

Lenguaje SQL

⚠ EN DESARROLLO ⚠

Introducción

SQL, que significa "*Structured Query Language*" (Lenguaje de Consulta Estructurado), es un lenguaje de programación utilizado para gestionar y manipular bases de datos relacionales. Fue desarrollado en la década de 1970 en IBM basándose en el trabajo de Edgar Codd y se ha convertido en el estándar de facto para interactuar con sistemas de gestión de bases de datos relacionales (SGBDR).

La primera versión de SQL normalizada por ANSI data de 1986. La norma SQL2 o SQL92 es la más importante y la mayoría de los SGBDR existentes implementan esta versión.

Características clave de SQL:

- **Declarativo:** SQL es un lenguaje de programación declarativo, lo que significa que describe el resultado deseado sin especificar el método para alcanzarlo. Esto permite a los usuarios centrarse en lo que quieren obtener, en lugar de cómo lograrlo.
- **Gestión de datos relacional:** SQL se utiliza principalmente en entornos de bases de datos relacionales, donde la información se organiza en tablas con relaciones entre ellas. Esto facilita la gestión y recuperación eficiente de datos.
- **Manipulación de datos:** SQL permite realizar operaciones fundamentales sobre datos, como la inserción (INSERT), la actualización (UPDATE), la eliminación (DELETE), y la recuperación (SELECT) de información de la base de datos.
- **Creación y modificación de estructuras de datos:** SQL también se utiliza para definir y modificar la estructura de las bases de datos, mediante la creación de tablas (CREATE TABLE), la alteración de tablas (ALTER TABLE), y la eliminación de tablas (DROP TABLE).
- **Consulta de datos:** La sentencia SELECT es esencial en SQL y se utiliza para recuperar datos de una o varias tablas. Permite filtrar, ordenar y agrupar información según las necesidades del usuario.
- **Integridad de datos:** SQL garantiza la integridad de los datos mediante restricciones como claves primarias, claves foráneas y otros mecanismos que aseguran la consistencia y fiabilidad de la información almacenada.
- **Transacciones:** SQL ofrece soporte para transacciones, permitiendo agrupar varias operaciones en una unidad atómica. Esto asegura que todas las operaciones se realicen con éxito o ninguna de ellas.

SQL se ha convertido en una herramienta esencial para cualquier persona involucrada en el desarrollo de software, administración de bases de datos, análisis de datos y otras disciplinas relacionadas con el manejo de información estructurada. Con su amplia adopción, el conocimiento de SQL es valioso en diversos contextos profesionales.

El lenguaje SQL se divide en cuatro subconjuntos:

- **DDL (*Data Definition Language*)**, que agrupa todos los comandos utilizados para crear, modificar o eliminar las estructuras de la base de datos (tablas, índices, vistas, etc.). Se trata principalmente de los comandos:
 - CREATE
 - ALTER
 - DROP

- **DML** (*Data Manipulation Language*), que agrupa los comandos utilizados para manipular los datos contenidos en la base de datos. Se trata principalmente de los comandos:
 - `SELECT`
 - `INSERT`
 - `DELETE`
 - `UPDATE`
- **DCL** (*Data Control Language*), que agrupa los comandos utilizados para administrar la seguridad de acceso a los datos. Se trata principalmente de los comandos:
 - `GRANT`
 - `REVOKE`
- **TCL** (*Transaction Control Language*), que agrupa los comandos utilizados para administrar la confirmación o no de actualizaciones realizadas sobre la base de datos. Se trata principalmente de los comandos:
 - `COMMIT`
 - `ROLLBACK`

Las palabras clave SQL **no distinguen entre mayúsculas y minúsculas**. Por ejemplo, `SELECT` es lo mismo que `select`.

Algunos sistemas de bases de datos **requieren un punto y coma (;)** al final de cada instrucción SQL. Es la forma estándar de separar cada sentencia SQL en sistemas de bases de datos que permiten ejecutar **más de una sentencia SQL** en la misma llamada al servidor.

⚠ Sección introductoria generada por ChatGPT

Modelo relacional

El modelo relacional fue creado por **Edgar Codd**, un investigador que trabajaba en IBM a principios de los años 70. Su trabajo pionero, "*A Relational Model of Data for Large Shared Data Banks*", sentó las bases de este modelo.

El modelo relacional se basa en el concepto de **conjunto**. Esquemáticamente, el modelo relacional se puede representar a través de una **tabla**, que en términos relacionales es una representación de una relación. Cada fila en la tabla corresponde a una **tupla (o registro)**, y cada columna representa un **atributo**.

En el modelo relacional, el término **relación** se utiliza para referirse a una tabla, mientras que el término **tupla** se utiliza para referirse a una fila. De forma similar, el término **atributo** se refiere a una columna de la tabla.

Este modelo presenta los datos de forma lógica, independientemente del modelo físico. El proveedor de la base de datos decide el modo de almacenamiento físico de las tablas. Esta independencia entre lo lógico y lo físico es una de las mayores ventajas de las bases de datos relacionales.

Una vez definidas las tablas, es necesario un lenguaje para manipularlas, lo cual se logra a través del **álgebra relacional**. Este formalismo matemático permite realizar operaciones sobre conjuntos de datos, como selecciones, proyecciones y uniones.

SQL implementa el álgebra relacional, y los sistemas de gestión de bases de datos relacionales (SGBDR) son los encargados de implementar el modelo relacional en la práctica.

El objeto principal gestionado por el modelo relacional es la **relación**, que está asociada a los conceptos de **dominio** y de **atributo**.

