



Bases de
Datos



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS
FUNDAMENTOS DE BASES DE DATOS



Normalización de Bases de Datos

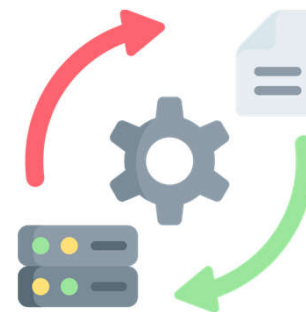
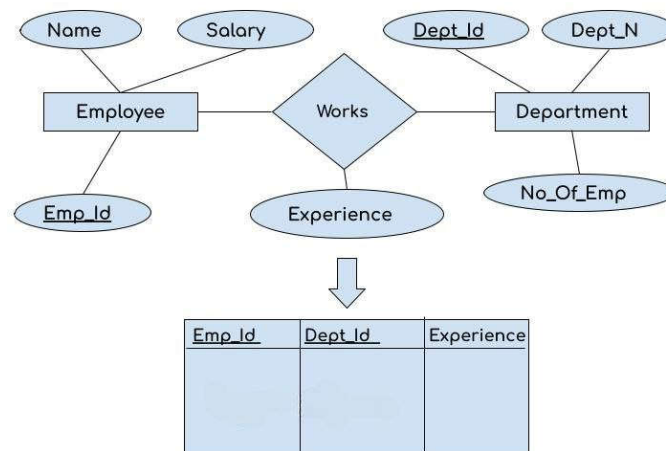
Gerardo Avilés Rosas

✉ gar@ciencias.unam.mx

INTRODUCCIÓN



- Al **diseñar** una **base de datos**, el **objetivo principal** es crear una **representación precisa** de los **datos involucrados** en el problema, las **relaciones entre los datos**, las **reglas de negocio** y las **restricciones** sobre los datos.
- Para lograr este objetivo, se suele tomar el **documento de requerimientos** de la base de datos y traducirlo a un **modelo gráfico**, independiente de la **tecnología de bases de datos** (**Modelo Entidad-Relación**) → **Modelo Conceptual**, que deberá **traducirse** a un formalismo que entienda el **Sistema Manejador de Bases de Datos (SMBD)**.



...INTRODUCCIÓN



- Transformar del **Modelo Lógico** al **Modelo Físico** puede resultar **demasiado arriesgado**, ya que el mapeo del **esquema conceptual** al **esquema lógico** no garantiza que este último **esté libre de redundancia y/o inconsistencias**.
 - ❑ **Redundancia de datos:** cuando **un hecho** (dato) se almacena **en más de un lugar**, de forma innecesaria.
 - ❑ **Inconsistencia de datos:** cuando **los mismos datos** tienen **diferentes formatos de representación** en varios lugares.



...INTRODUCCIÓN



- Que en una base de datos **exista redundancia** puede **causar inconsistencias en los datos**, que se traduce en **información poco confiable** y conduce a un **mayor costo de almacenamiento y acceso**.
- Una de las causas más comunes para **agregar redundancia** a un modelo de datos, es **tratar de incluir** en la **misma relación** atributos **univaluados** y **multivaluados**.

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-17	\$200,000
Copilco	Coyoacán	\$420 M	Laura Gómez	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Ana López	P-15	\$300,000
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo	P-14	\$300,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	P-11	\$180,000
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez	P-29	\$240,000
San Fernando	Tlalpan	\$740 M	Leticia López	P-16	\$260,000
Centro	Cuauhtémoc	\$1,800 M	Raúl González	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Hernán Ríos	P-25	\$500,000

ANOMALÍAS EN BASES DE DATOS



- Una **anomalía de base de datos** es un **error o inconsistencia**, que puede resultar cuando un usuario intenta realizar alguna **operación de mantenimiento de datos** en una relación que contiene **datos redundantes**.

- ☐ **Almacenamiento redundante:** parte de la información se almacena repetidamente.
- ☐ **Anomalía de actualización.** si se **actualiza una copia** de dichos datos repetidos, se crea una **inconsistencia** a menos que **todas las copias se actualicen** de manera similar.

```
UPDATE clienteprestamo SET alcaldia = 'Miguel Hidalgo'
```

```
WHERE cliente = 'Carlos Santos';
```

- ☐ **Anomalía de inserción:** es posible que **no se pueda almacenar** cierta información a menos que **también se almacene otra información no relacionada**.

```
INSERT INTO clienteprestamo(sucursal, alcaldia, activo)
```

```
VALUES('LORETO', 'Álvaro Obregón', '$60 M');
```

