



Tema 3-3: El Modelo de Datos Relacional

Gonzalo Cerruela García
ISCBD



Objetivos

- **La Teoría Relacional.**
- **El Modelo de Datos Relacional:**
 - **Terminología del modelo relacional**
 - **Consistencia de la representación lógica relacional.**
 - **Normalización de Relaciones:**
 - **Dependencias funcionales**
 - **Reglas de normalización.**
 - **Otros Tipos de Dependencias: la cuarta forma normal, la quinta forma normal.**



Bibliografía

1. **Bases de Datos: Desde Chen hasta Codd con Oracle-i, Luque Ruiz, I., Gómez-Nieto, M.A., López Espinosa, E., Cerruela García, G., Ra-Ma, 2001. (Capítulo 3)**
1. **T. CONNOLLY, C. BEGG, A. STRACHAN. Database Systems. A Practical Approach to Design, Implementation and Management. Segunda edición. Addison-Wesley. 1998 (Capítulo 3)**
1. **R. A. Elmasri, S. B. Navathe. Fundamentos de Sistemas de Bases de Datos, Addison-Wesley, 2002.(Capítulo 6)**
1. **C.J. Date (2001). Introducción a los Sistemas de Bases de Datos, Addison-Wesley, 2001. (Capítulos 4 y 5)**



La Teoría Relacional

La complejidad del esquema va a depender tanto de:

1. La complejidad del problema del mundo real.
1. La calidad del análisis de ese problema (su representación).
1. La evolución del problema y de los requisitos funcionales.



Fue E.F. Codd quien desarrolló en IBM-San José (California) el modelo de datos relacional. Este modelo está basado en conceptos muy sencillos teniendo asociada la teoría de normalización de relaciones.



Terminología del Modelo Relacional

Tablas:

Una tabla es una matriz rectangular que puede ser descrita de forma simple matemáticamente y que posee las siguientes propiedades:

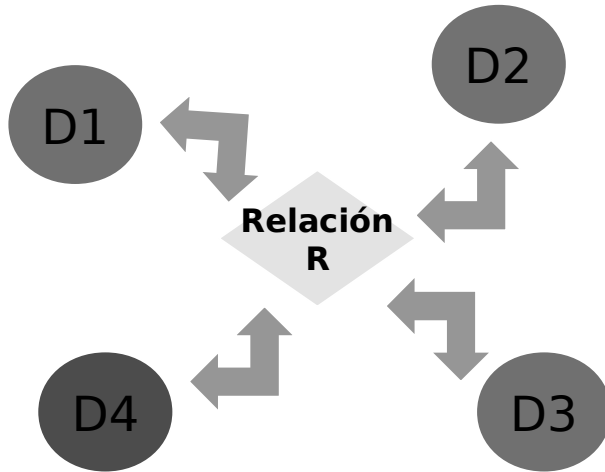
- I. Cada elemento de la matriz rectangular, representa a un ítem de datos elemental.
- II. Una tabla es homogénea por columnas.
- III. Cada columna de la tabla tiene asignado un nombre único.
- IV. Para una tabla todas las filas son diferentes, no se admiten filas duplicadas.
- V. Tanto las filas como las columnas pueden ser consideradas en cualquier secuencia.



Relación

R es una **relación** entre estos n conjuntos si es un conjunto de n tuplas no ordenadas $(d1, d2, d3, ..., dn)$ tales que $d1 \in D1, d2 \in D2, d3 \in D3, ..., dn \in Dn..$

producto cartesiano



**Dominio ?
Cardinalidad ?
Grado ?**

Alumno	<i>matricula#</i>	<i>nombre</i>	<i>apellidos</i>	<i>curso</i>	<i>nota</i>
	3456	José	Pérez de la Lastra	1	5.25
	0101	María	Antunez Sastre	2	7.80
	8743	Lourdes	Sánchez Argote	1	4.50
	1234	Antonio	Soria Madrid	3	6.35



DOMINIOS Y ATRIBUTOS

- Un **atributo** representa el uso de un dominio para una determinada relación; es decir, un atributo aporta un significado semántico a un dominio.
- Un **dominio** es un conjunto homogéneo definido mediante el uso de la abstracción en base a otro conjunto.

