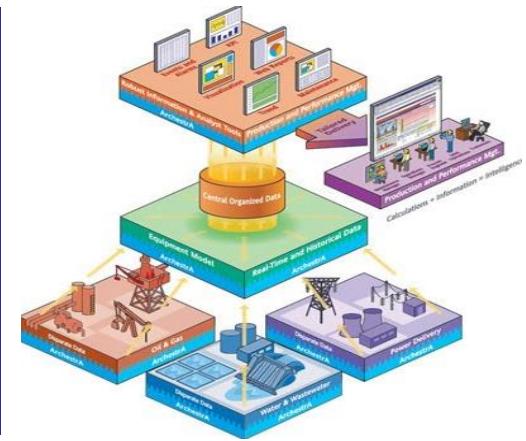
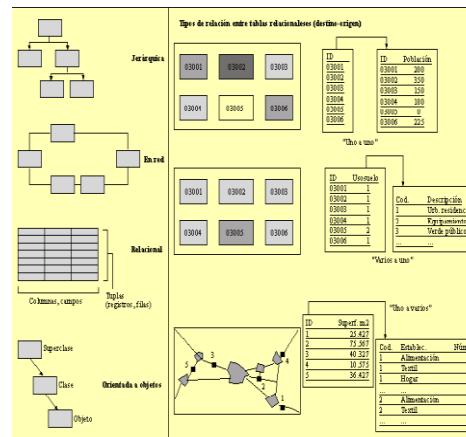


Tema 3

El Modelo de Datos Relacional

Parte 3/3

Grado en Ingeniería Informática



**Bases de
Datos**
2021/22

Departamento de Tecnologías de la Información
Universidad de Huelva

3.4.2 Cálculo relacional

- La parte manipulativa del modelo relacional consiste en el álgebra relacional y en el cálculo relacional
- El álgebra y el cálculo relacional son dos alternativas para establecer una base formal de la parte manipulativa del modelo
- El cálculo es un lenguaje formal de consulta basado en expresiones **declarativas**
 - Especifican **qué** debe obtenerse, no **cómo** debe hacerse
 - Se considera un lenguaje **no procedimental**
 - El álgebra ofrece un conjunto de operaciones explícitas que pueden utilizarse para generar la respuesta a la consulta
 - Basado en la **lógica de predicados de primer orden** (LPPO)
 - Álgebra y cálculo relacional tienen el **mismo poder de expresión** \Rightarrow para cada expresión del álgebra existe una expresión equivalente en el cálculo y viceversa
- Existen dos propuestas de cálculo: el cálculo relacional de tuplas y el cálculo relacional de dominios

3.4.2.1 Cálculo relacional de tuplas

- Una característica fundamental del CR es el concepto de **variable tupla**. Una variable tupla es una variable cuyos únicos valores permitidos son tuplas de una relación o bien expresiones del cálculo de tuplas
 - Por ejemplo, si s es una variable tupla definida sobre la relación R , en cualquier momento dado, s representará una tupla de R
- Propuesto por Codd en 1972
- Forma de las consultas:
 $\{ t \mid P(t) \} \rightarrow$ conjunto de todas las tuplas tales que el predicado P es cierto para t
- Notación:
 - $t.A \rightarrow$ valor de la tupla t en el atributo A
 - $r(t) \rightarrow$ la tupla t está en la relación r

Consultas de ejemplo

- Obtener una lista de las asignaturas de 4.5 créditos

$$\{ t \mid \text{asignatura}(t) \wedge t.\text{créditos} = 4.5 \}$$

- Si sólo se desea obtener el nombre de la asignatura, en vez de todos los atributos
 - Hay que escribir una expresión para una nueva relación: (*nombre*)
 - Se necesitan las tuplas de la relación (*nombre*) tales que hay una tupla en *asignatura* con el atributo *créditos* = 4.5
 - Se utiliza el operador de cuantificación \exists de la LPPO
 - Notación

$$(\exists t) (r(t) \wedge Q(t)) \rightarrow \text{existe una tupla } t \text{ en la relación } r \text{ tal que el predicado } Q(t) \text{ es verdadero}$$

