

Facultad de Ingeniería



Normalización

Tema V

Semestre 2023-1



UNAM
UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

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?

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

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

EmpID	Last Name	First Name	Project1	Time	Project	Time	Project	Time
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		

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		

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

- Descomponer atributos compuestos**
- Atributos con el mismo dominio**
- Evitar atributos que se tengan el mismo nombre**

¿Qué vamos a conseguir al aplicar la 1FN?

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

- **Evitar anomalías de actualización y pérdidas de información**

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...

¿Cómo conseguir 1FN en una tabla?

- **Opción 1: Crear una nueva tabla, la cual contendrá el atributo multivaluado (PK)* y la llave primaria de la tabla que se está normalizando (FK)**

Ejemplo



PK

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

FK

PK

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

¿Cumple la 1FN?

¿Cómo conseguir 1FN en una tabla?

- **Opción 2: Crear una llave primaria compuesta, con el atributo multivaluado y el atributo(s) más cercano a ser llave primaria**

Ejemplo



PK			PK	
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?

Ejemplo



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

Ejemplo



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

- ¿Hay atributos multivaluados o grupos de repetición?
- ¿Existe llave primaria?

Ejemplo



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

EN1-26	SEAN...	30-452-T3	0.25
EN1-26	SEAN...	30-457-T3	0.40
EN1-26	SEAN...	32-244-T3	0.30
EN1-33	AMY	30-452-T3	0.05
EN1-33	AMY	30-382-TC	0.35
EN1-33	AMY	32-244-T3	0.60
...

Tarea 12



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

$$DF : X \Rightarrow Y$$

$$t_1 X = t_2 X$$

$$t_1 Y = t_2 Y$$

Dependencias funcionales



num_Cuenta	nombre	calificacion	academia	materia
1	Juan	7	BD	Bases de datos
2	Maria	6	Redes	Criptografía
3	Juan	7	BD	Bases de datos avanzadas
4	Maria	6	Redes	Arq. C/S
5	Carlos	8	Software	Sistemas embebidos
6	Karla	8	PM	Admon. de proyectos

num_cuenta -> nombre
nombre -> num_cuenta
num_cuenta -> calificacion
academia -> materia
materia -> academia
{num_cuenta, nombre} -> calificacion
nombre -> calificacion
{nombre, calificacion} -> academia

num_cuenta -> {nombre, calificacion}
{academia, materia} -> nombre
{num_cuenta, calificacion} ->
academia
nombre -> materia
{nombre, calificacion, academia} ->
num_cuenta

Notación en el proceso de normalización:

- Notación de dependencia funcional:

