

Maestría en Bioinformática

Bases de Datos y Sistemas de Información



Fundamentos de Normalización

Ing. Alfonso Vicente, PMP
alfonso.vicente@logos.com.uy

Agenda

Introducción

Definiciones

Formas normales

- ¿Qué es la normalización?
- ¿Por qué es importante?
- Anomalías de actualización

Agenda

Introducción

Definiciones

Formas normales

- Dependencia funcional
- Atributos primos y no-primos

Agenda

Introducción

Definiciones

Formas normales

- Primera forma normal
- Segunda forma normal
- Tercera forma normal
- Forma normal de Boyce-Codd
- Otras formas normales

Agenda

Introducción

Definiciones

Formas normales

- ¿Qué es la normalización?
- ¿Por qué es importante?
- Anomalías de actualización

Introducción

¿Qué es la normalización?

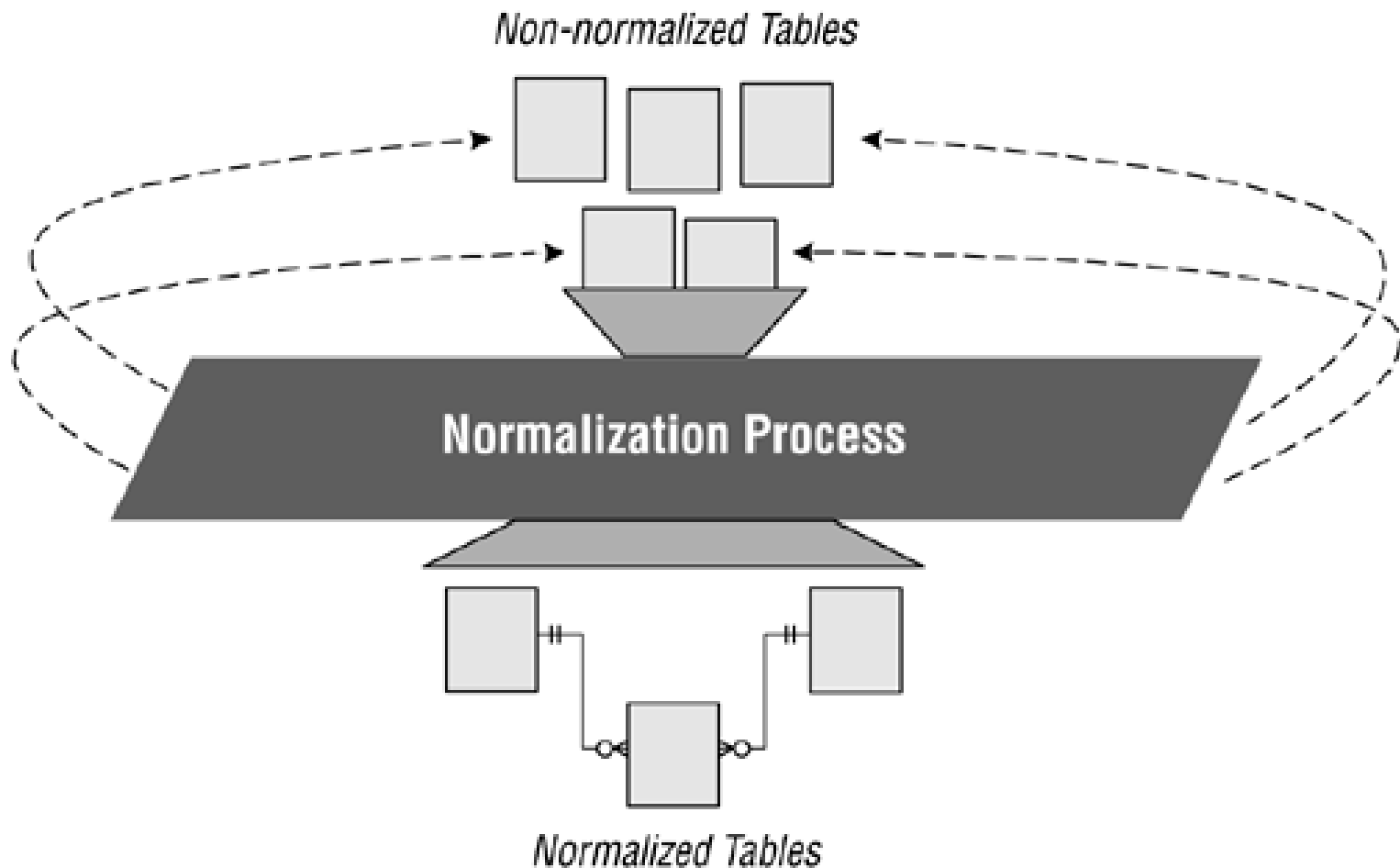
- Es un proceso que se realiza sobre el modelo lógico para garantizar que el modelo cumple ciertas reglas



