



BASES DE DATOS PARA IA/DATA SCIENCE

PABLO OCTAVIANO
JUAN FRANCISCO PINELLI



BASES DE DATOS RELACIONALES

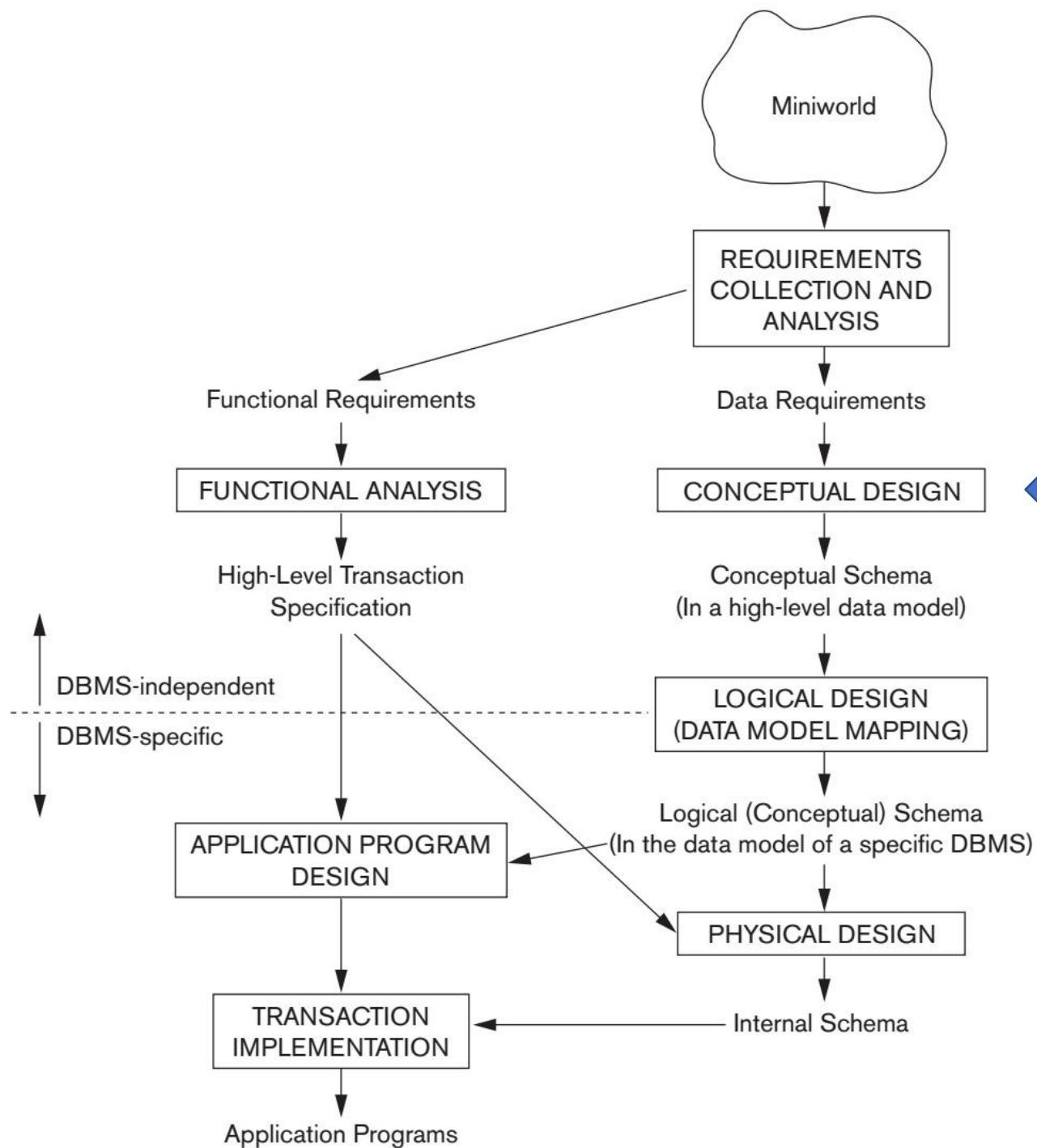
Modelo lógico,
relaciones, relaciones
y más relaciones



AGENDA

- Relaciones
- Diagramas de entidad relación
- Llaves
- Modelo Relacional
- Normalización