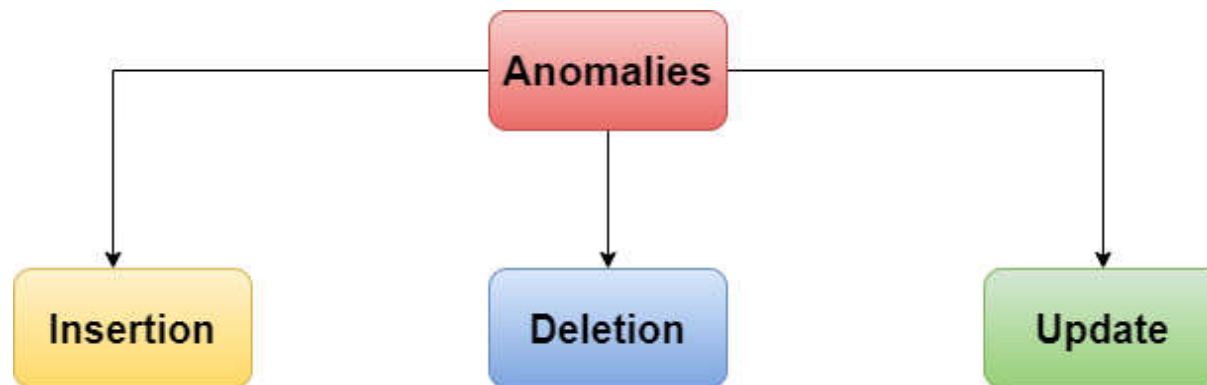
ANOMALÍAS EN BASES DE DATOS



- ❑ **Anomalía de eliminación.** es posible que **no pueda eliminar cierta información sin perder también otra información no relacionada.**

`DELETE FROM clienteprestamo`

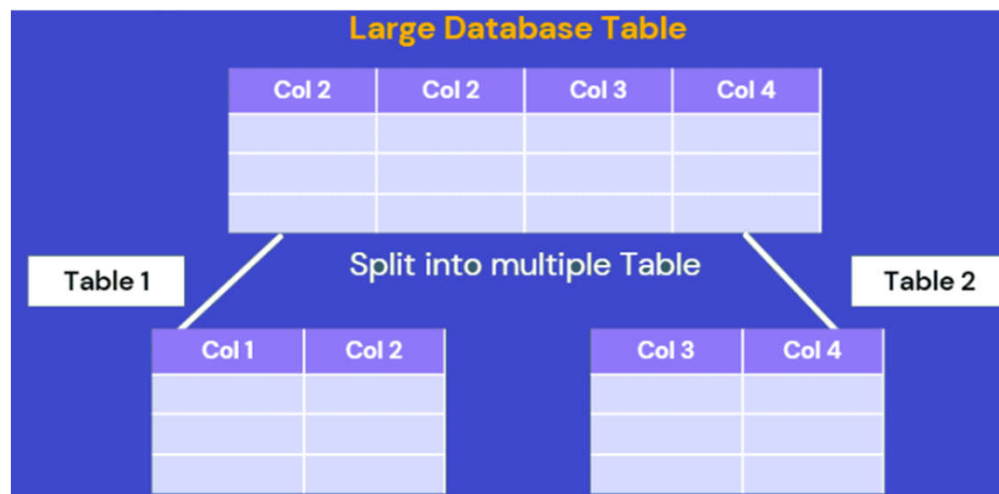
`WHERE numprestamo = 'P-11';`



NORMALIZACIÓN DE BASES DE DATOS



- Una **forma de acabar con anomalías** como la redundancia es a través de la **descomposición de relaciones** → **Proceso de descomponer relaciones con anomalías** (sucesivamente) para producir **relaciones pequeñas y bien estructuradas**:
 - ❑ El proceso **comprueba el cumplimiento** de una **serie de reglas**, de manera que al cumplirse una regla se aumenta el grado de normalización.
 - ❑ Cuando una **regla no se cumple**, la relación **debe descomponerse** en **varios esquemas** que **si la cumplan**.



...NORMALIZACIÓN DE BASES DE DATOS



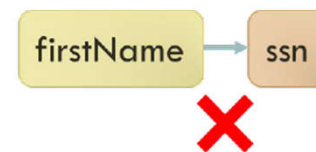
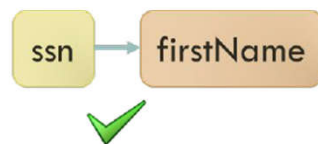
- La **Normalización** es una **técnica de diseño de bases de datos** que comienza **examinando** las **relaciones que existen entre los atributos**, denominadas **dependencias funcionales**.
 - ❑ Utiliza una **serie de reglas (formas normales)** que permiten **identificar el conjunto adecuado de relaciones** que **respalden los requerimientos de datos** de una organización.
- **Conjunto adecuado de relaciones**, aquel que cumple con las **siguientes características**:
 - ❑ **Número mínimo de atributos** necesarios para **respaldar los requerimientos** de datos.
 - ❑ **Atributos estrecha relación lógica** (dependencia funcional) deben **estar en la misma relación**.
 - ❑ **Redundancia mínima** en cada atributo representado una sola vez, **exceptuando a los atributos que forman parte o son la llave foránea**.

DEPENDENCIA FUNCIONAL



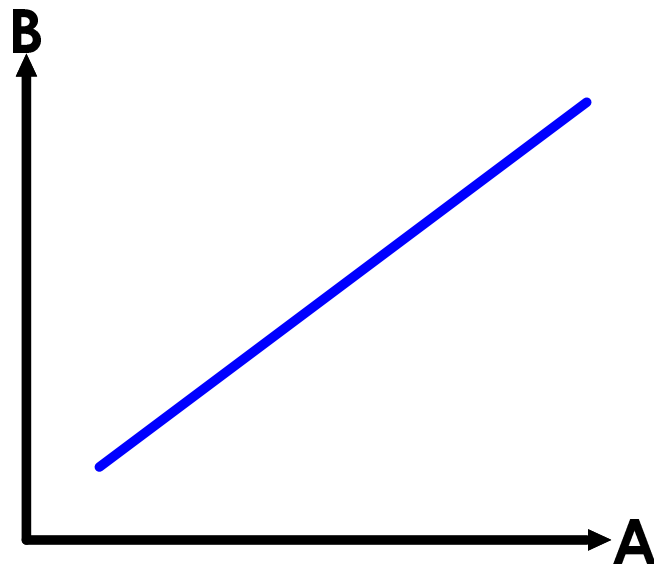
- Es **una restricción** que ocurre entre **dos conjuntos de atributos** de la base de datos. Suponer el **esquema** de base de datos relacional R tiene n **atributos**, $R(A_1, A_2, A_3, \dots, A_n)$ y que la **base de datos completa** está almacenada en la **relación** R .
 - Una **dependencia funcional**, denotada por $X \rightarrow Y$, ocurre entre **dos conjuntos de atributos** X e Y que son **subconjuntos de** R especifica **una restricción** sobre las **tuplas posibles** que pueden formar un **estado de la relación** R .