Paginas 63 y 64

Define Dominio	<i>expediente</i>	entero (4)	Fin definición;
Define Dominio	<i>primer-nombre</i>	carácter (15)	Fin definición;
Define Dominio	<i>final-nombre</i>	carácter (40)	Fin definición;
Define Dominio	<i>estudios</i>	entero (2)	Fin definición;
Define Dominio	<i>nota</i>	real (4)	Fin definición;

Define
Relación *Alumno*

(matricula#:	Dominio	<i>expediente,</i>
nombre:	Dominio	<i>primer-nombre,</i>
apellidos:	Dominio	<i>final-nombre,</i>
curso:	Dominio	<i>estudios,</i>
nota:	Dominio	<i>nota,</i>

Define
Relación *Alumno*

(matricula#	Dominio	<i>expediente,</i>
nombre:	Dominio	<i>primer-nombre,</i>
apellidos:	Dominio	<i>final-nombre,</i>
curso:	Dominio	<i>estudios,</i>
nota:	Dominio	<i>nota,</i>



INTENCIÓN Y EXTENSIÓN DE RELACIONES

La **intención** de una relación define todas las posibles extensiones admisibles para la misma. La intención de una relación define dos aspectos:

1. Una estructura de datos nominada en la que tanto la estructura — el objeto — como los ítems de datos que la componen —
2. Un conjunto de restricciones de integridad.

La **extensión** de una relación representa a cada uno de los objetos (tuplas), pertenecientes a un mismo tipo (relación), existentes en el dominio del problema en un momento dado.



CLAVES DE LAS RELACIONES

El conjunto de atributos que forman parte de la intención de una relación específica, uno o un conjunto de ellos tengan la propiedad de tomar valores únicos en el dominio del problema para cualquier extensión de esa relación.

Por ejemplo, en la relación *Alumno* utilizada anteriormente:

- ☐ El atributo *matricula#* tiene la propiedad de identificación única.
- ☐ Que sobre la base de la extensión de la relación *Alumno* presentada la agregación de los atributos *apellidos* y *nombre* también tiene esta propiedad.



INTEGRIDAD DE LOS ESQUEMAS RELACIONALES

- **Integridad de la clave:** *Ningún atributo que forme parte de la clave candidata de una relación podrá tomar valores nulos para ninguna tupla de esa relación.*
- **Integridad de referencia:** *Sea D un dominio, y sea R1 una relación con un atributo R1.a definido sobre el dominio D, entonces, en cualquier instante dado, cada valor de R1.a en R1 debe ser nulo o bien igual a algún valor V, el cual existe, en ese instante, para un atributo R2.b definido en el mismo dominio D sobre la relación R2 y en la cual está definido como clave primaria.*



Otras restricciones: En la definición del esquema relacional pueden imponerse, en teoría (dependerá del SGBD. relacional utilizado), otra serie de restricciones.

- Los valores permitidos para los atributos que forman parte de las relaciones existentes en el esquema. Por ejemplo, valor máximo y mínimo, lista de valores, etc.



NORMALIZACIÓN DE RELACIONES

Utilizando la Teoría de Normalización de Relaciones, existen un conjunto de Reglas de Normalización para satisfacer los siguientes objetivos:

1. No-existencia de redundancias superfluas.
2. Aumentar el desempeño de las operaciones de actualización de la base de datos.
3. Representar de forma coherente los objetos y relaciones.
4. Garantizar la fiabilidad sobre la información mantenida en la base de datos.



Dependencias Funcionales

La normalización de relaciones está basada en otra teoría, la *teoría de las dependencias*.

Dada una relación R , se dice que el atributo $R.y \in R$ es funcionalmente dependiente de otro atributo $R.x \in R$, y se expresa de la forma $R.x \rightarrow R.y$ si, y sólo si, cada valor de $R.x$ tiene asociado a él exactamente un valor de $R.y$ para cualquier extensión de la relación R .

- El concepto de dependencia funcional tiene en cuenta
 - Atributos de una misma relación y no a atributos de relaciones diferentes.
 - Pueda existir entre parejas de atributos de una relación.
 - Es intrínseca a la intención de una relación.
 - pueden ser simples o estar formados por la agregación de varios atributos.