Proceso de Creación de Base de datos



Usted está aquí



RELACIONES

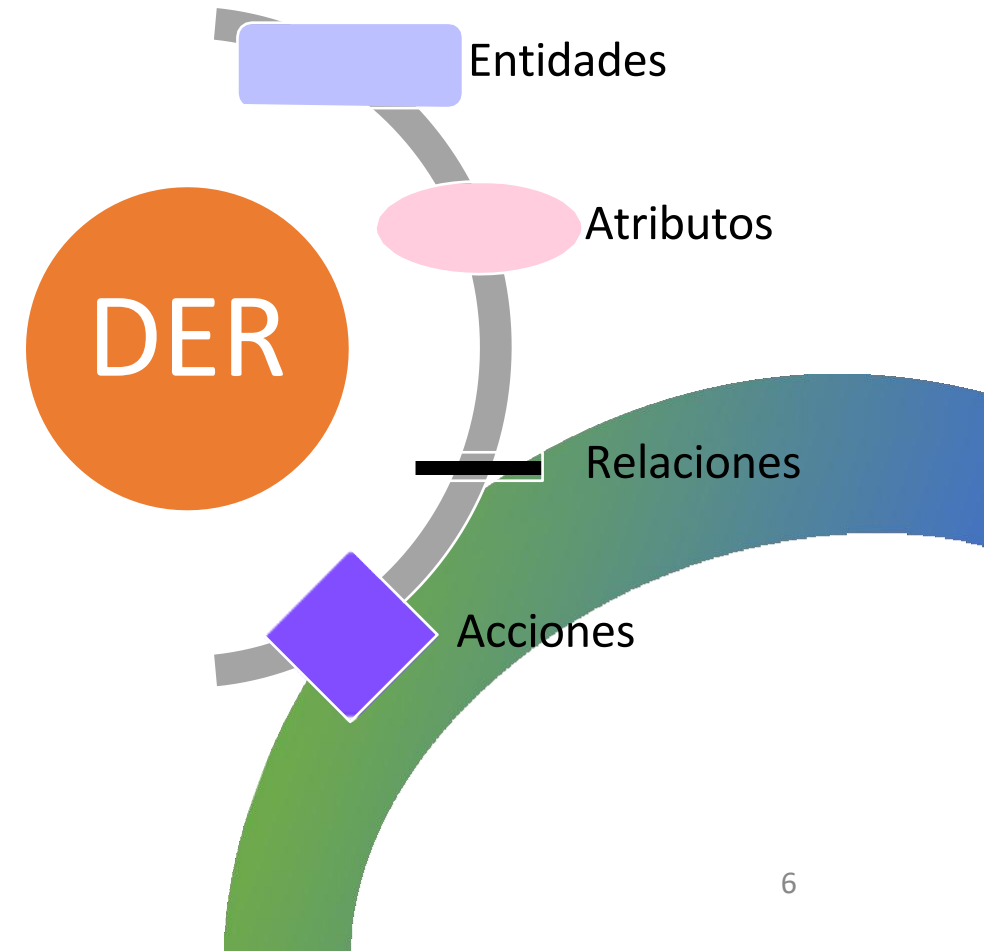
Conceptos básicos

Diagrama Entidad Relación (DER/ERD)

El modelo Entidad-Relación (E-R) es una manera de representar nuestra percepción del sistema que vamos a modelar.

Este consiste en un conjunto de objetos básicos:

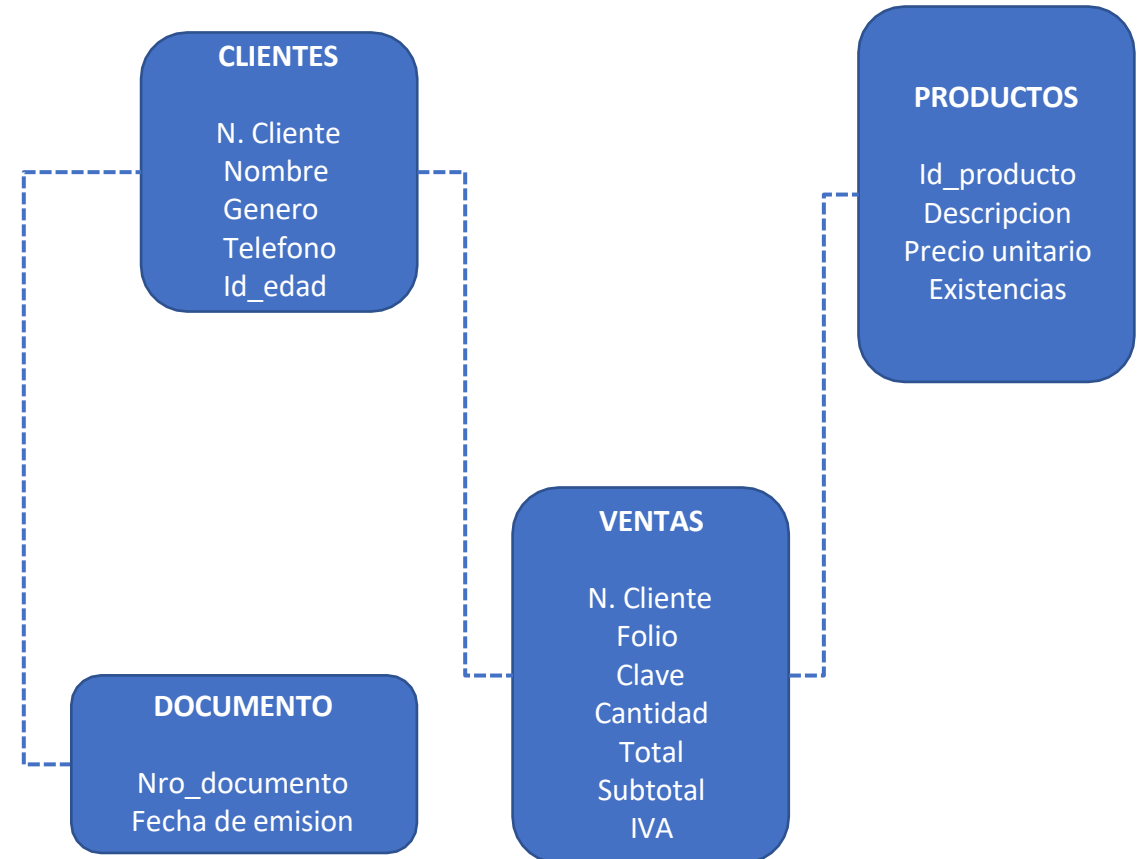
- **Entidades**
- **Atributos**
- **Interrelaciones**
- **Acciones**



Repasamos!

