Facultad de Ingeniería



Normalización

Tema V

Semestre 2022-2



Objetivo



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.



Recordando MR



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



Definición



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.



Definición



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



Ventajas



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



Desventajas



- 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



Desventajas



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





SalesStaff							
EmployeeID	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?





Not null	Not null
11011111	HOLLIGIE

SalesStaff							
EmployeeID	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			

X Los Angeles

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





SalesStaff							
EmployeeID	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

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



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?



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

¿Cumple la 1FN?



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...





PK

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

PK

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

123 55667788

¿Cumple la 1FN?



Ejercicio



PK: {DNI, Codigo_Tienda, fecha}

¿Cumple 1FN?

Partiendo de:

- no hay atts. multivaluados
- no hay grupos de repetición

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

$${X, Y} -> Z$$

 $X -> Z$

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



Sea la tabla estudiante_Proyecto

PK

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

S01 P40 Laura Redes

- 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 -> Y

{id_Estudiante, id_proyecto} ->
nom_Est, nom_Proy





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

X -> Y
{id_Estudiante, id_proyecto} ->
nom_Est, nom_Proy
id_Estudiante -> nom_Est
id_Proy -> nom_Proy



S4612

B004



Α	В	C	D	E	F
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

Dave Sinclair

PK: {A, B} {A,B} -> C, D, E, F A -> D, E B -> C {A, B} -> F

16 - 14th Avenue, Seattle, WA 98128

¿Cumple 2FN?
-¿Hay dependencia funcional parcial?
Sí

Assistant

10

La tabla mostrada no cumple 2FN

¿Cumple 1FN?

- ¿Hay atributos multivaluados?: No
- ¿Hay grupos de repetición? No

La tabla mostrada cumple 1FN

Ejercicio



Solución:

staffNo	name	position
S4555	Ellen Layman	Assistant
S4612	Dave Sinclair	Assistant

staffNo	ffNo branchNo hoursPer		
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





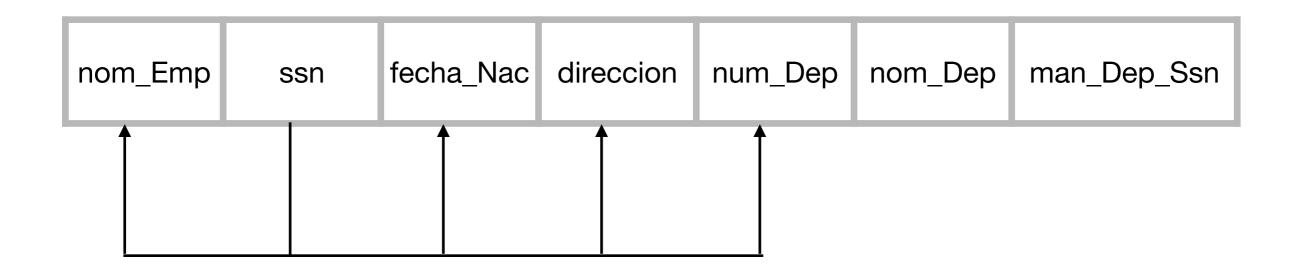
Notación en el proceso de normalización:

 Notación de dependencia funcional:



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

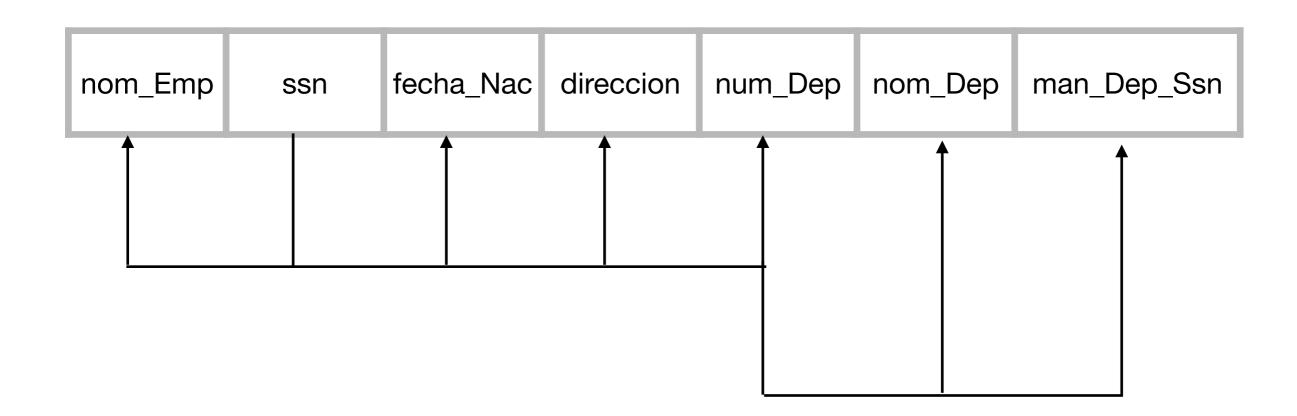






nom_Emp	ssn	fecha_Nac	direccion	num_Dep	nom_Dep	man_Dep_Ssn







A B C D E F G

nom_Emp	ssn	fecha_Nac	direccion	num_Dep	nom_Dep	man_Dep_Ssn



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

- Se encuentra en 1FN
- Ninguno de sus atributos noprincipales son funcionalmente dependientes en una parte de una llave primaria/candidata





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

 \mathbf{X} \mathbf{Y} \mathbf{Z} \mathbf{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 ¿Cumple la 2FN? No, porque existen dependencias funcionales parciales





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

{id_Empleado, id_Depto} -> {}

id_Depto -> Ubicacion

id_Empleado -> Nombre





Normalizando:

COLABORA

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

ID Depto	Ubicacion
DEP BD	Yucatán
DEP SD	CDMX
DEP CB	Guanajuato
DEP_CT	Torreón

DEPARTAMENTO

ID Empleado	Nombre
EMP-01	Guillermo
EMP-02	Fernando
EMP-03	Aldo
EMP-04	Carlos

EMPLEADO





Sea la tabla EMPLEADO:

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

 ${x, y} -> z$

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





Sea la tabla EMPLEADO:

¿Cumple la 1FN? -> Sí, ya que todos los alores son atómicos y no hay grupos de repetición ¿Cumple la 2FN? -> Sí, porque no hay

dependencias funcionales parciales





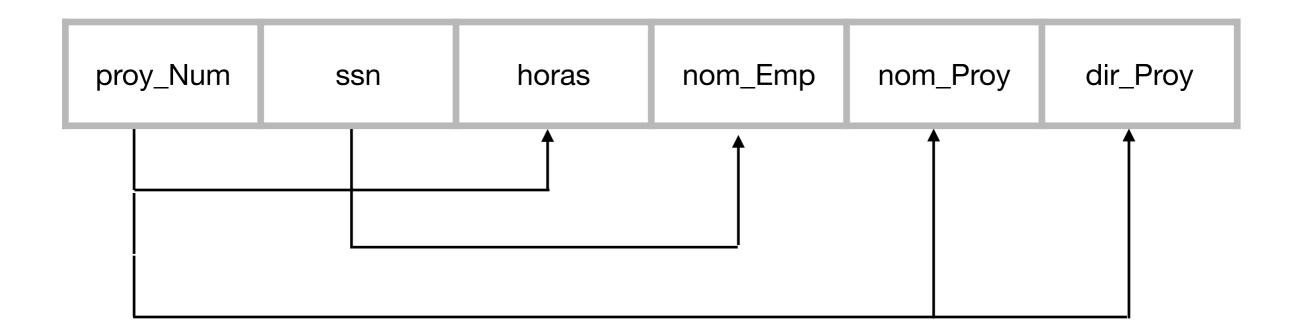
- Diagramas de dependencias:

proy_Num ssn	horas	nom_Emp	nom_Proy	dir_Proy
--------------	-------	---------	----------	----------





- Diagramas de dependencias:





Partiendo de:

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



Dependencias transitivas:



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_Estudiante 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:



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



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 redundacia de información a una o más tablas.





Técnica de optimización en bases de datos, en la cual, de manera premeditada, se agrega redundacia 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?



EMP_NAME	ADDRESS	DEPT_ID	PROJ_ID	DEPT_NAME
Joseph	Clinton Town	10	206	Accounting
Rose	Fraser Town	20	205	Quality
Mathew	Lakeside Village	10	206	Accounting
Stewart	Troy	30	204	Design
William	Holland	30	202	Design
	Joseph Rose Mathew Stewart	EMP_NAMEADDRESSJosephClinton TownRoseFraser TownMathewLakeside VillageStewartTroy	EMP_NAMEADDRESSDEPT_IDJosephClinton Town10RoseFraser Town20MathewLakeside Village10StewartTroy30	EMP_NAMEADDRESSDEPT_IDPROJ_IDJosephClinton Town10206RoseFraser Town20205MathewLakeside Village10206StewartTroy30204

DEPARTMENT	
DEPT_ID	DEPT_NAME
10	Accounting
20	Quality
30	Design
	Ĭ



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





Partiendo de:

A B ...

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

PK: {A}

PK: {A, E}





Partiendo de:

A B ...

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

Caso 1: PK: {A}

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





Caso 1: PK: {A}

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

1FN

- ¿hay atributos multivaluados? NO
- ¿hay gpos de repetición? NO Cumple 1FN

2FN - ¿La pk es simple? SÍ Cumple 2FN

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



Caso 1:

PK: {A}

1FN

- ¿hay atributos multivaluados? NO
- ¿hay gpos de repetición? NO Cumple 1FN

2FN

- ¿La pk es simple? SÍ Cumple 2FN

3FN

- ¿Hay transitividad entre atts no principales?
 (M,1) Sí hay transitividad -> No cumple 3FN

(1,M)

(1,1)

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

EMPLEADO



SUCURSAL

 $E -> \{F, G\}$



Caso 2:

PK: {A, E}

1FN

- ¿Hay atts multivaluados? NO
- ¿Hay gpos de repetición? NO Cumple 1FN

2FN

La pk es simple? NO
Hay que analizar si hay dependencias parciales

(M,M)

No cumple 2 FN ya que hay dependencias parciales

(1,M)

(1,M)

 $A \to \{B, C, D\}$

EMPLEADO

SUCURSAL



La PK es id_Orden

- Dibujar el diagrama de dependencias

Normalizar hasta 3FN



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

CASO 1:PK: A

CASO 2: PK : {A, F}

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

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

 $C -> \{D, E\}$

 $F -> \{G, I\}$

 ${A, F} \rightarrow H$



Sea la relacion Ventas (Num Venta, No Reg Carro, No Vendedor, No Cliente, Precio, Modelo, Marca, Nombre Cliente, Nombre Vendedor, Fechade Venta) 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





num_Cuenta	id_Curso	nombre_Estudi ante	nombre_Curso	calificacion	profesor	email_Profesor
12345	10	Javier	BD	8	Fernando	prof1@
90908	10	Elena	BD	7	Fernando	prof1@
12345	20	Javier	POO	9	Luis	prof2@
90908	20	Elena	POO	10	Luis	prof2@
67700	10	Sonia	BD	7	Fernando	prof1@
45645	10	Luis	BD	5	Fernando	prof1@
45645	30	Luis	SO	8	Laura	prof3@