



# Ingeniería de Datos

## Normalización

Nadia Rodríguez

# Agenda

---

- Primera forma normal
- Segunda forma normal
- Tercera forma normal



# Normalización

---

- Valida y mejora el diseño lógico que satisface ciertas restricciones, el cual *evita la duplicación innecesaria de datos.*
- Proceso de descomposición de relaciones con anomalías para producir relaciones *bien-estructuradas* más pequeñas.
- No siempre provoca el mejor rendimiento en los RDBMS.

# Ejemplo Redundancia de Datos

VUELO

vuelo#	fecha	aerolinea	avion#
DL242	10/23/08	Delta	k-yo-33297
DL242	10/24/08	Delta	t-up-73356
DL242	10/25/08	Delta	o-ge-98722
AA121	10/24/08	American	p-rw-84663
AA121	10/25/08	American	q-yg-98237
AA411	10/28/08	American	h-fe-65748

- **Redundancia:** La aerolínea se repite para el mismo número de vuelo.
- **Inconsistencia:** Cuando el nombre de la aerolínea cambia para un número de vuelo, ésta debe ser cambiada en varias filas.

# Relaciones bien estructuradas

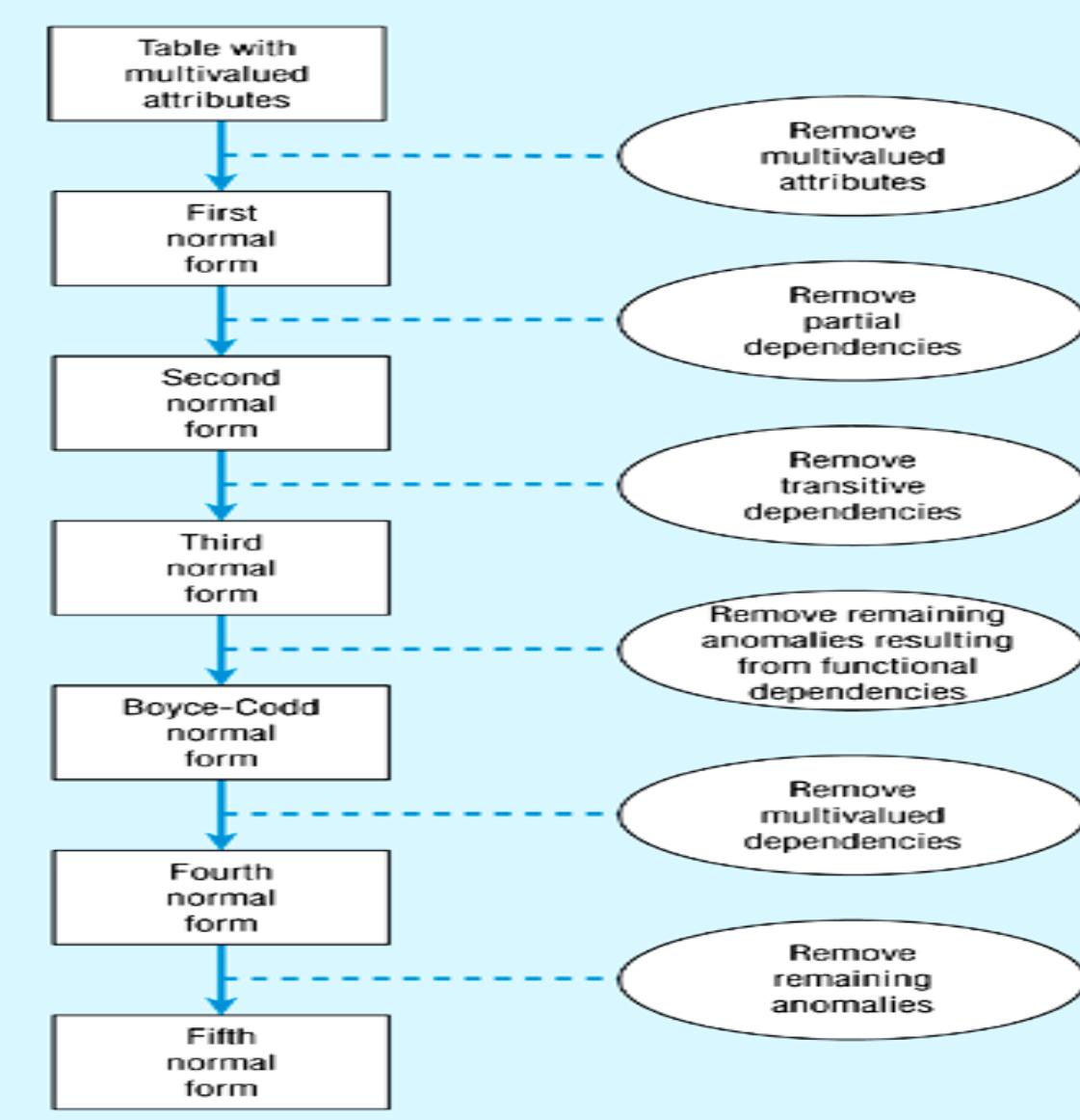
---

- Contiene mínima redundancia y permite que los usuarios añadan, eliminen y actualicen registros (filas) sin causar inconsistencias.
- La meta es evitar anomalías
  - **Anomalías de inserción** – la adición de nuevas filas obliga a los usuarios a crear datos duplicados o simplemente no es posible
  - **Anomalías de eliminación** – puede causar la pérdida de datos que deben ser requeridos para otras filas futuras
  - **Anomalías de modificación** – los datos cambiados obligan a cambiar otros causados por la duplicación

# Ejemplos de Anomalías

Codigo_Docente	Curso_Capacitacion	Fecha_Ingreso	Salario
42000	Excel Avanzado	20/11/2007	1,500
42000	PowerPoint Avanzado	10/07/2007	1,500
43000	Metodología Enseñanza	30/01/2008	3,200
44000	SQL Basico	31/03/2008	1,800
44000	DB2	25/10/2007	1,800
45000	Excel Avanzado	12/08/2007	2,700
45000	Java	19/09/2007	2,700

- **Anomalías de inserción:** Que pasa si quiero insertar un código de docente pero el docente todavía no ha tomado ningún curso de capacitación?
- **Anomalías de eliminación:** Que pasa si se borra la fila o registro del código de docente 43000?