Recordamos que a una tabla también la podemos llamar **relación**. Sin embargo, este término es mucho más utilizado para describir las interrelaciones entre tablas.

Una relación entre tablas supone que existen referencias entre llaves de una tabla.

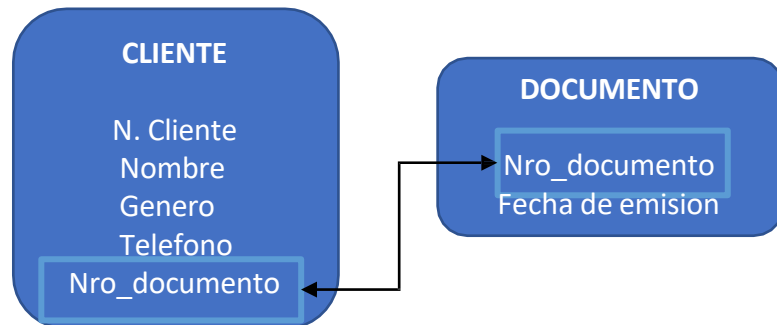


Relaciones

Uno a uno

One-to-one

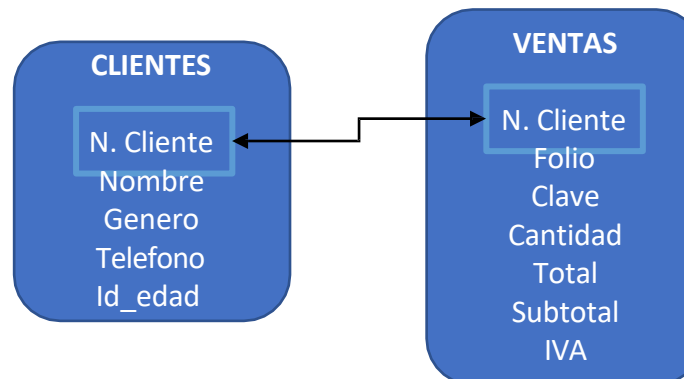
- Las relaciones uno a uno son aquellas que para cada registro de ambas tablas existe una relación univoca.
- Esto significa que cada valor en las tablas aparece solo una vez en la tabla.



Uno a muchos

One-to-many

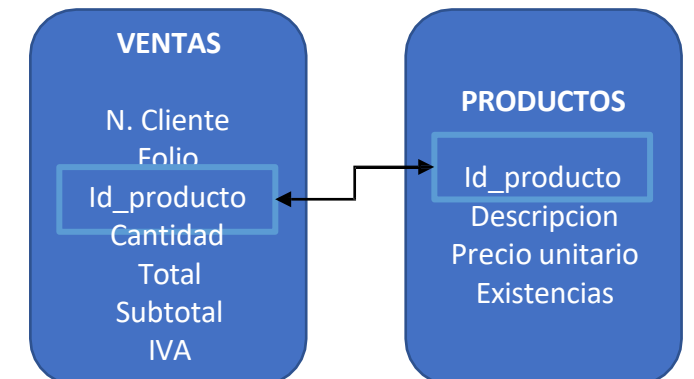
- En estos casos, consideramos que los datos de una tabla pueden aparecer múltiples veces en la siguiente.
- La restricción es que en una de ellas el registro debe ser único



Muchos a muchos

Many-to-Many

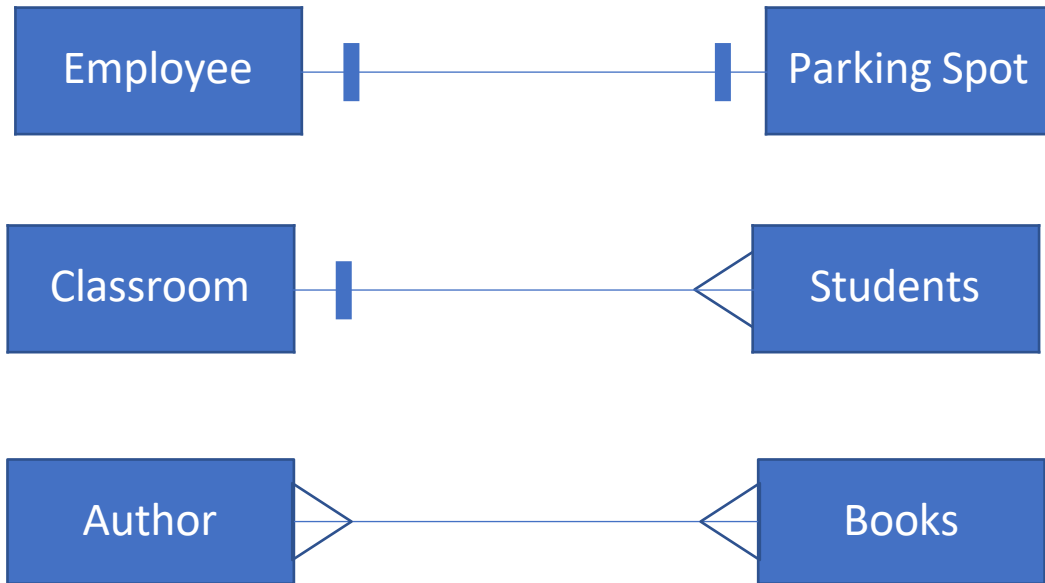
- Es un tipo de relación que ocurre cuando múltiples registros de una tabla se relacionan con muchos elementos de la otra tabla



