

álgebra relacional

Índice

- Selección σ
 - Proyección π
 - Unión \cup
 - Diferencia $-$
 - Intersección \cap
 - Producto cartesiano \times
 - Join natural \bowtie
 - Join externo $=\bowtie$, $\bowtie=$, $=\bowtie=$
-

- El **álgebra relacional** se trata de un **conjunto de operaciones** utilizadas para manipular relaciones.
- **Toman relaciones** como operandos y **devuelven relaciones** que pueden ser de nuevo manipuladas.
- Se compone de las siguientes operaciones básicas:
 - Unión, diferencia e intersección:** Con el significado usual en conjuntos pero aplicado a relaciones.
 - Selección:** Selecciona ciertas tuplas de una relación.
 - Proyección:** Selecciona ciertas columnas de una relación.
 - Productos y joins:** Composición de relaciones.
 - Renombrado** de relaciones y atributos.
 - División, agrupación y ordenación** de relaciones.
 - Modificación de tuplas:** Inserción, actualización y eliminación.

Selección σ *predicado*:

Objetivo: Selecciona tuplas que satisfacen una sentencia.

Donde:

- **Predicado:** Se trata de una expresión que incluye cualquier operador lógico.

- **Operandos:** Atributos o constantes.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

$\sigma_{\text{sueldo} \geq 8000 \wedge \text{ingreso} \geq 01/01/2003}(\text{Empleado})$

Aquí estamos diciendo que queremos seleccionar aquellas tuplas de mi tabla *Empleado* y que el predicado a cumplirse es que el **sueldo** del empleado debe de ser **mayor** a 8000 y su **fecha de ingreso** tiene que ser **mayor o igual** a 01/01/2003. A continuación, esa operación nos devuelve otra relación, como veremos, de cardinalidad 3 (número de tuplas):

RFC	Nombre	sueldo	depto	ingreso
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003

Ahora, hay ocasiones en las que no queremos proyectar TODOS los atributos (columnas), sino que queremos las necesarias. Para ello, existe la siguiente operación llamada Proyección.

Proyección π *lista_de_atributos*:

Objetivo: Obtener una tabla a partir de R (mi tabla original), eliminando los atributos nos especificados.

En la tabla resultante vemos los atributos en el mismo orden que en mi tabla original. Y algo importante a tener en cuenta; **los renglones duplicados se eliminan**.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

$\pi_{\text{sueldo, depto}}(\text{Empleado})$

Como vemos en nuestra expresión, solo queremos proyectar en nuestra nueva tabla las columnas (atributos) de **sueldo** y **depto** de nuestra tabla *Empleado*. Algo importante a tener en cuenta es que la proyección **no evalúa predicados**.

sueldo	depto
\$ 6,000.00	A1
\$ 8,000.00	A2
\$ 10,000.00	A2
\$ 8,000.00	A1
\$ 12,000.00	A1
\$ 15,000.00	A3
\$ 6,000.00	A2
\$ 10,000.00	A3

Unión $R \cup S$:

Objetivo: Obtener una tabla que contiene las tuplas de la primera relación además de las tuplas de la segunda relación. Además, a tener en cuenta, **los renglones duplicados se eliminan**.

Condiciones a cumplir: Al adaptar los operadores de conjuntos a las relaciones se debe asegurar que exista compatibilidad entre ellas:

Deben de tener el mismo **grado**, los atributos deben de tener el mismo **nombre** y el dominio del atributo **i** de **R** debe de ser el mismo que el atributo **i** de **S** $\forall i$.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

```
r = σsueldo >= 10000 (Empleado)
```

```
s = σdepto = A1 (Empleado)
```

```
πnombre (r ∪ s)
```

Lo que vemos como $=$ es que estamos haciendo una **relación temporal**, como si de una variable se tratase. Nosotros en la variable **r** guardamos el resultado de nuestra selección que demanda a los empleados de **sueldo** mayor a 10000, y en nuestra variable **s** guardamos el resultado de aquellos empleados que pertenezcan al **depto** A1. Posteriormente realizamos la **unión** de esas relaciones y proyectamos solamente el atributo nombre.