A esta relación se le aplican **reglas**:

- **Coherencia**: los datos en cualquier columna deben pertenecer al tipo definido por el dominio de ese atributo. La coherencia se asegura cuando cada valor pertenece al dominio sobre el que está definido el atributo.
- **Unicidad de las filas**: no pueden existir dos tuplas (filas) idénticas dentro de una misma relación (tabla). Cada tupla debe ser única.
- **Aceptación de valores nulos**: el concepto de valor nulo es válido y representa datos desconocidos o inaplicables en una columna.
 - **Restricción de entidad**: cualquier valor que forme parte de una clave primaria debe ser diferente de NULL, ya que las claves primarias deben identificar de manera única cada fila.
- **Atomicidad de los valores**: Los datos almacenados en cada columna de una tabla deben ser atómicos (monovalorados). Es decir, cada valor no puede ser dividido en valores más pequeños. Esto garantiza que la relación esté normalizada.
- **Independencia del orden**
 - El orden de los atributos (columnas) en una relación no tiene importancia.
 - El orden de las tuplas (filas) tampoco importa; las tuplas no están ordenadas.
- **Nombre único**:
 - Cada columna de una relación debe tener un nombre único dentro de esa tabla.
 - Cada relación (tabla) debe tener un nombre específico y diferente al resto de las relaciones en la base de datos.
- **Identificador**: atributo o conjunto de atributos que permiten caracterizar de forma unívoca cada elemento de la relación.
 - **Clave primaria**: identificador mínimo de una relación, que asegura la unicidad de cada tupla. Por ejemplo, en una tabla `clientes`, el campo `id_cliente` puede ser la clave primaria.
 - **Clave secundaria**: otros atributos que también pueden identificar de manera única una tupla, pero no son el identificador principal. Un ejemplo podría ser el `nif` de un cliente en la misma tabla `clientes`.
- **Integridad referencial**: esta regla impone que un atributo o conjunto de atributos de una relación aparezca como clave primaria en otra relación. Esto garantiza que las relaciones entre tablas sean válidas y consistentes.
 - **Clave externa**: atributo o conjunto de atributos que referencia la clave primaria de otra tabla, asegurando la integridad referencial. Por ejemplo, en una tabla `pedidos`, el campo `id_cliente` puede ser una clave externa que referencia a la clave primaria `id_cliente` en la tabla `clientes`.

Dominio

El dominio es el **conjunto de posibles valores** que un atributo puede tomar, representado por un nombre. Por ejemplo, el atributo `edad` en una tabla de personas puede tener como dominio el conjunto de números enteros positivos ($\text{edad} \geq 0$). Si otro atributo es `nombre`, el dominio puede ser el conjunto de todas las cadenas de texto válidas.

Si tenemos un atributo `estado_civil`, su dominio podría ser {"soltero", "casado", "divorciado", "viudo"}.

Cardinalidad

El número de elementos en un dominio se denomina **cardinalidad**. En el ejemplo anterior, la cardinalidad del dominio `estado civil` es 4.

Relación

Una relación definida sobre los dominios D_1, D_2, \dots, D_n es un subconjunto del producto cartesiano de estos dominios, caracterizado por un nombre. En términos prácticos, una **relación se representa como una tabla** en una base de datos relacional.

Una tabla `personas` puede ser una relación sobre los dominios `nombre`, `edad` y `estado civil`, donde cada fila (tupla) representa una persona.

Atributo

Un **atributo es una columna de la relación**, caracterizada por un nombre. Cada atributo está asociado con un dominio que define los valores posibles para esa columna.

En la relación `personas`, los atributos podrían ser `nombre`, `edad` y `estado civil`.

Grado

En una relación, el **grado es el número de atributos** que tiene. Dicho de otro modo, el número de columnas de una tabla.

Si la tabla `personas` tiene los atributos `nombre`, `edad` y `estado civil`, entonces el grado de la relación es 3.

Álgebra relacional

El **álgebra relacional** es un conjunto de operaciones que permite extraer y manipular datos de bases de datos relacionales. Se basa en la **teoría de conjuntos** y permite combinar, filtrar y transformar tablas (relaciones) existentes para generar nuevas tablas de resultados. Estos resultados pueden ser usados de **manera inmediata** o como base para otras operaciones.

Los operadores fundamentales basados en la teoría de conjuntos:

- **Selección (σ)**: este operador se usa para filtrar filas de una tabla que cumplen una condición específica. El resultado es un subconjunto de la tabla original.
- **Proyección (π)**: este operador selecciona columnas específicas de una tabla, eliminando las columnas que no son necesarias en la nueva tabla de resultados.
- **Unión (\cup)**: combina las filas de dos tablas que tienen la misma estructura (mismo número de columnas y dominios compatibles). La tabla de resultado contiene todas las filas de ambas tablas, eliminando duplicados.
- **Intersección (\cap)**: devuelve las filas que son comunes a dos tablas que tienen la misma estructura.
- **Diferencia ($-$)**: devuelve las filas que están en la primera tabla, pero no en la segunda, para tablas con la misma estructura.
- **Producto cartesiano (\times)**: combina todas las filas de dos tablas, generando todas las posibles combinaciones de sus tuplas.
- **Combinación (\bowtie)**: combina las filas de dos tablas basándose en una condición que involucra una o más columnas de cada tabla. Existen diferentes tipos de *join*:
 - **Theta Join (θ)**: combina tablas usando una condición cualquiera.
 - **Equijoin**: un tipo específico de *join* donde las columnas relacionadas deben tener valores iguales.

- **Natural Join:** Similar al *equijoin*, pero elimina las columnas repetidas en las tablas relacionadas.

Selección (σ)

La **selección o restricción** (σ) es una operación del álgebra relacional que permite **extraer filas (tuplas) de una relación (tabla)** que cumplen con una **condición específica**. La tabla resultante tendrá la misma estructura (número de columnas y dominios) que la relación original, pero solo contendrá las filas que satisfacen el criterio de selección.

La notación es $R_x = \sigma(\text{condición})(R)$, donde R es la relación de la cual se realiza la selección y R_x es la nueva relación resultante.

Las condiciones se pueden concatenar utilizando los operadores lógicos **AND** y **OR**, además de utilizar operadores de comparación como **=**, **>**, **<**, **>=**, **<=**, entre otros.

El término **"selección"** del álgebra relacional tiene un **significado diferente** del que se utiliza en SQL, que resulta de un hecho histórico desafortunado. En álgebra relacional, el término **"selección"** corresponde a lo que en SQL se corresponde con la cláusula **where**.

Como ejemplo de una selección, si tenemos una tabla **CLIENTES** y deseamos seleccionar todos los clientes que pertenecen a la región "Centro":

- **Tabla CLIENTES**

ID	Nombre	Región
1	Juan Pérez	Centro
2	Ana Gómez	Oeste
3	Luis Rodríguez	Centro
4	Marta López	Este
5	Carlos Ruiz	Oeste
6	Laura Sánchez	Centro

Esta operación se expresa como $CLIENTES_CENTRO = \sigma(\text{región} = \text{"Centro"})(CLIENTES)$. Esto generará una nueva tabla **CLIENTES_CENTRO** que incluirá únicamente las filas de **CLIENTES** donde la columna **región** tiene el valor "Centro".

