

Facultad de Ingeniería



Normalización

Tema V

Semestre 2020-2

El alumno comprenderá y aplicará los conceptos del proceso de Normalización de Bases de Datos con la finalidad de implementar mejores diseños, estableciendo un equilibrio entre los niveles de redundancia y desempeño establecidos en los requerimientos no funcionales de casos de estudio.

Propiedades:

- **No pueden existir dos relaciones que se llamen igual**
- **No pueden existir tuplas iguales**
- **No pueden existir atributos que tengan el mismo nombre**
- **No hay orden en tuplas ni en atributos**
- **Los valores de los atributos deben ser atómicos**

Concepto introducido por Frank Codd en su artículo “A relational model of data for large shared data banks”, motivado por las anomalías que se presentaban en los conjuntos de datos.

Técnica empleada en el diseño de bases de datos que busca reducir redundancias e inconsistencias en los datos -> Organización en los datos

- **Datos más consistentes**
- **Reducir datos duplicados
(reducir tamaño)**
- **Integridad de la base de datos**
- **Tablas más pequeñas**
- **Buenas prácticas**

- **Hay que tener nociones sólidas del concepto para llevar a cabo con éxito este proceso**
- **Se van a presentar más datos para ser unidos (joins entre tablas), lo que puede tomar tiempos considerables en ciertos casos -> Rendimiento**

- **Las tablas contendrán códigos en vez de datos reales, lo que implica buscar ese código en otra(s) tabla(s)**

SalesStaff						
<u>EmployeeID</u>	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

¿Qué problema(s) notan?

SalesStaff						
<u>EmployeeID</u>	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

¿Qué problema(s) pueden pasar cuando se agregue información?

SalesStaff						
<u>EmployeeID</u>	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

¿Qué problema(s) pueden surgir al actualizar información?

SalesStaff						
<u>EmployeeID</u>	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

¿Qué problema(s) pueden surgir al borrar información?

SalesStaff						
<u>EmployeeID</u>	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

¿Qué problema(s) pueden surgir al buscar información?

Un grupo de repetición es un conjunto de valores que ocurren varias veces en un registro.

EmployeeID	Name	Project	Time
EN1-26	Sean O'Brien	30-452-T3, 30-457-T3, 32-244-T3	0.25, 0.40, 0.30
EN1-33	Amy Guya	30-452-T3, 30-382-TC, 32-244-T3	0.05, 0.35, 0.60
EN1-35	Steven Baranco	30-452-T3, 31-238-TC	0.15, 0.80
EN1-36	Elizabeth Roslyn	35-152-TC	0.90
EN1-38	Carol Schaaf	36-272-TC	0.75
EN1-40	Alexandra Wing	31-238-TC, 31-241-TC	0.20, 0.70

EmpID	Last Name	First Name	Project1	Time1	Project2	Time2	Project3	Time3
EN1-26	O'Brien	Sean	30-452-T3	0.25	30-457-T3	0.40	32-244-T3	0.30
EN1-33	Guya	Amy	30-452-T3	0.05	30-382-TC	0.35	32-244-T3	0.60
EN1-35	Baranco	Steven	30-452-T3	0.15	31-238-TC	0.80		
EN1-36	Roslyn	Elizabeth	35-152-TC	0.90				
EN1-38	Schaaf	Carol	36-272-TC	0.75				
EN1-40	Wing	Alexandra	31-238-TC	0.20	31-241-TC	0.70		

¿Cómo identificar la llave primaria?

¿Cómo identificar llaves candidatas?

Una relación está en 1FN si no presenta grupos de repetición y cada columna contiene valores atómicos.

¿Qué vamos a conseguir al aplicar la 1FN?

- **Identificar la PK de cada tabla**
- **Evitar atributos multivaluados o atributos que representan lo mismo**
- **Evitar duplicidad de registros**

Sea la tabla CLIENTE:

ID_Cliente	nombre	apellido	telefono
123	Juan	López	12123212, 43556786
547	Irma	Arriaga	56567364
490	Pablo	Juárez	12334657, 65784532

¿Cumple la 1FN?