Patient		
ssn	firstName	lastName
235-14-7854	Sandra	Smith
192-48-0924	John	Moore
821-13-2108	Laura	Turner
874-72-0093	John	Moore

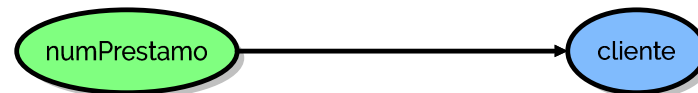


DEPENDENCIA FUNCIONAL

- La **restricción** indica que los **valores del atributo Y** de una tupla en **R**, **están determinados** por los **valores del atributo X**; además, los **valores del atributo X** de la tupla, **determinan de forma única** (funcional) **los valores del atributo Y**.

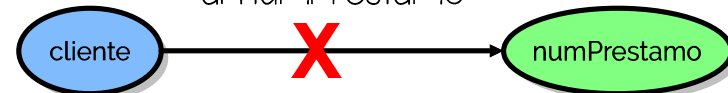


numPrestamo determina
funcionalmente al cliente



P-17 → Carlos Santos

cliente no determina funcionalmente
al numPrestamo



Carlos Santos → P-17
Carlos Santos → P-93

...DEPENDENCIA FUNCIONAL



- Las **dependencias funcionales** son una **propiedad de la semántica** de los atributos, de manera que los **diseñadores de la base de datos** deben tener una **comprensión clara del significado** de los atributos de R , para **especificar las DFs que deberían cumplirse** en R .
- Una **característica adicional** de las DFs que es **útil para la normalización** es que sus **determinantes** deben tener el **número mínimo de atributos** necesarios para mantener la DF con los **atributos del lado derecho**.
 - ❑ Una **dependencia funcional completa** indica que, si A y B son atributos de una relación, B **depende completa y funcionalmente** de A , si B **depende funcionalmente** de A , pero no de **ningún subconjunto propio** de A .
- Otro tipo de DF que es **importante definir**, ya que, **de existir** en un esquema de base de datos, **puede causar anomalías de actualización**, se denominan **dependencias funcionales transitivas**.
 - ❑ Una **dependencia funcional transitiva** se presenta cuando A , B y C son **atributos de una relación** tales que si $A \rightarrow B$ y $B \rightarrow C$, entonces C **depende transitivamente** de A a través de B (siempre que A **no dependa funcionalmente** de B o C).

...DEPENDENCIA FUNCIONAL



- Por ejemplo, vamos a suponer que tenemos la **relación**:
***Pedidos** (idCliente, nombre, paterno, materno, calle, ciudad, estado, CP, teléfono, noPedido fechaOrden, noArticulo, producto, precio, ¿enviado?)*
- Se espera, por ejemplo, que:
 - idCliente* → nombre, paterno, materno, calle, ciudad, estado, CP, teléfono
 - noPedido* → idCliente, fechaOrden
 - noArticulo* → producto, precio
 - noPedido + noArticulo* → ¿enviado?
- Aunque los valores de los atributos pueden cambiar, en cualquier momento, sólo hay uno.
- Las **DFs** se utilizan para:
 - ☐ Especificar restricciones sobre el conjunto de relaciones.
 - ☐ Examinar las relaciones y determinar si son legales bajo un conjunto de dependencias funcionales dado.

LLAVES DE LAS RELACIONES



- **Llave: conjunto de atributos $A_1, A_2, A_3, \dots, A_n$ tales que:**
 - i. **Determinan funcionalmente** cualquier otro atributo de R (**dos tuplas** distintas **no pueden coincidir** en todos los atributos $A_1, A_2, A_3, \dots, A_n$).
 - ii. **Ningún subconjunto propio** de $A_1, A_2, A_3, \dots, A_n$ **determina funcionalmente** a otros atributos de R , es decir, debe ser **mínimo**.
- **Llave candidata: atributo o combinación de atributos**, que **identifica de manera única** una tupla en una relación y debe cumplir las siguientes propiedades:
 - i. **Identificación única.** Para cada tupla, el valor de la llave **debe identificar** de **forma única** esa tupla. Esta propiedad implica que **cada atributo no llave** depende funcionalmente de esa llave.
 - ii. **No redundancia.** Ningún atributo en la llave se puede **eliminar** sin destruir la propiedad de **identificación única**.
- **Superllave** es un **conjunto de atributos** que **contiene una llave**. Cada llave es una superllave, pero algunas **superllaves no son mínimas**.

