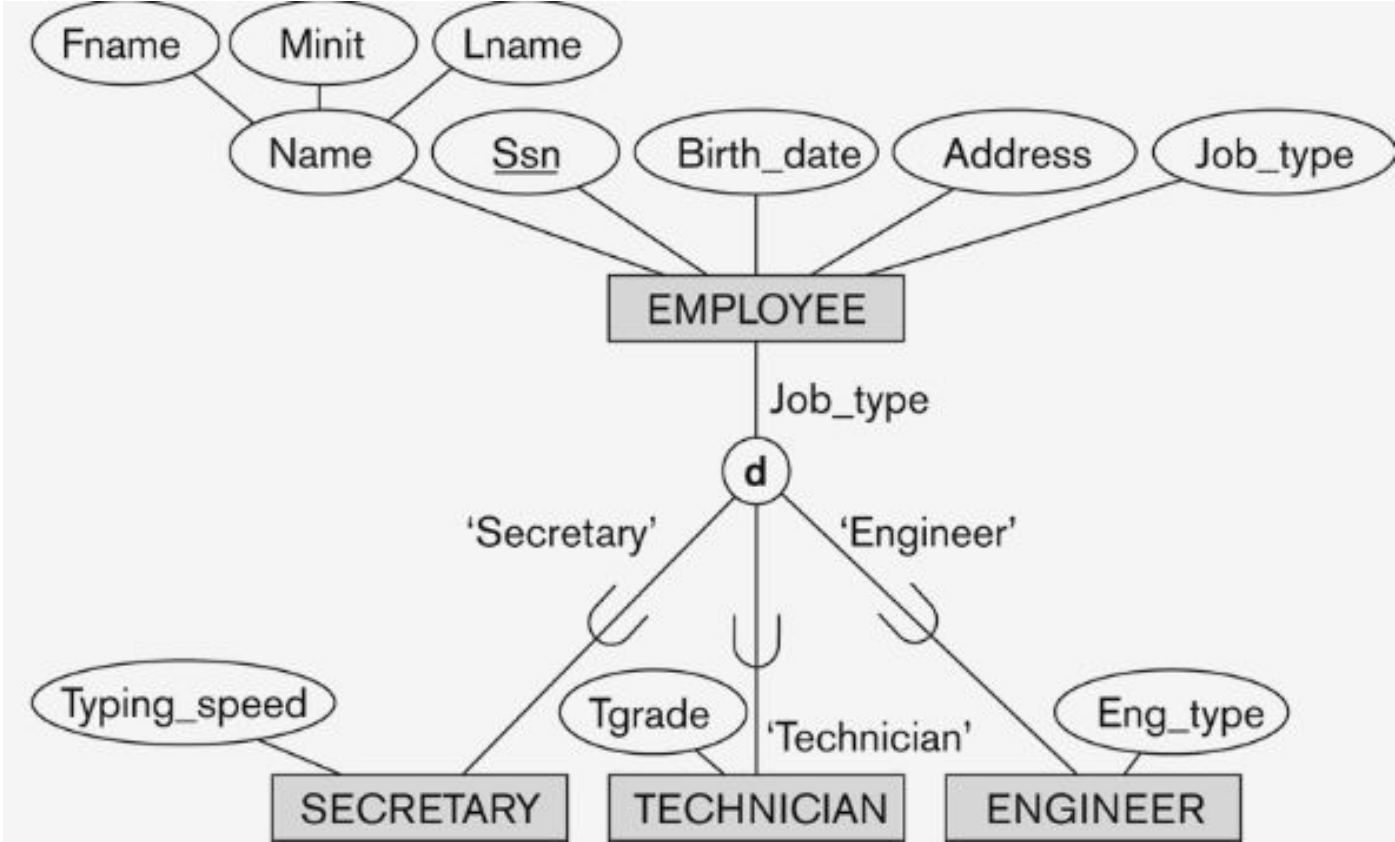


-Además de las entidades y las relaciones, en un DER se puede modelar los atributos de cada una de las entidades y señalar la clave primaria subrayándola.

- Las relaciones también pueden tener atributos. Por ejemplo un actor participa como reparto de una determinada película haciendo específicamente un personaje.

En un DER también podemos querer modelar qué relaciones “heredan” de otras. Esta situación se da cuando tenemos tipos y subtipos de algún objeto donde algunas características son comunes a todos y otras son específicas para cada subtipo.





Una compañía aérea necesita una base de datos para registrar la información de sus vuelos. Los vuelos tienen un identificador único. Además, cada vuelo tiene asignado un aeropuerto de origen y uno de destino (se asume que no hay escalas).

Los aeropuertos están identificados por unas siglas únicas (por ejemplo: VLC-Valencia, BCN-Barcelona, MAD-Madrid). Además, de cada aeropuerto se guarda el nombre de la ciudad en la que está situado y el país.

Cada vuelo es realizado por un avión. Los aviones tienen una matrícula que los identifica, el fabricante, un modelo e información sobre su capacidad (número máximo de pasajeros) y autonomía de vuelo (en horas). La asignación de aviones a vuelos no es única, así que es necesario saber la fecha en la que un avión realizó cada uno de los vuelos asignados.

NORMALIZACIÓN DE BASES DE DATOS