- **Resultado:** $CLIENTES_CENTRO = \sigma(\text{región} = \text{"Centro"})(CLIENTES)$

ID	Nombre	Región
1	Juan Pérez	Centro
3	Luis Rodríguez	Centro
6	Laura Sánchez	Centro

Otras condiciones podrían ser:

- $\sigma(\text{región} = \text{'Centro'} \text{ AND } \text{nombre} = \text{'Juan Pérez'}})(CLIENTES)$
- $\sigma(\text{región} = \text{'Centro'} \text{ OR } \text{región} = \text{'Oeste'}})(CLIENTES)$
- $\sigma(\text{id} > 2)(CLIENTES)$

- $\sigma(\text{id} \leq 3)(\text{CLIENTES})$

Traducción a SQL

La **selección o restricción** (σ) en álgebra relacional permite extraer únicamente las filas que cumplen una condición específica.

La operación $\sigma_s(\text{condición})$ se traduce de la siguiente manera:

```
SELECT * FROM s WHERE condición;
```

Por ejemplo, la restricción que selecciona los pedidos con el número de pedido igual a 100, representada en álgebra relacional como $\text{PED} = \sigma_{\text{NUMPED} = 100}(\text{PEDIDOS})$ se traduce a SQL de la siguiente manera:

```
SELECT * FROM PEDIDOS WHERE NUMPED = 100;
```

Proyección (π)

La **proyección** (π) es una operación del álgebra relacional que permite **extraer columnas específicas de una relación**, generando así una nueva relación que contiene únicamente los atributos seleccionados. Este operador es útil cuando se desea obtener una vista reducida de los datos, centrándose solo en los atributos relevantes.

La notación es $R_x = \pi(A_1, A_2, \dots, A_n)(R)$, donde A_1, A_2, \dots, A_n son los atributos a proyectar, R es la relación de la cual se realiza la proyección y R_x es la relación resultante.

La operación de proyección no solo reduce la cantidad de datos visualizados, sino que también elimina duplicados en los resultados, asegurando que cada combinación de valores de las columnas proyectadas sea única.

Por ejemplo, dada la tabla `CLIENTES`, si queremos extraer únicamente los nombres y las regiones de los clientes, aplicaríamos la proyección de la siguiente manera:

- **Tabla** `CLIENTES`

ID	Nombre	Región
1	Juan Pérez	Centro
2	Ana Gómez	Oeste
3	Luis Rodríguez	Centro
4	Marta López	Este
5	Carlos Ruiz	Oeste
6	Laura Sánchez	Centro

- **Resultado:** `CLIENTES_NOMBRES_REGIONES = $\pi(\text{Nombre}, \text{Región})(\text{CLIENTES})$`

Nombre	Región
Juan Pérez	Centro
Ana Gómez	Oeste

Nombre	Región
Luis Rodríguez	Centro
Marta López	Este
Carlos Ruiz	Oeste
Laura Sánchez	Centro

Traducción a SQL

La **proyección** (π) en álgebra relacional tiene por objetivo eliminar las columnas no necesarias. En SQL, esto se realiza enumerando únicamente aquellas columnas requeridas en la instrucción `SELECT`.

Por ejemplo, la proyección sobre una tabla de clientes: π CLIENTES(NUM, NOMBRE, PROVINCIA) se traduce a SQL de la siguiente manera:

```
SELECT NUM, NOMBRE, PROVINCIA FROM CLIENTES;
```

Las **proyecciones de agrupación** se pueden efectuar con la ayuda de dos posibles sintaxis:

- `SELECT DISTINCT {* | lista de columnas} FROM tabla;`
- `SELECT lista de columnas FROM tabla GROUP BY lista de columnas;`

La primera sintaxis (`DISTINCT`) permite mostrar solo una fila en caso de que la consulta devuelva varias filas idénticas, mostrando así solo valores únicos.

La segunda sintaxis (`GROUP BY`) se utiliza cuando se desea realizar una proyección de grupo, calculando valores agregados sobre las filas agrupadas. Agrupar primero las filas permite aplicar funciones agregadas, aunque generalmente implica un mayor uso de recursos del servidor.

Unión (\cup)

La **unión** (\cup) es una operación del álgebra relacional que permite **combinar dos relaciones** para obtener una nueva relación que contenga **todos los elementos distintos de ambas**, eliminando duplicados. Esta operación solo es válida para relaciones que tienen la misma estructura, es decir, deben tener el mismo número de columnas (mismo grado) y los mismos dominios para cada columna.

La notación es $R_x = R_1 \cup R_2$, donde R_x es la nueva relación resultante.

La operación de unión es útil en la manipulación de bases de datos para consolidar información de diferentes fuentes o categorías. Por ejemplo, puede ser utilizada para combinar listas de clientes de diferentes regiones o periodos de tiempo en una sola lista.

Por ejemplo, dadas dos tablas diferentes `CLIENTES_OESTE` y `CLIENTES_CENTRO`, la unión puede representarse como `CLIENTES = CLIENTES_OESTE \cup CLIENTES_CENTRO` y contendrá todas las filas de ambas tablas, eliminando los duplicados y proporcionando un conjunto de clientes que pertenecen a cualquiera de las dos regiones.

- Tabla** `CLIENTES_OESTE`

ID	Nombre	Región
1	Juan	Oeste
2	María	Oeste
3	Pedro	Oeste

- **Tabla** CLIENTES_CENTRO

ID	Nombre	Región
3	Pedro	Centro
4	Laura	Centro
5	Elena	Centro

- **Resultado:** CLIENTES = CLIENTES_OESTE \cup CLIENTES_CENTRO

ID	Nombre	Región
1	Juan	Oeste
2	María	Oeste
3	Pedro	Oeste
3	Pedro	Centro
4	Laura	Centro
5	Elena	Centro

Otra aplicación sería, por ejemplo, utilizar la selección sobre una misma tabla. Como ya se ha comentado, aplicar un operador de álgebra relacional genera una tabla resultante que puede ser utilizada. Por lo tanto, aplicar dos selecciones sobre una tabla genera dos tablas resultantes.

