

# Introducción a las Bases de Datos Relacionales

---

LCC SONIA VALERY

# TEMA 2

---

EL MODELO DE DATOS RELACIONAL

# Introducción

---

- ❑ El modelo relacional fue introducido en 1970 por Codd -quien fallecería en 2003 a los 79 años-, sobre una estructura de datos sencilla y uniforme: la relación.
- ❑ Codd fue quien propuso que los sistemas de datos deberían representarse en estructuras llamadas “relaciones”, definidas como un conjunto de tuplas, donde el orden no es importante.
- ❑ Sus preocupaciones se centraron en la parte de consulta, no en la de almacenamiento.

# Introducción

---

- Este modelo es relativamente nuevo y se ha establecido como el principal modelo de datos para aplicaciones comerciales de procesamiento de datos, fundamentalmente debido a que existen en el mercado muchos SGBD (Sistemas Gestores de Base de Datos) relacionales comerciales.

# Introducción

---

- ❑ Principalmente, el modelo relacional intenta describir una base de datos mediante un conjunto de predicados y una colección de atributos que describen las distintas restricciones sobre las combinaciones o valores posibles. Su consistencia viene dada por la inclusión de limitaciones añadidas en el diseño de bases de datos, denominado “esquema lógico”.
- ❑ Cabe destacar sobre el modelo relacional que fue el primer modelo de base de datos que se representó con términos matemáticos.

# Relaciones y sus propiedades

---

- ❑ El modelo relacional representa una base de datos como un conjunto de relaciones, esto es, tablas, formadas por filas y columnas.
- ❑ Los nombres de las tablas y columnas ayudan a la interpretación del significado de los valores que se encuentran en las filas de la tabla.
- ❑ En conclusión, puede afirmarse que una base de datos relacional es un conjunto de tablas con nombres únicos, en los que las filas representan hechos y las columnas representan propiedades.

# Relaciones y sus propiedades

---

- ❑ La relación es el elemento básico del modelo relacional que se puede representar como una tabla formada por filas y columnas.
- ❑ En ella se distingue su nombre, un conjunto de columnas denominadas “atributos” (propiedades de la tabla identificadas por nombre) y un conjunto de filas llamadas “tuplas”, las cuales contienen los valores de cada atributo.

# Relaciones y sus propiedades

## RELACION ALUMNO

### Atributos

Tuplas



NUM_CTA	APELLIDO _1	APELLIDO _2	NOMBRE_1	Nombre_2
123	Martínez	Álvarez	Juan	José
456	Jiménez	Collado	Luis	Ramon
789	González	Pérez	Verónica	
012	Suárez	Treviño	Cristina	María



# Relaciones y sus propiedades

---

- ❑ Dentro del modelo relacional, una fila se denomina “tupla”, una cabecera de columna es un atributo y una tabla es una relación.
- ❑ Se caracteriza, una vez representada como tabla, por no admitir filas duplicadas;
- ❑ Las filas y columnas no están ordenadas y la representación es plana, es decir, en el cruce de una fila y columna sólo puede haber un valor.



¿Qué diferencia existe entre las tuplas, los atributos y la relación?

# Propiedades de las relaciones

## Grado de una relación:

- Viene dado por el número de columnas que forman la tabla.

## Cardinalidad de una relación:

- Número de filas que forman parte de cada tabla.