- En este proceso, las relaciones insatisfactorias se descomponen repartiendo sus atributos entre relaciones más pequeñas

Introducción

¿Qué es la normalización?



Introducción

¿Por qué es importante la normalización?

- Para garantizar la integridad de los datos
- Para evitar redundancia
- Para evitar anomalías de actualización

¿Qué problema tiene la siguiente instancia de relación?

<u>cédula</u>	nombre	apellido	ced_jefe	nom_jefe	ape_jefe
4.200.300-5	Juan	Pérez	1.637.922-2	Severo	García
3.181.348-6	María	González	1.637.922-2	Severo	García
3.642.936-1	Juan	Martínez	1.637.922-2	Severo	García

Introducción

Anomalías de actualización – inserción

¿Qué problemas puede haber al insertar un nuevo empleado?

<u>cédula</u>	nombre	apellido	id_dep	nom_dep
4.200.300-5	Juan	Pérez	1	Administración
3.181.348-6	María	González	2	Ventas
3.642.936-1	Juan	Martínez	2	Ventas

¿Qué nos impide insertar la tupla:

(‘2.032.287-4’, ‘Martín’, ‘Ramírez’, 2, ‘Sistemas’) ?

¿Cómo agregamos un departamento sin empleados?

Introducción

Anomalías de actualización – eliminación

¿Qué problemas puede haber al eliminar a Juan Pérez?

<u>cédula</u>	nombre	apellido	id_dep	nom_dep
4.200.300-5	Juan	Pérez	1	Administración
3.181.348-6	María	González	2	Ventas
3.642.936-1	Juan	Martínez	2	Ventas

Introducción

Anomalías de actualización – modificación

¿Qué problemas puede haber al transferir a Juan Pérez a Ventas?

<u>cédula</u>	nombre	apellido	id_dep	nom_dep
4.200.300-5	Juan	Pérez	1	Administración
3.181.348-6	María	González	2	Ventas
3.642.936-1	Juan	Martínez	2	Ventas

Agenda

Introducción

Definiciones

Formas normales

- Dependencia funcional
- Atributos primos y no-primos

Definiciones

Dependencia funcional

Una dependencia funcional es una restricción entre conjuntos de atributos de una relación (tabla).

Sea R una relación, A y B conjuntos de atributos de R

Decimos que $A \rightarrow B$ (A determina funcionalmente a B , o B depende funcionalmente de A) si y sólo si para cada tupla, los valores iguales de A corresponden a valores iguales de B

Ejemplos:

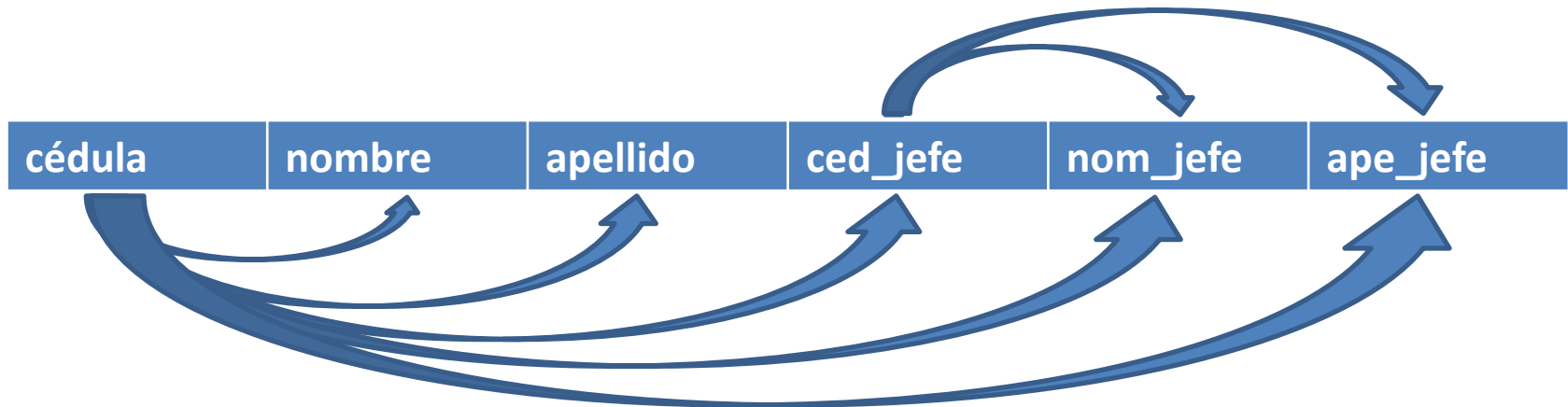
$\{\text{cedula}\} \rightarrow \{\text{nombre}\}$

$\{\text{cedula}\} \rightarrow \{\text{nombre, apellido}\}$

Definiciones

Dependencia funcional

En nuestro ejemplo, podemos identificar varias dependencias funcionales:









Note que $\{ced_jefe\}$ no es una clave, sin embargo identificamos dos DFs: $\{ced_jefe\} \rightarrow \{nom_jefe\}$ y $\{ced_jefe\} \rightarrow \{ape_jefe\}$

Definiciones

Atributos primos y no-primos

Recordemos que una clave candidata es una superclave minimal.

- Son primos los atributos de cualquier clave candidata
- Son no-primos los atributos que no forman parte de ninguna clave candidata

cédula	credencial	nombre	apellido	dirección	teléfono
					
<i>primo</i>	<i>primo</i>	<i>no-primo</i>	<i>no-primo</i>	<i>no-primo</i>	<i>no-primo</i>

Agenda

Introducción

Definiciones

Formas normales

- Primera forma normal
- Segunda forma normal
- Tercera forma normal
- Forma normal de Boyce-Codd
- Otras formas normales

Formas normales

Primera forma normal

Una relación está en primera forma normal (1NF) si los dominios de los atributos sólo incluyen valores atómicos.

Algunos puristas, como Chris Date, tienen una definición más estricta de 1NF, que incluso prohíbe la existencia de NULLs.

Por ejemplo, la siguiente relación no está en 1NF:

<u>cédula</u>	nombre	apellido	teléfonos
4.200.300-5	Juan	Pérez	5551122, 5558935
3.181.348-6	María	González	5551965
3.642.936-1	Juan	Martínez	5553521, 5552460, 5551035

Formas normales

Primera forma normal

La primera forma normal está asociada a la no existencia de “grupos repetidos”, pero hay diferentes escuelas sobre lo que significan “grupos repetidos”.