Dada la tabla CLIENTES queremos seleccionar los clientes de la región "Centro" y "Oeste":

- **Tabla** CLIENTES

ID	Nombre	Región
1	Juan Pérez	Centro
2	Ana Gómez	Oeste
3	Luis Rodríguez	Centro
4	Marta López	Este
5	Carlos Ruiz	Oeste
6	Laura Sánchez	Centro

- **Resultado:** CLIENTES = $\sigma(\text{región} = \text{"Centro"})$ (CLIENTES) \cup $\sigma(\text{región} = \text{"Oeste"})$ (CLIENTES)

ID	Nombre	Región
1	Juan Pérez	Centro
2	Ana Gómez	Oeste
3	Luis Rodríguez	Centro
5	Carlos Ruiz	Oeste
6	Laura Sánchez	Centro

Traducción a SQL

Esta operación permite combinar los resultados de dos o más consultas, obteniendo un conjunto de filas que cumplen con el mismo formato (mismo nombre de columnas, mismo tipo y en el mismo orden).

La operación `S U T` se expresa de la siguiente manera:

```
SELECT lista de columnas FROM S {UNION / UNION ALL} SELECT lista de columnas FROM T;
```

El operador `UNION` elimina las filas duplicadas del resultado, mientras que `UNION ALL` permite extraer todas las filas procedentes de las consultas, incluyendo las duplicadas.

Supongamos que tenemos dos tablas `CLIENTES_2022` y `CLIENTES_2023`, y queremos obtener la lista de todos los clientes de ambos años:

```
SELECT nombre FROM CLIENTES_2022
UNION
SELECT nombre FROM CLIENTES_2023;
```

Si usamos `UNION ALL`, obtendríamos todos los nombres, incluyendo los que están duplicados en ambas tablas:

```
SELECT nombre FROM CLIENTES_2022
UNION ALL
SELECT nombre FROM CLIENTES_2023;
```

Intersección (\cap)

La **intersección (\cap)** es una operación del álgebra relacional que permite obtener el conjunto de **elementos que son comunes a dos relaciones**. Esta operación solo es válida para relaciones que tienen la misma estructura, es decir, deben tener el mismo número de columnas (mismo grado) y los mismos dominios para cada columna.

La notación es `Rx = R1 \cap R2`, donde `Rx` es la nueva relación resultante.

La operación de intersección es útil en la manipulación de bases de datos para encontrar elementos que comparten características específicas en dos conjuntos diferentes. Por ejemplo, puede ser utilizada para identificar clientes que han realizado compras en dos diferentes periodos o categorías.

Es importante notar que, aunque la intersección es una operación válida y útil en álgebra relacional, algunos textos mencionan que su función se puede expresar **usando el operador diferencia (-)**. Es decir, si queremos encontrar los elementos comunes

entre dos relaciones, podemos calcularlo utilizando la diferencia de cada relación con la unión de ambas, lo que puede hacer que la intersección sea conceptualmente redundante en algunos contextos.

La fórmula para expresar esto sería: $R1 \cap R2 = R1 - (R1 - R2)$

Como ejemplo de una intersección (\cap), dadas las tablas `CLIENTES_OESTE` y `CLIENTES_CENTRO` la intersección puede representarse como `CLIENTES = CLIENTES_OESTE \cap CLIENTES_CENTRO` y contendrá todas las filas de ambas tablas, eliminando los duplicados y proporcionando un conjunto de clientes que pertenecen a las dos regiones.

• **Tabla** `CLIENTES_OESTE`

ID	Nombre	Región
1	Juan	Gijón
2	María	Zamora
3	Pedro	Mérida

• **Tabla** `CLIENTES_CENTRO`

ID	Nombre	Región
3	Pedro	Mérida
4	Laura	Madrid
5	Elena	Toledo

• **Resultado:** `CLIENTES = CLIENTES_OESTE \cap CLIENTES_CENTRO`

ID	Nombre	Región
3	Pedro	Mérida

Traducción a SQL

Esta operación permite obtener filas que están presentes en ambas consultas, siempre que estas tengan el mismo formato (mismo nombre de columnas, mismo tipo y en el mismo orden).

La operación `S \cap T` se expresa de la siguiente manera:

```
SELECT lista de columnas FROM S INTERSECT SELECT lista de columnas FROM T;
```

Por ejemplo, si tenemos dos tablas `CLIENTES_2022` y `CLIENTES_2023`, y queremos obtener los clientes que han estado en ambas tablas, podríamos usar:

```
SELECT nombre FROM CLIENTES_2022 INTERSECT SELECT nombre FROM CLIENTES_2023;
```

Diferencia (–)

La **diferencia** ($-$) es una operación del álgebra relacional que permite obtener una nueva relación **al eliminar de una relación (R1) los elementos que también están presentes en otra relación (R2)**. Esta operación solo es válida para relaciones que tienen la misma estructura, es decir, deben tener el mismo número de columnas (mismo grado) y los mismos dominios para cada columna.

La notación es $R_x = R_1 - R_2$, donde R_x es la nueva relación resultante.

La operación de diferencia es útil para identificar elementos que son exclusivos de una relación y que no están presentes en otra. Por ejemplo, puede ser utilizada para encontrar clientes que no han realizado compras en un periodo determinado.

Por ejemplo, dada la tabla `CLIENTES`, queremos seleccionar aquellos clientes existentes en el año 2022 pero que ya no son clientes en 2023:

• **Tabla** `CLIENTES`

ID	Nombre	Región	Año
1	Juan Pérez	Centro	2022
2	Ana Gómez	Oeste	2022
3	Luis Rodríguez	Centro	2022
4	Marta López	Este	2022
3	Luis Rodríguez	Oeste	2023
6	Laura Sánchez	Centro	2023
7	Elena Torres	Oeste	2023

• **Resultado:** `CLIENTES = $\sigma(\text{año} = 2022)(\text{CLIENTES}) - \sigma(\text{año} = 2023)(\text{CLIENTES})$`

ID	Nombre	Región	Año
1	Juan Pérez	Centro	2022
2	Ana Gómez	Oeste	2022
4	Marta López	Este	2022

En 2022, los clientes con ID 1, 2 y 4 están registrados, mientras que en 2023 no están presentes, lo que los convierte en clientes exclusivos de 2022.

Traducción a SQL

Esta operación permite obtener filas que están en una consulta pero no en otra, siempre que ambas consultas usen el mismo formato (mismo nombre de columnas, mismo tipo y en el mismo orden).

