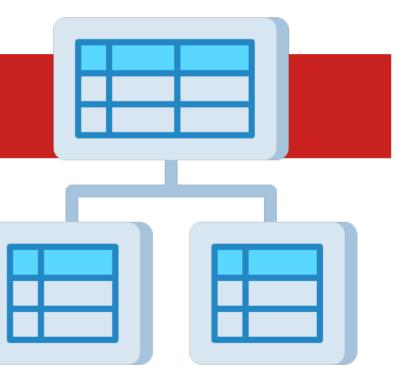
Bases de Datos I Unidad VI

Esquemas anómalos, dependencias funcionales y normalización



Esquemas relacionales anómalos ¿Por qué?

Se realizó un diseño erróneo:

- Mala percepción del universo del discurso.
- Errores en el diseño conceptual
- Errores en el diseño lógico

No se realizó ningún diseño:

- Base de datos heredada.
- Modelo relacional directo
- Las relaciones pueden contener anomalías.

¿Qué son las anomalías?

- Incapacidad para almacenar ciertos hechos.
- Ambigüedades.
- Información redundante.
- Existencia de valores inaplicables (nulos).
- Estados de la BD no válidos en el mundo real.
 - Anomalías de inserción
 - Anomalías de modificación
 - Anomalías de borrado

Ejemplo de anomalía: redundancia

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	HTT001	Renault	Clio
99.888.888	Perez	Jorge	ОТО333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- Los datos de una persona que tiene varios vehículos, se encuentran repetidos.
- Los datos de un vehículo que pertenece a varias personas, se encuentran repetidos.

Ejemplo de anomalía: inserción

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	OTO333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- No es posible agregar una persona, si no disponemos los datos de su(s) vehículo(s).
- No es posible agregar un vehículo, sin no disponemos los datos de su(s) dueño(s).

Ejemplo de anomalía: modificación

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Miguel	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	ОТО333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- Dada la redundancia de la información, es posible modificar los datos de una persona en una tupla, y dejar las otras con datos inconsistentes.
- Lo mismo para los vehículos.

Ejemplo de anomalía: eliminación

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	OTO333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- Al borrar la información de un vehículo, eliminamos los datos del dueño asociado al mismo.
- De igual forma al revés.

¿Cómo detectamos las anomalías?

 Estudiando ciertos vínculos o relaciones que se establecen entre los atributos de una relación o de un esquema relacional.

Estos vínculos se conocen como Dependencias Funcionales.

 Permiten abordar un análisis matemático de un esquema relacional, a través de la aplicación de distintos algoritmos.

¿Cómo corregimos las anomalías?

 Detectando distintas Dependencias Funcionales no deseables (Formas Normales), que dan lugar a las anomalías.

 Luego de la detección, procederemos a la corrección del esquema relacional a través de un proceso de descomposición que se conoce con el nombre Normalización.

Dependencias funcionales: conceptos previos

- Descriptor: Denominamos descriptor a un subconjunto X de atributos de una relación R(A).
- Por ejemplo:
 - La relación:Persona(dni, apellido, nombre)
 - Dos descriptores X e Y:
 X = {dni, apellido, nombre}
 Y = {dni, nombre}

Dependencias funcionales: conceptos previos

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	OTO333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- ¿Porque esta relación presenta anomalías?
- ¿Existe algún vínculo entre los distintos atributos de la misma?

- Sean X e Y dos descriptores.
- Se dice que existe una dependencia funcional entre X e Y de forma X → Y si y sólo si, se cumple que:
 - para dos tuplas cualquiera de la relación R, u y v tales que u[X] = v[X],
 - entonces necesariamente $\mathbf{u}[Y] = \mathbf{v}[Y]$.

- Una dependencia funcional es una relación existente entre uno o más atributos, en la cual al conocer el valor de un atributo se puede determinar el valor de otro.
- Especifica una restricción de las tuplas posibles que se pueden formar dentro de la relación.
- Formalmente se puede indicar que un atributo Y es dependiente del atributo X si el valor de X determina el valor de Y, por lo tanto si se conoce el valor de X se conocerá el valor de Y.
- Una dependencia funcional se denota de la siguiente manera: X → Y y se lee "X determina funcionalmente a Y". Que X determine funcionalmente a Y implica que para cada tupla de la relación donde los valores de X coincidan, también deberán hacerlos los de Y.

Dada una dependencia funcional entre X e Y de forma X → Y:

- Llamamos a X determinante o implicante.
- Llamamos a Y determinado o implicado.