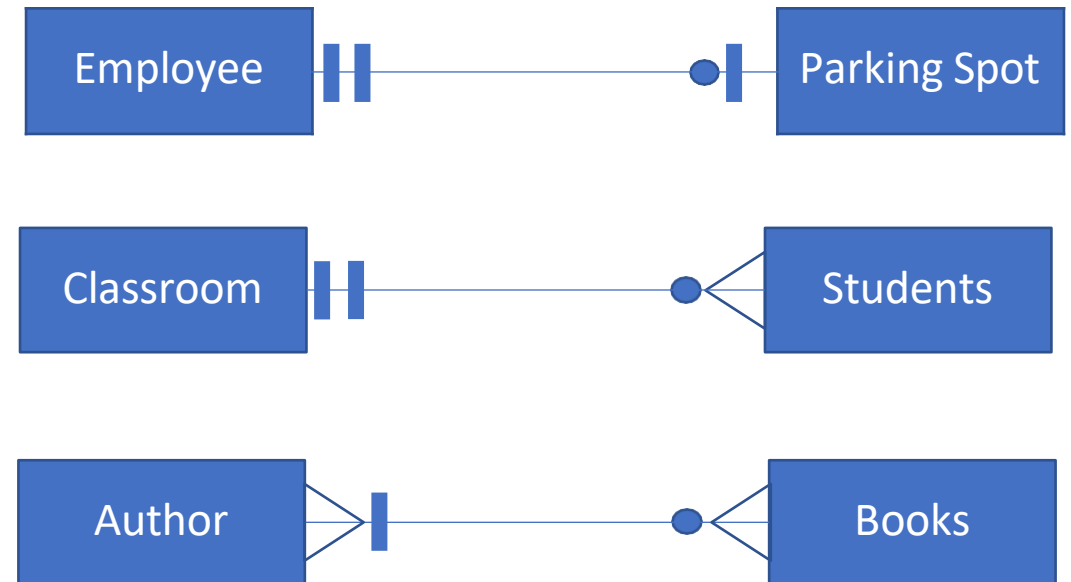
Relaciones

representación – notación Crow

Cardinalidad



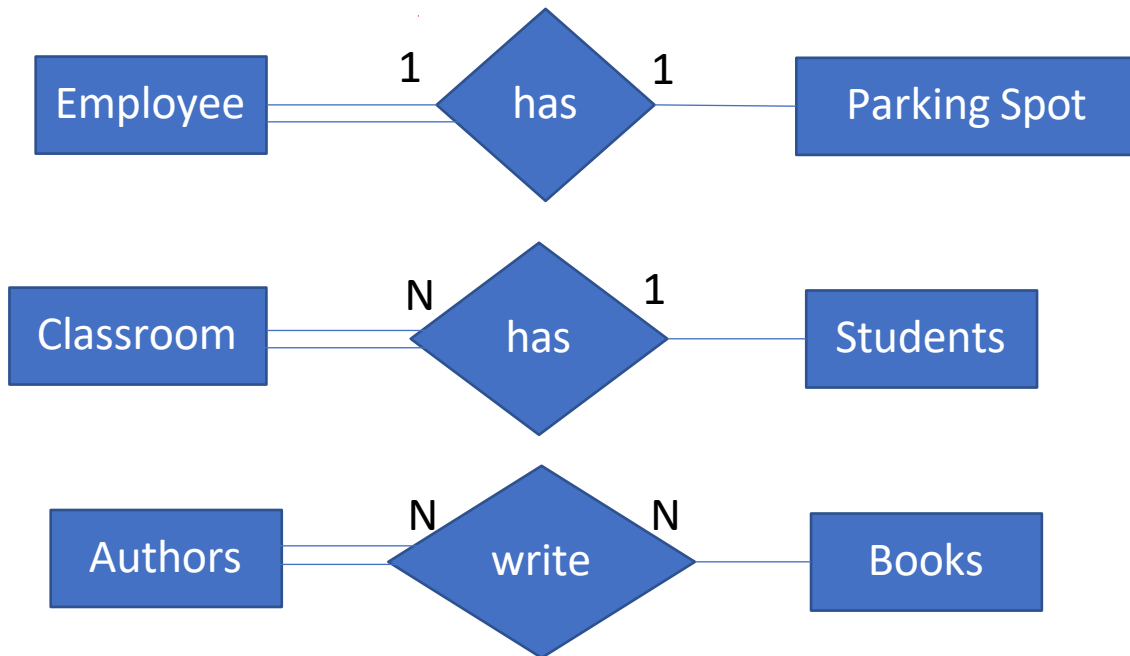
Participación



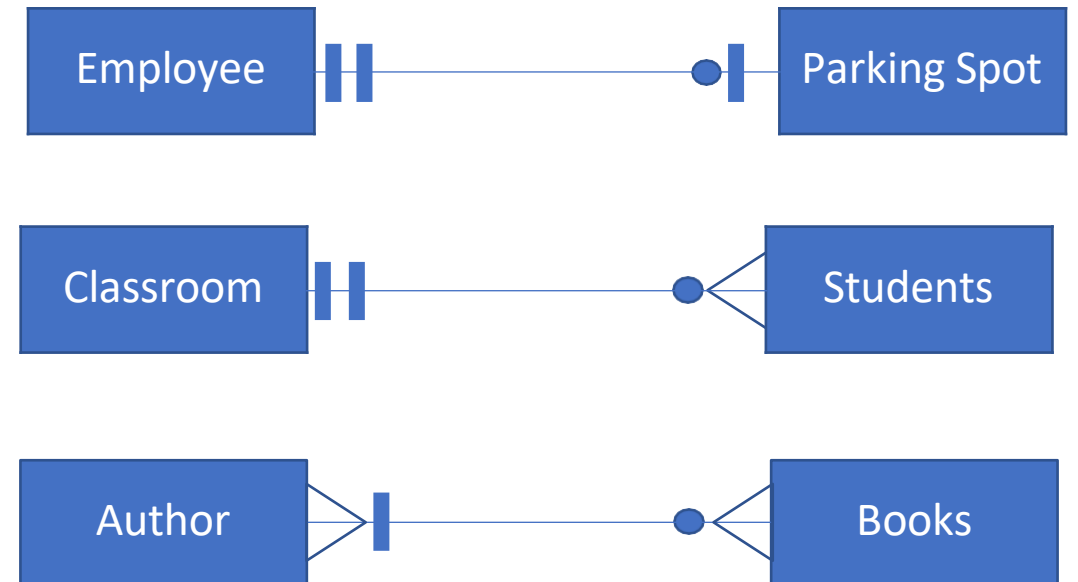
Relaciones

representación – notación de Chen

Chen



Crow





DER



BASES DE DATOS PARA INTELIGENCIA ARTIFICIAL

Base de datos corporativa

- 1) Una compañía se organiza en **departamentos**, cada departamento **tiene** nombre único, y **un empleado en particular** que maneja el departamento
- 2) Un **departamento controla** cierto número de **proyectos**, donde **cada proyecto** no cuenta con un **nombre unívoco** y un lugar determinado