La operación $S - T$ se expresa de la siguiente manera:

```
SELECT lista de columnas FROM S MINUS SELECT lista de columnas FROM T;
```

Por ejemplo, si tenemos dos tablas `CLIENTES_2022` y `CLIENTES_2023`, y queremos obtener los clientes que solo estuvieron en 2022, podríamos usar:

```
SELECT nombre FROM CLIENTES_2022 MINUS SELECT nombre FROM CLIENTES_2023;
```

Producto cartesiano (×)

El **producto cartesiano (×)** es una operación del álgebra relacional que permite **combinar todas las tuplas de una relación (R1) con todas las tuplas de otra relación (R2)**, generando una nueva relación con todas las combinaciones posibles de tuplas de ambas relaciones. El resultado es una relación que tiene el número total de columnas de R1 más las columnas de R2.

La notación es $R_x = R_1 \times R_2$, donde R_x es la nueva relación resultante de combinar las dos relaciones.

El producto cartesiano es útil cuando queremos combinar todas las posibles combinaciones de dos relaciones, pero en la práctica, suele ir acompañado de una operación selección para reducir el conjunto de resultados. Es una de las operaciones fundamentales que puede llevar a un `join` si luego se aplica una condición de filtrado.

Por ejemplo, dada la tabla `CLIENTES` y la tabla `PEDIDOS`, podemos generar todas las combinaciones posibles de clientes con sus pedidos:

- **Tabla** `CLIENTES`

ID	Nombre	Región
1	Juan Pérez	Centro
2	Ana Gómez	Oeste
3	Luis Rodríguez	Este

- **Tabla** `PEDIDOS`

Pedido_ID	Producto
101	Laptop
102	Smartphone

- **Resultado:** `PEDIDOS_CLIENTES = CLIENTES × PEDIDOS`

ID	Nombre	Región	Pedido_ID	Producto
1	Juan Pérez	Centro	101	Laptop
1	Juan Pérez	Centro	102	Smartphone
2	Ana Gómez	Oeste	101	Laptop
2	Ana Gómez	Oeste	102	Smartphone
3	Luis Rodríguez	Este	101	Laptop
3	Luis Rodríguez	Este	102	Smartphone

Traducción a SQL

El producto cartesiano permite **asociar cada fila de una tabla con cada fila de otras tablas**, generando todas las combinaciones posibles de filas. Esto puede resultar en un número de filas igual al producto del número de filas de las tablas

involucradas.

La operación $R \times T$ se expresa en SQL de la siguiente manera:

```
SELECT lista de columnas FROM S, T;
```

Si las columnas tienen el mismo nombre en ambas tablas, el nombre de la columna va precedido por el nombre de la tabla. Por lo tanto, el nombre completo de la columna es `nombretabla.nombrecolumna`.

Si tenemos dos tablas, `AULAS` y `PROFESORES`, el producto cartesiano podría verse así:

```
SELECT AULAS.nombre_aula, PROFESORES.nombre_profesor FROM AULAS, PROFESORES;
```

Esto devolvería una lista que combina cada aula con cada profesor, generando así todas las posibles combinaciones.

Combinaciones

La **combinación (JOIN)** es una operación del álgebra relacional que permite **combinar filas de dos relaciones basándose en una condición**. Esta operación une las filas de las dos tablas donde los valores de las columnas especificadas cumplen una condición determinada.

Existen varios tipos de combinaciones, siendo las más comunes: **combinación natural**, **combinación interna (inner join)** y **combinación externa (outer join)**.

- **Combinación interna (INNER JOIN)**: Devuelve las filas donde existe una coincidencia en ambas tablas, es decir, solo las filas que cumplen la condición especificada.

La notación para una combinación interna es $R1 \bowtie \text{condición } R2$.

- **Combinación externa (OUTER JOIN)**: Devuelve todas las filas de una tabla, y cuando no hay coincidencia en la otra tabla, los valores de esta se rellenan con NULL. Hay tres tipos principales:
 - **LEFT JOIN**: Devuelve todas las filas de la primera tabla y las coincidencias de la segunda, si las hay.
 - **RIGHT JOIN**: Devuelve todas las filas de la segunda tabla y las coincidencias de la primera.
 - **FULL JOIN**: Devuelve todas las filas de ambas tablas, con NULL donde no haya coincidencias.
- **Combinación natural (NATURAL JOIN)**: Une dos tablas automáticamente basándose en las columnas con el mismo nombre y tipo de dato.

Traducción a SQL

La combinación es una restricción sobre el producto cartesiano que vincula cada fila de una tabla con filas de otra tabla, de acuerdo a una condición dada. Existen varios tipos de combinaciones que se utilizan para extraer datos relacionados de diferentes tablas:

- **Combinación interna (INNER JOIN)**: retorna solo las filas que tienen coincidencias en ambas tablas.

La operación `S INNER JOIN T ON (condición)` se expresa en SQL de la siguiente manera:

```
SELECT lista de columnas FROM S INNER JOIN T ON condición;
```

- **Combinación externa (`OUTER JOIN`)**: devuelve todas las filas de una tabla y las filas coincidentes de otra tabla. Si no hay coincidencias, se llenan con valores nulos (`NULL`).
 - **Combinación externa izquierda (`LEFT OUTER JOIN`)**: retorna todas las filas de la tabla izquierda (S) y las filas coincidentes de la tabla derecha (T).

```
SELECT lista de columnas FROM S LEFT OUTER JOIN T ON condición;
```

- **Combinación externa derecha (`RIGHT OUTER JOIN`)**: retorna todas las filas de la tabla derecha (T) y las filas coincidentes de la tabla izquierda (S).

```
SELECT lista de columnas FROM S RIGHT OUTER JOIN T ON condición;
```

- **Combinación externa completa (`FULL OUTER JOIN`)**: retorna todas las filas de ambas tablas, con coincidencias donde existan, y valores nulos donde no haya coincidencias.

```
SELECT lista de columnas FROM S FULL OUTER JOIN T ON condición;
```

- **Combinación natural (`NATURAL JOIN`)**: se basa en columnas con el mismo nombre en ambas tablas y combina automáticamente las filas donde estas columnas coinciden.

La operación `S NATURAL JOIN T` se expresa en SQL de la siguiente manera:

```
SELECT lista de columnas FROM S NATURAL JOIN T;
```

Campos calculados elementales

Proyección sobre una relación asociada a un **cálculo** que se realiza sobre cada línea para **crear uno o varios atributos nuevos**.

