Facultad de Ingeniería



Normalización

Tema V

Semestre 2025-1



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 -> Mejor representación y 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?





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						
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?





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	Project	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 tengan el mismo nombre





¿Qué vamos a conseguir al aplicar la 1FN?

- Identificar la PK de cada relación
- 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



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



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





¿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)





PK

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

FK	P	K
----	---	---

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



¿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





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





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



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?





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



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



$$DF:X\Longrightarrow Y$$

$$t_n X = t_m X$$

$$t_n Y = t_m Y$$

Donde *n* y *m* son las *tuplas* donde hay coincidencia en X





$cod_{-}Provedor$	nombre	telefono	$\operatorname{cod}_{\operatorname{-Articulo}}$	descripcion	precio
p50	Juan	5576581	A10	tuerca	20
p32	Pedro	3349631	A20	tornillo	30
p50	Juan	5576581	A20	tornillo	32
p32	Pedro	3349631	A10	tuerca	18
p38	José	5567543	A49	martillo	45



Dependencia funcional trivial

 Dependencia funcional no trivial



Notación



Notación en el proceso de normalización:

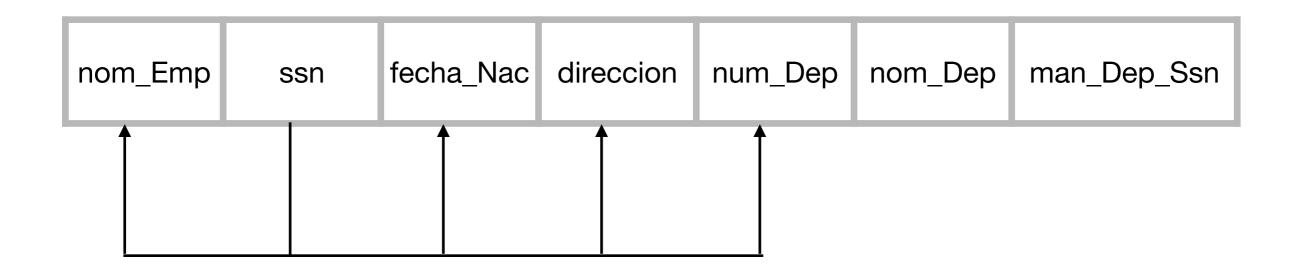
 Notación de dependencia funcional:



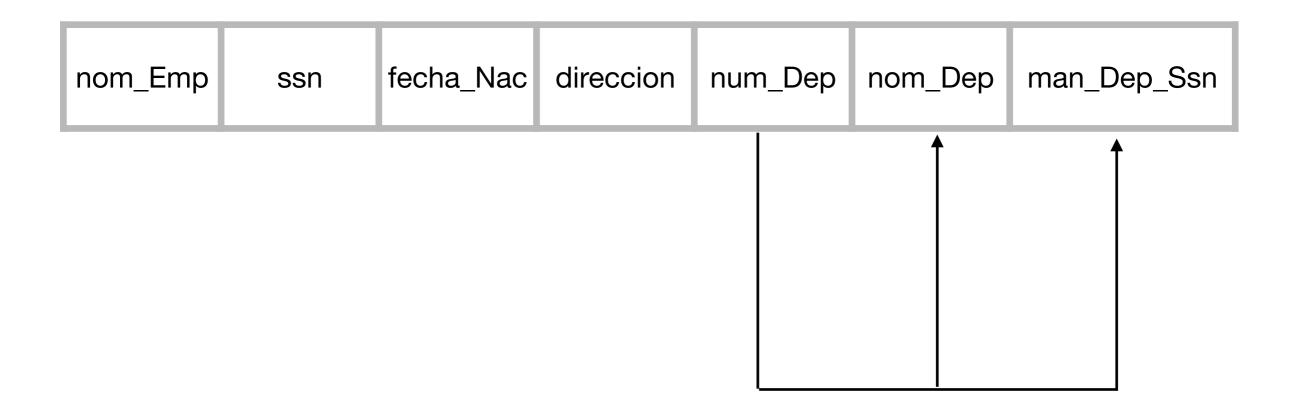


nom_Emp	ssn	fecha_Nac	direccion	num_Dep	nom_Dep	man_Dep_Ssn
---------	-----	-----------	-----------	---------	---------	-------------