Grado y  
Cardinalidad

# Propiedades de las relaciones

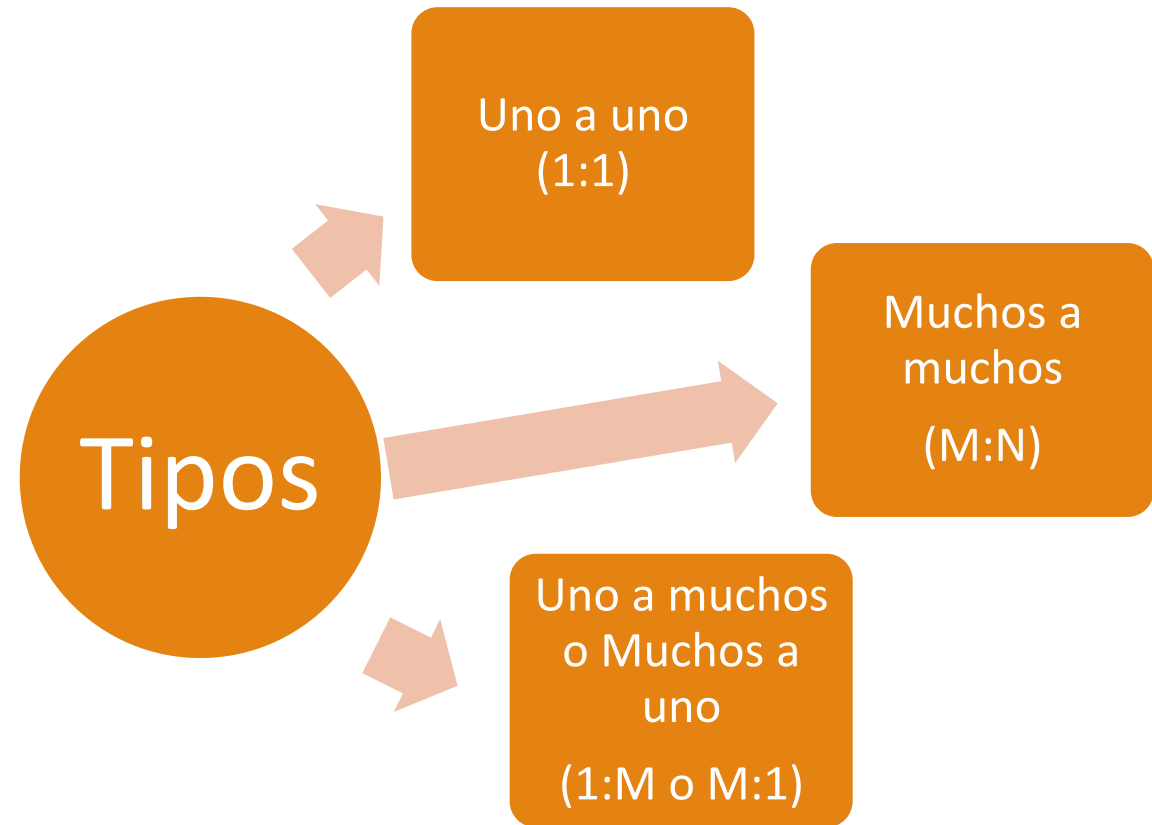
- ❑ Partiendo de la siguiente tabla o relación: DATOS\_ALUMNOS(NUM\_CTA\_NOMBRE, EDAD)

ALUMNOS		
NUM_CTA	NOMBRE	EDAD
123	MARÍA	18
456	SONIA	20
789	JUAN	20
011	Pedro	19

- ❑ Grado de relación sería igual a 3
- ❑ Cardinalidad sería igual a 4.

# Cardinalidad de un relación

Recordemos que en una relación el número de ocurrencias de una fila (tupla) se define como cardinalidad.



# Propiedades de las relaciones

## Nominadas

- **Persistentes:** aquellas relaciones cuya definición o esquema de relación permanece en la base de datos, borrándose solamente mediante una acción explícita del usuario.
- **Temporales:** desaparece de la base de datos en un cierto momento, sin necesidad de realizar ninguna acción explícita.

## Sin nombre

- **Resultados de consultas** que no se materializan, sino que entregan al usuario que ha realizado la consulta.

# Esquema de una Relación

---

- ❑ Un esquema contiene la definición estructural de relaciones o tablas de una base de datos
- ❑ Determina la identidad de la relación y qué tipo de información podrá ser almacenada dentro de ella;
- ❑ Contiene los metadatos (datos acerca de los datos) de la relación.

# Esquema de una Relación

---

Todo esquema constará de:

- ❑ Nombre de la relación (su identificador)
- ❑ Nombre de los atributos (o campos) de la relación
- ❑ El dominio de un atributo o campo define los valores permitidos para el mismo, equivalente al tipo de dato por ejemplo character, integer, date, string...



# Esquema de una Relación

---

Los Dominios pueden ser de 2 tipos:

- ❑ Dominio predefinido: son los tipos de datos que proporcionan los lenguajes de bases de datos.
- ❑ Dominio definido por el usuario: tipos de datos que definen los usuarios indicando el nombre del dominio y los valores que forman parte de él.

# Claves

---

Toda la información que contiene una base de datos debe poderse identificar de alguna forma.