Ejemplo



ID_Cliente	nombre	apellido	telefono1	telefono2
123	Juan	López	12123212	43556786
547	Irma	Arriaga	56567364	56567364
490	Pablo	Juárez	12334657	65784532

¿Cumple la 1FN?

Ejemplo



ID_Cliente	nombre	apellido	telefono
123	Juan	López	12123212
123	Juan	López	43556786
547	Irma	Arriaga	56567364
490	Pablo	Juárez	12334657
490	Pablo	Juárez	65784532

¿Cumple la 1FN?
Sí, pero...

Ejemplo



ID_Cliente	nombre	apellido
123	Juan	López
547	Irma	Arriaga
490	Pablo	Juárez

ID_Cliente	telefono
123	12123212
123	43556786
547	56567364
490	12334657
490	65784532

¿Cumple la 1FN?

Partiendo de:

DNI	Nombre	Codigo_Tienda	Direccion_Tienda	turno	fecha
33445566	Paola Martin	100A	Transmisiones Miliars 70	M	02/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	M	02/01/2020
86923456	Daniel Diaz	100A	Transmisiones Miliars 70	T	02/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	T	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	03/01/2020
45678367	Francisco Monte	200B	Periférico Norte 80	M	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	04/01/2020
45678367	Francisco Monte	100A	Transmisiones Miliars 70	M	04/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	T	04/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	M	05/01/2020

Relaciones parciales:

Una dependencia funcional (DF) que ocurre en una relación es parcial cuando la eliminación de uno de los atributos determinantes genera una DF que sigue ocurriendo en la relación

Relaciones parciales:

$$\{X, Y\} \rightarrow Z$$

$$X \rightarrow Z$$

**Lo anterior implica que Z
depende parcialmente de {X, Y}**

Sea la tabla estudiante_Proyecto

id_Estudiante	id_Proyecto	nom_Est	nom_Proj
S01	P10	Laura	Bases de datos
S02	P32	Juan	Clúster

- **Identificar PK y candidatas**
- **Validar las dependencias con los atributos restantes**

Sea la tabla estudiante_Proyecto

id_Estudiante	id_Proyecto	nom_Est	nom_Proy
S01	P10	Laura	Bases de datos
S02	P32	Juan	Clúster

¿Qué tipo de dependencia hay?

Notación en el proceso de normalización:

- Notación de dependencia funcional:

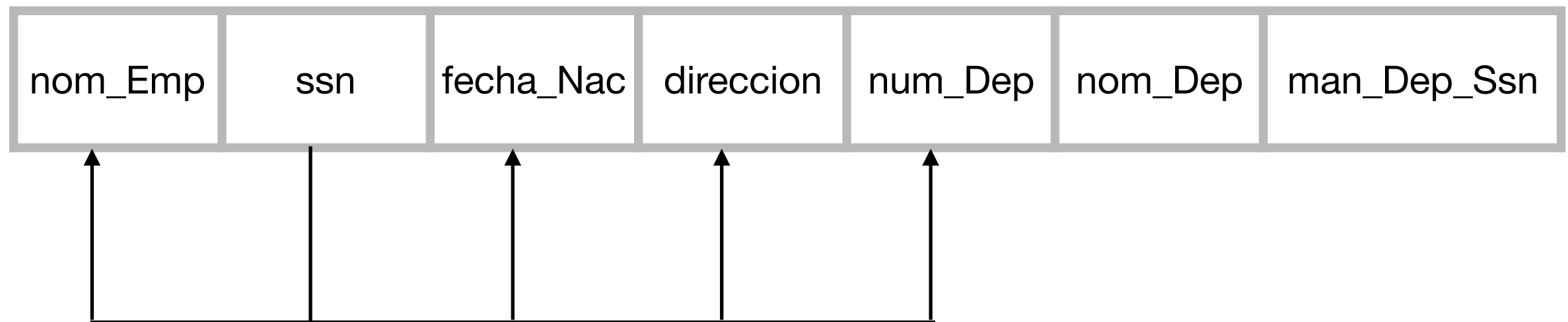
$\{A, B\} \rightarrow C$

$A \rightarrow C$

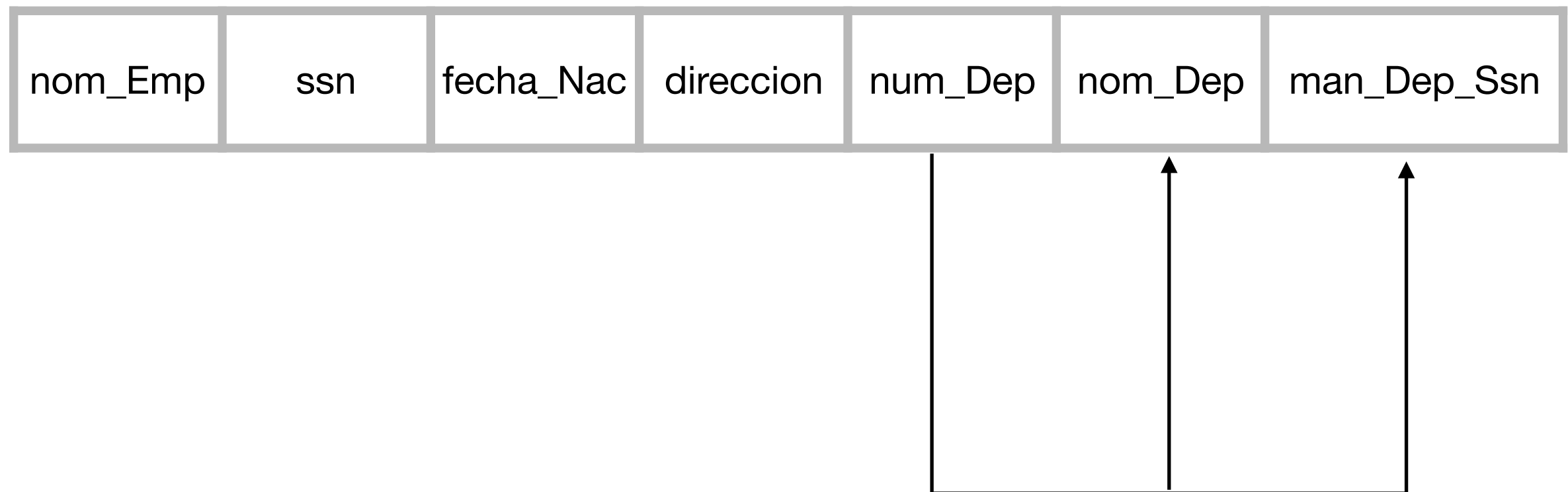
- Diagramas de dependencias:

nom_Emp	ssn	fecha_Nac	direccion	num_Dep	nom_Dep	man_Dep_Ssn
---------	-----	-----------	-----------	---------	---------	-------------

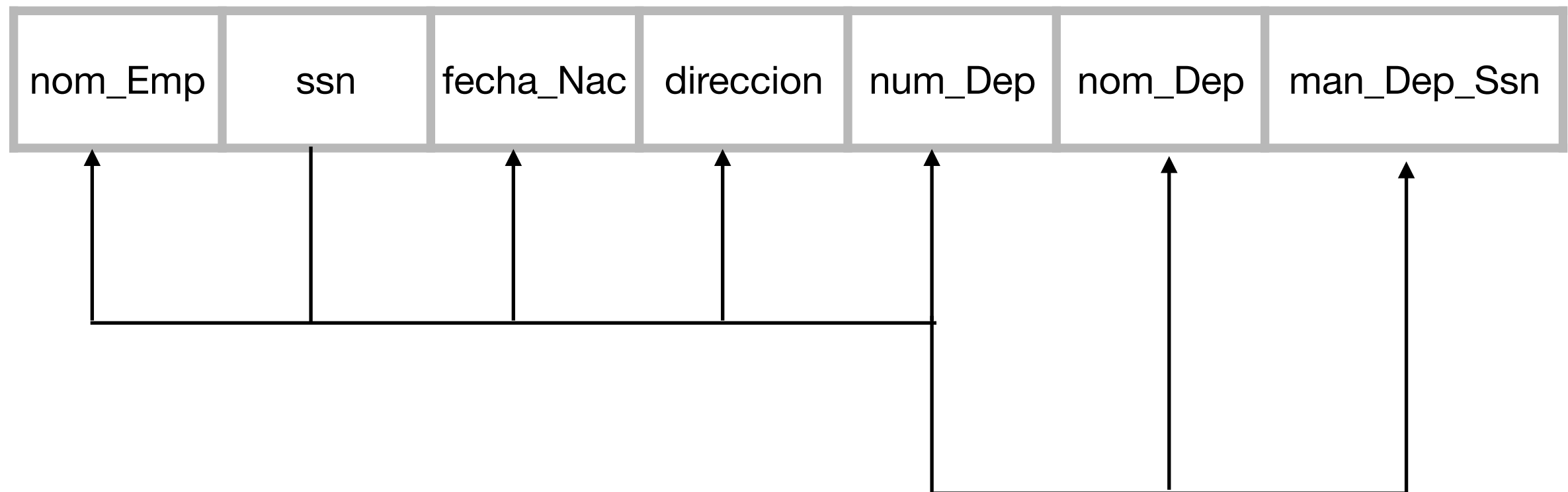
- Diagramas de dependencias:



- Diagramas de dependencias:



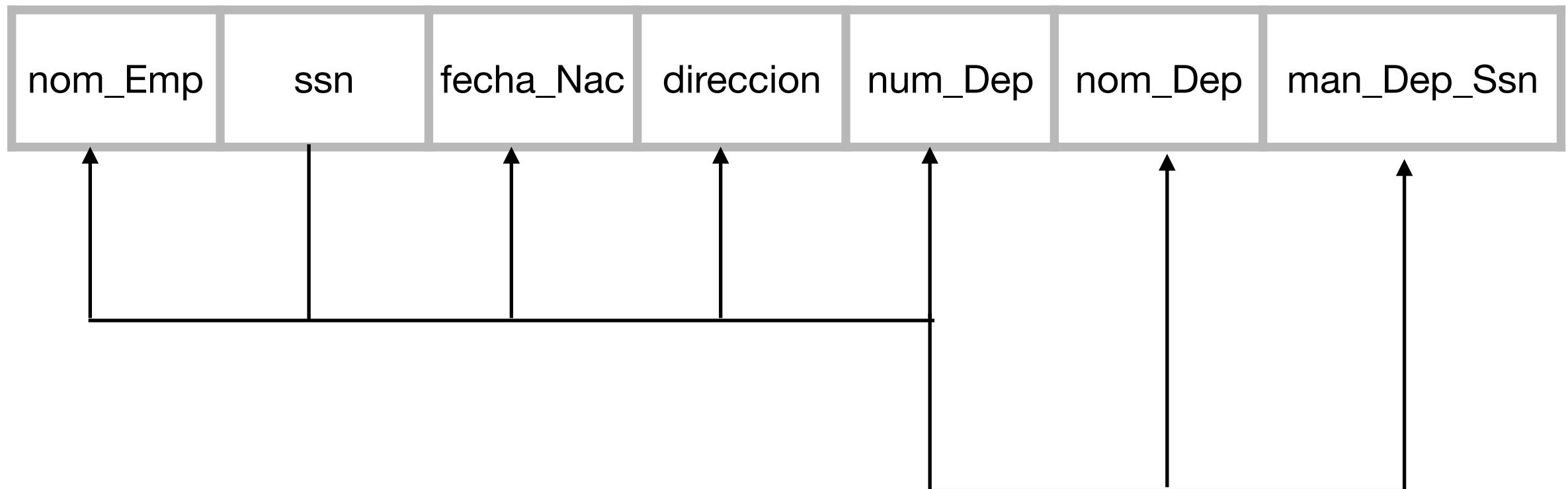
- Diagramas de dependencias:



A	B	C	D	E	F	G
---	---	---	---	---	---	---

B -> {A, C, D, E}

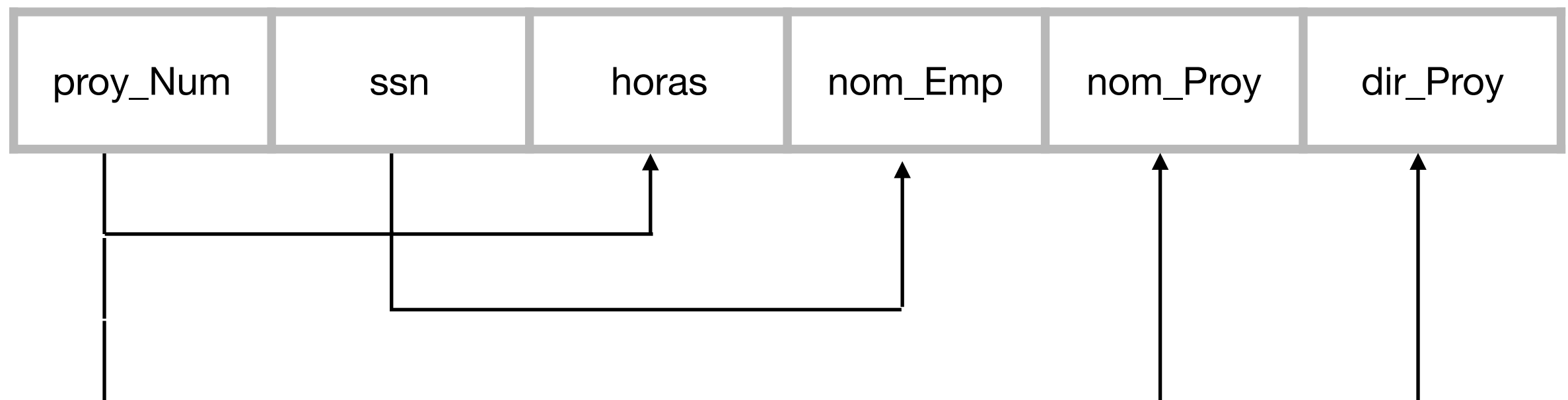
E -> {F, G}



- Diagramas de dependencias:

proy_Num	ssn	horas	nom_Emp	nom_Proj	dir_Proj
----------	-----	-------	---------	----------	----------

- Diagramas de dependencias:



Una tabla se encuentra en 2FN si y sólo si:

- **Se encuentra en 1FN**
- **Ninguno de sus atributos no-principales son funcionalmente dependientes en una parte de una llave primaria/candidata**

Sea la tabla EMPLEADO:

ID_Empleado	ID_Depto	Ubicacion
EMP-01	DEP_BD	Yucatán
EMP-02	DEP_SD	CDMX
EMP-03	DEP_CB	Guanajuato
EMP-04	DEP_CT	Torreón

¿Cumple la 1FN?

{id_Empleado, id_Depto} -> Ubicacion

id_Depto -> Ubicacion

Normalizando:

ID_Empleado	ID_Depto
EMP-01	DEP_BD
EMP-02	DEP_SD
EMP-03	DEP_CB
EMP-04	DEP_CT

ID_Depto	Ubicacion
DEP_BD	Yucatán
DEP_SD	CDMX
DEP_CB	Guanajuato
DEP_CT	Torreón

La siguiente tabla contiene información de fabricantes de cepillos de dientes:

Fabricante	Modelo	Nombre completo del modelo	País del fabricante
Forte	X-Prime	Forte X-Prime	Italia
Forte	Ultraclean	Forte Ultraclean	Italia
Dent-o-Fresh	EZBrush	Dent-o-Fresh EZBrush	USA
Kobayashi	ST-60	Kobayashi ST-60	Japón
Hoch	Toothmaster	Hoch Toothmaster	Alemania
Hoch	Contender	Hoch Contender	Alemania

¿Cumple la 2FN?

Partiendo de:

DNI	Nombre	Codigo_Tienda	Direccion_Tienda	turno	fecha
33445566	Paola Martin	100A	Transmisiones Miliars 70	M	02/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	M	02/01/2020
86923456	Daniel Diaz	100A	Transmisiones Miliars 70	T	02/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	T	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	03/01/2020
45678367	Francisco Monte	200B	Periférico Norte 80	M	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	04/01/2020
45678367	Francisco Monte	100A	Transmisiones Miliars 70	M	04/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	T	04/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	M	05/01/2020