Nombre
Jorge Rojas
Germán Flores
Noé Pérez
Carlos Hernández
María López
Verónica Sánchez
Oscar Romero

Diferencia **R - S**:

Objetivo: Crear una tabla con las tuplas que están en la relación R pero no en S.

Es una operación válida entre relaciones **compatibles**, así como sucedía con la operación de la unión. Y al igual que como venimos viendo, los renglones duplicados se eliminan.

Empleado

RFC	Nombre	sueldo	depto	ingreso	
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004	
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003	✓
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003	✓✓
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002	
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002	✓
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002	✓
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003	✓
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002	✓
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003	✓
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002	✓

```
r = σsueldo >= 10000 (Empleado)
```

```
s = odepto = A2 (Empleado)
πnombre, sueldo (r-s)
```

En la imagen anterior vemos nuestra tabla, de allí **extraemos** la relación **R** y la relación **S**. Las tildes de **color rojo** (a modo de ilustración) pertenecen a la relación **S** y las tildes de **color azul** pertenecen a la relación **R**. Para poder hacer la operación de la diferencia, tienen que quedar **todas las tuplas de R que no estén en S**, es básicamente como matemáticas, como la operación de conjuntos de matemáticas. Vemos dos tildes en la tercera tupla, esa tupla queda **totalmente anulada** ya que está en ambas relaciones, y tiene que estar **solamente en R**. El resultado de la operación es la siguiente:

Nombre	sueldo
Germán Flores	\$ 12,000.00
Noé Pérez	\$ 15,000.00
Carlos Hernández	\$ 12,000.00
María López	\$10,000.00

Intersección $R \cap S$:

Objetivo: Se trata de una relación con las tuplas que están en **R** y en **S** también. Es decir, esos valores de atributos que comparten, tal como matemáticas y la intersección de conjuntos. Es una operación válida entre relaciones **compatibles**, como vimos con **unión** y **diferencia**. Además, los renglones duplicados se eliminan.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

```
r = πsueldo(odepto = A1 (Empleado))
s = πsueldo(odepto = A2 (Empleado))
r ∩ s
```

Como vemos, hace una selección de **sueldos** en donde el **depto** sea igual a **A1** y luego la otra relación hace lo mismo pero con el depto igual a **A2**. Carlos Hernández y Patricia Ruiz cobran 8000, pero solo un valor prevalece, ya que **se eliminan los duplicados**.

sueldo
\$ 6,000.00
\$ 8,000.00

Producto cartesiano $R \times S$:

Permite combinar información de **cualquier** par de relaciones en dónde $R \times S = \{tq \mid t \in R \text{ y } q \in S\}$

Si R y S tienen atributos en común, es necesario renombrarlos para evitar ambigüedad.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

odepto = A1 (Empleado) x Departamento

Departamento

depto	nombre	fecha
A1	Sistemas	01/03/2002
A2	Mercadotécnica	01/01/2002
A3	Ventas	01/01/2001
A4	Recursos Humanos	01/01/2003

Se realizan todas las combinaciones posibles con las 4 tuplas seleccionadas:

RFC	Empleado.Nombre	sueldo	depto	ingreso	depto	departamento.nombre	fecha
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004	A1	Sistemas	01/03/2002
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002	A1	Sistemas	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002	A1	Sistemas	01/03/2002
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002	A1	Sistemas	01/03/2002
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004	A2	Mercadotécnica	01/01/2002
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002	A2	Mercadotécnica	01/01/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002	A2	Mercadotécnica	01/01/2002
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002	A2	Mercadotécnica	01/01/2002
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004	A3	Ventas	01/01/2001
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002	A3	Ventas	01/01/2001
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002	A3	Ventas	01/01/2001
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002	A3	Ventas	01/01/2001
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004	A4	Recursos Humanos	01/01/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002	A4	Recursos Humanos	01/01/2003
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002	A4	Recursos Humanos	01/01/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002	A4	Recursos Humanos	01/01/2003

Como vemos se hace una combinación de **todos contra todos**, pero cuidado, ya que quizás hay datos que no queremos y, que si no filtramos, ésta relación puede verse como **ilógica**. Las únicas **combinaciones** que me interesarían serían aquellas que corresponden con el **depto A1**.