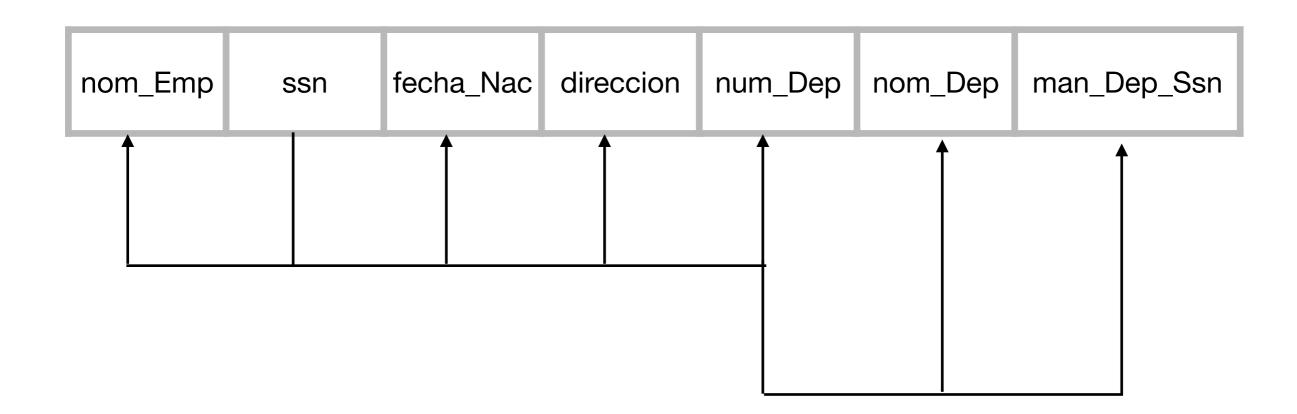














A B C D E F G

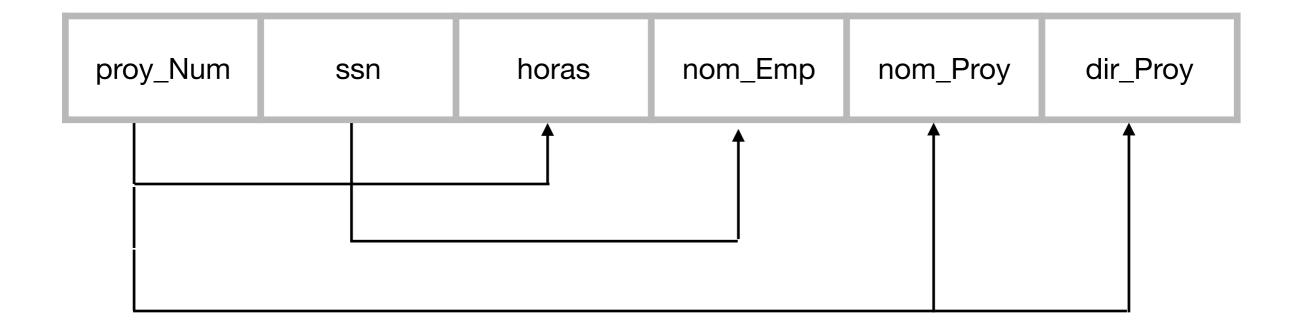
nom_Emp	ssn	fecha_Nac	direccion	num_Dep	nom_Dep	man_Dep_Ssn



- Indicar dependencias

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





Definiciones



- Llave
- Super llave (SK)
- Llave candidata (CK)
- Llave primaria (PK)



La cerradura de un conjunto de atributos X es el conjunto de aquellos atributos que pueden ser funcionalmente dependientes de X.

$$X^+$$





Procedimiento:

1) Agregar los atributos que componen a X como conjunto resultante de X^+





Procedimiento:

2) Agregar aquellos atributos que pueden ser funcionalmente dependientes de los atributos que ya estén contenidos en X^+





Procedimiento:

3) Repetir el paso 2 hasta que no puedan agregarse más atributos al conjunto X^+





Determinar:

$$A^+$$
 AD^+

$$B^+$$
 CD^+



Sea R(A, B, C, D, E) DF: {A->B, D->E}

Determinar:

$$A^{+}$$
 $ABCDE^{+}$ $ABDE^{+}$

 $ACDE^+$ ACD^+

¿Cuál(es) es SK? ¿Cuál(es) es CK?