El Modelo de Datos Relacional: CRT

- ❑ Obtener el nombre de aquellas asignaturas que tengan 4.5 créditos

$$\{ t \mid (\exists s) (\text{asignatura}(s) \wedge s.\text{créditos} = 4.5 \wedge t.\text{nomAsig} = s.\text{nombre}) \}$$

- La variable de tupla t sólo se define para el atributo nomAsig , ya que es el único atributo para el que se especifica una condición para $t \Rightarrow$ el resultado es una relación de (nomAsig)
 - Forma de cambiar el nombre a un atributo en la relación resultado
-
- ❑ Obtener el nombre de las asignaturas de segundo curso junto con el nombre del profesor responsable
- Involucra dos relaciones: asignatura y profesor
 - Es necesario utilizar dos cuantificadores existenciales relacionados por \wedge

$$\{ t \mid (\exists s) (\text{asignatura}(s) \wedge s.\text{curso} = 2 \wedge t.\text{nomAsig} = s.\text{nombre} \wedge (\exists u) (\text{profesor}(u) \wedge s.\text{prof} = u.\text{nPr} \wedge t.\text{nomProf} = u.\text{nombre})) \}$$

El Modelo de Datos Relacional: CRT

- Obtener el nombre de todas las personas (alumnos o profesores) de la escuela
 - ✓ Es necesario utilizar dos cuantificadores existenciales relacionados por \vee

$$\{ t \mid (\exists s) (\text{alumno}(s) \wedge t.\text{nombre} = s.\text{nombre}) \vee (\exists u) (\text{profesor}(u) \wedge t.\text{nombre} = u.\text{nombre}) \}$$

- Obtener el nombre de los alumnos matriculados en el año 2002 en la asignatura A004 pero no en la A011

$$\{ t \mid (\exists s) (\text{alumno}(s) \wedge t.\text{nombre} = s.\text{nombre} \wedge (\exists u) (\text{matrícula}(u) \wedge u.\text{año} = 2002 \wedge u.\text{idAsig} = 'A004' \wedge u.\text{alum} = s.\text{nAl}) \wedge \neg(\exists v) (\text{matrícula}(v) \wedge v.\text{año} = 2002 \wedge v.\text{idAsig} = 'A011' \wedge v.\text{alum} = s.\text{nAl})) \}$$

- ✓ $(\exists u) (\text{matrícula}(u) \dots)$ → exige que el alumno haya estado matriculado en el año 2002 en A004
- ✓ $\neg(\exists v) (\text{matrícula}(v) \dots)$ → excluye a los matriculados ese mismo año en A011

- Averiguar el nombre de todos los alumnos matriculados en el año 2001 en todas las asignaturas cuyo responsable es el profesor José Manuel Ramos López
 - ✓ En álgebra relacional este tipo de consultas se resuelve mediante el operador \div
 - ✓ Es necesario utilizar el operador de cuantificación \forall de la LPPO
 - ✓ Notación

$(\forall t) (r(t) \wedge Q(t)) \rightarrow Q(t)$ es verdadero para todas las tuplas t de la relación r

- ✓ $P \Rightarrow Q \equiv \neg P \vee Q$

$$\{ t \mid (\exists s) (\text{alumno}(s) \wedge t.nombre = s.nombre \wedge (\exists u) (\text{profesor}(u) \wedge u.nombre = 'José Manuel Ramos López' \wedge (\forall v) (\text{asignatura}(v) \wedge v.prof = u.nPr \Rightarrow (\exists w) (\text{matrícula}(w) \wedge w.idAsig = v.idAsig \wedge w.año = 2001 \wedge w.alum = s.nAl))) \}$$

Definición formal

- Las expresiones del cálculo relacional de tuplas son de la forma

$\{ t \mid F(t)\}$

F es una fórmula

En una fórmula pueden aparecer varias variables de tupla

- **Libre:** variable no cuantificada mediante \exists o \forall
- **Ligada:** variable incluida en un cuantificador

- Las fórmulas se construyen mediante **átomos**, que pueden ser de la forma:

$r(s)$, donde s es una variable de tupla y r es una relación

$s.x \theta t.y$, donde s y t son variables de tupla, x e y son nombres de atributos y θ es un operador de comparación ($<$, \leq , $=$, \neq , $>$, \geq)