Partiendo de:

staffNo	branchNo	branchAddress	name	position	hoursPerWeek
S4555	B002	City Center Plaza, Seattle, WA 98122	Ellen Layman	Assistant	16
S4555	B004	16 – 14th Avenue, Seattle, WA 98128	Ellen Layman	Assistant	9
S4612	B002	City Center Plaza, Seattle, WA 98122	Dave Sinclair	Assistant	14
S4612	B004	16 – 14th Avenue, Seattle, WA 98128	Dave Sinclair	Assistant	10



Solución:

Ejercicio

Solución:

staffNo	name	position
S4555	Ellen Layman	Assistant
S4612	Dave Sinclair	Assistant

staffNo	branchNo	hoursPerWeek
S4555	B002	16
S4555	B004	9
S4612	B002	14
S4612	B004	10

branchNo	branchAddress
B002	City Center Plaza, Seattle, WA 98122
B004	16 – 14th Avenue, Seattle, WA 98128

Relaciones transitivas:

$X \rightarrow Y$

$Y \rightarrow Z$

$\text{¿}X \rightarrow Z\text{?}$

Relaciones transitivas:

Supongamos que los estudiantes solo pueden estar matriculados en un solo curso y supongamos que los profesores solo pueden dar un curso

Relaciones transitivas:

Entonces tenemos que ID_Estudiente determina a Curso_Tomado y el Curso_Tomado determina a Profesor_Asignado, indirectamente podemos saber a través del ID_estudiante el Profesor_Asignado.

Una tabla se encuentra en 3FN si y sólo si:

- **Se encuentra en 2FN**
- **Cualquier atributo no-principal de la tabla sea no transitivamente dependiente de cada clave candidata de la misma**

Student Id	Student Name	Subject Id	Subject	Address
1DT15ENG01	Alex	15CS11	SQL	Goa
1DT15ENG02	Barry	15CS13	JAVA	Bengaluru
1DT15ENG03	Clair	15CS12	C++	Delhi
1DT15ENG04	David	15CS13	JAVA	Kochi

Identificar PK: student_ID

Identificar candidatas: no hay

Identificar TODAS las dependencias:

A -> {B, C, D, E}

C -> D

Tanto C como D son atributos no-principales, por lo que la tabla no esta en 3FN

$A \rightarrow \{B, C, D, E\}$

$C \rightarrow D$

Quitamos la dependencia transitiva

Student Id	Student Name	Subject Id	Subject	Address
1DT15ENG01	Alex	15CS11	SQL	Goa
1DT15ENG02	Barry	15CS13	JAVA	Bengaluru
1DT15ENG03	Clair	15CS12	C++	Delhi
1DT15ENG04	David	15CS13	JAVA	Kochi

Sin perder información ni la relación con la tabla base

Subject Id	Subject
15CS11	SQL
15CS13	JAVA
15CS12	C++
15CS13	JAVA

Partiendo de:

staffNo	name	position	salary	branchNo	branchAddress	telNo
S1500	Tom Daniels	Manager	46000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0003	Sally Adams	Assistant	30000	B001	8 Jefferson Way, Portland, OR 97201	503-555-3618
S0010	Mary Martinez	Manager	50000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S3250	Robert Chin	Supervisor	32000	B002	City Center Plaza, Seattle, WA 98122	206-555-6756
S2250	Sally Stern	Manager	48000	B004	16 – 14th Avenue, Seattle, WA 98128	206-555-3131
S0415	Art Peters	Manager	41000	B003	14 – 8th Avenue, New York, NY 10012	212-371-3000

Partiendo de:

DNI	Nombre	Codigo_Tienda	Direccion_Tienda	turno	fecha
33445566	Paola Martin	100A	Transmisiones Miliars 70	M	02/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	M	02/01/2020
86923456	Daniel Diaz	100A	Transmisiones Miliars 70	T	02/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	T	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	03/01/2020
45678367	Francisco Monte	200B	Periférico Norte 80	M	03/01/2020
12234456	Emiliano López	300C	Av. Universidad 3000	M	04/01/2020
45678367	Francisco Monte	100A	Transmisiones Miliars 70	M	04/01/2020
44552345	Laura Sanz	100A	Transmisiones Miliars 70	T	04/01/2020
33445566	Paola Martin	200B	Periférico Norte 80	M	05/01/2020

