

Álgebra Relacional

Capitulo 3

Objetivos

- Álgebra Relacional
- Proyección
- Selección
- Operaciones entre tablas(relaciones)
 - Unión
 - Diferencia
 - Producto Cruz
 - Intersección
 - Join(*)
 - División
 - Renombramiento

Definición de Tupla

- Se llama **tupla** a cada **fila** de una tabla.

<u>idEmpleado</u>	Puesto	Salario
100	Secretaria	3100
150	Gerente	3200
200	Secretaria	3100
250	Subgerente	3200
300	Secretaria	3100

Lenguajes Query Relacionales

- **Lenguajes Query:** Permite manipulación y acceso a los datos de una base de datos.
- Los lenguajes Query != Lenguajes de Programación
 - No se los utiliza para cálculos complejos
 - Soporta el acceso fácil y eficiente a grandes sets de datos
- *Entender el algebra relacional es la clave para entender SQL, procesamiento de queries!*

Un query es aplicado a las instancias de una tabla y el resultado de un query también es una instancia de una tabla.

Ejemplo de Instancias de las tablas

Botes
Marineros
Reserva

Reserva

<u>mid</u>	<u>bid</u>	<u>fecha</u>
22	101	10/10/96
58	103	11/12/96

Botes

<u>bid</u>	bnombre	color
101	Interlake	Azul
102	Interlake	Rojo
103	Clipper	Verde
104	Marine	Rojo

Marineros

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Ejemplo de Instancias

<i>R1</i>	<u>mid</u>	<u>bid</u>	<u>dia</u>
	22	101	10/10/96
	58	103	11/12/96

<i>s1</i>	<u>mid</u>	mnombre	rating	edad
	22	dustin	7	45.0
	31	lubber	8	55.5
	58	rusty	10	35.0

<i>s2</i>	<u>mid</u>	mnombre	rating	edad
	28	yuppy	9	35.0
	31	lubber	8	55.5
	44	guppy	5	35.0
	58	rusty	10	35.0

Algebra Relacional: 5 Operaciones basicas

- Proyección (π) Retiene solo las *columnas* requeridas de una relación (vertical).
- Selección (σ) Selecciona un subconjunto de *filas* de una relación (horizontal).
- Producto-cruz (\times) Nos permite combinar dos relaciones.
- Diferencia ($-$) Las tuplas en r1, pero no en r2.
- Union (\cup) Tuplas en r1 y/o en r2.

Como cada operación retorna una relación, *las operaciones pueden ser compuestas!*

Proyección

- Crea una nueva relación a partir de otra, pero incluyendo sólo algunas de las columnas
- Se representa con el símbolo π

$$\pi_{edad}(S2) \quad \pi_{snombre, rating}(S2)$$

- El operador proyección elimina duplicados
- *Esquema* resultante:
 - Exactamente los campos en la lista de proyección, con los mismos nombres que tenían en la relación inicial.

Proyección

<u>mid</u>	mnombre	rating	edad
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty s ₂	10	35.0

mnombre	rating
yuppy	9
lubber	8
guppy	5
rusty	10

$\pi_{mnombre, rating}^{(S2)}$

edad
35.0
55.5

$\pi_{edad}^{(S2)}$

Selección

- Selecciona las filas que satisfacen una condición de selección.
- El resultado es una relación.
- Se representa con el símbolo σ
- *Esquema* resultante:
 - El mismo de la relación inicial.

<u>mid</u>	mnombre	rating	edad
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

mnombre	rating
yuppy	9
rusty	10

$\sigma_{rating > 8}(S2)$
 $\pi_{mnombre, rating}(\sigma_{rating > 8}(S2))$

Escriba los siguientes queries en álgebra relacional:

- Las fechas en la que se ha reservado el bote #103.
- El id de los marineros que reservaron el bote #102
- Los nombres de los botes de color rojo.
- La edad de los marineros con rating mayor a 7.
- El id y color de los botes de nombre Interlake.
- El nombre y id de los marineros con 55 años de edad.

Reserva

<u>mid</u>	<u>bid</u>	<u>fecha</u>
22	101	10/10/96
58	103	11/12/96

Botes

<u>bid</u>	bnombre	color
101	Interlake	Azul
102	Interlake	Rojo
103	Clipper	Verde
104	Marine	Rojo

Marineros

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Unión

- $R \cup S$, la unión de R y S es el conjunto de elementos que están en R o S o ambos. Un elemento solo aparece una sola vez.
- Estas operaciones toman dos tablas que deben ser compatibles en unión, esto quiere decir:
 - El mismo número de columnas
 - Las columnas correspondientes tienen el mismo tipo

Unión

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

s1

<u>mid</u>	mnombre	rating	edad
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

s2

mid	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0
44	guppy	5	35.0
28	yuppy	9	35.0