Dependencias funcionales: ejemplo

Sea R(A, DF)

- R = Persona
 A = {Documento, Apellido, Nombre}
 DF = {X → Y}
 X = {Documento}
 Y = {Apellido, Nombre}
- O sea {Documento} → {Apellido, Nombre}

Dependencias funcionales: ejemplo

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	ОТО333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

- ¿Que dependencias existen en esta relación?
- ¿De que depende la existencia de dependencias entre los datos?

• Las dependencias funcionales son **restricciones semánticas**, más específicamente restricciones de **integridad de los datos**.

 Tener conocimiento de las dependencias funcionales en el esquema relacional permite evitar la redundancia en los datos.

 Son invariantes en el tiempo, siempre que no varíe el mundo real que se esta modelando, por lo tanto se han e cumplir para cualquier extensión del esquema de relación.

 No es posible deducir la existencia de una dependencia a partir de la observación de una extensión del esquema de relación.

 A partir de una extensión válida de un esquema de relación, sí será posible comprobar que una dependencia no se cumple para ese esquema.

Documento	Apellido	Nombre	Patente	Marca	Modelo
99.888.888	Perez	Jorge	HTT001	Renault	Clio
44.555.555	Alvarez	Maria	IZO121	Toyota	Kamry
99.888.888	Perez	Jorge	OTO333	Fiat	Uno
55.666.666	Fernandez	Jorgelina	JTJ777	Ford	Fiesta
99.888.888	Perez	Jorge	KKK456	Fiat	Uno

Con dependencias:

{Patente→ Documento, Marca, Modelo; Documento→ Apellido, Nombre}

¿Es posible agregar?

- 1. (99.888.888, Perez, Jorge, HTP886, Fiat, Siena)
- 2. (99.888.888, Diaz, Jorge, IZO120, Fiat, Spazio)
- 3. (11.222.222, Diaz, Maria, OTO333, Fiat, Uno)

- Son incluidas en la definición del esquema de relación.
- Sea el esquema de relación R(A, DF)

- Donde A es el conjunto de atributos de la relación.
- Y DF es el conjunto de dependencias funcionales existentes entre los atributos de la relación.

Dependencias funcionales: encontrando

- 1. Existe una gran cantidad de hoteles (H) en la ciudad.
- Cada hotel (H) tiene una categoría (C) y una dirección (D), donde se encuentra ubicado. Cada hotel (H) tiene varios números telefónicos (T) para poder comunicarse con él. Cada uno de los hoteles (H), tiene una cantidad de habitaciones (O).
- 3. Cada habitación (A), de un hotel (H) en particular, tiene una determinada capacidad (N) de personas. Además, cada habitación (A) de un hotel (H), tiene un número de interno (I) para poder comunicarse con ella.
- 4. Cada habitación (A) de un hotel (H), tiene un precio (P) determinado. El precio (P) depende en realidad, unicamente de la capacidad (N) de la habitación (A), sin importar de que hotel (H) sea la misma.

Dependencias funcionales: encontrando

Sea R(H, C, D, T, O, A, N, I, P)

- Punto 1 {}
- Punto 2 {H → CD, D→ H, T → H, H → O}
- Punto 3 {HA → N, HA → I, HI → A}
- Punto 4 {HA → P, N → P}

Dependencias funcionales: encontrando

Podemos clasificarlas en:

- Triviales y no triviales
- Plenas y parciales
- Elementales y no elementales
- Transitivas

Dependencia funcional trivial

 Una dependencia funcional X → Y se dice que es trivial, si Y es un subconjunto de X.



Dependencia funcional trivial: ejemplo

Sea R(AT, DF)

$$AT = \{A, B, C\}$$

$$DF = \{A \rightarrow C, AB \rightarrow B, C \rightarrow C\}$$

- AB → B es una dependencia funcional trivial, ya que B es subconjunto de AB.
- A → C es una dependencia funcional no trivial, dado que C no es un subconjunto de A.
- C → C es una dependencia funcional trivial ya que C = C.

Dependencia funcional plena o completa

- Se dice que Y tiene dependencia funcional plena o completa de X, si depende funcionalmente de X, pero no de ningún subconjunto estricto de X.
- Si la DF no es plena, se dice que es parcial.

$$X \Rightarrow Y sii \not\exists X' \subset X/X' \rightarrow Y$$

Dependencia funcional plena o completa: ejemplo

Sea R(AT, DF)
 AT = {A, B, C}
 DF = {A → C, AB → C, C → B}

- AB → C es una dependencia funcional parcial, ya que existe A
 → C y A es subconjunto de AB.
- C → B es una dependencia funcional plena, dado que no existe X subconjunto de C, con X → B.

