

ÁLGEBRA RELACIONAL

Servando Campillay

Introducción

- Un modelo de datos debe incluir un conjunto de operaciones para manipular la base de datos junto con los conceptos necesarios para la definición de su estructura y restricciones.
- El álgebra relacional es el conjunto de operaciones básicas del modelo relacional, el cual permite al usuario especificar las peticiones fundamentales de recuperación.
- El resultado de una recuperación es una nueva relación, la cual puede estar constituida por una o más relaciones. Una secuencia de operaciones de álgebra relacional conforma una expresión de álgebra relacional, cuyo resultado será también una nueva relación que representa el resultado de una consulta a la base de datos (o una petición de recuperación).

Introducción (cont.)

El álgebra relacional es muy importante por varias razones:

- La primera, porque proporciona un fundamento formal para las operaciones del modelo relacional.
- La segunda razón, y quizá la más importante, es que se utiliza como base para la implementación y optimización de consultas en los RDBMS (Sistemas de administración de bases de datos relacionales, *Relational DataBase Management Systems*).
- o Tercera, porque algunos de sus conceptos se han incorporado al lenguaje estándar de consultas SQL para los RDBMS.

NOTA

• El álgebra relacional se inspira en la teoría de conjuntos para especificar consultas en una base de datos relacional.

DEFINICIÓN DEL ALGEBRA RELACIONAL

• El Algebra Relacional (AR) se define como una 7-upla:

$$\Re = \langle U, D, dom, \sigma, \mathcal{A}_{\sigma}, \theta, \beta \rangle$$

Donde:

- U es un conjunto de atributos llamado **Esquema de la Relación Universal**, $U = \{A_1, A_2, A_n\}$. A_i le corresponde un dominio D_i de donde toma sus valores.
- 2. Des el conjunto formado por todos los dominios $D = \{D_1, D_2, \dots, D_N\}$.
- 3. **dom** es una función, dom : $U \to D$ tal que dom $(A_i) \subseteq D$.
- 4. σ es un conjunto de esquemas de relación, $\sigma = \{R_1, R_2, \dots, R_p\}$, donde $U = \bigcup_{i=1}^p R_i$ σ es llamado descomposición de U y conforma el esquema de la base de datos.

Definición del Algebra Relacional (cont.)

- 5. $\mathcal{A}_{\mathcal{O}}$ es un conjunto de instancias de relaciones $\mathcal{A}_{\mathcal{O}} = \{r_1, r_2, \dots, r_p\}$, donde cada r_i es una instancia de relación con esquema de relación R_i , y conforma la instancia de la base de datos.
- 6. Θ es un conjunto de operaciones relacionales que permite comparar valores del dominio que sean compatibles. Se utilizan los siguientes operadores relacionales: $<,>,\leq,\geq,=,\neq$.
- 7. β es un conjunto de operadores del AR cuyos operandos son las relaciones.

OPERACIONES DEL ÁLGEBRA RELACIONAL

- o Según algunos criterios se clasifican en 3:
 - 1. En termino de otras operaciones:
 - a. Operaciones primitivas
 - b. Operaciones no primitivas
 - 2. Según el numero de relaciones
 - a. Operaciones unarias
 - b. Operaciones binarias
 - 3. Según teoría de conjunto
 - a. Operaciones conjuntistas
 - b. Operaciones específicamente relacionales

OPERACIONES CONJUNTISTAS

• Las operaciones conjuntistas del álgebra relacional son la unión, la intersección, la diferencia y el producto cartesiano.

Unión

- La unión es una operación que, a partir de dos relaciones, obtiene una nueva relación formada por todas las tuplas que están en alguna de las relaciones de partida.
- Sea r y s instancias de relación con esquema R. denotamos con $r \cup s$ a la unión de tuplas de ambas relaciones.
- Nota: para poder aplicar la unión a dos relaciones, es preciso que las dos relaciones sean compatibles

UNIÓN (CONT.)

Ejemplo

RESULTADO ← RESULTADO1 U RESULTADO2.

