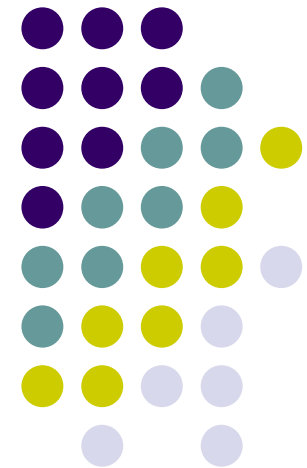


Diseño de Bases de Datos

Normalización



Un ejemplo



Proveedores (cod-proveedor, nom-proveedor, cod-insumo, precio)

CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	103	70
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	390	70

Un ejemplo

Proveedores (cod-proveedor, nom-proveedor, cod-insumo, precio)
se divide en:

DatosProveedor (cod-proveedor, nom-proveedor)

Suministros (cod-proveedor, cod-insumo, precio)



CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	103	70
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	390	70

CodProv	NomProv
#1	Silva
#2	Morales
#3	Gallardo

CodProv	CodInsumo	Precio
#1	100	200
#1	103	70
#2	201	200
#3	305	100
#3	390	70

Un ejemplo

Para la consulta: *“nombre del proveedor que vende el insumo del producto con código 103”*

- Join Natural entre **DatosProveedor** y **Suministros** teniendo en cuenta si hay atributos iguales

CodProv	NomProv
#1	Silva
#2	Morales
#3	Gallardo

CodProv	CodInsumo	Precio
#1	100	200
#1	103	70
#2	201	200
#3	305	100
#3	390	70

CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	103	70
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	390	70



Sin embargo ...



Una división mal hecha puede traer problemas:

IP (cod-insumo, precio)

NDP (cod-proveedor, nom-proveedor, precio)

CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	103	70
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	390	70

PI

CodInsumo	Precio
100	200
103	70
201	200
305	100
390	70

NDP

CodProv	NomProv	Precio
#1	Silva	200
#1	Silva	70
#2	Morales	200
#3	Gallardo	100
#3	Gallardo	70

Quiero hacer la misma consulta...



PI

CodInsumo	Precio
100	200
103	70
201	200
305	100
390	70

NDP

CodProv	NomProv	Precio
#1	Silva	200
#1	Silva	70
#2	Morales	200
#3	Gallardo	100
#3	Gallardo	70

Unión entre IP
y NDP
considerando
atributos con el
mismo nombre

CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	201	200
#1	Silva	103	70
#1	Silva	390	70
#2	Morales	100	200
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	103	70
#3	Gallardo	390	70

Hay más información
que en la tabla original

Otro ejemplo: Proyección

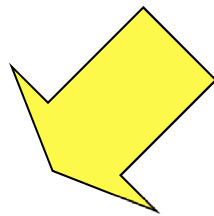


Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	rojo

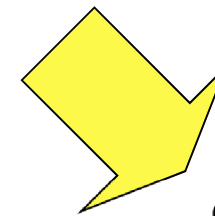
Genero una relación con patente y marca
Genero una relación marca, modelo y color

Se eliminan todas salvo una de las tuplas repetidas (Ej, *<Toyota, corollaXL, blanco>*)



R

Patente	Marca
MBO34L	Ford
LDA75K	Toyota
ADA89A	Fiat
LBF78G	Toyota
XSA67D	Ford



Q

Marca	Modelo	Color
Ford	Ka	verde
Fiat	Siena	gris
Toyota	CorollaXL	blanco
Ford	Ka	rojo

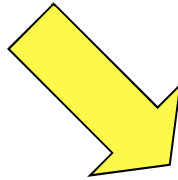
Regeneramos Auto



Auto = R * Q con atributos del mismo nombre

R

Patente	Marca
MBO34L	Ford
LDA75K	Toyota
ADA89A	Fiat
LBF78G	Toyota
XSA67D	Ford

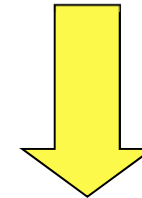


Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	rojo

Q

Marca	Modelo	Color
Ford	Ka	verde
Fiat	Siena	gris
Toyota	CorollaXL	blanco
Ford	Ka	rojo



Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
MBO34L	Ford	Ka	rojo
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	verde
XSA67D	Ford	Ka	rojo

Descomposición



- Es el reemplazo de una relación $R(A_1, A_2, \dots, A_n)$, por una colección de relaciones R_1, R_2, \dots, R_n obtenidas de las **proyecciones** de R y tal que la relación resultado de los productos naturales de $R_1 * R_2 * \dots * R_n$ tiene el mismo esquema que R .