BASES DE DATOS PARA INTELIGENCIA ARTIFICIAL

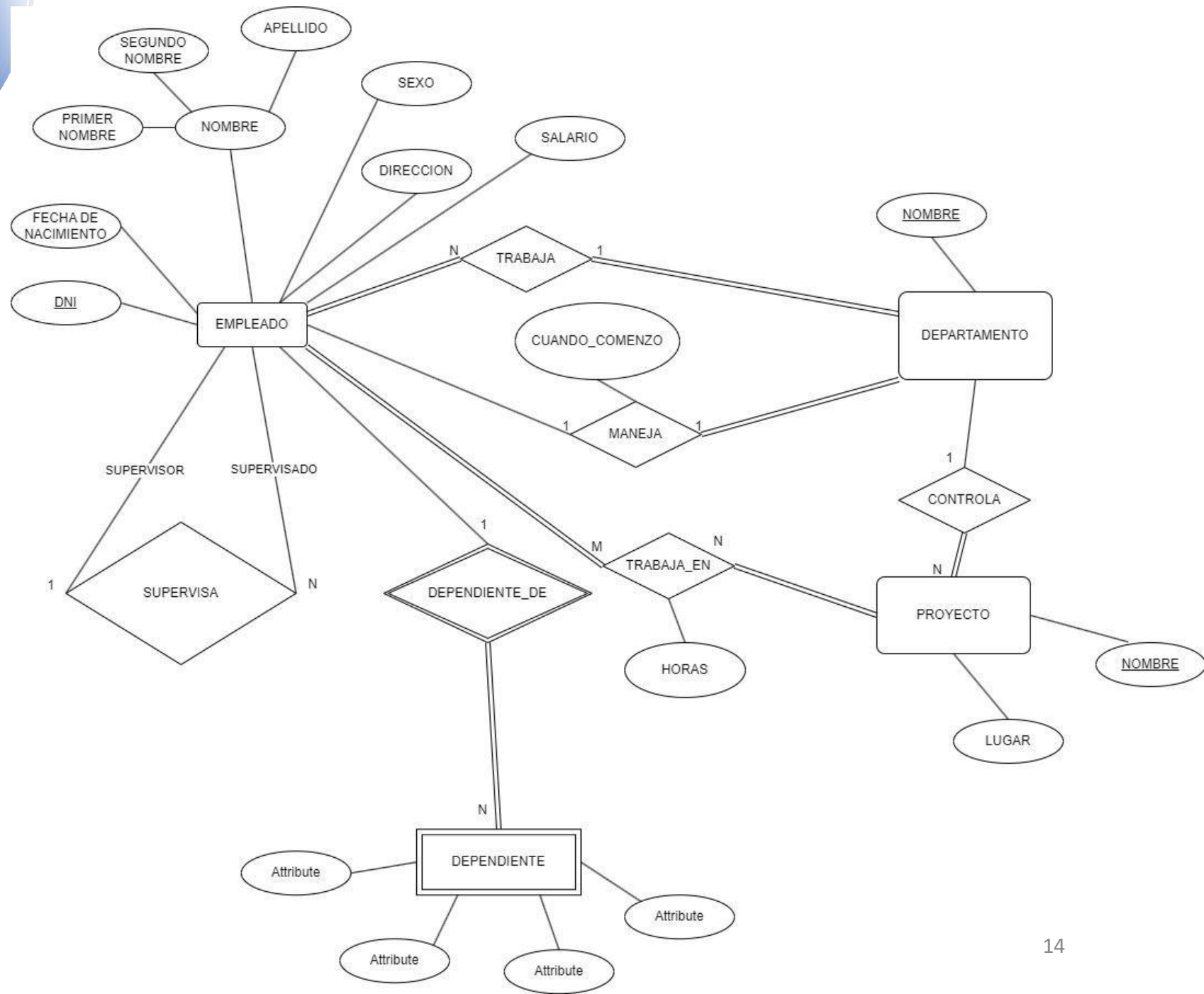
Base de datos corporativa

3) La base de datos guarda **el nombre** de cada empleado, el dni, domicilio, salario, sexo, y fecha de nacimiento. El **empleado** es asignado a un departamento, pero puede trabajar en **múltiples proyectos**.

4) Se requiere llevar un track del **número de horas por semana que cada empleado trabaja por proyecto**.

5) La base de datos también debería guardar información acerca de **cada persona que dependa** directamente del **empleado**, por temas de seguro. Deberá contar con el primer nombre, el sexo, la fecha de nacimiento y la relación que lleva con el empleado.

Solución planteada



Algoritmo de mapeo DER -> MR