La notación es $R_x = \pi S (A_1, \dots, N_1 = \text{expresión calculada} \dots)$, donde π (pi) indica la proyección y se aplica un cálculo a cada fila para generar nuevas columnas.

La expresión calculada puede ser:

- una operación aritmética
- una función matemática
- una función de cadena

Por ejemplo, podemos calcular el total de un pedido multiplicando el precio unitario por la cantidad pedida: `TOTAL_PEDIDO = π S PEDIDO (PEDIDONUM, REF, LINEA = PRECIO * CANTIDAD)`

Traducción a SQL

Los campos calculados elementales permiten obtener columnas calculadas para cada fila.

En SQL, los campos calculados se generan utilizando la cláusula `SELECT` con operaciones. La traducción de la operación $\pi S (A_1, \dots, N_1 = \text{expresión calculada} \dots)$ se traduce como:

```
SELECT col, ..., expresión FROM S;
```

Por ejemplo, podemos calcular el valor de los productos de un almacén:

```
ALMACEN =  $\pi$  ARTICULOS(REF, DESCRIPCION, COSTO = (PRECIO * VALOR))
```

Esto se puede traducir a SQL como:

```
SELECT REF, DESCRIPCION, (PRECIO * VALOR) FROM ARTICULOS;
```

Para asignar un nombre más claro a la columna calculada `(PRECIO * VALOR)`, se utiliza el alias `AS` :

```
SELECT REF, DESCRIPCION, (PRECIO * VALOR) AS VALOR_TOTAL FROM ARTICULOS;
```

Campos de valores agregados

Proyección sobre una relación asociada con uno o varios valores agregados que se calculan sobre un atributo para todos los elementos de la relación o de la agrupación vinculada a la proyección, con el fin de **crear uno o varios atributos nuevos**.

La notación es $R_x = \pi S (A_1, \dots, N_1 = \text{función estadística}(A_x), \dots)$

Las funciones estadísticas son:

- **COUNT(*)**
- **COUNT(atributo)**
- **SUM(atributo)**
- **MAX(atributo)**
- **MIN(atributo)**

Por ejemplo, podemos calcular el número de clientes que hay en una tabla `NUMCLIENTES = π CLIENTES(N = COUNT(*))` . También podemos calcular el precio más alto, el más bajo y el precio medio por categoría de artículos `STATS = π ARTICULOS(CATEGORIA, CARO=MAX(PRECIO), BARATO=(PRECIO), MEDIO=(PRECIO))` .

Traducción a SQL

Las **proyecciones del cálculo de valores agregados** permiten realizar cálculos estadísticos sobre los grupos especificados mediante `GROUP BY` .

La operación $\pi S (col, \dots, nvcol = \text{cálculo estadístico})$ se traduce a SQL de la siguiente manera:

```
SELECT lista de columnas, función_agregada FROM S GROUP BY lista de columnas;
```

La lista de columnas proyectadas debe ser **idéntica** a la lista de columnas agrupadas con `GROUP BY` .

Si en la consulta se utilizan expresiones calculadas en las columnas del `SELECT` , la agrupación debe realizarse sobre las mismas expresiones. Por ejemplo, si se calcula el valor total de ventas con `precio * cantidad` , la agrupación también debe incluir

`precio * cantidad` en lugar de solo precio o cantidad, para que los resultados reflejen correctamente los cálculos proyectados.

```
SELECT producto, precio * cantidad AS total_venta
FROM ventas
GROUP BY producto, precio * cantidad;
```

Referencias a las funciones de agregación en diferentes SGBDR:

- [Aggregate Functions - MySQL](#)
- [Aggregate Functions - PostgreSQL](#)
- [Aggregate Functions - Oracle Database](#)

SQL

Select

La sentencia `SELECT` se utiliza para seleccionar datos de una base de datos.

```
SELECT column1, column2, ... FROM table_name;
```

Para devolver todas las columnas sin especificar cada nombre de columna, se utiliza el asterisco (`*`) con el `SELECT` :

```
SELECT * FROM table_name;
```

La sentencia `SELECT DISTINCT` se utiliza para devolver solo valores distintos (diferentes). Dentro de una tabla, una columna a menudo contiene muchos valores duplicados y en ocasiones sólo se necesita listar los valores diferentes (distintos).

```
SELECT DISTINCT column1, column2, ... FROM table_name;
```

Al utilizar la palabra clave `DISTINCT` en una función `COUNT()` , podemos devolver el número de registros diferentes:

```
SELECT COUNT(DISTINCT column_name) FROM table_name;
```

Esta sentencia no funciona en Microsoft Acces. Una solución alternativa sería:

```
SELECT COUNT(*) AS column_name FROM (SELECT DISTINCT column_name FROM table_name);
```

Where

La cláusula `WHERE` se utiliza para filtrar registros, extrayendo aquellos registros que cumplen una condición específica.

```
SELECT column1, column2, ... FROM table_name WHERE condition;
```


La cláusula `WHERE` no solo se utiliza en sentencias `SELECT` sino que también se emplea en sentencias `UPDATE` , `DELETE` , etc.

Text Fields vs. Numeric Fields

SQL requiere **comillas simples alrededor de los valores de texto** (la mayoría de los sistemas de bases de datos también permiten comillas dobles).

Sin embargo, **los campos numéricos no deben ir entre comillas**.

```
-- Usando comillas simples
SELECT * FROM clientes WHERE nombre = 'Carlos';

-- Usando comillas dobles (si el sistema lo permite)
SELECT * FROM clientes WHERE nombre = "Carlos";

-- Valor numérico sin comillas
SELECT * FROM productos WHERE precio = 100;

-- Incorrecto: el valor numérico entre comillas puede causar errores inesperados
SELECT * FROM productos WHERE precio = '100';
```

Operators

Los operadores se utilizan para filtrar la búsqueda.