Para algunos, esta sería otra forma de violar 1NF:

<u>cedula</u>	nombre	apellido	teléfono1	teléfono2	teléfono3
4.200.300-5	Juan	Pérez	5551122	5558935	
3.181.348-6	María	González	5551965		
3.642.936-1	Juan	Martínez	5553521	5552460	5551035

Formas normales

Primera forma normal

La forma de solucionar el problema es descomponer la relación en dos relaciones, con los “grupos repetidos” en su propia relación:

PERSONAS

<u>cedula</u>	nombre	apellido
4.200.300-5	Juan	Pérez
3.181.348-6	María	González
3.642.936-1	Juan	Martínez

TELEFONOS_PERSONA

<u>cedula</u>	<u>teléfono</u>
4.200.300-5	5551122
4.200.300-5	5558935
3.181.348-6	5551965
3.642.936-1	5553521
3.642.936-1	5552460
3.642.936-1	5551035

TELEFONOS_PERSONA(CEDULA) REFERENCES PERSONAS(CEDULA)

Formas normales

Segunda forma normal

Una relación R está en segunda forma normal (2NF) si está en 1NF y si todo atributo no-primario de R depende funcionalmente de manera total de la clave primaria.

La siguiente relación está en 1NF pero no está en 2NF, porque nombre y apellido dependen sólo de cedula:

- La PK es {cedula, proyecto}
- {cedula} \rightarrow {nombre, apellido}

<u>cedula</u>	<u>proyecto</u>	nombre	apellido	horas
4.200.300-5	1	Juan	Pérez	52
3.181.348-6	1	María	González	20
3.181.348-6	2	María	González	32

Formas normales

Segunda forma normal

La forma de solucionar esto es separar los atributos que dependen de toda la PK de los que no

PERSONAS

<u>cedula</u>	nombre	apellido
4.200.300-5	Juan	Pérez
3.181.348-6	María	González

DEDICACIONES

<u>cedula</u>	<u>proyecto</u>	horas
4.200.300-5	1	52
3.181.348-6	1	20
3.181.348-6	2	32

DEDICACIONES(CEDULA) REFERENCES PERSONAS(CEDULA)

Formas normales

Tercera forma normal

Una relación R está en tercera forma normal (3NF) si está en 2NF y si no existe ninguna dependencia funcional transitiva entre atributos que no son clave

La siguiente relación está en 2NF, pero no está en 3NF porque región depende de país:

- La PK es {cedula}
- {cedula} \rightarrow {país} y {país} \rightarrow {región}

<u>cedula</u>	nombre	apellido	país	región
4.200.300-5	Juan	Pérez	Uruguay	Sudamérica
3.181.348-6	María	González	España	Europa

Formas normales

Tercera forma normal

La forma de solucionar esto es separar en otra relación la DF que viola 3NF: {país} → {región}

PERSONAS

<u>cedula</u>	nombre	apellido	país
4.200.300-5	Juan	Pérez	Uruguay
3.181.348-6	María	González	España

PAISES

<u>país</u>	región
Uruguay	Sudamérica
España	Europa

PERSONAS(PAÍS) REFERENCES PAISES(PAÍS)

Formas normales

Forma normal de Boyce-Codd

Una relación R está en forma normal de Boyce-Codd (BCNF) si está en 2NF y si para cada dependencia $A \rightarrow B$, el determinante A es una clave candidata de R

BCNF es muy parecida a 3NF, pero es un poco más estricta

Este es un ejemplo de relación en 3NF, pero no en BCNF. Note que $\{\text{país}\} \rightarrow \{\text{tipo_doc}\}$ y $\{\text{país}\}$ no es una clave de la relación.

<u>tipo_doc</u>	<u>nro_doc</u>	nombre	apellido	país
CI	4.200.119-2	Juan	Pérez	Uruguay
DNI	10563145-8	María	González	Argentina

Formas normales

Otras formas normales

Existen otras formas normales que no veremos (4NF, 5NF) cada una más restrictiva que la anterior.

Siempre intentaremos generar modelos lógicos en los que cada relación esté al menos en 3NF (si hicimos bien el MER y el pasaje MER \rightarrow MR esto está asegurado).

NF	¿Qué significa?
1NF	No hay grupos repetidos
2NF	No hay dependencias parciales
3NF	No hay dependencias transitivas
BCNF	No hay DFs con determinantes no-claves