Dependencia funcional elemental

 Decimos que una dependencia X → Y es elemental, si es una dependencia funcional plena y no trivial, en la que el implicado es simple (tiene un solo atributo).

$$(\nexists Y' \subset Y) \land (Y \not\subseteq X) \land (\nexists X' \subset X/X' \rightarrow Y)$$

Dependencia funcional elemental: ejemplo

Sea R(AT, DF)

$$AT = \{A, B, C\}$$

 $DF = \{A \rightarrow C, B \rightarrow AC\}$

 A → C es una dependencia funcional elemental, ya que es plena (no existe X → C con X subconjunto de A), no es trivial (C no es subconjunto de A) y el implicado es un único atributo(sólo es C).

Dependencia funcional transitiva

- Para analizar la transitividad se necesitan mínimamente tres conjuntos de atributos X, Y, Z.
- Una dependencia es transitiva cuando el implicado de una dependencia actúa en otra como implicante.
- Es decir X → Y y a su vez Y → Z, por lo tanto, por transitividad, puede deducirse que X → Z.

$$X \to Y \to Z$$

$$X \to Y \to Z$$

Dependencia funcional transitiva: ejemplo

Sea R(AT, DF)

$$AT = \{A, B, C, D, E, F\}$$

$$DF = \{A \rightarrow B, B \rightarrow C, A \rightarrow DE, DE \rightarrow F\}$$

- $A \rightarrow C$, ya que $A \rightarrow B$ y $B \rightarrow C$.
- $A \rightarrow F$, ya que $A \rightarrow DE$ y $DE \rightarrow F$.

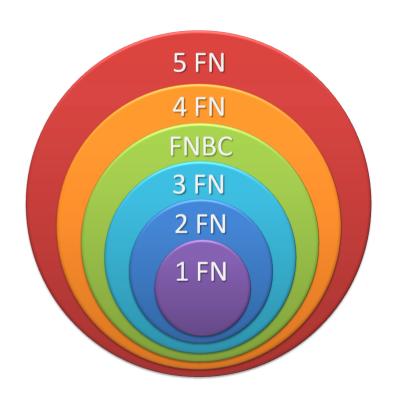
Reglas de inferencia de dependencias funcionales

- El conocimiento de ciertas dependencias funcionales nos puede llevar a inferir la existencia de otras, que no se encontraban en el conjunto inicial.
- Es posible deducir nuevas dependencias que sean una consecuencia lógica del conjunto de partida (Axiomas de Armstrong).
- Es posible encontrar claves candidatas.
- Eliminar dependencias funcionales redundantes.

Formas normales

 Las formas normales nos van a permitir detectar si una relación presenta anomalías.

 Cuanto mayor es la forma normal de una relación, menor número de anomalías presenta.



Primeras Formas Normales

 Edgar Frank Codd propuso en 1970 tres formas normales basadas en las dependencias funcionales:

- Primera Forma Normal (1FN),
- Segunda Forma Normal (2FN)
- Tercera Forma Normal (3FN).

Primeras Formas Normales

 Debido a que en tercera forma normal (3FN) aún persisten algunos problemas en las relaciones, en 1974 Boyce y Codd introdujeron una definición más restrictiva de la tercera formal normal (3FN), que se denominó:

Forma Normal de Boyce-Codd (FNBC).

Formas Normales avanzadas

- Fagin introdujo nuevas formas normales, que no están basadas en dependencias funcionales:
 - En 1977 introdujo la cuarta forma normal (4FN), basada en el concepto de dependencia multivaluada.
 - En 1979 introdujo la quinta formal normal (5FN), basada en el concepto de dependencia de combinación.
 - En **1981** introdujo la forma normal de dominios y claves **(FNDC)**, que no está basada en **ningún tipo de dependencias**.

Primera Forma Normal (1FN)

 La primera forma normal (1FN) es una restricción inherente al modelo relacional, por lo que su cumplimiento es obligatorio.

 Una relación que no está en 1FN, no pertenece al modelo relacional.

Primera Forma Normal (1FN)

Establece que:

- Los dominios de los atributos deben incluir solo valores atómicos.
- El **valor** de un **atributo** en una tupla, debe ser un valor **monovalente** (individual) proveniente del dominio de ese atributo.