$S1 \cup S2$

Diferencia

- $R - S$, la diferencia de R y S , es el conjunto de elementos que están en R pero no en S .
- Es importante resaltar que $R - S$ es diferente a $S - R$.
- También deben ser compatibles en unión

Diferencia

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S1

<u>mid</u>	mnombre	rating	edad
28	yuppy	9	35.0
31	lubber	8	55.5
44	guppy	5	35.0
58	rusty	10	35.0

S2

mid	mnombre	rating	edad
22	dustin	7	45.0

$S1 - S2$

<u>mid</u>	mnombre	rating	edad
28	yuppy	9	35.0
44	guppy	5	35.0

$S2 - S1$

Producto Cruz

- $R \times S$, los esquemas de ambas relaciones se mezclan y unen.
- Las filas en R se unen con las filas en S.
- No necesitan ser compatibles en unión.
- *Esquema resultante* un registro por cada registro de S1 y R1, con los nombres de registro heredados si es posible.
 - *Pueden haber conflictos de nombres: Ambos, S1 y R1 tienen un registro con el mismo nombre.*
 - *En este caso, se puede usar el operador de renombre:*

$$\rho (C(1 \rightarrow sid1, 5 \rightarrow sid2), S1 \times R1)$$

Producto Cruz

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

S1

<u>mid</u>	<u>bid</u>	<u>dia</u>
22	101	10/10/96
58	103	11/12/96

R1

S1 X R1 =

mid	mnombre	rating	edad	mid	bid	dia
22	dustin	7	45.0	22	101	10/ 10/ 96
22	dustin	7	45.0	58	103	11/ 12/ 96
31	lubber	8	55.5	22	101	10/ 10/ 96
31	lubber	8	55.5	58	103	11/ 12/ 96
58	rusty	10	35.0	22	101	10/ 10/ 96
58	rusty	10	35.0	58	103	11/ 12/ 96

Funciones de agregación

- Funciones que toman una colección de valores y devuelven como resultado un único valor.
- Por ejemplo: la función de agregación **sum** toma un conjunto de valores y devuelve la suma de los mismos.

Sum
Avg
Count
Min
Max

Por tanto, la función **sum** aplicada a la columna edad de la instancia A devuelve 24.

La función de agregación **avg** devuelve la media de los valores. Cuando se aplica al conjunto anterior devuelve el valor 4.

La función de agregación **count** devuelve el número de elementos del conjunto, y devolvería 6 en el caso anterior.

La función de agregación **min** devuelve 1.

La función de agregación **max** devuelve 11.

A	
	edad
	1
	1
	3
	4
	4
	11

- Si se desea obtener la suma de los ratings de los marineros:

$G_{\text{sum(rating)}}$ Marineros

La operación del álgebra relacional G significa que se debe aplicar agregación, y el subíndice indica la operación de agregación a aplicar.

Marineros

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

El resultado de la expresión anterior es una relación con un único atributo, que contiene una sola fila con un valor correspondiente a la suma de los ratings de los marineros.

Sum(rating)

25

Distinct

Hay casos en los que se deben borrar los valores repetidos antes de calcular una función de agregación.

Si se desean borrar los valores repetidos hay que utilizar los mismos nombres de funciones que antes, con la cadena de texto «distinct» añadida al final del nombre de la función (por ejemplo, count distinct).

```
COUNT ( [DISTINCT] A)  
SUM ( [DISTINCT] A)  
AVG ( [DISTINCT] A)  
MAX (A)  
MIN (A)
```

A

edad
1
1
3
4
4
11

trabajo-por-horas

<i>nombre-empleado</i>	<i>nombre-sucursal</i>	<i>sueldo</i>
González	Centro	1.500
Díaz	Centro	1.300
Jiménez	Centro	2.500
Catalán	Leganés	1.600
Cana	Leganés	1.500
Cascallar	Navacerrada	5.300
Fernández	Navacerrada	1.500
Ribera	Navacerrada	1.300

- Si se quiere averiguar el número de sucursales que aparecen en la instancia trabajo-por-horas:

$G_{\text{count(nombre-sucursal) trabajo-por-horas}}$

Retorna 8, cuando en realidad sólo hay 3 sucursales.

$G_{\text{count distinct(nombre-sucursal) (trabajo-por-horas)}}$

- En este caso, el nombre de cada sucursal sólo se cuenta una vez, independientemente del número de empleados que trabajen en la misma.

Quiero saber cuántos marineros hay.

$G_{\text{count}(\text{mid})}(\text{Marineros})$

Reserva

<u>mid</u>	<u>bid</u>	<u>fecha</u>
22	101	10/10/96
58	103	11/12/96

Quiero saber la edad promedio de los marineros con rating=10.

$G_{\text{avg}(\text{edad})}(\sigma_{\text{rating}=10} \text{Marineros})$

Marineros

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Quiero saber la edad del marinero más joven con rating mayor a 7.