A continuación definimos:

- ☐ Superclave
- ☐ Clave Candidata
- ☐ Clave Primaria
- ☐ Clave Alternativa de una relación

# Superclave

---

- Una Superclave de una relación es un subconjunto de los atributos del esquema tal que NO puede haber dos tuplas que tengan la **misma combinación de valores**.



# Ejemplo de Superclave

En la relación: EMPLEADOS(RFC, NSS, Nombre, Apellido, Telefono)



Las **superclaves** de la relación serían los siguientes subconjuntos de atributos:



{RFC, NSS, Nombre, Apellido, Teléfono}

{RFC, Apellido}

{RFC}

{NSS}

# Superclave

---

- ❑ Toda relación tiene, por lo menos, una Superclave, que es la formada por todos los atributos de su esquema.
- ❑ Esto se debe a la propiedad que cumple toda relación de no tener tuplas repetidas.

{RFC, NSS, nombre, apellido, teléfono}

# Clave Candidata

---

- Una Clave Candidata de una relación es: Una Superclave C que cumple que ningún subconjunto de C es Superclave.
- La clave candidata permite identificar cualquier tupla de una relación, de manera que no sobre ningún atributo para hacer la identificación.

Clave Candidata = **Superclave MINIMA**

{RFC} {NSS}



# Clave Candidata

- En la relación EMPLEADOS(RFC, NSS, Nombre, Apellido, Telefono)

Hay dos claves candidatas, {RFC} y {NSS}, se puede elegir como clave primaria {RFC}.

- Subrayaremos el atributo RFC en el esquema de la relación:

EMPLEADOS(RFC, NSS, Nombre, Apellido, Telefono)

- La clave {NSS} será una clave alternativa de EMPLEADOS.

# Clave Primaria

---

- ❑ Una Clave Primaria de una relación es: **una de las claves candidatas**.
- ❑ Es la clave candidata cuyos valores se utilizarán para identificar las tuplas de la relación.
- ❑ Las claves candidatas no elegidas como primaria se denominan **Claves Alternativas**.



# Clave Compuesta

---

- ❑ Es posible que una clave candidata o una clave primaria conste de más de un atributo (compuesta)
- ❑ En la relación: DESPACHOS(EDIFICIO, NÚMERO, superficie) la clave primaria está formada por los atributos EDIFICIO Y NÚMERO
- ❑ En este caso, podrá ocurrir que dos despachos diferentes estén en el mismo edificio, o bien que tengan el mismo número, pero nunca pasará que tengan la misma combinación de valores para edificio y número.

# Claves Foráneas

---

- ❑ Hasta ahora hemos estudiado las relaciones de forma individual
- ❑ Una BDR normalmente contiene más de una relación, para poder representar distintos tipos de hechos que suceden en el mundo real.
- ❑ Una Clave Foránea es aquella que es Clave Primaria en otra tabla que participa en una relación.



# Claves Foráneas

- ❑ Por ejemplo, podríamos tener una pequeña base de datos que contuviese dos relaciones: EMPLEADOS Y DESPACHOS
- ❑ EMPLEADOS, que almacenaría datos de los empleados de una empresa
- ❑ DESPACHOS, que almacenaría los datos de los despachos que tiene la empresa.

EMPLEADOS			
RFC	NOMBRES	APELLIDOS	SUELDO

DESPACHOS		
EDIFICIO	NUMERO	SUPERFICIE



# Claves Foráneas

- En el modelo relacional, para reflejar este tipo de vínculos, tenemos la posibilidad de expresar conexiones entre las distintas tuplas de las relaciones.
- Por ejemplo, en la base de datos anterior, que tiene las relaciones EMPLEADOS y DESPACHOS, puede ser necesario conectar tuplas de EMPLEADOS con tuplas de DESPACHOS para indicar qué despacho tiene asignado cada empleado.

EMPLEADOS					
RFC	NOMBRES	APELLIDOS	SUELDO	EDIFICIO	NUMERO



DESPACHOS		
EDIFICIO	NUMERO	SUPERFICIE



# Claves Foráneas

- En ocasiones, incluso puede ser necesario reflejar lazos entre tuplas que pertenecen a una misma relación.
- Por ejemplo, en la misma base de datos anterior puede ser necesario conectar determinadas tuplas de EMPLEADOS con otras tuplas de EMPLEADOS para indicar, para cada empleado, quién actúa como su jefe.

