**ETL con SQL (solución)**

1. Crear una tabla llamada **clientes** con las siguientes columnas: id\_empleado, nombre, apellido y departamento.
2. Insertar en la tabla los sig. datos:
3. (1, 'Juan', 'Pérez', 'Ventas'),
4. (2, 'Ana', 'García', 'Finanzas'),
5. (3, 'Luis', 'Martínez', 'Marketing'),
6. (4, 'Sofía', 'Rodríguez', 'Ventas'),
7. (5, 'Juan', 'Pérez', 'Ventas')
8. (6,’nombre\_alumno/a’,’apellido\_alumno/a’,’Externo’);
9. Crear un código para identificar datos duplicados.
10. Eliminar filas duplicadas utilizando ROW\_NUMBER().
11. Eliminar registros duplicados con la cláusula DELETE y JOIN.

--1. Crear una tabla de empleados--

CREATE TABLE empleados (

id\_empleado INT NOT NULL,

nombre varchar(50),

apellido varchar(50),

departamento varchar(50),

PRIMARY KEY (id\_empleado)

);

--2. Insertar datos en la tabla empleado--

INSERT INTO empleados (id\_empleado,nombre,apellido,departamento)

VALUES (1, 'Juan', 'Pérez', 'Ventas'),

(2, 'Ana', 'García', 'Finanzas'),

(3, 'Luis', 'Martínez', 'Marketing'),

(4, 'Sofía', 'Rodríguez', 'Ventas'),

(5, 'Juan', 'Pérez', 'Ventas');

--3. Identificar registros duplicados usando la clausula GROUP BY y HAVING

SELECT nombre, apellido, departamento, COUNT(\*)

FROM empleados

GROUP BY nombre, apellido, departamento

HAVING COUNT(\*) > 1;

--4. Eliminar duplicados usando la función ROW\_NUMBER()--

WITH tabla\_cte AS (

SELECT \*, ROW\_NUMBER() OVER (PARTITION BY nombre, apellido, departamento ORDER BY id\_empleado) AS fila\_numero

FROM empleados

)

DELETE FROM tabla\_cte

WHERE fila\_numero > 1;

--Verificar si se eliminaron duplicados--

SELECT\*FROM empleados

--5. Otra forma de eliminar registros duplicados

--volver a ingresar registro duplicado--

INSERT INTO empleados (id\_empleado, nombre, apellido, departamento)

VALUES (5, 'Juan', 'Pérez', 'Ventas');

SELECT\*FROM empleados

--Eliminar registros duplicados mediante clausulas DELETE y JOIN

DELETE e1

FROM empleados e1 --intersección del conjunto empleados

INNER JOIN empleados e2

ON e1.nombre = e2.nombre AND e1.apellido = e2.apellido AND e1.departamento = e2.departamento

WHERE e1.id\_empleado < e2.id\_empleado; --criterio para decidir cuál de los registros duplicados debe eliminarse

SELECT\*FROM empleados