Definición algo más amplia de dependencia funcional:

Dada una relación R, se dice que el atributo R.y \in R es funcionalmente dependiente de otro atributo R.x \in R, y se expresa en la forma $R.x \rightarrow R.y$ si, y sólo si, siempre que dos o más tuplas de R coincidan en sus valores de R.x, también, para esas tuplas, existirá una coincidencia en los valores del atributo R.y.

Representación textual de las dependencias funcionales

Formato general

$R.x \rightarrow (R.a, R.b, \dots R.n)$

Relación Alumno

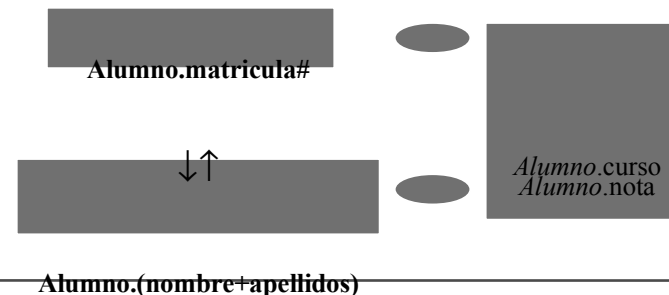
<i>Alumno.matricula#</i>	\rightarrow	<i>(Alumno.nota, Alumno.curso)</i>
<i>Alumno.matricula#</i>	\rightarrow	<i>(Alumno.nombre, Alumno.apellidos)</i>
<i>Alumno.matricula#</i>	\leftrightarrow	<i>(Alumno.(nombre + apellidos))</i>

Representación gráfica de las dependencias funcionales

Formato general

$R.x \rightarrow R.a, R.b, \dots, R.n$

Relación Alumno



Dependencia Funcional Completa

Se dice que el atributo $R.y \in R$ es funcionalmente dependiente y de forma completa de otro atributo $R.x \in R$ si, y sólo si, depende funcionalmente de $R.x$ y no de ningún subconjunto de los atributos que formen parte del atributo $R.x$.

Según esta definición se puede apreciar:

- Que si el atributo $R.x$ es un agregado formado por la concatenación de varios atributos pertenecientes a la relación, entonces la dependencia funcional $R.x \rightarrow R.y$, puede no ser completa, ya que puede existir una dependencia funcional de la forma $R.z \rightarrow R.y$, para $R.z \subset R.x$.
- Que el atributo $R.y$ sea simple o compuesto no tiene relevancia para que la dependencia funcional $R.x \rightarrow R.y$ sea completa o no.



PROPIEDADES DE LAS DEPENDENCIAS FUNCIONALES

Reflexiva: *Dados los atributos **a** y **b** de una relación **R**, para los que se cumple que $R.b \subseteq R.a$ entonces, en la relación **R**, está presente una dependencia funcional de la forma $R.a \rightarrow R.b$.*

Aumento: *Dados los atributos **a** y **b** de una relación **R** en la que está presente la dependencia funcional $R.a \rightarrow R.b$, entonces también estará presente la dependencia funcional $R.(a+c) \rightarrow R.(b+c)$, siendo **c** cualquier otro atributo que forme parte de la intención de la relación **R**.*

Transitiva: *Dados los atributos **a**, **b** y **c** de una relación **R** en la que están presentes las dependencias funcionales $R.a \rightarrow R.b$ y $R.b \rightarrow R.c$, entonces también estará presente la dependencia funcional $R.a \rightarrow R.c$.*



Unión: *Dados los atributos **a**, **b** y **c** de una relación **R** en la que están presentes las dependencias funcionales $R.a \rightarrow R.b$ y $R.a \rightarrow R.c$, entonces también estará presente la dependencia funcional $R.a \rightarrow R.(b+c)$.*

Pseudo-transitiva: *Dados los atributos **a**, **b**, **c** y **d** de una relación **R** en la que están presentes las dependencias funcionales $R.a \rightarrow R.b$ y $R.(b+c) \rightarrow R.d$, entonces también estará presente la dependencia funcional $R.(a+c) \rightarrow R.d$.*

Descomposición: *Dados los atributos **a**, **b** y **c** de una relación **R** en la que está presente la dependencia funcional $R.a \rightarrow R.b$, y se cumple que $R.c \subseteq R.b$, entonces también estará presente la dependencia funcional $R.a \rightarrow R.c$.*



Reglas de normalización

La teoría de la normalización está basada en la aplicación de una serie de reglas a las que se les denomina *Reglas de normalización*.