RESULTADO1

Dni
123456789
333445555
666884444
453453453

RESULTADO2

Dni
333445555
888665555

RESULTADO

Dni
123456789
333445555
666884444
453453453
888665555

UNIÓN (CONT.)

EMPLEADOS_ADM				
<u>DNI</u>	nombre	apellido	edificiodesp	númerodesp
40.444.255	Juan	García	Marina	120
33.567.711	Marta	Roca	Marina	120

EMPLEADOS_PROD				
<u>DNI</u>	nombreemp	númerodesp		
33.567.711	Marta	Roca	Marina	120
55.898.425	Carlos	Buendía	Diagonal	120
77.232.144	Elena	Pla	Marina	230
21.335.245	Jorge	Soler	NULO	NULO
88.999.210	Pedro	González	NULO	NULO

$R \coloneqq EMPLEADOS_ADM \cup EMPLEADOS_PROD.$

R				
DNI	nombre	apellido	edificiodesp	númerodesp
40.444.255	Juan	García	Marina	120
33.567.711	Marta	Roca	Marina	120
55.898.425	Carlos	Buendía	Diagonal	120
77.232.144	Elena	Pla	Marina	230
21.335.245	Jorge	Soler	NULO	NULO
88.999.210	Pedro	González	NULO	NULO

Intersección

- La intersección es una operación que, a partir de dos relaciones, obtiene una nueva relación formada por las tuplas que pertenecen a las dos relaciones de partida.
- o Sean R un esquema de relación, r y s instancias de relaciones con esquema R. Denotamos con r ∩ s a la intersección de tuplas de ambas relaciones. La relación resultante tiene esquema R, y su instancia esta conformada por el conjunto de tuplas que pertenecen a r y a s.
- Nota: La intersección, como la unión, sólo se puede aplicar a relaciones que tengan tuplas similares. Para poder hacer la intersección de dos relaciones, es preciso, pues, que las relaciones sean compatibles.

Intersección (cont.)

EMPLEADOS_ADM				
<u>DNI</u>	nombre	apellido	edificiodesp	númerodesp
40.444.255	Juan	García	Marina	120
33.567.711	Marta	Roca	Marina	120

EMPLEADOS_PROD				
<u>DNI</u>	nombreemp	apellidoemp	edificiodesp	númerodesp
33.567.711	Marta	Roca	Marina	120
55.898.425	Carlos	Buendía	Diagonal	120
77.232.144	Elena	Pla	Marina	230
21.335.245	Jorge	Soler	NULO	NULO
88.999.210	Pedro	González	NULO	NULO

 $R := EMPLEADOS_ADM \cap EMPLEADOS_PROD.$

		R		
DNI	nombre	apellido	edificiodesp	númerodesp
33.567.711	Marta	Roca	Marina	120

Intersección (cont.)

Prov1:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S1	Smith	20	Londres
S2	Jones	10	París
S3	Blake	30	París
S4	Clark	20	Londres
S5	Adams	30	Atenas

Prov2:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S6	Jones	40	Londres
S2	Jones	10	París
S7	Adams	10	Roma

$Prov1 \cap Prov2$:

#Proveedor	Nombre_P	Categoría_P	Ciudad_ P
S2	Jones	10	París

DIFERENCIA

- La diferencia es una operación que, a partir de dos relaciones, obtiene una nueva relación formada por todas las tuplas que están en la primera relación y, en cambio, no están en la segunda. La diferencia es una operación binaria, y la diferencia entre las relaciones T y S se indica como T-S.
- Sean R un esquema de relación, r y s instancias de relación con esquema R. Denotamos con r-s a la diferencia de tuplas de ambas relaciones. La relación resultante tiene esquema R, y su instancia esta conformada por el conjunto de tuplas que pertenecen a r y no pertenecen a s.

DIFERENCIA (CONT.)