Muchas combinaciones para encontrar las posibles llaves candidatas -> La complejidad es proporcional a la cantidad de atributos en la tabla/relación





Sea R(A, B, C, D, E) DF: {A->B, D->E}

Encontrar todas las CKs y la PK





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





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:

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



Considere los siguientes datos:

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?





Considere los siguientes datos:

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

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



¿Cumple la 1FN? -> Sí, ya que no hay valores multivaluados ni grupos de rep. ¿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
EMP-03	DEP_DB
<u>EMP-04</u>	DEP CT

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

EMPLEADO

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

DEPARTAMENTO





Sea la tabla EMPLEADO:

ID Empleado	ID Depto	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





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
EMP-04	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



repetición



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



Ejercicio 1_5



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 1_5



Solución:

staffNo	name	position	
S4555	Ellen Layman	Assistant	
S4612	Dave Sinclair	Assistant	

staffNo	branchNo	hoursPerWeek	
S4555	B002	16	
S4555	B004	9	
84612 B002		14	
S4612	B004	10	

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



Ejercicio 2_5



Sea R(A, B, C, D)
DF: {A->B, B->C, C->A}

Encontrar todas las CKs
Determinar el conjunto de APs
Indicar la PK

Ejercicio 3_5



Sea R(A, B, C, D)
DF: {AB->CD, D->B, C->A}

Encontrar todas las CKs
Determinar el conjunto de APs
Indicar la PK



Dependencias transitivas:



Dependencias 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.





Dependencias 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



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++

Ejercicio 4_5



- Encontrar todas las CKs
- Determinar el conjunto de APs
- Indicar la PK
- Normalizar hasta 3FN



Ejercicio 4_5



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



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?



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
	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



FNBC



Su objetivo es que las dependencias funcionales sólo dependan de la clave primaria o de cualquier clave candidata



FNBC



Partimos de la relación R(A, B, C, D) y las dependencias:

FNBC



Separamos en dos relaciones:

R1(A, C, D)

R2(D, B)



Su objetivo es que no haya dependencias múltiples.





¿Cómo identificar dependencias múltiples?

- La tabla contiene al menos, tres atributos
- A determina valores múltiples de B
- A determina valores múltiples de C
- By C son independientes





Sea R(A, B, C)

$$A ->-> B$$

Entonces:

R1 (A, B) y R2(A, C)



s_id	course	hobby
1	Science	Cricket
1	Maths	Hockey
2	C#	Cricket
2	Php	Hockey

Tabla con información de alumnos, los cursos que toman y sus hobbies





R1

s_id	course
1	Science
1	Maths
2	C#
2	Php

R2

s_id	hobby
1	Cricket
1	Hockey
2	Cricket
2	Hockey



Su objetivo es asegurar que, cuando separamos tablas para eliminar redundancia, al hacer una operación de natural join, no se agregue ni pierda información.





SUBJECT	LECTURER	CLASS
Mathematics	Alex	SEMESTER 1
Mathematics	Rose	SEMESTER 1
Physics	Rose	SEMESTER 1
Physics	Joseph	SEMESTER 2
Chemistry	Adam	SEMESTER 1

Si queremos agregar a alguien de tercer semestre, pero no sabemos qué materia a va a impartir... ¿Qué pasa?





			5NF
SUBJECT	LECTURER	CLASS	LECTURER
Mathematics	Alex	SEMESTER 1	Alex
Mathematics	Rose	SEMESTER 1	Rose
Physics	Rose	SEMESTER 1	Rose
Physics	Joseph	SEMESTER 2	Joseph
Chemistry	Adam	SEMESTER 1	Adam

CLASS	SUBJECT	
SEMESTER 1	Mathematics	
SEMESTER 1	Physics	
SEMESTER 1	Chemistry	
SEMESTER 2	Physics	

Ejercicio 5_5



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

Ejercicio 5_5



 ¿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