LLAVES DE LAS RELACIONES



- Si X es una **llave** de R esto implica que $X \rightarrow Y$ para cualquier subconjunto de atributos Y de R .
- Si $X \rightarrow Y$ en R , esto no implica que $Y \rightarrow X$
- Una **DF** es una **propiedad semántica** de los atributos, la cual se debe cumplir para la **extensión** en una relación.
- Las *extensiones* que satisfacen las restricciones de DFs se denominan **extensiones legales** o **estados legales** debido a que obedecen las restricciones de la DF.

A	B	C	D
a1	b1	c1	d1
a1	b2	c1	d2
a2	b2	c2	d2
a2	b3	c2	d3
a3	b3	c2	d4

¿ $A \rightarrow B$?

¿ $A \rightarrow C$?

¿ $AB \rightarrow CD$?

REGLAS DE INFERENCIA DE ARMSTRONG



- Fueron desarrolladas por **William W. Armstrong** en **1974**: ***Dependency Structures of Data Base Relationships.***
- Se trata de un conjunto de reglas que permiten deducir **todas** las dependencias funcionales que tienen lugar en un conjunto de atributos dados, como **consecuencia** de aquellas que se asumen como **ciertas**, a partir del conocimiento del problema.
- Este resultado permite establecer un conjunto de algoritmos sencillos para:
 1. **Encontrar el conjunto cerrado de un conjunto de dependencias funcionales.**
 2. **Encontrar la equivalencia lógica de esquemas.**
 3. **Deducir dependencias funcionales.**
 4. **Calcular las llaves de un esquema.**

...REGLAS DE INFERENCIA DE ARMSTRONG



1. Regla de la **reflexividad**: Si $Y \subseteq X$, entonces $X \rightarrow Y$
 2. Regla del **aumento**: $\{X \rightarrow Y\} \Rightarrow XZ \rightarrow YZ$
 3. Regla de la **transitividad**: $\{X \rightarrow Y, Y \rightarrow Z\} \Rightarrow X \rightarrow Z$
 4. Regla de la **descomposición**: $\{X \rightarrow YZ\} \Rightarrow X \rightarrow Y$ y $X \rightarrow Z$
 5. Regla de la **unión**: $\{X \rightarrow Y, X \rightarrow Z\} \Rightarrow X \rightarrow YZ$
 6. Regla de la **pseudo-transitividad**: $\{X \rightarrow Y, WY \rightarrow Z\} \Rightarrow WX \rightarrow Z$
-
- Por ejemplo, sean $R = (A, B, C, G, H, I)$ y $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$
 - Algunos miembros de F^+ serían:
 - ☐ **Unión**:
 - ☐ **Pseudo-transitividad**:
 - ☐ **Aumento**:
 - ☐ **Transitividad**:

CERRADURA DEL CONJUNTO DE ATRIBUTOS



- Para crear las dependencias funcionales de un esquema, requerimos:
 - ❑ Especificar a partir de la semántica de los atributos de R , el conjunto F de DFs .
 - ❑ Usando las reglas de inferencia de Armstrong obtener otras DFs .
- Para obtener todas las dependencias funcionales de manera sistemática:
 - ❑ Determinar el conjunto de atributos X del lado izquierdo de alguna DF en F .
 - ❑ Determinar el conjunto $X +$ de todos los atributos que son dependientes de X .
- Dado $\{A_1, A_2, \dots, A_n\}$ un conjunto de atributos y F un conjunto de DFs , la **cerradura del conjunto de atributos** $(\{A_1, A_2, \dots, A_n\} +)$ bajo las dependencias en F es el conjunto de **atributos** B tales que cada relación que satisface todas las dependencias en F también satisface:

$$A_1, A_2, \dots, A_n \rightarrow B$$

ALGORITMO PARA CALCULAR X^+ BAJO F



Repetir

anterior X^+ = X^+ ;

Para cada $Y \rightarrow Z$ en F hacer

Sí $Y \subseteq X^+$ entonces $X^+ = X^+ \cup Z$;

hasta que anterior X^+ = X^+ ;

Ejemplos:

1. Sean $R = \{A, B, C, G, H, I\}$ y $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$, ¿cuál sería la cerradura AG ?
2. Sea $R = \{A, B, C, D, E, F\}$ y $F = \{AB \rightarrow C, BC \rightarrow D, D \rightarrow E, CF \rightarrow B\}$, ¿cuál sería la cerradura AB ?

DEPENDENCIAS DEDUCIDAS



Si queremos probar si una **dependencia funcional** $A_1, A_2, \dots, A_n \rightarrow B$, se **deduce** de un conjunto de dependencias F , debemos calcular $\{A_1, A_2, \dots, A_n\}^+$, si B está ahí, entonces la **DF** si es deducida del conjunto F , en caso contrario no es deducida de F .