$G_{\text{min}(\text{edad})}(\sigma_{\text{rating} > 7} \text{Marineros})$

Botes

<u>bid</u>	bnombre	color
101	Interlake	Azul
102	Interlake	Rojo
103	Clipper	Verde
104	Marine	Rojo

trabajo-por-horas

<i>nombre-empleado</i>	<i>nombre-sucursal</i>	<i>sueldo</i>
González	Centro	1.500
Díaz	Centro	1.300
Jiménez	Centro	2.500
Catalán	Leganés	1.600
Cana	Leganés	1.500
Cascallar	Navacerrada	5.300
Fernández	Navacerrada	1.500
Ribera	Navacerrada	1.300

- Si se desea hallar la suma total de sueldos de todos los empleados en cada sucursal del banco por separado, en lugar de hallar la suma de sueldos de todo el banco.
- Para ello hay que dividir la instancia trabajo-por- horas en grupos basados en la sucursal y aplicar la función de agregación a cada grupo.

La expresión siguiente obtiene el resultado deseado utilizando el operador de agregación G:

`nombre-sucursal Gsum(sueldo) (trabajo-por-horas)`

Nombre-sucursal	sum(sueldo)
Centro	5300
Leganés	3100
Navacerrada	8100

El atributo nombre-sucursal subíndice a la izquierda de G indica que la relación de entrada trabajo-por-horas debe dividirse en grupos de acuerdo con el valor de nombre-sucursal.

trabajo-por-horas

<i>nombre-empleado</i>	<i>nombre-sucursal</i>	<i>sueldo</i>
González	Centro	1.500
Díaz	Centro	1.300
Jiménez	Centro	2.500
Catalán	Leganés	1.600
Cana	Leganés	1.500
Cascallar	Navacerrada	5.300
Fernández	Navacerrada	1.500
Ribera	Navacerrada	1.300

- Si deseo averiguar el sueldo máximo de los empleados de cada oficina, además de la suma de los sueldos, habría que escribir la expresión:

nombre-sucursal $G_{\text{sum}(\text{sueldo}), \text{max}(\text{sueldo})}$ (trabajo-por-horas)

<i>nombre-sucursal</i>	<i>suma-sueldo</i>	<i>sueldo-máximo</i>
Centro	5.300	2.500
Leganés	3.100	1.600
Navacerrada	8.100	5.300

Encontrar la edad del marinero más joven para cada nivel de rating.

$\text{rating } G_{\min(\text{edad})}(\text{Marineros})$

Para cada rating encuentre la edad promedio de los marineros mayores a 18 años.

$\text{rating } G_{\text{avg}(\text{edad})}(\sigma_{\text{edad} \geq 18} \text{Marineros})$ **Marineros**

Reserva

<u>mid</u>	<u>bid</u>	<u>fecha</u>
22	101	10/10/96
58	103	11/12/96

<u>mid</u>	mnombre	rating	edad
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

Botes

<u>bid</u>	bnombre	color
101	Interlake	Azul
102	Interlake	Rojo
103	Clipper	Verde
104	Marine	Rojo

Resumen

- Álgebra relacional: Un pequeño grupo de operadores que mapean relaciones con relaciones
 - Operacional, en el sentido que uno especifica el orden explícito de las operaciones
 - Se pueden mezclar
- Las operaciones básicas incluyen: σ , π , \times , \cup , $-$
- Operadores compuestos importantes: \cap , \bowtie , $/$

Álgebra Relacional

Sean las relaciones:

ADMINISTRADOR

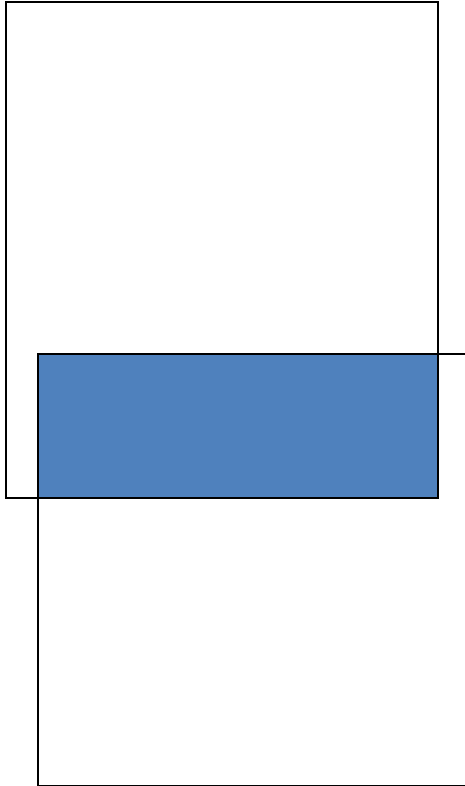
código	nombre	edad
1	Jorge Campos	33
2	Enrique Muñoz	25
3	Esteban Paz	21