- **Igual que (=)**

```
SELECT * FROM Products WHERE Price = 18;
```

- **Mayor o menor que (> , <)**

```
SELECT * FROM productos WHERE precio > 50;
```

- **Mayor o igual que (>=), menor o igual que (<=)**

```
SELECT * FROM empleados WHERE antigüedad >= 5;
```

- **Distinto de (<> o !=)**

```
SELECT * FROM pedidos WHERE estado <> 'completado';
```

- **Operador LIKE (para búsquedas de patrones)**

```
SELECT * FROM clientes WHERE nombre LIKE 'An%';
```

- **Operador BETWEEN (para rangos)**

```
SELECT * FROM productos WHERE precio BETWEEN 100 AND 200;
```

- Operador `IN` (para listas de valores)

```
SELECT * FROM empleados WHERE departamento_id IN (1, 2, 3);
```

Order BY

La palabra clave `ORDER BY` se utiliza para ordenar el conjunto de resultados en orden ascendente o descendente.

```
SELECT column1, column2, ... FROM table_name ORDER BY column1, column2, ... ASC|DESC;
```

En una cláusula de ejemplo como `ORDER BY column1, column2 ASC;`, el orden de los resultados funcionaría de la siguiente manera:

1. Primero, se ordena por `column1`. Los registros se ordenan en orden ascendente (de menor a mayor) según los valores en `column1`.
2. Luego, se ordena por `column2` en caso de empate. Si hay registros que tienen el mismo valor en `column1`, entonces esos registros se ordenarán según los valores de `column2`, también en orden ascendente (de menor a mayor).

Otro cláusula de ejemplo como `ORDER BY column1 DESC, column2 ASC;` funcionaría de la misma manera, cambiando únicamente si es ascendente o descendente.

La palabra clave `ORDER BY` ordena los registros **en orden ascendente por defecto**. Para ordenar los registros en orden descendente, se debe usar la palabra clave `DESC`.

Para los valores de texto, la palabra clave `ORDER BY` ordenará **alfabéticamente**.

And

El operador `AND` se utiliza para filtrar registros en función de dos o más condiciones. Se mostrará el registro **si todas las condiciones son verdaderas**.

```
SELECT column1, column2, ... FROM table_name WHERE condition1 AND condition2 AND condition3 ...;
```

Los operadores `AND` y `OR` se pueden combinar en la misma consulta. En ese caso se puede usar paréntesis para agrupar distintas cláusulas.

Or

La cláusula `WHERE` puede contener uno o más operadores `OR`.

Este operador se utiliza para filtrar registros en función de dos o más condiciones. Se mostrará el registro **si alguna de las condiciones es verdadera**.

```
SELECT column1, column2, ... FROM table_name WHERE condition1 OR condition2 OR condition3 ...;
```

Los operadores `AND` y `OR` se pueden combinar en la misma consulta. En ese caso se puede usar paréntesis para agrupar distintas cláusulas.

Not

El operador `NOT` se utiliza en combinación con otros operadores para obtener el resultado opuesto, también llamado resultado negativo.

```
SELECT column1, column2, ... FROM table_name WHERE NOT condition;
```

- `NOT LIKE`

```
-- Select customers that does not start with the Letter 'A':  
SELECT * FROM Customers WHERE CustomerName NOT LIKE 'A%';
```

- `NOT BETWEEN`

```
-- Select customers with a customerID not between 10 and 60:  
SELECT * FROM Customers WHERE CustomerID NOT BETWEEN 10 AND 60;
```

- `NOT IN`

```
-- Select customers that are not from Paris or London:  
SELECT * FROM Customers WHERE City NOT IN ('Paris', 'London');
```

- `NOT Greater Than`

```
-- Select customers with a CustomerId not greater than 50:  
SELECT * FROM Customers WHERE NOT CustomerID > 50;
```

- `NOT Less Than`

```
-- Select customers with a CustomerID not Less than 50:  
SELECT * FROM Customers WHERE NOT CustomerId < 50;
```

Insert Into

La instrucción `INSERT INTO` se utiliza para insertar nuevos registros en una tabla. Es posible escribir la instrucción `INSERT INTO` de varias maneras:

- De **forma declarativa**, especificando tanto los nombres de las columnas como los valores a insertar. Es la forma menos propensa a errores:

```
INSERT INTO table_name (column1, column2, column3, ...) VALUES (value1, value2, value3, ...);
```

- **De forma implícita**, ya que si se están agregando valores para **todas las columnas** de la tabla, no es necesario especificar los nombres de las columnas en la consulta SQL. Sin embargo, hay que asegurarse de que el orden de los valores sea el mismo que el de las columnas en la tabla. En este caso, la sintaxis de `INSERT INTO` sería la siguiente:

```
INSERT INTO table_name VALUES (value1, value2, value3, ...);
```

- **Mediante una subconsulta** `SELECT` , para insertar registros desde otra tabla o consulta. Esto permite copiar datos de una fuente existente o generar datos dinámicamente. En este caso, la sintaxis de `INSERT INTO` sería la siguiente:

```
INSERT INTO table_name (column1, column2, column3, ...)
SELECT value1, value2, value3, ... FROM other_table WHERE condition;
```

También es posible insertar **varias filas en una sola instrucción**. Para insertar múltiples filas de datos, se utiliza la misma instrucción `INSERT INTO` , pero con varios conjuntos de valores:

```
INSERT INTO table_name (column1, column2, column3, ...)
VALUES (value1_1, value2_1, value3_1, ...),
       (value1_2, value2_2, value3_2, ...),
       (value1_3, value2_3, value3_3, ...);
```

Hay que asegurarse de separar cada conjunto de valores con una coma (,).

Null Values

Un campo con un valor `NULL` es un **campo sin valor**.

Si un campo en una tabla es opcional, es posible insertar un nuevo registro o actualizar un registro sin agregar un valor a este campo. En ese caso, el campo se guardará con un valor `NULL` .

Un valor `NULL` es diferente de un valor cero o de un campo que contiene espacios. Un campo con un valor `NULL` es aquel que ha sido dejado en blanco durante la creación del registro.

No es posible probar los valores `NULL` con operadores de comparación, como `=` , `<` o `<>` . En su lugar, se debe usar los operadores `IS NULL` e `IS NOT NULL` .

```
-- IS NULL Syntax
SELECT column_names FROM table_name WHERE column_name IS NULL;

-- IS NOT NULL Syntax
SELECT column_names FROM table_name WHERE column_name IS NOT NULL;
```

Update

La instrucción `UPDATE` se utiliza para modificar los registros existentes en una tabla.

```
UPDATE table_name SET column1 = value1, column2 = value2, ... WHERE condition;
```

Hay que tener **especial atención** al actualizar registros en una tabla. La cláusula `WHERE` especifica qué registro(s) deben ser actualizados. Si se omite la cláusula `WHERE` , ¡todos los registros en la tabla serán actualizados!