Join Natural $R \bowtie S$:

Empareja y combina automáticamente filas de dos tablas basándose en **columnas con el mismo nombre** y tipos de datos **compatibles**. Devuelve un resultado sin columnas duplicadas redundantes.

Empleado

RFC	Nombre	suelo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

$\pi_{RFC, nombre, depto, nom_depto}(Empleado \bowtie Departamento)$

Como vemos, hacemos un join entre **Empleado** y **Departamento**, fijándose qué atributo tienen en **común**; viendo que tienen el atributo **depto** como común, podremos unir ambas tablas y realizar la selección solicitada.

RFC	Nombre	depto	nom_depto
SACV750524	Verónica Sánchez	A1	Sistemas
HEFC790926	Carlos Hernández	A2	Mercadotécnica
ROPJ081001	Jorge Rojas	A2	Mercadotécnica
ROAO740114	Oscar Romero	A1	Sistemas
FORG760206	Germán Flores	A1	Sistemas
PEVN780704	Noé Pérez	A3	Ventas
JUBF910410	Fabián Juárez	A2	Mercadotécnica
HEVC950826	Carlos Hernández	A1	Sistemas
RUZP810512	Patricia Ruíz	A2	Mercadotécnica
LOMM750919	María López	A3	Ventas

Join externo

Aquí se intenta mantener **toda la información** de los operandos, incluso para aquellas tuplas que no participan en el join.

Se rellenan con **valores nulos** las tuplas que no tienen correspondencia en el join. Por ejemplo, vemos que tenemos tres tipos:

- **Left (\bowtie):** Se tienen en cuenta todas las filas del primer operando
- **Right (\bowtie):** Se tienen en cuenta todas las filas del segundo operando.
- **Full (\bowtie):** Se tienen en cuenta todas las filas de ambos operandos

Renombrar ρ

La notación se compone de:

- ρ NuevoNombre (R)
- ρ NuevoNombre \leftarrow atributo (R)

Asigna un nuevo nombre a la relación y/o atributos. Ojo, no se obtiene una nueva relación.

Por ejemplo:

```
p D (Departamento)
```

D

depto	nombre	fecha
A1	Sistemas	01/03/2002
A2	Mercadotécnica	01/01/2002
A3	Ventas	01/01/2001
A4	Recursos Humanos	01/01/2003

Como vemos, la relación es la misma, sólo que pasamos a cambiarle su nombre.

Agrupación y

La notación es la siguiente:

```
yatributo_agrupación;funcion_agregacion → alias (R)
```

Como todas las columnas de la relación tienen un nombre, utilizamos un **alias** para poder diferenciar esta nueva columna que contendrá esta operación.

Esta operación me permitirá generar grupos de valores en función de un atributo determinado.

Algunas de las funciones de agregación son **sum**, **avg**, count, min, max. Las dos primeras están obligadas a trabajar con dominios numéricos.

Empleado

RFC	Nombre	sueldo	depto	ingreso
SACV750524	Verónica Sánchez	\$ 6,000.00	A1	01/01/2004
HEFC790926	Carlos Hernández	\$ 8,000.00	A2	01/01/2003
ROPJ081001	Jorge Rojas	\$ 10,000.00	A2	01/10/2003
ROAO740114	Oscar Romero	\$ 8,000.00	A1	01/03/2002
FORG760206	Germán Flores	\$ 12,000.00	A1	01/01/2002
PEVN780704	Noé Pérez	\$ 15,000.00	A3	01/06/2002
JUBF910410	Fabián Juárez	\$ 6,000.00	A2	01/11/2003
HEVC950826	Carlos Hernández	\$ 12,000.00	A1	01/06/2002
RUZP810512	Patricia Ruíz	\$ 8,000.00	A2	01/04/2003
LOMM750919	María López	\$ 10,000.00	A3	01/10/2002

```
ydepto;avg(sueldo)→sueldoprom (Empleado)
```

Aquí estoy diciendo que quiero generar grupos por **depto**, y determino el promedio del **sueldo** de cada, y posteriormente llamando ésta nueva columna con un **alias 'sueldoprom'**.

depto	sueldoprom
A1	\$ 9,500.00
A2	\$ 8,000.00
A3	\$ 12,500.00

- Sujeto a cambios -