	EMPLEADOS_ADM					
<u>DNI</u>	nombre	apellido	edificiodesp	númerodesp		
40.444.255	Juan	García	Marina	120		
33.567.711	Marta	Roca	Marina	120		

EMPLEADOS_PROD						
<u>DNI</u>	nombreemp	apellidoemp	edificiodesp	númerodesp		
33.567.711	Marta	Roca	Marina	120		
55.898.425	Carlos	Buendía	Diagonal	120		
77.232.144	Elena	Pla	Marina	230		
21.335.245	Jorge	Soler	NULO	NULO		
88.999.210	Pedro	González	NULO	NULO		

$R := EMPLEADOS_ADM - EMPLEADOS_PROD$

R					
DNI	nombre	apellido	edificiodesp	númerodesp	
40.444.255	Juan	García	Marina	120	

DIFERENCIA (CONT.)

Prov1:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S1	Smith	20	Londres
S2	Jones	10	París
S3	Blake	30	París
S4	Clark	20	Londres
S5	Adams	30	Atenas

Prov2:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S6	Jones	40	Londres
S2	Jones	10	París
S7	Adams	10	Roma

Prov1 - Prov2:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S1	Smith	20	Londres
S3	Blake	30	París
S4	Clark	20	Londres
S5	Adams	30	Atenas

PRODUCTO CARTESIANO

- El producto cartesiano es una operación que, a partir de dos relaciones, obtiene una nueva relación formada por todas las tuplas que resultan de concatenar tuplas de la primera relación con tuplas de la segunda. El producto cartesiano es una operación binaria. Siendo T y S dos relaciones que cumplen que sus esquemas no tienen ningún nombre de atributo común, el producto cartesiano de T y S se indica como $T \times S$.
- Producto cartesiano (cartesian product), conocida también como producto cruzado (cross product) o Concatenación cruzada (cross join), que se identifica por x. Se trata también de una operación de conjuntos binarios, aunque no es necesario que las relaciones en las que se aplica sean una 'unión compatible.

PRODUCTO CARTESIANO (CONT.)

EDIFICIOS_EMP			
<u>edificio</u>	supmediadesp		
Marina	15		
Diagonal	10		

DESPACHOS					
<u>edificio</u>	<u>número</u>	superficie			
Marina	120	10			
Marina	230	20			
Diagonal	120	10			
Diagonal	440	10			

$R := EDIFICIOS \times DESPACHOS.$

R						
nombreedificio	supmediadesp	edificio	número	superficie		
Marina	15	Marina	120	10		
Marina	15	Marina	230	20		
Marina	15	Diagonal	120	10		
Marina	15	Diagonal	440	10		
Diagonal	10	Marina	120	10		

PRODUCTO CARTESIANO (CONT.)

Prov:

#Proveedor	Nombre_P	Categoría_P	Ciudad
S1	Smith	20	Londres
S2	Jones	10	París
S3	Blake	30	París

Proy:

#Proyecto	Nombre	Ciudad
J1	Ordenador	París
J2	Perforadora	Roma

$\mathbf{Prov} \times \mathbf{Proy}$:

#Proveedor	Nombre_P	Categoría_P	Ciudad	#Proyecto	Nombre	Ciudad'
S1	Smith	20	Londres	J1	Ordenador	París
S2	Jones	10	París	J1	Ordenador	París
S3	Blake	30	París	J1	Ordenador	París
S1	Smith	20	Londres	J2	Perforadora	Roma
S2	Jones	10	París	J2	Perforadora	Roma
S3	Blake	30	París	J2	Perforadora	Roma

Notar que:

 $esq(Prov \times Proy) = \{ \#Proveedor, \ Nombre_P, \ Categor\'ia, \ Ciudad, \ \#Proyecto, \ Nombre, \ Ciudad' \} \\ \square$

Sea $r \times s = v$, luego tenemos los siguientes casos extremales:

- $esq(r) = \emptyset$ entonces esq(v) = esq(s) = S y $v = \emptyset$.
- $r = \emptyset$ entonces $v = \emptyset$.

OPERACIONES ESPECÍFICAMENTE RELACIONALES

• Las operaciones específicamente relacionales son la selección, la proyección y la combinación.