- Ej: Si tenemos:

R_1 = Una relación de Auto con patente, modelo y marca

R_2 = Un relación de Auto con modelo y marca

...resulta que:

$R_1 * R_2 = \text{Auto ?}$

R * Q ≠ Auto



¿¿¿ Auto = R * Q ???

Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
MBO34L	Ford	Ka	rojo
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	verde
XSA67D	Ford	Ka	rojo

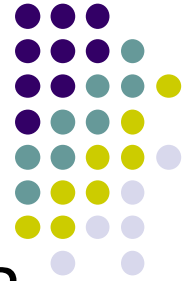
La relación original

¿Qué sucedió aquí?

Auto

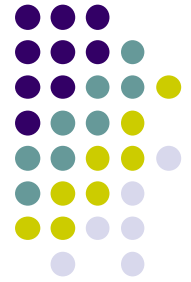
Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	rojo

Descomposición sin Pérdida



- Es la descomposición de una relación R en R_1, R_2, \dots, R_n tal que **para toda extensión de R** se tiene que $R = R_1 * R_2 * \dots * R_n$.
- El problema de la concepción de bases de datos relacionales se reduce a la descomposición sin pérdida de las relaciones universales con todos sus atributos, en subrelaciones que no tengan anomalías

Descomposición sin Pérdida



¿A dónde nos puede llevar una mala descomposición?

¿Y una descomposición insuficiente?

Anomalías

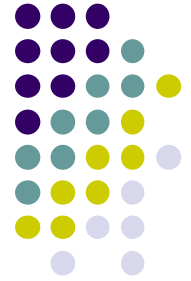


DNI	Nombre_Prof	Fecha_Nac	Cod_Depto	Nombre_Depto
9.980.623	Pedro Perez	1/6/73	01	Computación
10.334.890	Maria Lopez	1/6/76	02	Control
12.334.222	Mario Lobo	1/6/77	01	Computación
13.434.122	Laura Gonzalez	1/6/78	03	Investigación
13.566.002	Josefina Garcia	1/12/78	NULL	NULL
17.544.672	Miguel Bravo	1/6/84	02	Control
18.244.670	Andres Leon	1/6/85	01	Computación

¿Qué problemas o anomalías se pueden producir en esta relación?

¿Qué “cosas malas” pueden ocurrir?

Normalización



¿Qué es normalizar una base de datos?

Es encontrar una descomposición adecuada de la “relación universal” de la base de datos que nos permite cumplir con los criterios de eficacia, ausencia de redundancia, evolución, comprensión, flexibilidad enunciados anteriormente

Normalización



NUEVAMENTE:

¿Qué es normalizar una base de datos?

Es llevar el esquema de la base de datos a alguna de las formas normales . . .

. . . y para eso, necesitamos conocer y comprender el concepto de “formas normales” y “dependencia funcional”

Formas Normales Basadas en Clave Primaria



- Superclave, clave, clave candidata, atributo primo
- Formas normales sin DFs
 - 1 FN
- Formas normales con DFs
 - Dependencias Funcionales
 - 2 FN
 - 3 FN
 - 4 FN



Primera Forma Normal (1NF)

Una relación \mathcal{R} está en 1NF si los dominios utilizados sólo contienen valores atómicos o escalares

nss	apellido	puesto	salario	emails
111	Juan Perez	Jefe	3000	juanp@ecn.es , jefe2@ecn.es
222	José Sanchez	Administrativo	1500	jsanchez@ecn.es
333	Ana Diaz	Administrativo	1500	adiaz@ecn.es , ana32@gmail.com
...

Otro ejemplo: Clientes con deudas



rut_cliente	deuda
1111-1	(10.000,30/4/93) , (10.000,30/5/93) , (10.000,30/6/93)
9999-9	(20.000,30/4/93) , (10.000,30/5/93)

Otro Ejemplo: Una biblioteca



<i>CodLibro</i>	<i>Titulo</i>	<i>Autor</i>	<i>Editorial</i>	<i>NombreLector</i>	<i>FechaDev</i>
1001	Variable compleja	Murray Spiegel	McGraw Hill	Pérez Gómez, Juan	15/04/2005
1004	Visual Basic 5	E. Petroustsos	Anaya	Ríos Terán, Ana	17/04/2005
1005	Estadística	Murray Spiegel	McGraw Hill	Roca, René	16/04/2005
1006	Oracle University	Nancy Greenberg y Priya Nathan	Oracle Corp.	García Roque, Luis	20/04/2005
1007	Clipper 5.01	Ramalho	McGraw Hill	Pérez Gómez, Juan	18/04/2005