Para llegar al modelo relacional final, se necesitan una secuencia de pasos que definirán las formas de las tablas finales, al igual que los vínculos y llaves entre ellas.

- 1. Mapear entidades fuertes**
- 2. Mapear entidades débiles**

Algoritmo de mapeo DER -> MR

3. Mapear Relaciones del tipo 1:1

- Opción 1. Foreign keys, preferiblemente en entidades con participación *total*
- Se une a la entidad en una sola tabla, válido cuando la participación de ambos es *total*
- Referencias cruzadas (tabla de relaciones)

Algoritmo de mapeo DER -> MR

4. Mapear Relaciones del tipo 1:N

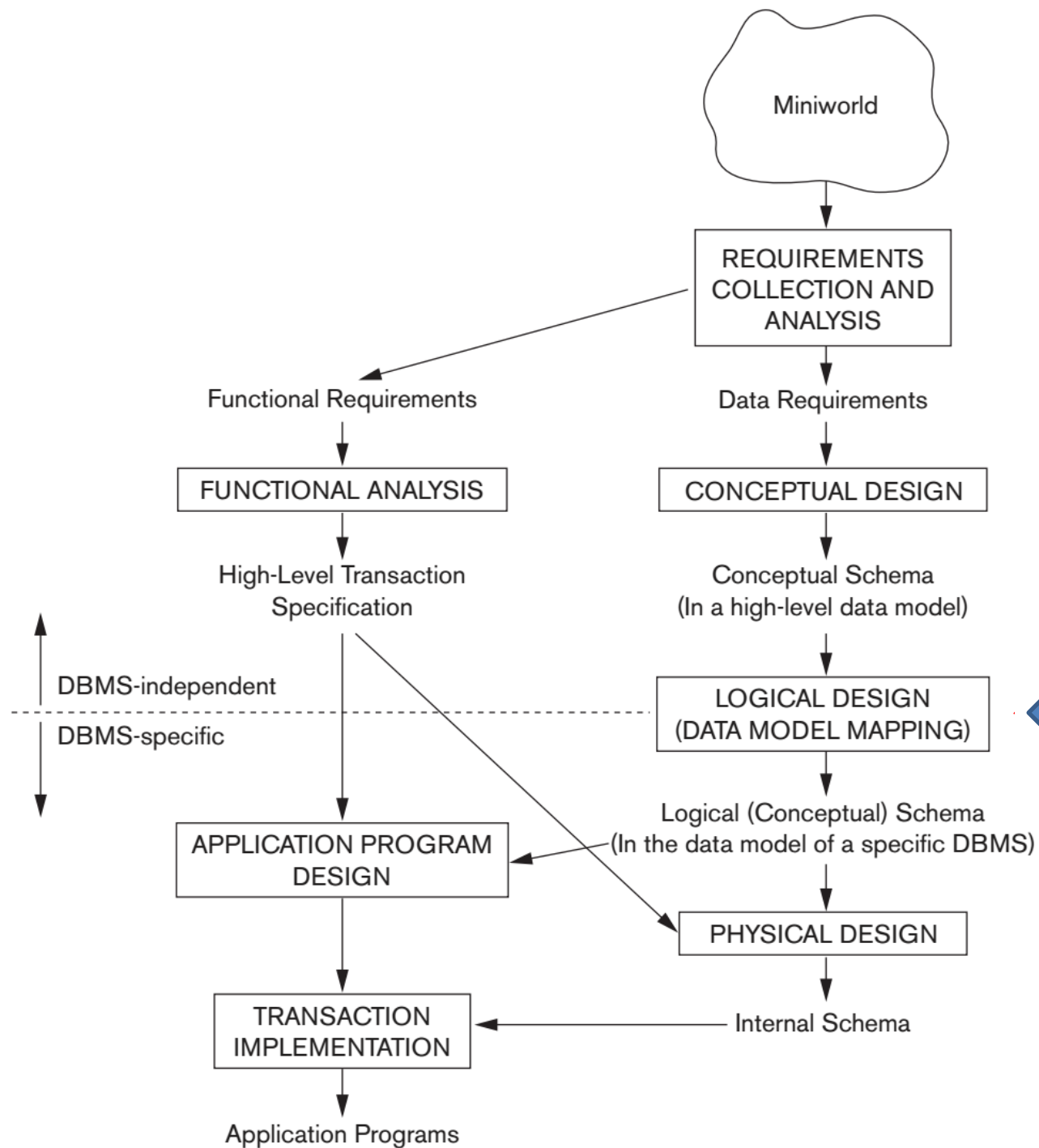
- Opción 1. Foreign keys, preferiblemente sobre la entidad que tenga cardinalidad N
- Referencias cruzadas (tabla de relaciones)