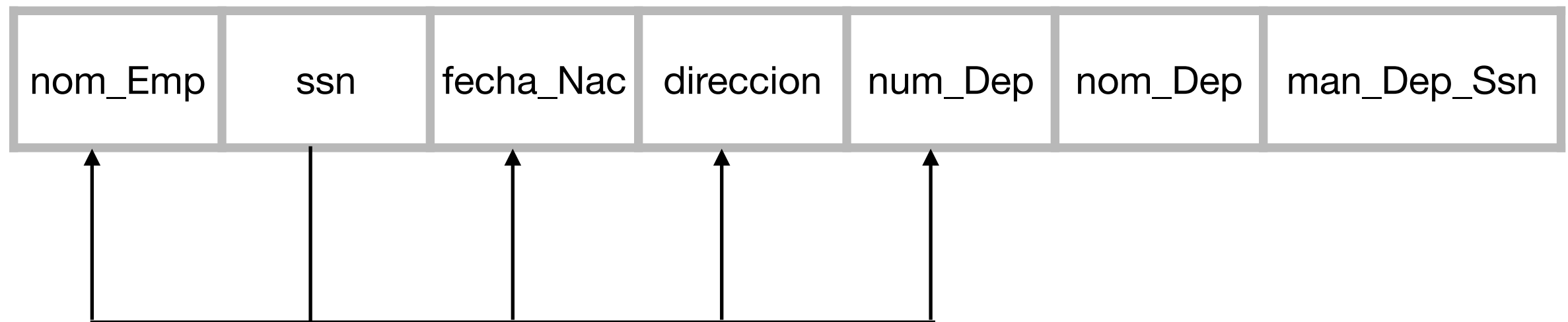
$\{A, B\} \rightarrow C$

$A \rightarrow C$

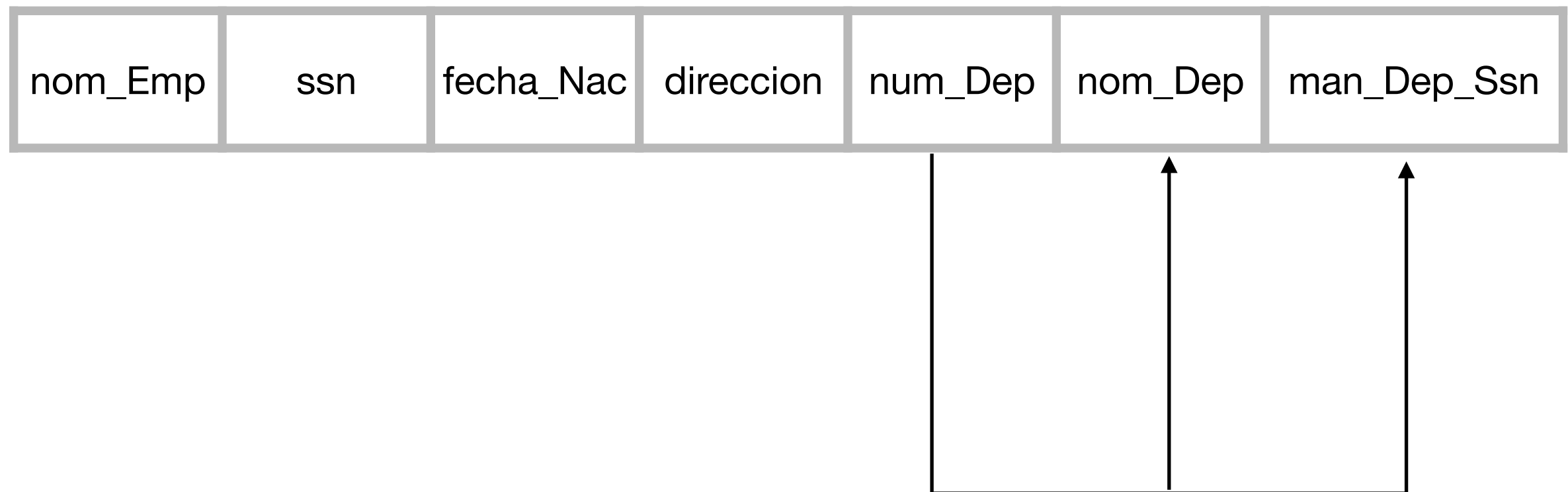
- Diagramas de dependencias:

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

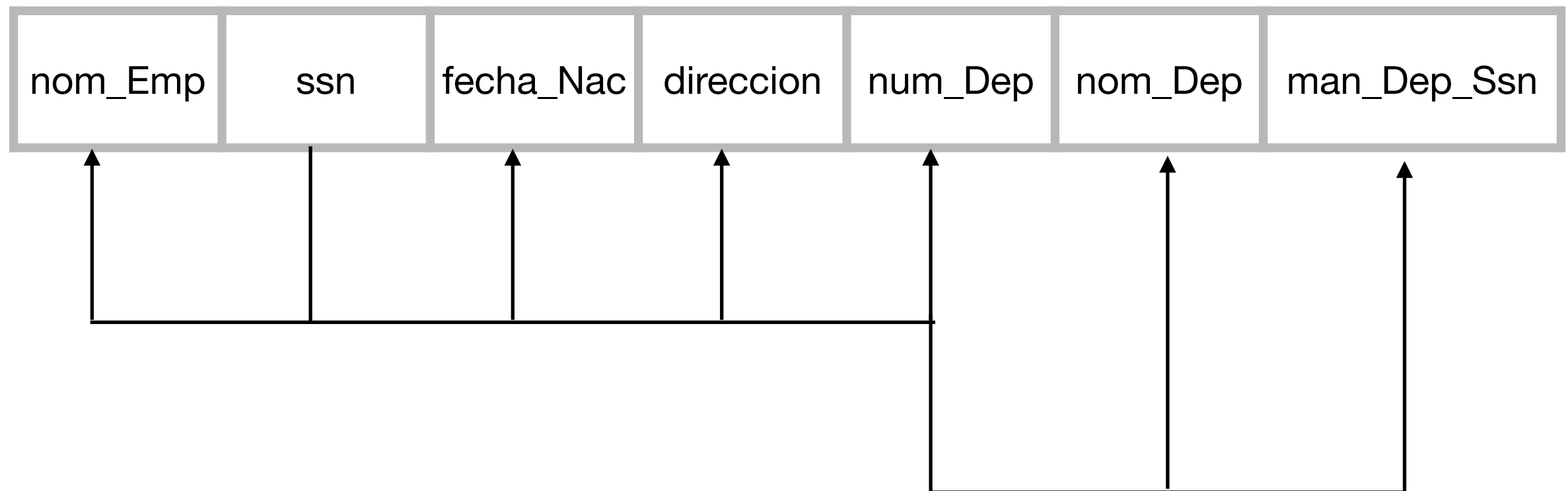
- Diagramas de dependencias:



- Diagramas de dependencias:



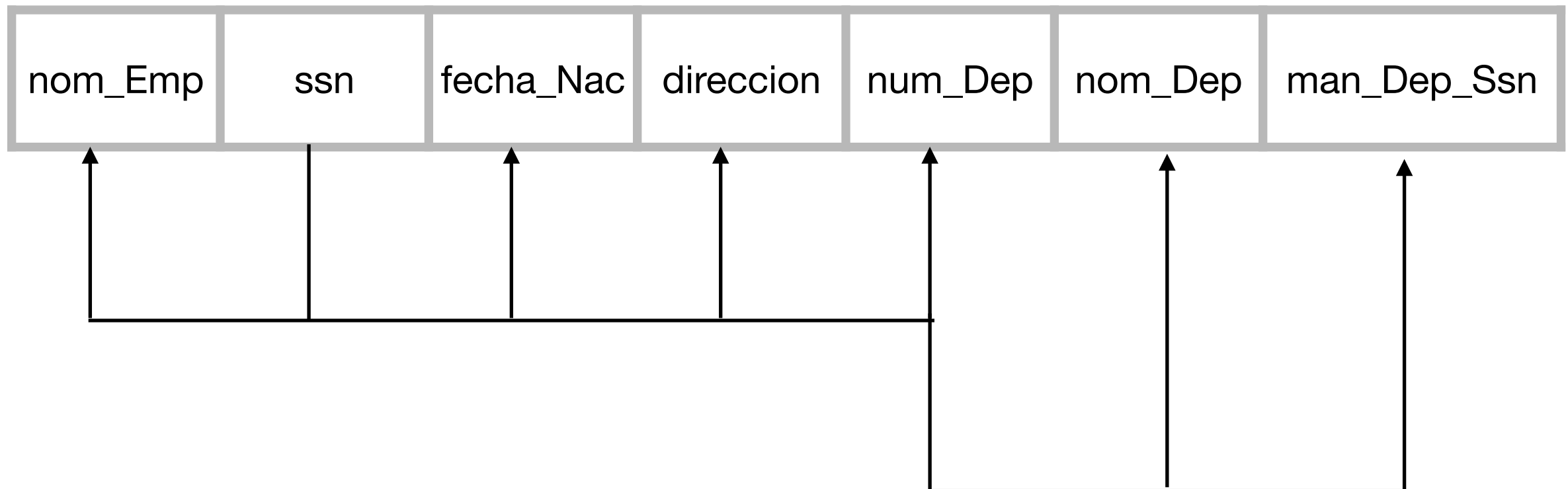
- Diagramas de dependencias:



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

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

$E \rightarrow \{F, G\}$



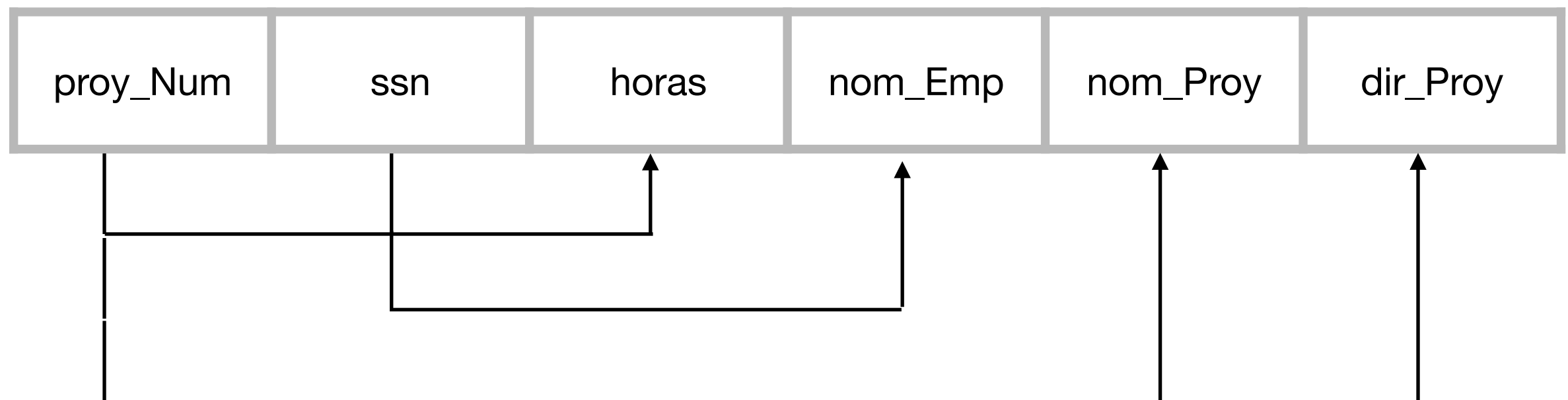
Ejemplo

- Indicar dependencias

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

Ejemplo

- 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**

Dependencias 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

Dependencias 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?

$X \rightarrow Y$

$\{\text{id_Estudiante, id_proyecto}\} \rightarrow$
 nom_Est, nom_Proy

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

X -> Y

{id_Estudiante, id_proyecto} ->

nom_Est, nom_Proj

id_Estudiante -> nom_Est

id_Proj -> nom_Proj

Sea la tabla EMPLEADO:

ID_Emplead	ID_Depto	Ubicacion	Nombre
EMP-01	DEP_BD	Yucatán	Guillermo
EMP-02	DEP_SD	CDMX	Fernando
EMP-03	DEP_CB	Guanajuato	Aldo
EMP-04	DEP_CT	Torreón	Carlos

¿Dependencias?

Sea la tabla EMPLEADO:

ID_Emplead	ID_Depto	Ubicacion	Nombre
EMP-01	DEP_BD	Yucatán	Guillermo
EMP-02	DEP_SD	CDMX	Fernando
EMP-03	DEP_CB	Guanajuato	Aldo
EMP-04	DEP_CT	Torreón	Carlos

X Y Z W

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

$Y \rightarrow Z$

$X \rightarrow W$

Sea la tabla EMPLEADO:

¿Cumple la 1FN? -> Sí, ya que no hay valores multivaluados ni grupos de rep.

