

Teórico 4 – Introducción al modelo relacional

Fundamentos

Un par ordenado es una pareja de elementos (a, b) donde el orden es importante, es decir, (a, b) es distinto de (b, a). El producto cartesiano de dos conjuntos A y B, que llamamos $A \times B$, es el conjunto de todos los pares ordenados (a, b) con $a \in A$ y $b \in B$.

Recordemos que una relación (binaria) de A en B, es un subconjunto del producto cartesiano $A \times B$. Las relaciones son de vital importancia en el modelo relacional de Codd, tanto que el término "relacional" de su teoría viene de las relaciones.

Un resultado importante que debemos recordar es que $\text{COUNT} (A \times B) = \text{COUNT} (A) * \text{COUNT} (B)$. Si A tiene 2 mil elementos y B cuatro mil, $A \times B$ tendrá 8 millones de elementos.

En la teoría relacional, suele simplificarse el término relación como sinónimo de tabla, ya que muchas veces se representan las relaciones en forma de tablas. Veamos la forma simplificada.

cedula	nombre	apellido	edad	sexo
1274539	Juan	Pérez	25	M
3483912	María	Gómez	12	F

De esta forma, una tabla puede verse como un conjunto (de "filas"). Cada "fila", además, tiene valores para ciertas variables que nos permiten predicar sobre las "filas" (las "filas" que tienen un valor de edad menor a 18, etc.).

Menores = { p : Personas | edad(p) < 18 }

Aquí se hace evidente el hecho de que un sistema de información pretende modelar algún aspecto del mundo real, una tabla de personas de este tipo podría representar la lista de clientes de una empresa.

Pensemos ahora en cómo modelar matemáticamente el contenido de las "filas". Los valores de las variables se pueden representar como pares ordenados, o lo que es lo mismo, como elementos de una relación (que podría ser una función). Imaginemos que para cada persona (o "fila") se define una relación:

- cedula, con codominio DECIMAL(7, 0)
- nombre, con codominio VARCHAR(20)
- apellido, con codominio VARCHAR(20)
- edad, con codominio el intervalo natural [0..100] ... o INTEGER
- sexo, con codominio el conjunto {'M', 'F'} ... o CHAR(1)

De esta forma, la tabla anterior se podría representar de la forma:

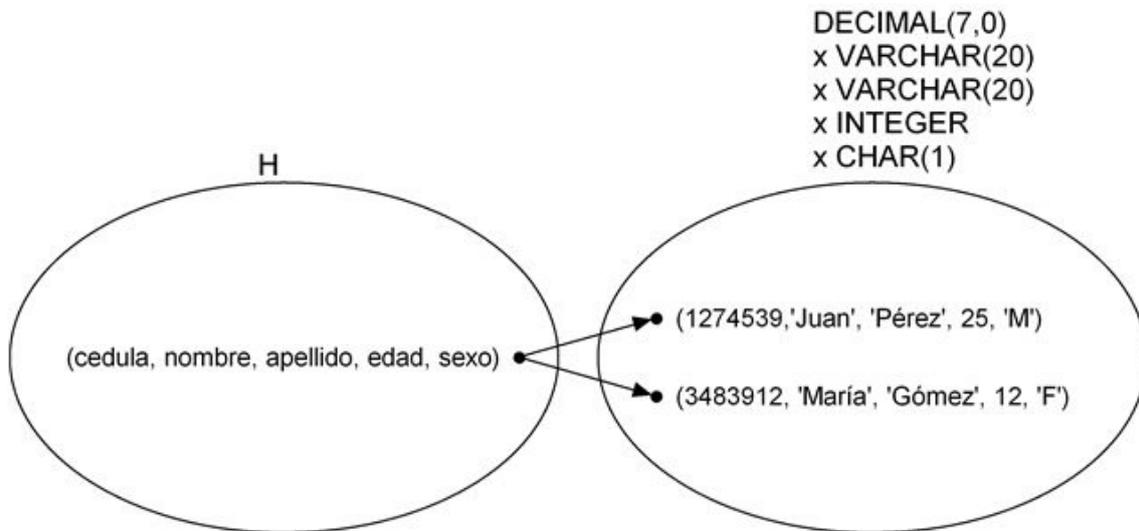
```
T :=
{ { (cedula, 1274539), (nombre, 'Juan'), (apellido, 'Pérez'), (edad, 25), (sexo, 'M') }
, { (cedula, 3483912), (nombre, 'María'), (apellido, 'Gómez'), (edad, 12), (sexo, 'F') }
}
```