5. Mapear Relaciones del tipo M:N

- Referencias cruzadas (tabla de relaciones)

6. Mapear Atributos

Normalización



Usted está aquí

Guía informal de diseño de esquemas

Asegurarnos que la semántica de los atributos sea la correcta

Reducir información redundante en las tablas

Reducir la cantidad de valores NULL en las tuplas

No permitir la posibilidad de generar tuplas espurias o anomalías

Anomalías por Inserción

Anomalías por Borrado

Anomalías por Modificación

SALES

ITEMS

PRICE

CREATED_AT

QUANTITY

FULL_NAME

ADDRESS

...

Normalización

Proceso de aplicar una serie de reglas para minimizar (o eliminar) la redundancia de datos

Evita anomalías sobre los datos

Separa grandes tablas en tablas mas pequeñas y las une a través de relaciones

Normalización

SALES
ITEMS
PRICE
CREATED_AT
QUANTITY
FULL_NAME
ADDRESS
...



ORDERS
ORDER_ITEMS_ID
CUSTOMER_ID
CREATED_AT
QUANTITY
...

ORDER ITEMS
ITEM
PRICE
CATEGORY
DESCRIPTION
ID
...

CUSTOMER
EMAIL
ID
FULL_NAME
ADDRESS
...

Normalización

1 FN

PRIMARY KEY

CADA ATRIBUTO DEBE

CONTENER UN SOLO VALOR

ATÓMICO

SIN GRUPOS REPETIDOS.

1FN

name	address	phone
Thomas Monero	12 Hangover Sq.	3456
Jessica Moos	12 Hangover Sq.	3456
Adam Hardy	Cerrito 333	1243
Maria Anders	Av. del Libertador 900	8906

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

1FN

Primary key

customer_id	name	address	phone
101	Thomas Monero	12 Hangover Sq.	3456
102	Jessica Moos	12 Hangover Sq.	3456
103	Adam Hardy	Cerrito 333	1243
104	Maria Anders	Av. del Libertador 900	8906

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

1FN

customer_id	name	address	phone
101	Thomas Monero	12 Hangover Sq.	3456
102	Jessica Moos	12 Hangover Sq.	3456
103	Adam Hardy	Cerrito 333	1243
104	Maria Anders	Av. del Libertador 900	8906

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

1FN

customer_id	first_name	last_name	address	phone
101	Thomas	Monero	12 Hangover Sq.	3456
102	Jessica	Moos	12 Hangover Sq.	3456
103	Adam	Hardy	Cerrito 333	1243
104	Maria	Anders	Av. del Libertador 900	8906

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

1FN

order_detail

order_id	product_id	price	quantity	new_product	new_product_price	new_product_quantity
101	1	45.6	30	2	18	12
102	3	12	61	4	43.9	10
103	3	10	47	1	18	22

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

1FN

order_detail

order_id	product_id	price	quantity
101	1	45.6	30
101	2	18	12
102	3	12	61
102	4	43.9	10
103	3	10	47
103	1	18	22

- PRIMARY KEY
- CADA ATRIBUTO DEBE CONTENER UN SOLO VALOR ATÓMICO
- SIN GRUPOS REPETIDOS.

Normalización

2 FN

CUMPLE CON 1 FN
SIN DATOS REDUNDANTES
LOS ATRIBUTOS QUE NO
FORMAN PARTE DE LA CLAVE
ESTAN COMPLETAMENTE
DEFINIDOS POR LA CLAVE

2FN

order_id	product_id	price	quantity	order_date	shipping_detail
101	1	45.6	30	2019-08-16	31 Hill Haven Drive
101	2	18	12	2019-08-16	31 Hill Haven Drive
102	3	12	61	2020-10-10	12 Breezewood Court
102	4	43.9	10	2020-10-10	12 Breezewood Court
103	3	10	47	2018-03-17	100 West Street
103	1	18	22	2018-03-17	100 West Street

CUMPLE CON 1 FN
SIN DATOS REDUNDANTES
LOS ATRIBUTOS QUE NO FORMAN
PARTE DE LA CLAVE ESTAN
COMPLETAMENTE DEFINIDOS POR
LA CLAVE