EMPLEADOS				
RFC	NOMBRES	APELLIDOS	SUELDO	RFC_JEFE



# Claves Foráneas

---

- ❑ El mecanismo que proporcionan las bases de datos relacionales para conectar tuplas son las claves foráneas de las relaciones.
- ❑ Las claves foráneas permiten establecer conexiones entre las tuplas de las relaciones.
- ❑ Los valores de una clave foránea deben estar presentes en la clave primaria correspondiente, o bien deben ser valores nulos.
- ❑ En caso contrario, la clave foránea representaría una referencia o conexión incorrecta.

# Claves Foráneas

---

- ❑ En la relación de esquema:

EMPLEADOS(RFC, NOMBRES, APELLIDOS, RFC-JEFE, EDIFICIO, NUMERO),

- ❑ La clave foránea {EDIFICIO, NUMERO} referencia la relación DESPACHOS(EDIFICIO, NÚMERO, SUPERFICIE).
- ❑ De este modo, se cumple que todos los valores que no son nulos de los atributos EDIFICIO, NUMERO son valores que existen para los atributos EDIFICIO, NUMERO de DESPACHOS, tal y como se puede ver a continuación:



# Claves Foráneas

EMPLEADOS					
<u>RFC</u>	NOMBRES	APELLIDOS	RFC-JEFE	EDIFICIO	NUMERO
GALP720514	Juan	García	NULO	Marina	120
ROMA781028	Marta	Roca	GALP720514	Marina	120
BUCA730613	Carlos	Buendia	GALP720514	Diagonal	120
PLEL650205	Elena	Pla	GALP720514	NULO	NULO

DESPACHOS		
<u>EDIFICIO</u>	<u>NUMERO</u>	SUPERFICIE
Marina	120	10
Marina	122	15
Marina	230	20
Diagonal	120	10





¿Qué diferencia existe entre una superclave y una clave primaria?