SELECCIÓN

- Podemos ver la selección como una operación que sirve para elegir algunas tuplas de una relación y eliminar el resto. Más concretamente, la selección es una operación que, a partir de una relación, obtiene una nueva relación formada por todas las tuplas de la relación de partida que cumplen una condición de selección especificada.
- Selección se emplea para seleccionar un subconjunto de las tuplas de una relación que satisfacen una condición de selección. Se puede considerar esta operación como un filtro que mantiene sólo las tuplas que satisfacen una determinada condición. Selección puede visualizarse también como una partición horizontal de la relación en dos conjuntos de tuplas: las que satisfacen la condición son seleccionadas y las que no, descartadas.

SELECCIÓN (CONT.)

Prov:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S1	Smith	20	Londres
S2	Jones	10	París
S3	Blake	30	París
S4	Clark	20	Londres
S5	Adams	20	Atenas

 $\sigma_{\rm (Categoría_P=30)}({\rm Prov})$:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S3	Blake	30	París

 $\sigma_{(Categoria_P < 30 \ \land \ (Ciudad_P = Atenas \ \lor \ Ciudad_P = Paris))}(Prov):$

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S2	Jones	10	París
S5	Adams	20	Atenas

Sea $\sigma_F(r) = s$ luego tenemos los siguientes casos extremales:

- Ninguna tupla de r satisface F, entonces $s = \emptyset$
- \bullet Todas las tuplas satisfacen la fórmula F , entonces s=r

SELECCIÓN (CONT.)

EDIFICIOS_EMP		
<u>edificio</u>	supmediadesp	
Marina	15	
Diagonal	10	

DESPACHOS			
<u>edificio</u>	<u>número</u>	superficie	
Marina	120	10	
Marina	230	20	
Diagonal	120	10	
Diagonal	440	10	

 $R := DESPACHOS(edificio = Marina\ y\ superficie > 12).$

	R	
edificio	número	superficie
Marina	230	20

PROYECCIÓN

- Podemos considerar la proyección como una operación que sirve para elegir algunos atributos de una relación y eliminar el resto. Más concretamente, la proyección es una operación que, a partir de una relación, obtiene una nueva relación formada por todas las (sub)tuplas de la relación de partida que resultan de eliminar unos atributos especificados.
- Si pensamos en una relación como en una tabla, la operación Selección elige algunas de las filas de la tabla a la vez que descarta otras. Por otro lado, Proyección selecciona ciertas columnas de la tabla y descarta otras. Si sólo estamos interesados en algunos atributos de una relación, usamos la operación Proyección para planear la relación sólo sobre esos atributos.

Proyección (cont.)

EMPLEADOS_ADM				
<u>DNI</u>	nombre	apellido	edificiodesp	númerodesp
40.444.255	Juan	García	Marina	120
33.567.711	Marta	Roca	Marina	120

 $R := EMPLEADOS_ADM[nombre, apellido].$

R		
nombre	apellido	
Juan	García	
Marta	Roca	

PROYECCIÓN (CONT.)

 $Proveedores = \{ \#Proveedor, Nombre_P, Categoría, Ciudad_P \}$

Prov:

#Proveedor	Nombre_P	Categoría_P	Ciudad_P
S1	Smith	20	Londres
S2	Jones	10	París
S3	Blake	30	París
S4	Clark	20	Londres
S5	Adams	30	Atenas

 $\Pi_{\text{\#Proveedor,Nombre_P}}(\text{Prov})$:

# Proveedor	Nombre_P
S1	Smith
S2	Jones
S3	Blake
S4	Clark
S5	Adams

 $\Pi_{\text{Categoría,Ciudad_P}}(\text{Prov})$:

Categoría_P	Ciudad_P
20	Londres
10	París
30	París
30	Atenas

Sea $\Pi_X(r) = s$, luego tenemos los siguientes casos extremales:

- $X = \emptyset$ entonces $esq(s) = \emptyset$ y $s = \emptyset$.
- X = esq(r) = R entonces esq(s) = esq(r) = R y s = r.
- $r = \emptyset$ entonces esq(s) = X y $s = \emptyset$.