2FN

DEPENDENCIA IA PARCIAL



order_id	product_id	price	quantity	order_date	shipping_detail
101	1	45.6	30	2019-08-16	31 Hill Haven Drive
101	2	18	12	2019-08-16	31 Hill Haven Drive
102	3	12	61	2020-10-10	12 Breezewood Court
102	4	43.9	10	2020-10-10	12 Breezewood Court
103	3	10	47	2018-03-17	100 West Street
103	1	18	22	2018-03-17	100 West Street

PRIMARY
KEY
COMPUESTA

NON KEY
COLUMNS

CUMPLE CON 1 FN
SIN DATOS REDUNDANTES
LOS ATRIBUTOS QUE NO FORMAN
PARTE DE LA CLAVE ESTAN
COMPLETAMENTE DEFINIDOS POR
LA CLAVE

2FN

order_id	product_id	price	quantity
101	1	45.6	30
101	2	18	12
102	3	12	61
102	4	43.9	10
103	3	10	47
103	1	18	22

order_id	order_date	shipping_detail
101	2019-08-16	31 Hill Haven Drive
102	2020-10-10	12 Breezewood Court
103	2018-03-17	100 West Street

CUMPLE CON 1 FN
SIN DATOS REDUNDANTES
LOS ATRIBUTOS QUE NO FORMAN
PARTE DE LA CLAVE ESTAN
COMPLETAMENTE DEFINIDOS POR
LA CLAVE

Normalización

3 FN

CUMPLE CON 2 FN

SIN DATOS REDUNDANTES

NO TENER DEPENDENCIAS

TRANSITIVAS (NO INGESTEN

COSAS DE OTRAS TABLAS QUE

NO SEAN PRIMARY KEYS)

3FN

ProductID	ProductName	Supplier ID	Supplier Name	Supplier Address	UnitPrice
1	Rssle Sauerkraut	3	Plutzer	3200 ParkWay	45.6
2	Chartreuse verte	3	Plutzer	3200 ParkWay	18
3	Spegesild	10	Lyngbysild	131 Stadium Drive	12
4	Vegie-spread	10	Lyngbysild	131 Stadium Drive	43.9
5	Chocolate Bar	43	Cedar Choc	20 Barfield Lane	10
6	Choco Milk	43	Cedar Choc	20 Barfield Lane	18

- CUMPLE CON 2 FN
- SIN DATOS REDUNDANTES
- NO TENER DEPENDENCIAS TRANSITIVAS (NO INGESTEN COSAS DE OTRAS TABLAS QUE NO SEAN PRIMARY KEYS)

3FN

DEPENDENCIA
TRANSITIVA

DEPENDEN
CIA

ProductID	ProductName	Supplier ID	Supplier Name	Supplier Address	UnitPrice
1	Rssle Sauerkraut	3	Plutzer	3200 ParkWay	45.6
2	Chartreuse verte	3	Plutzer	3200 ParkWay	18
3	Spegesild	10	Lyngbysild	131 Stadium Drive	12
4	Vegie-spread	10	Lyngbysild	131 Stadium Drive	43.9
5	Chocolate Bar	43	Cedar Choc	20 Barfield Lane	10
6	Choco Milk	43	Cedar Choc	20 Barfield Lane	18

- CUMPLE CON 2 FN
- SIN DATOS REDUNDANTES
- NO TENER DEPENDENCIAS TRANSITIVAS (NO INGESTEN COSAS DE OTRAS TABLAS QUE NO SEAN PRIMARY KEYS)

3FN

ProductID	ProductName	Supplier ID	Supplier Name	Supplier Address	UnitPrice
1	Rssle Sauerkraut	3	Plutzer	3200 ParkWay	45.6
2	Chartreuse verte	3	Plutzer	3200 ParkWay	18
3	Spegesild	10	Lyngbysild	131 Stadium Drive	12
4	Vegie-spread	10	Lyngbysild	131 Stadium Drive	43.9
5	Chocolate Bar	43	Cedar Choc	20 Barfield Lane	10
6	Choco Milk	43	Cedar Choc	20 Barfield Lane	18

- CUMPLE CON 2 FN
- SIN DATOS REDUNDANTES
- NO TENER DEPENDENCIAS TRANSITIVAS (NO INGESTEN COSAS DE OTRAS TABLAS QUE NO SEAN PRIMARY KEYS)

3FN

ProductID	ProductName	Supplier ID	UnitPrice
1	Rssle Sauerkraut	3	45.6
2	Chartreuse verte	3	18
3	Spegesild	10	12
4	Vegie-spread	10	43.9
5	Chocolate Bar	43	10
6	Choco Milk	43	18

Supplier ID	Supplier Name	Supplier Address
3	Plutzer	3200 ParkWay
10	Lyngbysild	131 Stadium Drive
43	Cedar Choc	20 Barfield Lane

- CUMPLE CON 2 FN
- SIN DATOS REDUNDANTES
- NO TENER DEPENDENCIAS TRANSITIVAS (NO INGESTEN COSAS DE OTRAS TABLAS QUE NO SEAN PRIMARY KEYS)

DUDAS/CONSULTAS

pablo.jav.octaviano@gmail.com

jfranciscopinelli@gmail.com