- La aplicación de una regla de normalización toma una relación como argumento de entrada y da como resultado dos o más relaciones, y:

- 1.La relación objeto de la aplicación de la regla es desestimada.
- 2.No se introducen nuevos atributos en el esquema relacional resultante de la normalización.
- 3.Los atributos de la relación objeto de la normalización pasan a formar parte de la intención de una o más de las relaciones resultantes.
- 4.En la aplicación de la regla de normalización se ha debido eliminar, al menos, una dependencia existente entre los atributos de la relación.



LA PRIMERA FORMA NORMAL FN1

Una relación R satisface la primera forma normal (FN1) si, y sólo si, todos los dominios subyacentes de la relación R contienen valores atómicos.

LA SEGUNDA FORMA NORMAL FN2

Una relación R satisface la segunda forma normal (FN2) si, y sólo si, satisface la primera forma normal y cada atributo de la relación depende funcionalmente de forma completa de la clave primaria de esa relación.

Esquema-1

Matricula (dni, asignatura#,
apellidos, nombre,
nota, curso, aula, lugar)

Por tanto, se va a suponer que existe una dependencia funcional entre los atributos *asignatura#* y *curso*, de la forma:
Matricula.asignatura# → *Matricula.curso*.

Insertión ?

Borrado ?

Actualización ?



Proceso de descomposición de la relación *Matricula*

Esquema-2

Imparte (asignatura#, curso)

Matricula-2 (dni, **asignatura#**, apellidos, nombre, nota, aula, lugar)

- ❑ No se ha producido pérdida de información, puesto que el atributo *asignatura#* debe definirse como clave foránea de la relación *Imparte*. De esta forma, el atributo *asignatura#* en la relación *Matricula-2* para cualquier tupla, sólo podrá tomar valores existentes en alguna tupla de la relación *Imparte*.
- ❑ La relación *Imparte*, se encuentra en *FN2*.



El mismo razonamiento debe hacerse para la dependencia funcional no completa existente entre la clave de la relación *Matricula-2* y el agregado (*apellidos + nombre*).

Esquema-3

Imparte (asignatura#, curso)
Alumno-2 (dni, apellidos, nombre)
Matricula-3 (dni, asignatura#, nota, aula, lugar)



LA TERCERA FORMA NORMAL FN3

Una relación R satisface la tercera forma normal (FN3) si, y sólo si, satisface la segunda forma normal y cada atributo no primo de la relación no depende funcionalmente de forma transitiva de la clave primaria de esa relación. Es decir, no pueden existir dependencias entre los atributos que no forman parte de la clave primaria de la relación R.

En el ejemplo anterior, se observa que existe una dependencia funcional entre los atributos no primos *lugar* y *aula*, además de las dependencias funcionales completas entre los atributos *lugar* y *aula* con la clave de la relación *Matricula-3*.

Matricula-3 (dni, asignatura#, nota, aula, lugar)

Matricula-3.(dni, asignatura#) → Matricula-3.aula

Matricula-3.(dni, asignatura#) → Matricula-3.lugar

Matricula-3.(aula) → Matricula-3.lugar



La existencia de estas dependencias entre los atributos no primos:

- **Insertión de tuplas ?**
- **Borrado de tuplas ?**
- **Modificación de tuplas ?**

Matricula-3 (dni, asignatura#, nota, aula, lugar)

Matricula-3.(dni, asignatura#) → Matricula-3.aula

Matricula-3.(dni, asignatura#) → Matricula-3.lugar

Matricula-3.(aula) → Matricula-3.lugar

Esquema-4

Imparte (asignatura#, curso)

Alumno-2 (dni, apellidos, nombre)

Ubicacion (aula, lugar)

Matricula-4 (dni, asignatura#, nota, **aula**,

Matricula-4 se encuentra en FN3, puesto que todas las dependencias funcionales existentes son completas y entre atributos no primos y la clave de la relación.



LA FORMA NORMAL DE BOYCE-CODD *FNBC*

☞ Se denomina determinante funcional a uno o un conjunto de atributos de una relación R del cual depende funcionalmente de forma completa algún otro atributo de la misma relación.

Se puede expresar la forma normal de *Boyce-Codd* de la forma:

Una relación R satisface la forma normal de Boyce-Codd (FNBC) si, y sólo si, se encuentra en FN1, y cada determinante funcional es una clave candidata de la relación R .