COMBINACIÓN

• La combinación es una operación que, a partir de dos relaciones, obtiene una nueva relación formada por todas las tuplas que resultan de concadenar tuplas de la primera relación con tuplas de la segunda, y que cumplen una condición de combinación especificada.

COMBINACIÓN (CONT.)

EDIFICIOS_EMP		
<u>edificio</u>	supmediadesp	
Marina	15	
Diagonal	10	

DESPACHOS			
<u>edificio</u>	<u>número</u>	superficie	
Marina	120	10	
Marina	230	20	
Diagonal	120	10	
Diagonal	440	10	

 $R := EDIFICIOS[nombreedificio = edificio, supmediadesp \le superficie]$ DESPACHOS.

R				
nombreedificio	supmediadesp	edificio	número	superficie
Marina	15	Marina	230	20
Diagonal	10	Diagonal	120	10
Diagonal	10	Diagonal	440	10

OPERACIONES RELACIONALES BINARIAS

o Concatenación (join) Y División (división)

DIVISIÓN O COCIENTE

- \circ R \div S
- Define una relación sobre el conjunto de atributos C, incluido en la relación R, y que contiene el conjunto de valores de C, que en las tuplas de R están combinadas con cada una de las tuplas de S

Condiciones:

- grado(R) > grado(S)
- conjunto atributos de S sub conjunto de atributos de R
- Equivalencia con operadores básicos

$$X_1 = \Pi_C(R);$$
 $X_2 = \Pi_C((S X X_1) - R);$ $X = X_1 - X_2$

R1

E#	Proyecto
320	RX338A
320	PY254Z
•	RX338A
323	NC168T
•	PY254Z
•	PY254Z
324	NC168T

R2

Proyecto
RX338A
PY254Z

R1÷R2

E#
320
323

Unión Natural (Natural Join)

RMS 6 R * S

- El resultado es una relación con los atributos de ambas relaciones y se obtiene combinando las tuplas de ambas relaciones que tengan el mismo valor en los atributos comunes
 - Normalmente la operación de join se realiza entre los atributos comunes de dos tablas que corresponden a la clave primaria de una tabla y la clave foránea correspondiente de la otra tabla
 - Método
 - o Se realiza el producto cartesiano R x S
 - o Se seleccionan aquellas filas del producto cartesiano para las que los atributos comunes tengan el mismo valor
 - o Se elimina del resultado una ocurrencia (columna) de cada uno de los atributos comunes
- o Equivalencia con operadores básicos

$$R \bowtie_F S = \sigma_F(R XS)$$

JOIN

• Outer Join

- Es una variante del Join en la que se intenta mantener toda la información de los operandos, incluso para aquellas filas que no participan en el Join
- Se "rellenan con nulos" las tuplas que no tienen correspondencia en el Join
- Tres variantes
 - Left
 - o se tienen en cuenta todas las filas del primer operando
 - \circ Right
 - o se tienen en cuenta todas las filas del segundo operando
 - Full
 - o se tienen en cuenta todas las filas de ambos operandos

Join (cont.)

R1

E#	Nombre	D#
320	José	D1
322	Rosa	D3
•	María	D3
•	José	D5

R2

D#	Descrip
D1	Central
D3	I+D
D4	Ventas

R1 * R2

E#	Nombre	D#	Descrip
320	José	D1	Central
322	Rosa	D3	I+D
•	María	D3	I+D

R1 * LEFT R2

E#	Nombre	D#	Descrip
320	José	D1	Central
322	Rosa	D3	I+D
•	María	D3	I+D
•	José	D5	nul1

R1 * RIGHT R2

2	E#	Nombre	D#	Descrip
	•	José	D1	Central
	322	Rosa	D3	I+D
	•	María	D3	I+D
	nul1	nul1	D4	Ventas

R1 * FULL R2

E#	Nombre	D#	Descrip
320	José	D1	Central
322	Rosa	D3	I+D
•	María	D3	I+D
•	José	D5	nu11
nul1	null	D4	Ventas