Por ejemplo, sea $R = \{A, B, C, D, E, F\}$ y $F = \{AB \rightarrow C, BC \rightarrow D, D \rightarrow E, CF \rightarrow B\}$

- Probar que $AB \rightarrow D$

- Probar que $D \rightarrow A$



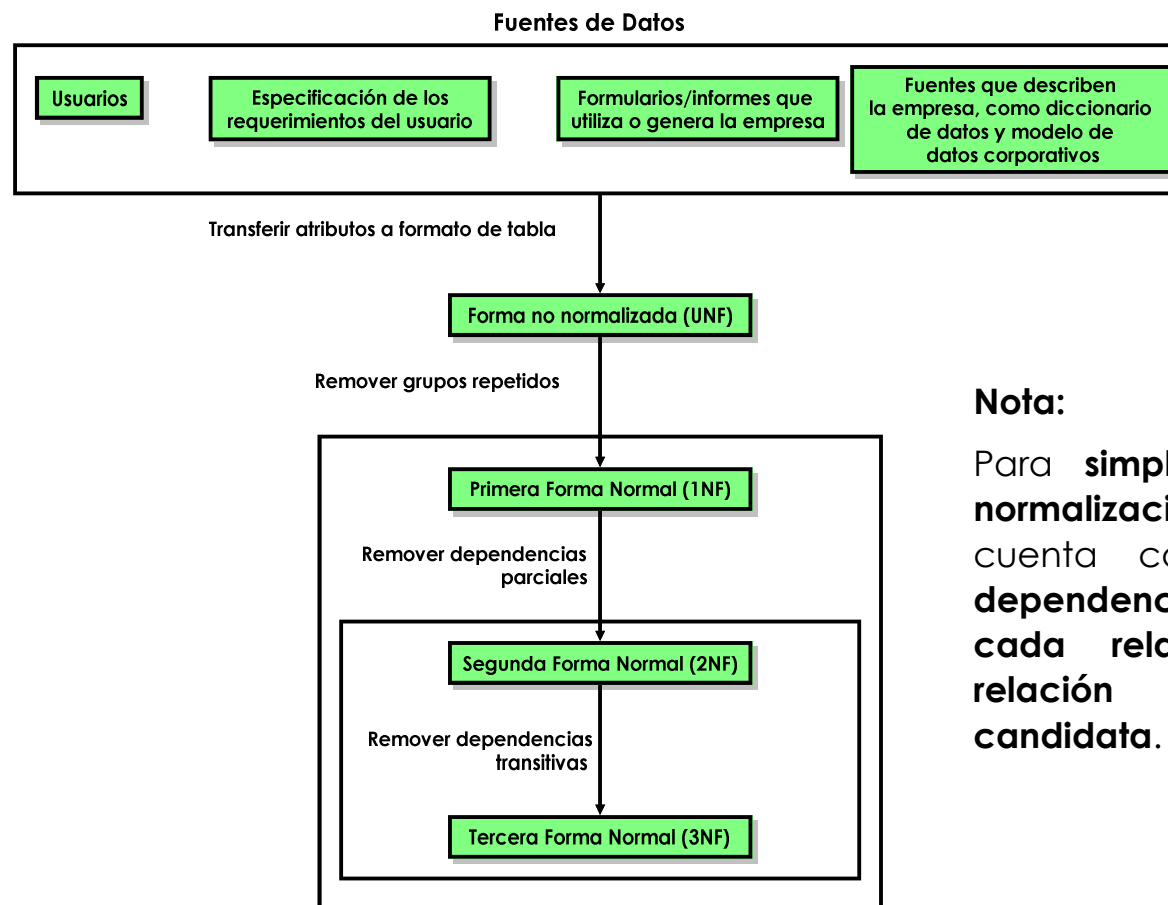
Y ahora...
¿Cómo normalizo
una Base de Datos?

NORMALIZACIÓN DE RELACIONES



- Es una **técnica formal** para **analizar relaciones basadas** en su **llave primaria** (o **llaves candidatas**) y **dependencias funcionales**, e involucra una **serie de reglas** que **se pueden usar para probar las relaciones individuales** y que una base de datos **se pueda normalizar** en cualquier grado.
 - ❑ Cuando **no se cumple un requisito**, se dice que la relación viola el requerimiento **debe descomponerse** en relaciones que **cumplan individualmente** las reglas de normalización.
- Inicialmente se propusieron **tres formas normales: 1NF, 2NF y 3NF**.
 - ❑ Posteriormente, **R. Boyce** y **E.F. Codd** introdujeron una **definición más fuerte** de la **3NF** que llamaron **Forma Normal de BoyceCodd (BCNF, 1974)**.
- Exceptuando la **1NF**, todas las demás **Formas Normales** se basan en las **dependencias funcionales** que tienen lugar en una relación (Maier, 1983).
- La **normalización** requiere una **serie de pasos**, donde **cada paso** corresponde a una **forma normal**. A medida que **avanza la normalización**, las relaciones se vuelven **progresivamente más restringidas** y **menos vulnerables** a las **anomalías**.
 - ❑ Para la **existencia** del **Modelo Relacional**, solo es **obligatoria** la **1NF** para poder hablar de relaciones; de manera que las **demás formas normales son opcionales**, pero se recomienda llegar por lo menos a **3NF**.

...NORMALIZACIÓN DE RELACIONES



Nota:

Para **simplificar** el **proceso de normalización**, se asumirá que se cuenta con un **conjunto de dependencias funcionales** para **cada relación** y que **cada relación** tiene alguna **llave candidata**.