Del *Esquema-4* se puede deducir que todas las relaciones se encuentran en *FNBC*:

- **Los únicos determinantes funcionales son las claves para cada una de las relaciones.**
- **Las únicas dependencias funcionales son entre los atributos no primos y la clave**



EJEMPLO

Esquema-5

Matricula-5 (dni, asignatura#, apellidos, nombre, nota, curso, aula, lugar)

En la cual existen dos determinantes funcionales, cada uno de ellos compuesto y formado por los agregados: dni, asignatura# y apellidos, nombre, asignatura#

Las dependencias existentes en la relación Matricula-5 son:

Matricula-5.asignatura# → Matricula-5.curso

Matricula-5.(dni, asignatura#) → Matricula-5.aula

Matricula-5.(dni, asignatura#) → Matricula-5.lugar

Matricula-5.(dni, asignatura#) → Matricula-5.nota

Matricula-5.(apellidos, nombre, asignatura#) → Matricula-5.aula

Matricula-5.(apellidos, nombre, asignatura#) → Matricula-5.lugar

Matricula-5.(apellidos, nombre, asignatura#) → Matricula-5.nota

Matricula-5.aula → Matricula-5.lugar

Matricula-5.(dni, asignatura#) ↔ Matricula-5.(apellidos, nombre, asignatura#)



Existen dependencias entre atributos que no son determinantes funcionales y que es necesario eliminar.

Esquema-6

Imparte	<i>(<u>asignatura#</u>, curso)</i>
Ubicacion	<i>(<u>aula</u>, lugar)</i>
Matricula-6	<i>(<u>dni, asignatura#, apellidos, nombre, nota</u>, aula)</i>

- ❑ Se puede observar que, sobre la base de la definición de la *FNBC*, en el *Esquema-6* todas las relaciones se encuentran en *FN3*
- ❑ *Imparte* y *Ubicacion* se encuentran además en *FNBC*, pues sólo existe un determinante funcional y un atributo dependiente del mismo de forma completa
- ❑ La relación *Matricula-6*, aunque se encuentra en la *FN3*, no satisface la *FNBC*.



Matricula-6.dni \leftrightarrow Matricula-6.(apellidos, nombre)

En las dos claves candidatas está presente un atributo común, el atributo *asignatura#*, el cual forma parte de los dos determinantes funcionales, por lo que ambas claves se encuentran traslapadas[1].

[1] Dos claves candidatas se dice que están traslapadas si cada una de ellas está formada por dos o más atributos y alguno de ellos es común a ambas.



La relación *Matricula-6* sería conveniente descomponerla en dos relaciones que satisfagan la *FNBC*, quedando el nuevo esquema de la forma:

La relación *Matricula-6* sería conveniente descomponerla en dos relaciones que satisfagan la *FNBC*, quedando el nuevo esquema de la forma:

Esquema-7

Imparte	<i>(<u>asignatura#</u>, curso)</i>
Ubicacion	<i>(<u>aula</u>, lugar)</i>
Alumno-3	<i>(<u>dni</u>, <u>apellidos</u>, <u>nombre</u>)</i>
Matricula-7	<i>(<u>dni</u>, <u>asignatura#</u>, <i>nota</i>, <i>aula</i>)</i>



La cuarta Forma Normal *FN4*

Dependencia Multievaluada

- Dada una relación R , se dice que el atributo $R.y \in R$ depende de forma multivaluada de otro atributo $R.x \in R$, o que $R.x$ multidetermina a $R.y$, y se expresa de la forma $R.x \twoheadrightarrow R.y$
 - Si, y sólo si, cada valor de $R.x$ tiene asignado un conjunto bien definido de valores de $R.y$ y este conjunto es independiente de cualquier valor que tome otro atributo $R.z \in R$, el cual depende del valor de $R.x$
-
- ✓ Al menos, debe estar formada por tres atributos.
 - ✓ Las dependencias multivaluadas representan la independencia existente entre dos conjuntos $R.y$ y $R.z$, la cual está correlacionada por la dependencia que tiene cada uno de estos conjuntos con el conjunto $R.x$ del cual dependen ambos de forma multivaluada.