- **Anomalías de modificación:** Que pasa si el docente con código 42000 recibe un aumento de salario?



# Dependencia Funcional y Claves

---

- Dependencia Funcional:
  - El valor de un atributo A (*determinante*) determina el valor de otro atributo B, o sea B es funcionalmente dependiente de A
$$A \rightarrow B$$
- Clave Candidata:
  - Identificador único. Una clave o claves que podrían convertirse en la clave primaria
    - Ej. Tanto el código de empleado como el DNI, ambas son claves candidatas de la entidad Empleado
    - Cada campo no clave es funcionalmente dependiente de cada clave candidata

# Ejemplo de Normalización

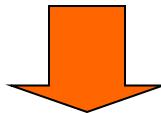
Factura RUC 30001 Nro 1240001	RUC _____	Nombre _____		
	Dirección _____			
Cod Artículo	Descripción Artículo	Cantidad	Precio Unitario	Precio Total
Total Factura				
Fecha / /				

# Atributos de la Entidad

RUC Empresa



Nro factura	RUC Cliente	Nombres	Dirección	Total	Fecha
Cod Artículo	Descripción Artículo	Cantidad	Precio Unitario	Precio Total	



Construir la gran tabla

Nro_Factura	RUC_Cliente	Nombre_Cliente	Direccion_Cliente	Total	Fecha	Cod_Articulo	Desc_Articulo	Cant	Prec_Unitario	Prec_Total
1240001	0000125	JUAN PEREZ PAZ	Arica 325. Lince	2250	20/08/2007	010	Lapicero Azul punta fina	100	2.5	250
1240001	0000125	JUAN PEREZ PAZ	Arica 325. Lince	2250	20/08/2007	753	Cuaderno rayado A4	200	10	2000

# Primera Forma Normal

---

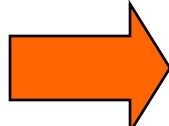
- Eliminar atributos multivaluados (grupos repetitivos)
- Cada atributo debe ser atómico

# Ejemplo de Atomicidad

HORARIO\_VUELOS

vuelo#	dias	aerolinea	hsalida	origen	hllegada	destino
DL242	MO WE FR	DELTA	10:40	ATL	12:30	BOS
SK912	SA SU	SAS	12:00	CPH	15:30	JFK
AA242	MO FR	AA	08:00	CHI	10:10	ATL

Atributos deben  
tener valores  
atómicos



HORARIO\_VUELOS

vuelo#	dias	aerolinea	hsalida	origen	hllegada	destino
DL242	MO	DELTA	10:40	ATL	12:30	BOS
DL242	WE	DELTA	10:40	ATL	12:30	BOS
DL242	FR	DELTA	10:40	ATL	12:30	BOS
SK912	SA	SAS	12:00	CPH	15:30	JFK
SK912	SU	SAS	12:00	CPH	15:30	JFK
AA242	MO	AA	08:00	CHI	10:10	ATL
AA242	FR	AA	08:00	CHI	10:10	ATL