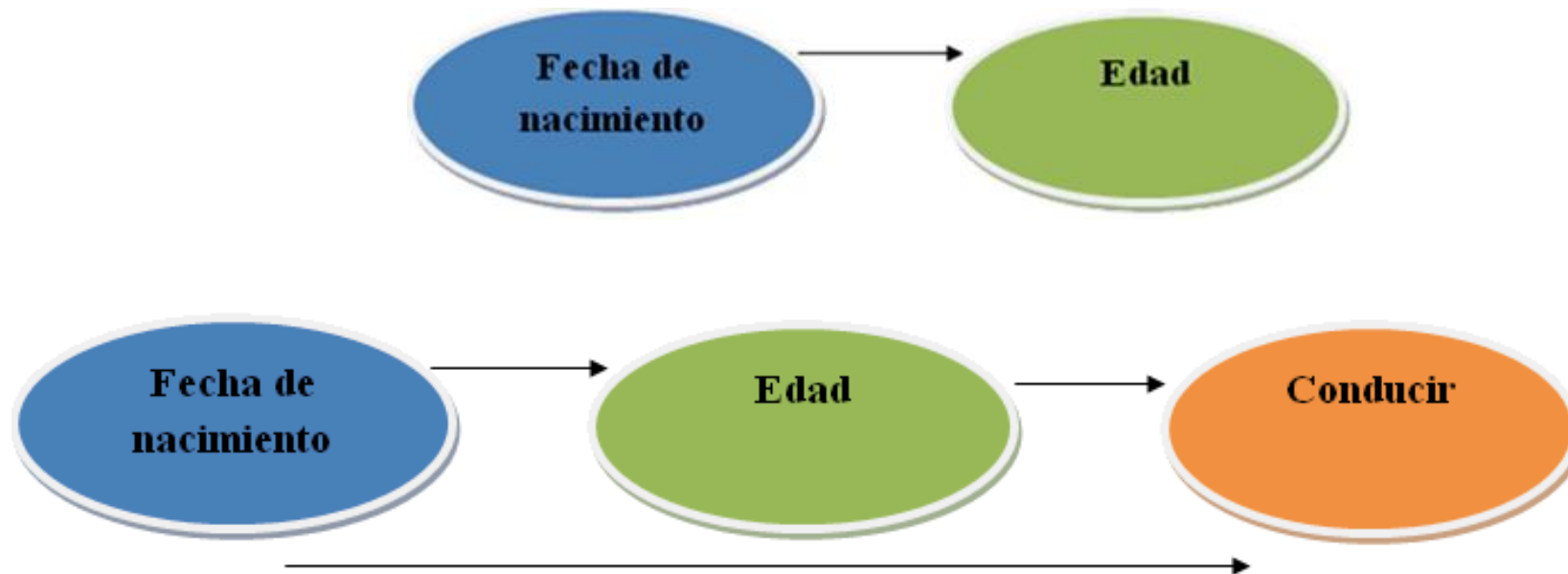
# Dependencias funcionales

- ❑ El concepto de dependencia funcional es de gran importancia para diseñar esquemas de bases de datos relacionales, puesto que es el REQUISITO BASE que deben cumplir los esquemas para realizar el proceso de normalización.
- ❑ Indica la relación entre atributos de una misma tabla en una base de datos.
- ❑ Cuando coinciden los valores de un atributo A coincidirán también los valores del atributo B.
- ❑ B es funcionalmente dependiente de A:





# Ejemplo





¿Cómo definirías con tus palabras una dependencia funcional?

# Normalización

---

- ❑ Proporciona una medida formal que nos indica o nos permite medir si una relación esta bien definida.
- ❑ Se enfoca en encontrar un conjunto óptimo de relaciones que de lugar a una estructura de fácil mantenimiento, evitando redundancias y errores en algunas operaciones.
- ❑ Consiste en analizar y descomponer las relaciones que forman la base de datos, dando lugar a mejores agrupaciones de atributos para la formación de tablas o relaciones.
- ❑ Evita una serie de problemas y anomalías que pueden proceder de un diseño erróneo.

# Normalización

---

- ❑ Las 3 primeras formas normales fueron definidas por Codd en 1970.
- ❑ En 1974, Boyce realizaría una revisión de la tercera forma normal y la llamaría "forma normal de Boyce-Codd".
- ❑ Después, surgirían la 5ta forma normal y 4ta forma normal, en 1977 y 1979, respectivamente, propuestas por Fagin.

# Normalización

---

- ❑ Los principales objetivos que se persiguen con la normalización son aquellos que evitan problemas como:
  - ❑ Ambigüedades: se definen así a aquellos datos que NO clarifican al registro que representan.
  - ❑ Redundancia: datos que se repiten continuamente por las tablas de las bases de datos.
  - ❑ Pérdida de restricciones de integridad: proporcionan una seguridad de la base de datos ante modificaciones, ya que se guardan en el diccionario de datos para poder ser consultado cuando se realicen actualizaciones.
  - ❑ Anomalías en operaciones de modificación de datos: evitar que si se elimina o inserta un solo elemento haya que eliminar o insertar varias tuplas.

# Normalización

---

- ❑ En el proceso de normalización se debe cumplir:
- ❑ Al aplicar una forma normal determinada se produzca descomposición sin pérdida de datos
- ❑ Se conserven las dependencias funcionales
- ❑ Las tablas resultantes de la descomposición conservan las dependencias funcionales en la relación original.



# Primera forma normal (1 FN)

---

- ❑ Una tabla está en Primera Forma Normal si:
- ❑ Los dominios de atributos solo deben incluir valores individuales (indivisibles), es decir los atributos son atómicos.
- ❑ No pueden existir relaciones anidadas y atributos multivaluados.
- ❑ La tabla contiene una clave primaria única.
- ❑ La clave primaria no contiene atributos nulos.

# Primera forma normal (1 FN)

---

- ☐ No debe existir variación en el número de columnas.
- ☐ Los Campos no clave deben identificarse por la clave (Dependencia Funcional)
- ☐ Debe Existir una independencia del orden tanto de las filas como de las columnas, es decir, si los datos cambian de orden no deben cambiar sus significados
- ☐ Esta forma normal elimina los valores repetidos dentro de una Base de Dato



# Ejemplo 1FN

❑ Se tiene una tabla que contiene los siguientes datos:

PROFESORES		
RFC	NOMBRE	DEPARTAMENTO
PEGL720312	LUIS	MATEMÁTICAS COMPUTACIÓN
RALJ700223	JUAN	QUÍMICA

❑ Esta tabla no cumple con la 1ra FN, puesto que cuenta con un valor multivaluado, es decir, que tiene varios valores un mismo campo.