¿Cumple la 2FN? No, porque existen dependencias funcionales parciales

Ejemplo



{id_Empleado, id_Depto} -> {Ubicacion, Nombre}

{id_Empleado, id_Depto} -> {}

id_Depto -> Ubicacion

id_Empleado -> Nombre

Normalizando:

COLABORA

ID Empleado	ID Depto
<u>EMP-01</u>	<u>DEP BD</u>
<u>EMP-02</u>	<u>DEP SD</u>
<u>EMP-03</u>	<u>DEP CB</u>
<u>EMP-04</u>	<u>DEP CT</u>

ID Depto	Ubicacion
<u>DEP BD</u>	Yucatán
<u>DEP SD</u>	CDMX
<u>DEP CB</u>	Guanajuato
<u>DEP CT</u>	Torreón

DEPARTAMENTO

ID Empleado	Nombre
<u>EMP-01</u>	Guillermo
<u>EMP-02</u>	Fernando
<u>EMP-03</u>	Aldo
<u>EMP-04</u>	Carlos

EMPLEADO

Sea la tabla EMPLEADO:

<u>ID Empleado</u>	<u>ID Depto</u>	Horas
FMP-01	DFP RD	10
<u>EMP-02</u>	<u>DEP SD</u>	5
<u>EMP-03</u>	<u>DEP CB</u>	15
<u>EMP-04</u>	<u>DEP CT</u>	13
<u>EMP-01</u>	<u>DEP SD</u>	7

Sea la tabla EMPLEADO:

<u>ID Empleado</u>	<u>ID Depto</u>	Horas
FMP-01	DFP RD	10
<u>EMP-02</u>	<u>DEP SD</u>	5
<u>EMP-03</u>	<u>DEP CB</u>	15
<u>EMP-04</u>	<u>DEP CT</u>	13
<u>EMP-01</u>	<u>DEP SD</u>	7

$\{x, y\} \rightarrow z$

¿Cumple la 1FN? \rightarrow Sí, ya que todos los valores son atómicos y no hay grupos de repetición

Sea la tabla EMPLEADO:

¿Cumple la 1FN? -> Sí, ya que todos los valores son atómicos y no hay grupos de repetición

¿Cumple la 2FN? -> Sí, porque no hay dependencias funcionales parciales

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

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

Tarea 13



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

Dependencias transitivas:

$X \rightarrow Y$

$Y \rightarrow Z$

$\text{¿}X \rightarrow Z\text{?}$

Relaciones transitivas:

Supongamos que Juan es hermano de Pedro. Además, Pedro es hermano de Javier, por lo que podemos concluir que Juan es hermano de Javier.

Relaciones transitivas:

Entonces tenemos que Juan determina a Pedro y Pedro determina a Javier, indirectamente podemos saber que hay relación entre Juan y Javier.

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

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

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

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

FK

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

Separar la transitividad**PK**

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

Técnica de optimización en bases de datos, en la cual, de manera premeditada, se agrega redundancia de información a una o más tablas.

Técnica de optimización en bases de datos, en la cual, de manera premeditada, se agrega redundancia de información a una o más tablas.

Métodos:

- **Agregar columnas redundantes**
- **Agregar columnas derivadas**
- **Juntar tablas**

EMPLOYEE				
EMP_ID	EMP_NAME	ADDRESS	DEPT_ID	PROJ_ID
100	Joseph	Clinton Town	10	206
101	Rose	Fraser Town	20	205
102	Mathew	Lakeside Village	10	206
103	Stewart	Troy	30	204
104	William	Holland	30	202

DEPARTMENT	
DEPT_ID	DEPT_NAME
10	Accounting
20	Quality
30	Design

Se te solicita generar un reporte de los empleados y el nombre de su departamento. ¿Qué harías?

Denormalización



EMPLOYEE					
EMP_ID	EMP_NAME	ADDRESS	DEPT_ID	PROJ_ID	DEPT_NAME
100	Joseph	Clinton Town	10	206	Accounting
101	Rose	Fraser Town	20	205	Quality
102	Mathew	Lakeside Village	10	206	Accounting
103	Stewart	Troy	30	204	Design
104	William	Holland	30	202	Design

DEPARTMENT	
DEPT_ID	DEPT_NAME
10	Accounting
20	Quality
30	Design

Ventajas:

- Obtención rápida de información**
- Se evita el uso de joins**
- Consultas más fáciles de leer y elaborar**

Desventajas:

- Específica a un requerimiento en particular**
- Implica mayor espacio de almacenamiento**
- Redundancia de información**
- Dificultad para agregar y actualizar información**
- Pérdida de información**

Ejercicio

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