DESCOMPOSICIÓN DE RELACIONES



- Una **descomposición de un esquema de relación R** consiste en **reemplazar el esquema de relación** por **dos (o más) esquemas** que contengan cada uno **un subconjunto de los atributos de R** y **juntos incluyen todos los atributos**.
- Si **no se tiene cuidado al descomponer** un esquema de relación, se **pueden generar más problemas** que soluciones, de ahí que será importante considerar **dos cuestiones**:
 1. **¿Es necesario descomponer una relación?** Se sabe que **sí un esquema de relación** está en alguna **forma normal**, es posible saber **cuáles tipos de problemas no pueden surgir**, entonces se puede utilizar **la información** que aporta la forma normal para **ayudar a decidir si se debe descomponer más o no**.
 2. **¿Qué problemas podría causar una descomposición dada?** Se deben considerar **dos propiedades** de la descomposición:
 - ❑ **Propiedad de reunión sin pérdida.** Permite recuperar cualquier instancia de la relación descompuesta de las instancias correspondientes de las relaciones más pequeñas.
 - ❑ **Propiedad de preservación de DFs.** Permite imponer cualquier restricción en la relación original en algunas de las relaciones más pequeñas, de manera que no es necesario realizar reunión de las relaciones más pequeñas para verificar si se viola una restricción en la relación original.

...DESCOMPOSICIÓN DE RELACIONES



Prestamo

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-17	\$200,000
Copilco	Coyoacán	\$420 M	Laura Gómez	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Ana López	P-15	\$300,000
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo	P-14	\$300,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	P-11	\$180,000
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez	P-29	\$240,000
San Fernando	Tlalpan	\$740 M	Leticia López	P-16	\$260,000
Centro	Cuauhtémoc	\$1,800 M	Raúl González	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Hernán Ríos	P-25	\$500,000

Si se descompone la relación en:

1. Una relación **Sucursal** (*Sucursal, Alcaldía, Activo, Cliente*) y
2. Una relación **Prestamo** (*Cliente, numPrestamo, Importe*)

...DESCOMPOSICIÓN DE RELACIONES



Sucursal

Sucursal	Alcaldia	Activo	Cliente
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos
Copilco	Coyoacán	\$420 M	Laura Gómez
Viveros	Coyoacán	\$340 M	Ana López
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo
Eugenia	Benito Juárez	\$80 M	Carlos Santos
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez
San Fernando	Tlalpan	\$740 M	Leticia López
Centro	Cuauhtémoc	\$1,800 M	Raúl González
Viveros	Coyoacán	\$340 M	Hernán Ríos

Prestamo

Cliente	numPrestamo	importe
Carlos Santos	P-17	\$200,000
Laura Gómez	P-23	\$400,000
Ana López	P-15	\$300,000
Pedro Toledo	P-14	\$300,000
Carlos Santos	P-93	\$100,000
Raúl Pérez	P-11	\$180,000
Alma Vázquez	P-29	\$240,000
Leticia López	P-16	\$260,000
Raúl González	P-23	\$400,000
Hernán Ríos	P-25	\$500,000

...DESCOMPOSICIÓN DE RELACIONES



1. Encontrar todas las sucursales que tienen préstamos con importe menor a **\$200,000**:

$\pi_{\text{Sucursal}} \sigma_{\text{importe} < 200000} (\text{Prestamo})$

Con el esquema original:

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-17	\$200,000
Copilco	Coyoacán	\$420 M	Laura Gómez	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Ana López	P-15	\$300,000
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo	P-14	\$300,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	P-11	\$180,000
San Ángel	Álvaro Obregon	\$60 M	Alma Vazquez	P-29	\$240,000
San Fernando	Tlalpan	\$740 M	Leticia López	P-16	\$260,000
Centro	Cuauhtémoc	\$1,800 M	Raúl González	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Hernán Ríos	P-25	\$500,000

Eugenia y Zapata

...DESCOMPOSICIÓN DE RELACIONES



Con el **esquema fraccionado**:

$\pi_{\text{Sucursal}} \sigma_{\text{importe} < 200000} (\text{Sucursal} \bowtie \text{Prestamo})$

Sucursal				Prestamo		
Sucursal	Alcaldia	Activo	Cliente	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	Carlos Santos	P-17	\$200,000
Copilco	Coyoacán	\$420 M	Laura Gómez	Laura Gómez	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Ana López	Ana López	P-15	\$300,000
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo	Pedro Toledo	P-14	\$300,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	Raúl Pérez	P-11	\$180,000
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez	Alma Vázquez	P-29	\$240,000
San Fernando	Tlalpan	\$740 M	Leticia López	Leticia López	P-16	\$260,000
Centro	Cuauhtémoc	\$1,800 M	Raúl González	Raúl González	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Hernán Ríos	Hernán Ríos	P-25	\$500,000

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-93	\$100,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	P-11	\$180,000

Eugenia, Zapata y Centro

...DESCOMPOSICIÓN DE RELACIONES



2. Indicar los préstamos que se tienen en cada sucursal el cliente **Carlos Santos**:

$\gamma_{\text{count}(\text{Sucursal}) \rightarrow \text{conteo}} \sigma_{\text{nom_cliente} = \text{'Carlos Santos'}} (\text{Prestamo})$

Con el esquema original:

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-17	\$200,000
Copilco	Coyoacán	\$420 M	Laura Gómez	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Ana López	P-15	\$300,000
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo	P-14	\$300,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez	P-11	\$180,000
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez	P-29	\$240,000
San Fernando	Tlalpan	\$740 M	Leticia López	P-16	\$260,000
Centro	Cuauhtémoc	\$1,800 M	Raúl González	P-23	\$400,000
Viveros	Coyoacán	\$340 M	Hernán Ríos	P-25	\$500,000

DESCOMPOSICIÓN DE RELACIONES

Con el **esquema fraccionado**:

$\gamma_{\text{count}(\text{Sucursal}) \rightarrow \text{conteo}} \sigma_{\text{nom_cliente} = \text{'Carlos Santos'}} (\text{Sucursal} \bowtie \text{Prestamo})$

Sucursal

Sucursal	Alcaldia	Activo	Cliente
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos
Copilco	Coyoacán	\$420 M	Laura Gómez
Viveros	Coyoacán	\$340 M	Ana López
Centro	Cuauhtémoc	\$1,800 M	Pedro Toledo
Eugenia	Benito Juárez	\$80 M	Carlos Santos
Zapata	Benito Juárez	\$1,600 M	Raúl Pérez
San Ángel	Álvaro Obregón	\$60 M	Alma Vázquez
San Fernando	Tlalpan	\$740 M	Leticia López
Centro	Cuauhtémoc	\$1,800 M	Raúl González
Viveros	Coyoacán	\$340 M	Hernán Ríos

Prestamo

Cliente	numPrestamo	importe
Carlos Santos	P-17	\$200,000
Laura Gómez	P-23	\$400,000
Ana López	P-15	\$300,000
Pedro Toledo	P-14	\$300,000
Carlos Santos	P-93	\$100,000
Raúl Pérez	P-11	\$180,000
Alma Vázquez	P-29	\$240,000
Leticia López	P-16	\$260,000
Raúl González	P-23	\$400,000
Hernán Ríos	P-25	\$500,000

Sucursal	Alcaldia	Activo	Cliente	numPrestamo	importe
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-17	\$200,000
Centro	Cuauhtémoc	\$1,800 M	Carlos Santos	P-93	\$100,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-17	\$200,000
Eugenia	Benito Juárez	\$80 M	Carlos Santos	P-93	\$100,000

...DESCOMPOSICIÓN DE RELACIONES



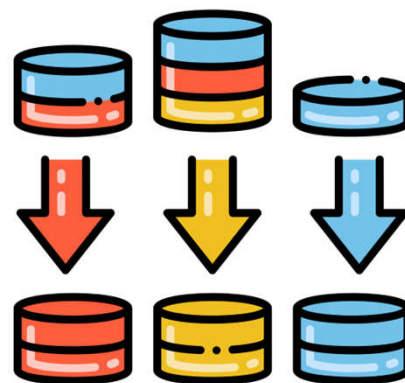
Ambas relaciones tienen a **Cliente** en común, así que para reunirlos se usa este atributo, que **no resulta adecuado** puesto que un cliente puede tener **varios préstamos** no necesariamente en la misma sucursal.



OBJETIVOS DE LA NORMALIZACIÓN



- Para **normalizar** una **Base de Datos** se desea:
 - ❑ **Minimizar la redundancia** de datos, **evitando anomalías** y **conservando espacio** de almacenamiento.
 - ❑ **Simplificar** el cumplimiento de las **restricciones de integridad referencial**.
 - ❑ Hacer **más fácil** el **mantenimiento** de datos (*insert, delete, update*)
 - ❑ Proporcionar un **mejor diseño** (*representación mejorada del mundo real*) y una **base sólida** para el crecimiento futuro.
 - ❑ Lograr que las relaciones fraccionadas tengan **JOIN sin pérdida**.
 - ❑ **Conservar las dependencias funcionales**.



- El **proceso de normalización** busca producir **relaciones pequeñas y bien estructuradas**.
 - ❑ Una **relación bien estructurada** es una relación que contiene **redundancia mínima** y permite a los usuarios realizar **operaciones de mantenimiento de datos** sobre las tuplas de una relación, **sin provocar errores o inconsistencias**.
- Una **forma normal** se refiere a los **estados de una relación** que requiere que ciertas **normas relativas a las relaciones** entre los atributos (o **dependencias funcionales**) **se satisfagan**
- Un **esquema de relación** se puede someter a los siguientes **estados de normalización** son:
 1. **Primera forma normal (1NF)**. Cualquier **atributo multivaluado** (incluso **grupos repetidos**) **han sido removidos**. Solo se permiten **valores atómicos y posiblemente nulos**.
 2. **Segunda forma normal (2NF)**. Cualquier **dependencia funcional parcial** **se han removido** (los atributos que no son llave se identifican por toda la llave primaria).
 3. **Tercera forma normal (3NF)**. Cualquier **dependencia transitiva** **se ha removido** (atributos que no son llave son identificados solo por la llave primaria).
 4. **Forma normal de Boyce-Codd (BCNF)**. **Anomalías restantes**, resultados de las **dependencias funcionales** **se han removido** (puede haber más de una llave primaria para los mismos atributos).

...FORMAS NORMALES

5. **Cuarta Forma Normal (4NF).** Cualquier **dependencia multivaluada** se han removido.
6. **Quinta Forma Normal (5NF).** **Anomalías que no se pudieron retirar** por las anteriores formas normales, se han removido.

	1NF	2NF	3NF	4NF	5NF
Decomposition of Relation	R	R ₁₁	R ₂₁	R ₃₁	R ₄₁
		R ₁₂	R ₂₂	R ₃₂	R ₄₂
Conditions			R ₂₃	R ₃₃	R ₄₃
				R ₃₄	R ₄₄
					R ₄₅
	Eliminate Repeating Groups	Eliminate Partial Functional Dependency	Eliminate Transitive Dependency	Eliminate Multi-values Dependency	Eliminate Join Dependency

PRIMERA FORMA NORMAL (1NF)



- Está considerada como parte de la **definición formal** de una **relación en Modelo Relacional** y se definió para **no permitir atributos multivaluados, atributos compuestos** y sus combinaciones.
- La **1NF** establece que el **dominio de un atributo debe incluir solo valores atómicos** (simples, indivisibles) y que el **valor de cualquier atributo en una tupla** debe ser un **valor único del dominio de ese atributo**.