Ejemplo 1FN

Documento	Apellido	Nombre	Teléfonos
11.222.333	Perez	Gustavo	444444, 434343
44.555.666	Alvarez	Marcela	424242
89.123.456	Gonzalez	Alvaro	414141, 432222

- El atributo teléfonos, es multivaluado.
- No es una relación, no pertenece al modelo relacional.
- No está en 1FN.

Ejemplo 1FN

Documento	Apellido	Nombre
11.222.333	Perez	Gustavo
44.555.666	Alvarez	Marcela
89.123.456	Gonzalez	Alvaro

Documento	Telefono
11.222.333	44444
11.222.333	434343
44.555.666	424242
89.123.456	414141
89.123.456	432222

- La transformamos a 1FN.
- Todos los atributos son monovalentes.

Ejemplo 1FN

 Siempre es posible transformar un esquema de relación que no esté en 1FN, en un esquema o esquemas de relación en 1FN, sin que se produzca pérdida de información o de dependencias.

Segunda Forma Normal (2FN): Conceptos previos

- Atributo principal o primo:
 - es aquel que **forma parte** del **alguna clave candidata** de la relación.

- Atributo no principal o no primo:
 - es aquel que **no forma parte** del **ninguna clave candidata** de la relación.

Segunda Forma Normal (2FN)

Está basada en:

- El concepto de dependencia plena o completa.
- En las interrelaciones que existen entre los atributos principales y no principales.
- Una relación está en Segunda Forma Normal (2FN) si y sólo si:
 - Está en **1FN.**
 - Cada atributo no principal tiene dependencia funcional plena respecto de cada una de las claves.
 - Un atributo no principal no debe depender parcialmente de la clave.

Segunda Forma Normal (2FN)

- La 2FN no se cumple cuando:
 - Algún atributo no principal, depende funcionalmente de algún subconjunto de una clave.
- Siempre están en 2FN las relaciones:
 - Binarias (relaciones de grado 2).
 - Con todos los atributos son principales, es decir, forman parte de alguna clave.
 - Con todas las claves candidatas simples, es decir, que contienen un sólo atributo.

Ejemplo 2FN

Sea R(AT, DF)

```
    AT = {Legajo, CArea, Nombre, Area, Desde}
    DF = {Legajo, CArea → Desde;
    Legajo → Nombre;
    CArea → Area}
```

Legajo	CArea	Nombre	Area	Desde
2121	101	Roberto	Contaduría	01/01/2001
4444	101	Rodolfo	Contaduría	02/01/2002
2424	102	María	Tesorería	01/03/2003
3333	102	Ricardo	Tesorería	05/05/2005

Ejemplo 2FN

 Se descompone en (Se crea una nueva tabla para cada dependencia parcial):

```
    R1(AT, DF)
    AT = {Legajo, Nombre}
    DF = {Legajo → Nombre}
    R2(AT, DF)
    AT = {CArea, Area}
    DF = {CArea → Area}
    R3(AT, DF)
    AT = {Legajo, CArea, Desde}
    DF = {Legajo, CArea → Desde}
```

Ejemplo 2FN

Legajo	Nombre
2121	Roberto
4444	Rodolfo
2424	María
3333	Ricardo

Legajo	CArea	Desde
2121	101	01/01/2001
4444	101	02/01/2002
2424	102	01/03/2003
3333	102	05/05/2005

CArea	Area
101	Contaduría
102	Tesorería

Segunda Forma Normal (2FN)

 Siempre es posible transformar un esquema de relación que no esté en 2FN, en esquemas de relación en 2FN, sin que se produzca pérdida de información o de dependencias.

Tercera Forma Normal (3FN)

- La tercera forma normal (3FN) está basada en:
 - El concepto de dependencia transitiva.
- Una relación está en 3FN si:
 - Está en **2FN.**
 - No existe ningún atributo no principal que dependa transitivamente de alguna clave de la relación.

Tercera Forma Normal (3FN)

La 3FN no se cumple cuando:

 Existen atributos no principales que dependen funcionalmente de otros atributos no principales.

Siempre están en 3FN las relaciones:

- Binarias.
- Con todos los atributos son principales, es decir, forman parte de alguna clave.
- Si esta en 2FN y tiene un único atributo no principal.