Formas Normales Basadas en Clave Primaria



- Superclave, clave, clave candidata, atributo primo
- Formas normales sin DFs
 - 1 FN
- Formas normales con DFs
 - Dependencias Funcionales
 - 2 FN
 - 3 FN
 - 4 FN

Dependencias Funcionales



- Son ***restricciones de integridad*** que permiten conocer que ***relaciones*** existen entre dos o más atributos del mundo real.
- Son propiedades ***inherentes*** al ***contenido semántico*** de los datos, que se han de cumplir para cualquier extensión del esquema de relación.
- Informalmente, Y **depende funcionalmente** de X si ***cada valor*** de X tiene asociado ***siempre el mismo*** valor de Y en una relación R que contiene a X y Y como atributos.

Dependencias Funcionales

Supongamos que un profesor puede trabajar para varios deptos . . .



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
9.980.623	Pedro Perez	1/6/73	M	01	Computación
10.334.890	Maria Lopez	1/6/76	F	01	Computación
18.244.670	Andres Leon	1/6/85	M	03	Investigación
12.334.222	Mario Lobo	1/6/77	M	02	Control
13.566.002	Josefina Garcia	1/12/78	F	02	Control
10.334.890	Maria Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	01	Computación
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
17.544.672	Miguel Bravo	1/6/84	M	02	Control
18.244.670	Andres Leon	1/6/85	M	01	Computación

Dependencias Funcionales



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
9.980.623	Pedro Perez	1/6/73	M	01	Computación
10.334.890	María Lopez	1/6/76	F	01	Computación
18.244.670	Andres Leon	1/6/85	M	03	Investigación
12.334.222	Mario Lobo	1/6/77	M	02	Control
13.566.002	Josefina Garcia	1/12/78	F	02	Control
10.334.890	María Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	01	Computación
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
17.544.672	Miguel Bravo	1/6/84	M	02	Control
18.244.670	Andres Leon	1/6/85	M	01	Computación

¿Qué características destacan en la información de los profesores?

¿Se repite la información de los profesores?

¿Que sucede con los atributos Fecha_Nac y Sexo con respecto al DNI?

¿Qué relaciones existen?

Otra forma de ver ...



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
10.334.890	Maria Lopez	1/6/76	F	01	Computación
10.334.890	Maria Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	02	Control
12.334.222	Mario Lobo	1/6/77	M	01	Computación
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	02	Control
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
17.544.672	Miguel Bravo	1/6/84	M	02	Control
18.244.670	Andres Leon	1/6/85	M	03	Investigación
18.244.670	Andres Leon	1/6/85	M	01	Computación
9.980.623	Pedro Perez	1/6/73	M	01	Computación

¿Qué características destacan en la información de los profesores?

¿Se repite la información de los profesores?

¿Que sucede con los atributos Fecha_Nac y Sexo con respecto al DNI?

¿Qué relaciones existen?

Dependencias Funcionales



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
9.980.623	Pedro Perez	1/6/73	M	01	Computación
10.334.890	Maria Lopez	1/6/76	F	01	Computación
18.244.670	Andres Leon	1/6/85	M	03	Investigación
12.334.222	Mario Lobo	1/6/77	M	02	Control
13.566.002	Josefina Garcia	1/12/78	F	02	Control
10.334.890	Maria Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	01	Computación
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
17.544.672	Miguel Bravo	1/6/84	M	02	Control
18.244.670	Andres Leon	1/6/85	M	01	Computación

¿Y con respecto a la información de los departamentos?

¿Se repite?

¿Qué ocurre con los atributos de las distintas filas?

Otra forma de ver ...



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
10.334.890	Maria Lopez	1/6/76	F	01	Computación
12.334.222	Mario Lobo	1/6/77	M	01	Computación
18.244.670	Andres Leon	1/6/85	M	01	Computación
9.980.623	Pedro Perez	1/6/73	M	01	Computación
10.334.890	Maria Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	02	Control
13.566.002	Josefina Garcia	1/12/78	F	02	Control
17.544.672	Miguel Bravo	1/6/84	M	02	Control
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
18.244.670	Andres Leon	1/6/85	M	03	Investigación

¿Y con respecto a la información de los departamentos?