Delete

La instrucción `DELETE` se utiliza para eliminar registros existentes en una tabla.

```
DELETE FROM table_name WHERE condition;
```

Hay que tener **especial atención** al eliminar registros en una tabla. La cláusula `WHERE` especifica qué registro(s) deben ser eliminados. Si se omite la cláusula `WHERE`, ¡todos los registros en la tabla serán eliminados!

Filter

Cuando trabajamos con bases de datos, a menudo necesitamos **limitar el número de registros** que una consulta `SELECT` devuelve. Esto es útil cuando solo queremos obtener una muestra de los datos o cuando estamos trabajando con tablas grandes y queremos optimizar el rendimiento.

En SQL, se pueden utilizar diferentes cláusulas para controlar la cantidad de registros devueltos, como `LIMIT` en MySQL o `SELECT TOP` en SQL Server. A continuación, veremos cómo se aplica en los principales sistemas de gestión de bases de datos:

- **SQL Server / MS Access Syntax:**

```
SELECT TOP number|percent column_name(s) FROM table_name WHERE condition;

-- With ORDER BY
SELECT TOP number column_name(s) FROM table_name WHERE condition ORDER BY column_name ASC|DESC;
```

- **MySQL | PostgreSQL Syntax:**

```
SELECT column_name(s) FROM table_name WHERE condition LIMIT number;

-- With ORDER BY
SELECT column_name(s) FROM table_name WHERE condition ORDER BY column_name(s) ASC|DESC LIMIT number;
```

- **Oracle 12 Syntax:**

```
SELECT column_name(s) FROM table_name ORDER BY column_name(s) FETCH FIRST number ROWS ONLY;
```

- **Older Oracle Syntax:**

```
SELECT column_name(s) FROM table_name WHERE ROWNUM <= number;
```

- **Older Oracle Syntax (with ORDER BY):**

```
SELECT * FROM (SELECT column_name(s) FROM table_name ORDER BY column_name(s)) WHERE ROWNUM <= number;
```

Aggregate Functions

TODO

Data Types

- [MySQL](#)

- [PostgreSQL](#)
- [Oracle Database](#)

Functions

- [MySQL](#)
- [PostgreSQL](#)
- [Oracle Database](#)

Práctica

MySQL en Docker con persistencia

Creación del contenedor

```
# Crear el contenedor
docker run --name mysql-container -e MYSQL_ROOT_PASSWORD=tu_contraseña -e MYSQL_DATABASE=mi_base_de_datos -v mysql_data:

# Comprobar que el contenedor está corriendo
docker ps
```

- `--name mysql-container` : Nombre del contenedor.
- `-e MYSQL_ROOT_PASSWORD=tu_contraseña` : Establece la contraseña para el usuario root.
- `-e MYSQL_DATABASE=mi_base_de_datos` : Crea una base de datos inicial llamada `mi_base_de_datos`.
- `-v mysql_data:/var/lib/mysql` : Crea un volumen llamado `mysql_data` para persistir los datos.
- `-p 3306:3306` : Expone el puerto 3306 (puerto predeterminado de MySQL) para conectarte desde tu máquina.
- `-d mysql:latest` : Utiliza la imagen más reciente de MySQL y ejecuta el contenedor en segundo plano.

Conectarse al contenedor

1. Conectarse desde cualquier cliente MySQL usando `localhost:3306` , el usuario `root` y la contraseña configurada.
2. Conectarse al contenedor de MySQL desde la terminal:

```
docker exec -it mysql-container mysql -uroot -p
```

- `docker exec` : Ejecuta un comando en un contenedor en ejecución.
- `-it` : Permite la interacción (interactiva) y asigna un pseudo-TTY.
- `mysql-container` : Nombre del contenedor que especificaste al crear el contenedor.
- `mysql` : Llama al cliente de MySQL.

- `-uroot` : Especifica que te conectas con el usuario root.
- `-p` : Te pedirá que ingreses la contraseña del usuario root (la que configuraste al crear el contenedor).

Reiniciar y parar el contenedor

```
# Reiniciar el contenedor
docker start mysql-container

# Parar el contenedor
docker stop mysql-container
```

PostgreSQL en Docker con persistencia

Creación del contenedor

```
docker run --name postgres-container -e POSTGRES_PASSWORD=tu_contraseña -e POSTGRES_DB=mi_base_de_datos -v postgres_data:
```

- `--name postgres-container` : Nombre del contenedor.
- `-e POSTGRES_PASSWORD=tu_contraseña` : Establece la contraseña para el usuario postgres.
- `-e POSTGRES_DB=mi_base_de_datos` : Crea una base de datos inicial llamada `mi_base_de_datos`.
- `-v postgres_data:/var/lib/postgresql/data` : Crea un volumen llamado `postgres_data` para persistir los datos.
- `-p 5432:5432` : Expone el puerto 5432 (puerto predeterminado de PostgreSQL) para conectarte desde tu máquina.
- `-d postgres:latest` : Utiliza la imagen más reciente de PostgreSQL y ejecuta el contenedor en segundo plano.

Conectarse al contenedor

Es posible conectarse al contenedor vía el cliente `psql` de PostgreSQL:

```
docker exec -it postgres-container psql -U postgres
```

Reiniciar y parar el contenedor

```
# Reiniciar el contenedor
docker start postgres-container

# Parar el contenedor
docker stop postgres-container
```

Referencias

MySQL

- <https://dev.mysql.com/doc/>

- <https://cheatsheets.zip/mysql>

PostgreSQL

- <https://www.postgresql.org/docs/>
- <https://wiki.postgresql.org/>
- <https://cheatsheets.zip/postgres>

Microsoft SQL

- <https://learn.microsoft.com/es-es/sql>

Oracle SQL

- <https://docs.oracle.com/en/database/oracle/oracle-database/index.html>

General

- <https://roadmap.sh/sql>
- <https://www.w3schools.com/sql/default.asp>
- <https://www.sqltutorial.org/>
- <https://www.sqlitetutorial.net/>
- <https://datalemur.com/sql-tutorial>
- <https://github.com/XD-DENG/SQL-exercise>
- https://www.sqlzoo.net/wiki/SQL_Tutorial
- <https://sqlbolt.com/>
- <https://bookdown.org/paranedagarcia/database/sql.html>
- <https://gestionbasesdatos.readthedocs.io/es/latest/index.html>

Licencia



Esta obra está bajo una [licencia de Creative Commons Reconocimiento-Compartir Igual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).