# Primera Forma Normal 1FN

## Grupos Repetitivos y valores atómicos

RUC 30001  
Nro 1240001

RUC: 0000125 Nombre: JUAN PEREZ PAZ  
Dirección: Arica 325 Lince

Cod Artículo	Descripción Artículo	Cant	Prec.Unit	Prec.Tot.
010	Lapicero Azul punta fina	100	2.50	250.00
753	Cuaderno rayado A4	200	10.00	2000.00

Fecha 20/08/2007

Total Factura

2250.00

# Resultados de la Primera Forma Normal

## Grupo No Repetitivo

Nro factura	RUC Cliente	Nombres	Dirección	Total	Fecha

Nro_Factura	RUC_Cliente	Nombre_Cliente	Direccion_Cliente	Total	Fecha
1240001	0000125	JUAN PEREZ PAZ	Arica 325. Lince	2250	20/08/2007

## Grupo Repetitivo

Nro factura	Cod Artículo	Descripción Artículo	Cantidad	Precio Unitario	Precio Total

Nro_Factura	Cod_Articulo	Desc_Articulo	Cant	Prec_Unitario	Prec_Total
1240001	010	Lapicero Azul punta fina	100	2.5	250
1240001	753	Cuaderno rayado A4	200	10	2000

# Segunda Forma Normal

---

- *Dependencias Funcionales*
  - Cada atributo no clave debe estar determinado por toda la clave primaria (dependencia funcional total) y no parte de ésta (dependencia funcional parcial)
  - Separar dependencias funcionales parciales

# Segunda Forma Normal

## Dependencias Funcionales

Grupo No Repetitivo

Nro factura	RUC Cliente	Nombres	Dirección	Total	Fecha
-------------	-------------	---------	-----------	-------	-------

Dependencias Parciales

Grupo Repetitivo

Nro factura	Cod Artículo	Descripción Artículo	Cantidad	Precio Unitario	Precio Total
-------------	--------------	----------------------	----------	-----------------	--------------

Dependencias Totales

# Resultados de la Segunda Forma Normal

Nro factura	RUC Cliente	Nombres	Dirección	Total	Fecha
-------------	-------------	---------	-----------	-------	-------

Nro_Factura	RUC_Cliente	Nombre_Cliente	Direccion_Cliente	Total	Fecha
1240001	0000125	JUAN PEREZ PAZ	Arica 325. Lince	2250	20/08/2007

Nro factura	Cod Artículo	Cantidad	Precio Total
-------------	--------------	----------	--------------

Nro_Factura	Cod_Articulo	Cant	Prec_Total
1240001	010	100	250
1240001	753	200	2000

Cod Artículo	Descripción Artículo	Precio Unitario
--------------	----------------------	-----------------

Cod_Articulo	Desc_Articulo	Prec_Unitario
010	Lapicero Azul punta fina	2.5
753	Cuaderno rayado A4	10

# Tercera Forma Normal

---

- *Dependencias transitivas*
  - Un atributo no clave determina funcionalmente a otro u otros atributos no clave

# Tercera Forma Normal

## Dependencias Transitivas

Dependencias Transitivas

Nro factura	RUC Cliente	Nombres	Dirección	Total	Fecha
-------------	-------------	---------	-----------	-------	-------

Nro factura	Cod Artículo	Cantidad	Precio Total
-------------	--------------	----------	--------------

Cod Artículo	Descripción Artículo	Precio Unitario
--------------	----------------------	-----------------

# Resultados de la Tercera Forma Normal

CLIENTE

RUC Cliente	Nombres	Dirección
-------------	---------	-----------

RUC_Cliente	Nombre_Cliente	Direccion_Cliente
0000125	JUAN PEREZ PAZ	Arica 325. Lince

FACTURA

Nro factura	RUC Cliente	Total	Fecha
-------------	-------------	-------	-------

Nro_Factura	RUC_Cliente	Total	Fecha
1240001	0000125	2250	20/08/2007

LINEA FACTURA

Nro factura	Cod Artículo	Cantidad	Precio Total
-------------	--------------	----------	--------------

Nro_Factura	Cod_Articulo	Cant	Prec_Total
1240001	010	100	250
1240001	753	200	2000

ARTICULO

Cod Artículo	Descripción Artículo	Precio Unitario
--------------	----------------------	-----------------

Cod_Articulo	Desc_Articulo	Prec_Unitario
010	Lapicero Azul punta fina	2.5
753	Cuaderno rayado A4	10

# Diagrama Entidad Relación

---