$s.x \theta c$, donde s es una variable de tupla, x es el nombre de un atributo, θ es un operador de comparación y c es una constante del dominio del atributo x

□ Reglas para formar fórmulas

- Un átomo es una fórmula
- Si F es una fórmula, también lo son $\neg F$ y (F)
- Si F_1 y F_2 son fórmulas, también lo son $F_1 \wedge F_2$, $F_1 \vee F_2$ y $F_1 \Rightarrow F_2$
- Si $F(s)$ es una fórmula que contiene una variable de tupla s libre, también lo son $(\exists s) (F(s))$ y $(\forall s) (F(s))$

□ Equivalencias entre operadores

- $F_1 \wedge F_2 \equiv \neg(\neg F_1 \vee \neg F_2)$
- $(\forall s) F(s) \equiv \neg (\exists s) (\neg F(s))$
- $F_1 \Rightarrow F_2 \equiv \neg F_1 \vee F_2$

Seguridad de las expresiones

- Las expresiones del cálculo relacional de tuplas pueden generar relaciones infinitas

$$\{ t \mid \neg \text{asignatura}(t) \}$$

- Hay infinitas tuplas que no están en *asignatura*
- La mayor parte de las tuplas ni siquiera aparecen en la base de datos
- Expresión insegura:** aquella que produce un número infinito de tuplas \Rightarrow no se desea permitir ese tipo de expresiones

- **Dominio** de una fórmula del cálculo relacional de tuplas \rightarrow conjunto de todos los valores que aparecen:

- Como valores constantes en la expresión, o
- Que existen en cualquier tupla de las relaciones a las que hace referencia en la expresión

□ Ejemplos

- ✓ $\text{dom}(\{\neg \text{asignatura}(t)\}) \rightarrow$ todos los valores que aparecen en alguna tupla de *asignatura*
- ✓ $\text{dom}((\exists s) (\text{asignatura}(s) \wedge s.\text{curso} = 2 \wedge t.\text{nomAsig} = s.\text{nombre} \wedge (\exists u) (\text{profesor}(u) \wedge s.\text{prof} = u.\text{nPr} \wedge t.\text{nomProf} = u.\text{nombre})) \rightarrow$ todos los valores que aparecen en alguna tupla de *asignatura* y *profesor*, junto con el valor 2

□ Expresión segura: si todos los valores de su resultado pertenecen al dominio de la expresión

- ✓ El resultado de $\{ t \mid \neg \text{asignatura}(t)\}$ es inseguro porque incluirá tuplas que no están en la relación *asignatura* y esos valores no pertenecen al dominio de la expresión
- ✓ El resto de ejemplos son expresiones seguras

3.4.2.2 Cálculo relacional de dominios

- ❑ La especificación formal del cálculo de dominios fue propuesta tras la creación del sistema **QBE** en el centro de investigación T. J. Watson de IBM (1977)
- ❑ Muy relacionado con el cálculo relacional de tuplas, pues también se basa en la **LPPO**
- ❑ Utiliza **variables de dominio** que toman sus valores del dominio de un atributo, en lugar de una tupla completa
- ❑ El álgebra relacional básica (sin las operaciones extendidas), el cálculo relacional de tuplas (expresiones seguras) y el cálculo relacional de dominios (expresiones seguras) tienen la **misma potencia expresiva**