Ejemplo 3FN

Sea R(AT, DF)

```
- AT = {<u>Documento</u>, Nombre, Curso, Aula}
```

Documento	Nombre	Curso	Aula
11.222.333	Juan Pablo	Informática	Α
22.333.444	Jose María	Inglés	В
33.444.555	Juan Jose	Informática	Α
44.555.666	Carlos Alberto	Inglés	В

Ejemplo 3FN

Se descompone en:

```
    R1(AT, DF)
    AT = {Documento, Nombre, Curso}
    DF = {Documento → Nombre, Curso}
    R2(AT, DF)
    AT = {Curso, Aula}
    DF = {Curso → Aula}
```

Ejemplo 3FN

Documento	Nombre	Curso
11.222.333	Juan Pablo	Informática
22.333.444	Jose María	Inglés
33.444.555	Juan Jose	Informática
44.555.666	Carlos Alberto	Inglés

Curso	Aula
Informática	Α
Inglés	В

Tercera Forma Normal (3FN)

 Siempre es posible transformar un esquema de relación que no esté en 3FN, en esquemas de relación en 3FN, sin que se produzca pérdida de información o de dependencias.

Un esquema relacional en 3FN se considera de buena calidad.
 3FN garantiza un buen equilibrio entre eficiencia y control de redundancia.

Forma Normal de Boyce-Codd (FNBC)

- Se trata de una redefinición más estricta de la 3FN, surge como solución a las anomalías de actualización que se producen cuando:
 - La relación tiene **varias** claves candidatas
 - Las claves candidatas son compuestas
 - Las claves candidatas se solapan o tienen atributos en común

- Una relación se encuentra en FNBC si:
 - Está en **3FN**.
 - Todo determinante es una clave.

Forma Normal de Boyce-Codd (FNBC)

- La FNBC no se cumple cuando:
 - Existe **algún determinante** que **no es clave** candidata.

- Siempre están en FNBC las relaciones:
 - Binarias.
 - Que están en **3FN** y sus **claves no se solapan**.

Ejemplo FNBC

Sea R(AT, DF)

```
- AT = {<u>Estudiante</u>, <u>Asignatura</u>, Profesor}
```

DF = {Estudiante, Asignatura → Profesor;Profesor → Asignatura}

Estudiante	Asignatura	Profesor
23.444.555	BD1	Ted Codd
24.333.222	BD1	Ted Codd
24.444.555	BD1	Peter Chen
23.333.222	BD2	Jhon Backus

Ejemplo FNBC

Se descompone en:

```
    R2(AT, DF)
    AT = {Estudiante, Profesor}
    DF = {}
    R1(AT, DF)
```

 $AT = \{ Profesor, Asignatura \}$

 $DF = \{Profesor \rightarrow Asignatura\}$

Ejemplo FNBC

Estudiante	Profesor
23.444.555	Ted Codd
24.333.222	Ted Codd
24.444.555	Peter Chen
23.333.222	Jhon Backus

Profesor	Asignatura
Ted Codd	BD1
Peter Chen	BD1
Jhon Backus	BD2

Forma Normal de Boyce-Codd (FNBC)

 No siempre es posible transformar un esquema de relación que no está en FNBC en esquemas de relación en FNBC sin que se produzca pérdida de dependencias funcionales.

 Sí se puede asegurar que la transformación se puede producir siempre sin pérdida de información.

Normalización

 La normalización de los datos es un proceso incremental y restrictivo, en el cual el esquema de una relación se descompone, repartiendo los atributos y dependencias en esquemas de relación más pequeños para:

- Minimizar datos redundantes
- Evitar anomalías.

Normalización

- La descomposición obtenida debe poseer las siguientes propiedades:
 - Conservación de la Información.
 - Conservación de las Dependencias Funcionales.

Conservación de la Información

- La finalidad de esta propiedad es conseguir que el proceso de normalización se lleve a cabo sin pérdida de la información existente.
- Para que se cumpla esta propiedad son precisas dos condiciones:

- Conservación de los atributos
- Conservación del contenido (tuplas)

Conservación de la Información

- La finalidad de esta propiedad es conseguir que el proceso de normalización se lleve a cabo sin pérdida de la información existente.
- Para que se cumpla esta propiedad son precisas dos condiciones:

- Conservación de los atributos
- Conservación del contenido (tuplas)

Métodos de Normalización

 Existen dos enfoques para la normalización de un esquema de relación:

- Análisis o descomposición (Top-Down)
- Síntesis (Bottom-Up)

A tener en cuenta

- Toda relación que está en FNBC está en 3FN.
- La mayoría de las relaciones que están en 3FN, también están en FNBC, pero no necesariamente.
- Es mejor tener los diseños en FNBC, pero si no se puede, entonces se dejan en 3FN.
- 2FN y 1FN no se consideran buenos diseños.

Bibliografía

 Adoración de Miguel Castaño y otros. Diseño de Bases de Datos Relacionales. 2000. Alfaomega y Ra-Ma.