¿Se repite?

¿Qué ocurre con los atributos de las distintas filas?

Dependencias Funcionales



DNI	Nombre_Prof	Fecha_Nac	Sexo	Cod_Depto	Nombre_Depto
9.980.623	Pedro Perez	1/6/73	M	01	Computación
10.334.890	Maria Lopez	1/6/76	F	01	Computación
18.244.670	Andres Leon	1/6/85	M	03	Investigación
12.334.222	Mario Lobo	1/6/77	M	02	Control
13.566.002	Josefina Garcia	1/12/78	F	02	Control
10.334.890	Maria Lopez	1/6/76	F	02	Control
12.334.222	Mario Lobo	1/6/77	M	01	Computación
13.434.122	Laura Gonzalez	1/6/78	F	03	Investigación
13.566.002	Josefina Garcia	1/12/78	F	03	Investigación
17.544.672	Miguel Bravo	1/6/84	M	02	Control
18.244.670	Andres Leon	1/6/85	M	01	Computación

Dependencias Funcionales



Son restricciones de integridad que permiten conocer que interrelaciones existen entre los atributos del mundo real

Otra vez, el Auto



Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	rojo

Las DFs de casos de Auto



Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
MBO34L	Ford	Ka	rojo
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	verde
XSA67D	Ford	Ka	rojo

Las DFs de casos de Auto



Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco
XSA67D	Ford	Ka	rojo

Las DFs de casos de Auto



Auto

Patente	Marca	Modelo	Color
MBO34L	Ford	Ka	verde
XXR34L	Chrysler	Ka	rojo
LDA75K	Toyota	CorollaXL	blanco
ADA89A	Fiat	Siena	gris
LBF78G	Toyota	CorollaXL	blanco

Dependencia Funcional



Sea \mathcal{R} una relación con atributos (a_1, a_2, \dots, a_n) , y sean \mathcal{X} e \mathcal{Y} dos subconjuntos de los atributos a_i . Se dice que \mathcal{Y} depende funcionalmente de \mathcal{X} y se anota

$$\mathcal{X} \rightarrow \mathcal{Y}$$

Si para todo par de tuplas t_1 y t_2 se cumple

$$t_1[\mathcal{X}] = t_2[\mathcal{X}] \Rightarrow t_1[\mathcal{Y}] = t_2[\mathcal{Y}]$$

Otra vez, los proveedores ...



CodProv	NomProv	CodInsumo	Precio
#1	Silva	100	200
#1	Silva	103	70
#2	Morales	201	200
#3	Gallardo	305	100
#3	Gallardo	390	70

cod-proveedor \rightarrow nom-proveedor
(cod-proveedor, cod-insumo) \rightarrow precio

Otra interpretación? Es válido?

cod-proveedor \rightarrow nom-proveedor
cod-insumo \rightarrow precio

Las dependencias funcionales

- sirven para capturar propiedades del mundo real
- dan semántica a las tablas
- definen restricciones

Clave



Un subconjunto \mathcal{K} de los atributos (a_1, a_2, \dots, a_n) , de una relación \mathcal{K} es **clave** si:

- $\mathcal{K} \rightarrow (a_1, a_2, \dots, a_n)$, y
- no existe $\mathcal{Y} \subset \mathcal{K}$ tal que $\mathcal{Y} \rightarrow (a_1, a_2, \dots, a_n)$

Nota: si existe $\mathcal{K} \subset \mathcal{X}$ y $\mathcal{X} \rightarrow (a_1, a_2, \dots, a_n)$ y

$$\mathcal{K} \rightarrow (a_1, a_2, \dots, a_n),$$

entonces \mathcal{X} es superclave o clave candidata

Superclave, clave, clave candidata, clave primaria

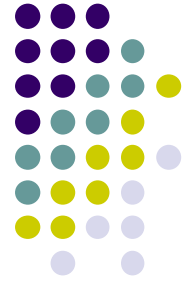


$$R = \{A_1, A_2, \dots, A_n\}$$

Una clave K es superclave y además si se le quita un atributo cualquiera deja de ser clave

- *Clave primaria* es un clave elegida arbitrariamente entre las *claves candidatas*
- *Atributo primo*, es un atributo que pertenece a alguna clave *candidata*

Propiedades Deseables de la Descomposición



- No pérdida de información
- Conservar las dependencias funcionales
- Perder dependencias funcionales significa perder restricciones.