Definición formal

- ❑ Las expresiones del cálculo relacional de dominios son de la forma

$$\{ x_1, x_2, \dots, x_m \mid F(x_1, x_2, \dots, x_n) \}$$

x_i representa una variable de dominio

F es una **fórmula**

- ❑ Las fórmulas se construyen mediante **átomos**, que pueden ser de la forma:
 - ✓ $R(x_1, x_2, \dots, x_n)$, donde R es el nombre de una relación de grado n y cada x_i , con $1 \leq i \leq n$, es una variable de dominio. Significado: una lista de valores $\langle x_1, x_2, \dots, x_n \rangle$ debe ser una tupla en la relación cuyo nombre es R , donde x_i es el valor del i -ésimo atributo de la tupla
 - ✓ $x \theta y$, donde x e y representan variables de dominio y θ es un operador de comparación ($<$, \leq , $=$, \neq , $>$, \geq)
 - ✓ $x \theta c$, donde x es una variable de dominio, θ es un operador de comparación y c es una constante del dominio del atributo para el que x es una variable de dominio
- ❑ **Reglas** para formar fórmulas
 - ✓ Un átomo es una fórmula
 - ✓ Si F es una fórmula, también lo son $\neg F$ y (F)
 - ✓ Si F_1 y F_2 son fórmulas, también lo son $F_1 \wedge F_2$, $F_1 \vee F_2$ y $F_1 \Rightarrow F_2$
 - ✓ Si $F(x)$ es una fórmula que contiene una variable de dominio x libre, también lo son $(\exists x) F(x)$ y $(\forall x) F(x)$
 - ✓ Como notación abreviada se escribe:
 $(\exists x, y, z) F(x, y, z)$ en lugar de $(\exists x) (\exists y) (\exists z) F(x, y, z)$

Consultas de ejemplo

- Obtener una lista de las asignaturas de 4.5 créditos

$$\{ i, n, c, cs, ct, e, p \mid \text{asignatura}(i, n, c, cs, ct, e, p) \wedge c = 4.5 \}$$

- Obtener el nombre de aquellas asignaturas que tengan 4.5 créditos

$$\{ n \mid (\exists i, c, cs, ct, e, p) (\text{asignatura}(i, n, c, cs, ct, e, p) \wedge c = 4.5) \}$$
Notación simplificada:

cuantificar sólo las variables que intervienen en condiciones:

$$\{ n \mid (\exists c) (\text{asignatura}(i, n, c, cs, ct, e, p) \wedge c = 4.5) \}$$
Notación alternativa:

$$\{ n \mid \text{asignatura}(i, n, 4.5, cs, ct, e, p) \}$$

- Obtener el nombre de las asignaturas de segundo curso junto con el nombre del profesor responsable

$$\{ na, np \mid (\exists p) (\text{asignatura}(i, na, c, 2, ct, e, p) \wedge (\exists pr) (\text{profesor}(pr, d, np, ds, a) \wedge p = pr)) \}$$
Alternativa:

$$\{ na, np \mid (\exists p) (\text{asignatura}(i, na, c, 2, ct, e, p) \wedge \text{profesor}(p, d, np, ds, a)) \}$$

El Modelo de Datos Relacional: CRD

- Obtener el nombre de todas las personas (alumnos o profesores) de la escuela

$$\{ n \mid \text{profesor}(p, dp, n, ds, an) \vee \text{alumno}(a, da, n, f, l, nh, o) \}$$

- Obtener el nombre de los alumnos matriculados en el año 2002 en la asignatura A004 pero no en la A011

$$\{ n \mid (\exists a) (\text{alumno}(a, da, n, f, l, nh, o) \wedge \text{matrícula}(a, 'A004', p, 2002, fj, s, d) \wedge \neg \text{matrícula}(a, 'A011', pp, 2002, ffj, ss, dd)) \}$$

- Averiguar el nombre de todos los alumnos matriculados en el año 2001 en todas las asignaturas cuyo responsable es el profesor José Manuel Ramos López

$$\{ n \mid (\exists a) (\text{alumno}(a, da, n, f, l, nh, o) \wedge (\exists p) (\text{profesor}(p, dp, 'José Manuel Ramos López', ds, an) \wedge (\forall ia) (\text{asignatura}(ia, na, c, cs, ct, e, p) \Rightarrow \text{matrícula}(a, ia, pr, 2001, fj, s, d)))) \}$$

Seguridad de las expresiones

- Las expresiones del cálculo relacional de dominio también pueden generar relaciones infinitas

$$\{ i, n, c, cs, ct, e, p \mid \neg \text{asignatura}(i, n, c, cs, ct, e, p) \}$$

- No es segura porque permite valores del resultado que no están en el dominio de la expresión

- Debe considerarse cuidadosamente la forma de las fórmulas dentro de los cuantificadores \exists y \forall

$$\{ x \mid (\exists z) (\neg r(x, z) \wedge P(z)) \}$$

- Para probar la fórmula hay que considerar los valores de z que no aparecen en r . Como las relaciones son finitas \Rightarrow no aparece en r un número infinito de valores
 - No es posible probar la fórmula