NoCliente	NombreC	idProp	Direccion	InicioRenta	FinRenta	Renta	IdDueño	NombreD
CR76	Carlos Pérez	PG4	Duraznos 196	01-07-2012	31-08-2013	3500	CO40	Ana Sánchez
		PG16	Faisanes 325	01-09-2013	01-09-2014	4500	CO93	Juan López
CR56	Sergio Ríos	PG4	Duraznos 196	01-07-2011	10-06-2012	3500	CO40	Ana Sánchez
		PG36	Tulipanes 825	10-10-2012	01-12-2013	3750	CO93	Juan López
		PG16	Cedros 456	01-11-2014	10-08-2015	4800	CO93	Juan López

NoCliente	NombreC	idProp	Direccion	InicioRenta	FinRenta	Renta	IdDueño	NombreD
CR76	Carlos Pérez	PG4	Duraznos 196	01-07-2012	31-08-2013	3500	CO40	Ana Sánchez
CR76	Carlos Pérez	PG16	Faisanes 325	01-09-2013	01-09-2014	4500	CO93	Juan López
CR56	Sergio Ríos	PG4	Duraznos 196	01-07-2011	10-06-2012	3500	CO40	Ana Sánchez
CR56	Sergio Ríos	PG36	Tulipanes 825	10-10-2012	01-12-2013	3750	CO93	Juan López
CR56	Sergio Ríos	PG16	Cedros 456	01-11-2014	10-08-2015	4800	CO93	Juan López

SEGUNDA FORMA NORMAL (2NF)



- Esta **forma normal** se basa en el concepto de **dependencia funcional total**, y se **aplica a relaciones con llaves compuestas**. Una **relación** con una **llave primaria** formada por **un solo atributo** está automáticamente en al menos **2NF**.
- Una relación está en **2NF** si está en **1NF** y cada atributo que **no es llave primaria depende funcionalmente** de la llave primaria.
- La **normalización** de las relaciones **1NF** a **2NF** implica la **eliminación de dependencias parciales**.
 - Si existe una **dependencia parcial**, se **eliminan los atributos parcialmente dependientes** de la relación **colocándolos en una nueva relación junto con una copia de su determinante**.

NoCliente	NombreC
CR76	Carlos Pérez
CR56	Sergio Ríos

NoCliente	idProp	InicioRenta	FinRenta
CR76	PG4	01-07-2012	31-08-2013
CR76	PG16	01-09-2013	01-09-2014
CR56	PG4	01-07-2011	10-06-2012
CR56	PG36	10-10-2012	01-12-2013
CR56	PG16	01-11-2014	10-08-2015

idProp	Direccion	Renta	IdDueño	NombreD
PG4	Duraznos 196	3500	CO40	Ana Sánchez
PG16	Faisanes 325	4500	CO93	Juan López
PG36	Tulipanes 825	3750	CO93	Juan López

TERCERA FORMA NORMAL (3NF)

- La normalización de las relaciones **2NF** a **3NF** implica la **eliminación de dependencias transitivas**.
 - Si existe una **dependencia transitiva**, se deben eliminar los **atributos transitivamente dependientes** de la relación **colocando los atributos en una nueva relación** junto con una **copia del determinante**.
- Una relación **R** está en **tercera forma normal (3NF)** con respecto a **DF**, si para toda **dependencia funcional no trivial** $A_1, A_2, A_3, \dots, A_n \rightarrow B$, se tiene que:
 - El lado izquierdo $\{A_1, A_2, A_3, \dots, A_n\}$ es una **superllave** o bien,
 - El lado derecho **B**, es miembro de alguna **llave candidata** de **R**.

- Algoritmo para obtener 3NF**

Hacer **F** mínimo

Para toda **dependencia funcional** en **F** mínimo:

a. **Crear una relación** que contenga **sólo los atributos** de cada **dependencia funcional**.

b. **Eliminar un esquema** si es **subconjunto** de otro.

Si no existen esquemas que contengan llaves candidatas, crear una relación con esos atributos.

FORMA NORMAL DE BOYCE-CODD (*BCNF*)



- *BCNF* se basa en **DFs** que toman en cuenta todas las **llaves candidatas** en una relación.
- Una relación ***R*** está en ***BCNF*** si y sólo si en toda **DF no trivial** $A_1, A_2, A_3, \dots, A_n \rightarrow B$ para ***R***, se tiene que $\{A_1, A_2, A_3, \dots, A_n\}$ es **superllave** de ***R***.
- **Algoritmo para obtener *BCNF***
 - ❑ Buscar una **dependencia funcional no trivial** $X \rightarrow B$ que viole ***BCNF***.
 - ❑ Calcular X^+ .
 - ❑ **Dividir *R*** en $R_1(X^+) \cup R_2((R - X^+) \cup X)$.
 - ❑ Encontrar las **dependencias funcionales** para las **nuevas relaciones**.

¡GRACIAS!

No estés muy orgulloso de haber comprendido estas notas. La habilidad para manejar los **conceptos de Normalización** es insignificante comparado con el poder de **la Fuerza**.