Otra forma de representar lo mismo es como una sola relación de H en el producto cartesiano de los codominios, siendo $H = \{ \text{cedula, nombre, apellido, edad, sexo} \}$

Nos queda entonces algo como lo siguiente:

```
T :=
{ ( (cedula, nombre, apellido, edad, sexo), (1274539, 'Juan', 'Pérez', 25, 'M') )
, ( (cedula, nombre, apellido, edad, sexo), (3483912, 'María', 'Gómez', 12, 'F') )
}
```

Gráficamente sería:



Una forma fácil de ver estas relaciones es a través de una tabla, como veíamos al principio.

A las filas las llamaremos tuplas, y formalmente, diremos que T es una tabla sobre H , si toda tupla de T es una función sobre H . Si bien esta definición no impone nada sobre el conjunto H , en la práctica H será un conjunto de nombres que representan atributos de alguna entidad del mundo real que queremos modelar.

El dominio de las funciones (tuplas) del conjunto (de la tabla) se denomina cabecera (header).

Para definir una tabla, debemos tener los nombres de las columnas (la cabecera H), y el codominio de la función sobre H .

```
TABLA PERSONAS (
  CEDULA      DECIMAL(7, 0),
  NOMBRE      VARCHAR(20),
  APELLIDO    VARCHAR(20),
  EDAD        INTEGER,
  SEXO        CHAR(1)
)
```

Para estos fines se debe contar con un lenguaje apropiado, que permita además realizar operaciones sobre las tuplas: consultas y ABM (Altas, Bajas y Modificaciones).

Algo como:

```
INSERTAR EN PERSONAS(CEDULA, NOMBRE, APELLIDO, EDAD, SEXO)
EL VALOR (1452956, 'Pedro', 'Rodríguez', 38, 'M')
```

```
ELIMINAR DE PERSONAS
LAS QUE CUMPLAN (EDAD < 18)
```

```
MODIFICAR EN PERSONAS
HACER APELLIDO = 'Rodríguez'
PARA LAS QUE CUMPLAN (APELLIDO = 'Rodríguez')
```

```
SELECCIONAR NOMBRE, APELLIDO DE PERSONAS
QUE CUMPLAN (EDAD > 30)
```

La sintaxis de estas operaciones se parece mucho al lenguaje SQL que se ha convertido en un estándar y veremos más adelante.

Ahora podemos adelantar un panorama del resto del curso

- Veremos cómo realizar el diseño conceptual de una base de datos, concretamente veremos una técnica para modelar los aspectos relevantes del mundo real, llamada Modelo Entidad-Relación (MER).

Es importante notar que la teoría relacional se llama relacional por la forma de modelar las tablas como relaciones entre conjuntos, y esto no tiene nada que ver con la 'R' del MER.

- Estudiaremos cómo traducir nuestros diseños conceptuales (que en definitiva son diagramas, dibujos) en el diseño lógico de la base de datos (creación de las tablas), e introduciremos algunos conceptos más sobre la teoría relacional como claves, restricciones y formas normales.
- Veremos un lenguaje (matemático) que nos permitirá realizar consultas sobre los datos, llamado álgebra relacional.
- Veremos y utilizaremos el lenguaje SQL para realizar consultas, así como para realizar ABM sobre las tuplas.
- Utilizaremos algún DBMS en particular para crear tablas y probar los conceptos aprendidos y SQL en la práctica.