Nota: Aquí se usa el atributo edad, pero en una BD se suele guardar mas bien la fecha de nacimiento

PRODUCTOR

código	nombre	edad
2	Enrique Muñoz	25
8	Jorge Arias	30
10	Juan Martínez	19

Álgebra Relacional



- Intersección: $R_1 \cap R_2$
 - Retorna una relación que contiene todas las tuplas que aparecen tanto en R_1 como en R_2
 - Las relaciones **deben tener cabecera idéntica** para que sean **compatibles para la intersección**

Álgebra Relacional

- ADMINISTRADOR \cap PRODUCTOR:

código	nombre	edad
2	Enrique Muñoz	25

Álgebra Relacional

→ No traducirla al español como unión

- Reunión (*join*): Tiene varias modalidades:
 - Reunión natural (*natural join*): $R_1 \bowtie R_2$
Sea $R_1(A,B)$ y $R_2(B,C)$. El resultado de $R_1 \bowtie R_2$ es una relación con cabecera (A,B,C) donde cada tupla de R_1 se combina con cada tupla de R_2 **siempre y cuando tengan el mismo valor** en el atributo en común (B).
A, B y C pueden ser atributos compuestos.

Álgebra Relacional

EMPLEADO

código	nombre	edad	depto
1	Jorge Campos	33	1
2	Enrique Muñoz	25	1
3	Esteban Paz	21	1
8	Jorge Arias	30	2
10	Juan Martínez	19	2
12	Anselmo Rodas	28	6

DEPARTAMENTO

depto	descripción
1	Administración
2	Producción
3	Ventas
4	Finanzas

Nota: En este ejemplo, el campo **depto** en EMPLEADO **no** es CF con respecto a la relación DEPARTAMENTO

Atributo de
join

Álgebra Relacional

- EMPLEADO \bowtie DEPARTAMENTO:

código	nombre	edad	depto	descripción
1	Jorge Campos	33	1	Administración
2	Enrique Muñoz	25	1	Administración
3	Esteban Paz	21	1	Administración
8	Jorge Arias	30	2	Producción
10	Juan Martínez	19	2	Producción

Álgebra Relacional

- Semi reunión (*semi join*): $R_1 \triangleright R_2$:
 - Similar a la reunión natural pero solo se muestran los atributos de la relación de la izquierda ☐ ¿Utilidad?

Álgebra Relacional

- EMPLEADO $\bowtie_{\text{depto} > \text{dep}}$ $\rho_{\text{DEPT}(\text{dep}, \text{desc})}$
 $(\pi_{(\text{depto}, \text{descripción})}(\text{DEPARTAMENTO}))$

código	nombre	edad	depto	dep	desc
8	Jorge Arias	30	2	1	Administración
10	Juan Martínez	19	2	1	Administración
12	Anselmo Rodas	28	6	1	Administración
12	Anselmo Rodas	28	6	2	Producción
12	Anselmo Rodas	28	6	3	Ventas
12	Anselmo Rodas	28	6	4	Finanzas

Álgebra Relacional

- División: $R_1 \div R_2$

Sean las relaciones $R_1(\mathbf{A}, \mathbf{B})$ y $R_2(\mathbf{B})$. El resultado de $R_1 \div R_2$ es una relación con el atributo **A**

- Un valor **a** $\in A$ hace parte del resultado de la división si **a** está en R_1 combinado con cada valor $\in B$ que aparece en R_2
- **A y B pueden ser atributos compuestos**

Álgebra Relacional

R1: ARTISTAYGENERO

codartista	nomgénero
45	Soul
45	Pop
45	Dance
8	Soul
8	Pop
23	Dance

R2: GÉNERO

nomgénero
Soul
Pop
Dance

Álgebra Relacional

- $\text{ARTISTAYGENERO} \div \text{GÉNERO}$

codartista
45

Ya que en este caso solo el artista 45 está combinado con todos los géneros que aparecen en la relación GÉNERO

Renombramiento

4.11 Renombramiento ρ

Renombrar una relación para facilitar la interacción con otras

$$\rho_s(R)$$

Ej.

$$\pi_{t.nombre} (\sigma_{s.nombre='carlos' \text{ and } t.curso='IS341'} (\rho_s(\text{PROFE}) \times \rho_t(\text{CURSO})))$$

Renombrar un atributo

Suponiendo $R(A,B,C)$

$\rho_{R(A,X,C)}(R)$	$\rho_{s(A,X,C)}(R)$	$\pi_{A, B \text{ as } X, C}(R)$	$\pi_{B \text{ as } X}(R)$
$= R(A,X,C)$	$= S(A,X,C)$	$= R(A,X,C)$	$= R(X)$