**VENTA: {num_Auto, fecha_Venta,
vendedor, comision, descuento}**

num_Auto -> fecha_Venta

num_Auto -> descuento

fecha_Venta -> descuento

vendedor -> comision

**Considerar que un auto puede ser
vendido por muchos vendedores**

Ejercicio



La siguiente tabla almacena información sobre las fechas y duración de las actividades deportivas que se realizan en un colegio

Ejercicio



cod_Act	nom_Act	id_Prof	nom_Pr	sede	fecha	hora_I	hora_Fin
1	pilates	12345	juan	pabellon	9/9/2007	10:00	11:00
2	fitness	76454	javier	pabellon	9/9/2007	10:00	11:00
1	pilates	12345	juan	pabellon	11/9/2007	9:30	11:00
1	pilates	12345	juan	pabellon	15/9/2007	12:00	13:00
3	yoga	76454	javier	multiusos	15/9/2007	9:00	10:00
4	gimnasia	12345	juan	multiusos	1/10/2007	12:00	13:00
3	yoga	76454	javier	multiusos	15/9/2007	11:00	12:00
4	gimnasia	88903	lidia	pabellon	1/10/2007	12:00	14:00
2	fitness	88903	lidia	pabellon	9/9/2007	10:00	11:00
1	pilates	76454	javier	multiusos	9/9/2007	10:00	12:00

- **Indicar un ejemplo de problema a la hora de borrar un dato**
- **Dibujar el diagrama de dependencias**
- **Normalizar hasta 3FN**

$\{A, C, E\} \rightarrow \{B, D, F, G, H\}$

Niveles de aislamiento en las BD's

A -> B

C -> D

{A, C, E} -> {F, G, H}

Niveles de aislamiento en las BD's

Ejercicio



Ordenes

Id_orden	Fecha	Id_cliente	Nom_cliente	Estado	Num_art	nom_art	cant	Precio
2301	23/02/11	101	Martin	Caracas	3786	Red	3	35,00
2301	23/02/11	101	Martin	Caracas	4011	Raqueta	6	65,00
2301	23/02/11	101	Martin	Caracas	9132	Paq-3	8	4,75
2302	25/02/11	107	Herman	Coro	5794	Paq-6	4	5,00
2303	27/02/11	110	Pedro	Maracay	4011	Raqueta	2	65,00
2303	27/02/11	110	Pedro	Maracay	3141	Funda	2	10,00

**Tabla1(id_Orden, fecha, id_Cliente,
nombre_Cliente, estado)**

**Tabla2 (id_orden, no_Articulo,
nombre_Articulo, cantidad, precio)**

**Tabla1(id_Orden, fecha, id_Cliente,
nombre_Cliente, estado)**

Tabla2 (id_orden, no_Articulo, cantidad)

**Tabla3(no_Articulo, nombre_Articulo,
precio)**

**Tabla1(id_Orden, fecha, id_Cliente,
nombre_Cliente, estado)**

Tabla2 (id_orden, no_Articulo, cantidad)

**Tabla3(no_Articulo, nombre_Articulo,
precio)**

Tabla1(id_Orden, fecha, id_Cliente)

Tabla4(id_Cliente, nombre_Cliente, estado)

Tabla2 (id_orden, no_Articulo, cantidad)

Tabla3(no_Articulo, nombre_Articulo, precio)

- **La PK es id_Orden**
- **Dibujar el diagrama de dependencias**
- **Normalizar hasta 3FN**

Sea la relacion Ventas(NumVenta, NoRegCarro, NoVendedor, NoCliente, Precio, Modelo, Marca, NombreCliente, NombreVendedor, FechadeVenta) de un sistema de captura de órdenes. Funciones de dependencia:

Numventa -> {NoRegCarro, NoVendedor, NoCliente, Precio, FechadeVenta }

NoRegCarro -> {Modelo, Marca}

NoVendedor -> NombreVendedor

NoCliente -> NombreCliente

- **¿Cuál de los atributos puede ser empleado como llave primaria y por qué?**
- **¿En qué forma normal se encuentra la relación ventas? Justificar respuesta**
- **Normalizar hasta 3FN**