# Ejemplo 1FN

## ❑ Tabla PROFESORES en 1ra FN

PROFESORES		
RFC	NOMBRE	DEPARTAMENTO
PEGL720312	LUIS	MATEMÁTICAS
PEGL720312	LUIS	COMPUTACIÓN
RALJ700223	JUAN	QUÍMICA

❑ Los datos ya están atomizados, pero hay cierta redundancia que es normal en la 1FN y que en la 2FN se optimiza.



# Actividad

□ Normalización a 1ra FN de la siguiente tabla:

BANCO		
RFC	NOMBRE	CUENTAS
GOPM751212	MARIA	CTA123
		CTA321
ARGP690703	PABLO	CTA456

# Segunda forma normal (2FN)

---

- ❑ Una tabla esta en segunda forma normal si:
  - ❑ Esta en 1FN
  - ❑ Posee una clave principal y el resto de atributos deben depender de ella
  - ❑ Si hay atributos que solo dependen de una parte de la clave deben formar otra tabla distinta.
- ❑ 2FN establece que todas las dependencias parciales se deben eliminar y separar dentro de sus propias tablas.
- ❑ Una dependencia parcial es un término que describe a aquellos datos que no dependen de la clave de la tabla para identificarlos.

# Segunda forma normal (2FN)

❑ Se tiene una tabla que contiene los siguientes datos:

ALUMNOS				
NUM_CTA	CURSO	NOMBRE	APELLIDO	CALIFICACIÓN
123	1	MARÍA	RAMIREZ	8
456	1	SONIA	GONZALEZ	9
456	4	SONIA	GONZALEZ	10
789	4	JUAN	LOPEZ	7
789	8	JUAN	LOPEZ	8

❑ Esta tabla no se encuentra en 2FN porque suponiendo que NUM\_CTA y CURSO forman la clave principal, NOMBRE y APELLIDO dependen solamente del NUM\_CTA, por lo tanto, deben realizarse dos tablas

# Segunda forma normal (2FN)

---

ALUMNOS		
<u>NUM_CTA</u>	NOMBRE	APELLIDO
123	MARÍA	RAMIREZ
456	SONIA	GONZALEZ
789	JUAN	LOPEZ

CALIFICACIONES		
<u>CURSO</u>	<u>NUM_CTA</u>	CALIFICACIÓN
1	123	8
1	456	9
4	456	10
4	789	7
8	789	8



# Segunda forma normal (2FN)

---

- ❑ Una relación se encuentra en segunda forma normal si está en primera forma normal y todos sus atributos dependen de la clave primaria completa.
- ❑ Al haber alcanzado la Segunda Forma Normal, se gana como ventaja que se pueden añadir nuevas columnas a una tabla sin afectar a las demás tablas.



□ ¿Cuáles son las diferencias entre la 1 FN y la 2fN?

# Tercera forma normal (3FN)

---

- ❑ Una relación esta en tercera forma normal cuando:
  - ❑ La tabla se encuentre en 2da FN
  - ❑ Ningún atributo dependa funcionalmente de otros atributos que no sean clave.
  - ❑ Cuando ningún atributo pueda depender transitivamente de la clave primaria.
- ❑ Se encuentra basada en el concepto de dependencia funcional transitiva.

# Tercera forma normal (3FN)

❑ Se tiene una tabla que contiene los siguientes datos:

ALUMNOS				
<u>NUM_CTA</u>	NOMBRE	APELLIDO	COD_POSTAL	POBLACIÓN
123	MARÍA	RAMIREZ	01710	ALVARO OBREGON
456	SONIA	GONZALEZ	09440	BENITO JUAREZ
789	JUAN	LOPEZ	14230	TLALPAN
987	PEDRO	JIMENEZ	01710	ALVARO OBREGON
564	CARMEN	PEREZ	09440	BENITO JUAREZ

❑ Esta tabla no se encuentra en tercera forma normal debido a que POBLACIÓN solo depende de COD\_POSTAL, no de NUM\_CTA, por lo tanto, hay que dividirla en dos tablas para que esté en 3ra FN.