Ejemplo

SUPUESTOS

Docencia (dni, asignatura#, aula)

1. Cada alumno (*dni*), puede estar matriculado en un conjunto de asignaturas.
2. Al alumno se le asigna un conjunto de aulas en las cuales recibirá la docencia con independencia de las asignaturas en las que esté matriculado o no.
3. Cada asignatura puede ser impartida en cualquiera de las aulas disponibles, con independencia de que haya alumnos matriculados en las mismas.

Se observa, en la relación *Docencia*, que existe una independencia entre:

- ☞ Las aulas en las que se les imparte docencia a los alumnos y las asignaturas (SUPUESTO 2).
- ☐ Las asignaturas que cursa cada alumno y el aula en la que se les imparte la docencia (SUPUESTO 1).



Dependencias Multivaluadas:

Docencia.dni → Docencia.asignatura#

Docencia.dni → Docencia.aula

Problemas:

- La relación *Docencia* presenta mucha redundancia, puesto que, por ejemplo, como un alumno puede recibir la docencia en cualquier aula independientemente de la asignatura no es necesario considerar la información de la asignatura para cada pareja de valores (*dni*, *aula*). El mismo razonamiento se puede hacer para las parejas de valores (*dni*, *asignatura*) con respecto a las aulas.
- Aunque el aula en la que un alumno recibe la docencia de las asignaturas que cursa sea independiente de las asignaturas, no se puede conocer hasta que no se matricule en esas asignaturas.

Esquema-8

Docencia-1 (dni, asignatura#)

Docencia-2 (dni, aula)



La Cuarta Forma Normal Definición

Una relación R está en FN4 si, y sólo si, siempre que exista una dependencia multivaluada en R de la forma $R.x \twoheadrightarrow R.y$, todos los demás atributos de R son funcionalmente dependientes de $R.x$ ($R.x \rightarrow R.z, \forall R.z \in R$).



La quinta forma normal *FN5*

En una relación R puede estar presente un caso especial de dependencias llamadas *Dependencias de Reunión*, las cuales son la base de la aplicación de la *FN5*.

Dada una relación R de esquema $R(a_1, a_2, \dots, a_n)$ se dice que existe una dependencia de reunión si, y sólo si, la relación R puede ser construida a partir de la reunión natural de las relaciones R_1, R_2, \dots, R_n obtenidas por la proyección de R sobre los atributos a_1, a_2, \dots, a_n , respectivamente.

$$\begin{array}{l} R_1 = \text{Project}(R/a_1) = \Pi(R/a_1) \\ R_2 = \text{Project}(R/a_2) = \Pi(R/a_2) \\ \dots = \dots = \dots : \\ R_n = \text{Project}(R/a_n) = \Pi(R/a_n) \end{array} \quad \Rightarrow \quad R = R_1 \quad R_2 \quad \dots \quad R_n$$



Ejemplo

Docencia (dni, asignatura#, aula)

Supuestos:

- Que un alumno puede estar matriculado en un conjunto de asignaturas.
- Que para cada asignatura existen una serie de aulas en las que se puede impartir la docencia.
- Que para cada asignatura el alumno recibe la docencia en todas las aulas asignadas a esa asignatura.



- ☞ En este caso, aunque la relación *Docencia* se encuentra en *FNBC* y además en *FN4* (no existen dependencias que violen estas reglas, pues ahora se ha considerado que las aulas no son independientes de las asignaturas en las que se matriculan los alumnos).
- Se observa que si existe dependencia de reunión entre los atributos *dni*, *asignatura#* y *aula*, esta dependencia da lugar a que exista una gran redundancia en esta relación.

Esquema-9

Docencia-1 (dni, asignatura#)

Docencia-2 (asignatura#, aula)

Docencia-3 (dni, aula)



El nuevo *Esquema-9* obtenido representa mucho más claramente el problema real en el que:

- Un alumno está matriculado en una serie de asignaturas, y en una asignatura pueden estar matriculados varios alumnos. Representado por la relación *Docencia-1*.
- A cada asignatura se le asigna un conjunto de aulas para la impartición de la docencia, y en un aula se puede impartir la docencia de varias asignaturas. Representado por la relación *Docencia-2*.
- Cada alumno recibe la docencia en un conjunto de aulas, y en cada aula se imparte la docencia para un conjunto de alumnos. Representado por la relación *Docencia-3*.