# Tercera forma normal (3FN)

---

ALUMNOS			
<u>NUM_CTA</u>	NOMBRE	APELLIDO	COD_POSTAL
123	MARÍA	RAMIREZ	01700
456	SONIA	GONZALEZ	09400
789	JUAN	LOPEZ	14200
987	PEDRO	JIMENEZ	01700
564	CARMEN	PEREZ	09400

UBICACION	
<u>COD_POSTAL</u>	POBLACIÓN
01700	ALVARO OBREGON
09400	BENITO JUAREZ
14200	TLALPAN

# ANOMALÍAS DE LA 3ra FN

□ Analicemos la siguiente tabla:

TUTORIAS		
NUM_CTA	ASIGNATURA	TUTOR
123	MATEMÁTICAS	LUIS
123	QUÍMICA	JUÁN
789	MATEMÁTICAS	LUIS
987	BIOLOGÍA	PEDRO
387	QUIMICA	MARIA
387	COMPUTACIÓN	SONIA

# ANOMALÍAS DE LA 3ra FN

---

- ❑ Supongamos que los requerimientos que sustentan esta tabla son:
  - ❑ Un estudiante (NUM\_CTA), pueda tener una o más asignaturas
  - ❑ Una asignatura puede tener varios tutores como consejeros
  - ❑ Un tutor sólo proporciona asesoría en una asignatura

# ANOMALÍAS DE LA 3ra FN

---

- ❑ Puesto que los estudiantes pueden tener varias asignaturas, NUM\_CTA, no determina la asignatura.
- ❑ Como los estudiantes pueden tener varios tutores, NUM\_CTA tampoco determina al tutor.
- ❑ NUM\_CTA por sí mismo no puede ser una clave, ESTO ES:
- ❑ La combinación (NUM\_CTA, ASIGNATURA) determina al tutor
- ❑ La combinación (NUM\_CTA, TUTOR), determina ASIGNATURA.



# ANOMALÍAS DE LA 3ra FN

---

- ❑ Cualquiera de las combinaciones puede ser una clave.
- ❑ Dos o más atributos o conjuntos de atributos que puedan ser una clave, se denominan claves candidatas.
- ❑ Cualquiera de las candidatas que se seleccione para ser la clave se denomina clave primaria.
- ❑ TUTORIAS está en primera forma normal. También está en segunda forma normal, pues cualquier atributo que no es clave depende de toda la clave (sin importar cual clave candidata se seleccione).
- ❑ También está en tercera forma normal porque no tiene dependencias transitivas.

# ANOMALÍAS DE LA 3ra FN

---

- ☐ A pesar de todo esto tiene anomalías por modificación.
- ☐ Suponga que estudiante 387 deja la escuela.
- ☐ Si quita la tupla de estudiante 387 se perderá el hecho de que Sonia imparte asesoría en COMPUTACION. Ésta es una anomalía de eliminación.
- ☐ En forma similar ¿Cómo se almacena el hecho de que Pablo es asesor en Física?
- ☐ No es posible, hasta que un estudiante se inscribe en tal materia. Esta es una anomalía de inserción.

# Forma normal de Boyce-Codd (FNBC)

- ❑ Situaciones como la anterior conducen a la definición de la forma normal de Boyce-Codd (BCNF): Una tabla está en BCNF si cada determinante es una clave candidata.
- ❑ TUTORIAS no está en BCNF puesto que tiene un determinante, TUTOR , que no es una clave candidata.
- ❑ Las siguientes tablas ya están en FNBC:

ALUMNOS_CON_TUTORIAS	
<u>NUM_CTA</u>	<u>TUTOR</u>
123	LUIS
123	JUÁN
789	LUIS
987	PEDRO
387	MARIA
387	SONIA

TUTORES	
<u>TUTOR</u>	ASIGNATURA
LUIS	MATEMÁTICAS
JUÁN	QUÍMICA
PEDRO	BIOLOGÍA
MARIA	QUIMICA
SONIA	COMPUTACIÓN

# Referencias

---

- Raymond 'Ray' Boyce (1947-1974) was an American computer scientist who was known for his research in relational databases. He is most known for his work co-developing the SQL database language and Boyce-Codd normal form.
- Ronald Fagin (born 1945) is an American mathematician and computer scientist, and IBM Fellow at the IBM Almaden Research Center. He is known for his work in database theory, finite model theory, and reasoning about knowledge