

Para cargar un conjunto de datos de clientes en una base OLTP y verificar si existen registros duplicados, sigue estos pasos:

1. Carga del Conjunto de Datos

- Si los datos están en un archivo CSV, puedes importarlos usando herramientas como COPY en PostgreSQL, LOAD DATA INFILE en MySQL o BULK INSERT en SQL Server.
- Asegúrate de que la tabla tenga restricciones adecuadas como claves primarias y restricciones UNIQUE para evitar duplicados.

2. Verificación de Registros Duplicados

Para detectar duplicados en la base de datos, puedes ejecutar una consulta SQL con GROUP BY y HAVING para encontrar registros repetidos.

Consulta para encontrar duplicados por nombre y correo electrónico:

```
SELECT nombre, correo, COUNT(*) AS total_duplicados
FROM clientes
GROUP BY nombre, correo
HAVING COUNT(*) > 1;
```

◆ Explicación:

- Agrupa los registros por nombre y correo (o cualquier combinación de columnas que deba ser única).
- Cuenta cuántas veces aparece cada combinación.
- Filtra aquellos que aparecen más de una vez.

Consulta para encontrar registros con el mismo correo electrónico:

```
SELECT correo, COUNT(*) AS total_duplicados
FROM clientes
GROUP BY correo
HAVING COUNT(*) > 1;
```

Esto es útil si el correo debe ser único.

Consulta para detectar duplicados considerando todas las columnas:

Si quieres detectar duplicados exactos en todas las columnas:

```
SELECT *, COUNT(*) AS total_duplicados
FROM clientes
GROUP BY nombre, correo, telefono, direccion
HAVING COUNT(*) > 1;
```

3. Opciones para Eliminar Duplicados

Si encuentras duplicados, puedes eliminarlos manteniendo solo una fila usando ROW_NUMBER() (en bases de datos como PostgreSQL, MySQL 8+, SQL Server):

```
WITH Duplicados AS (  
  SELECT id, ROW_NUMBER() OVER (PARTITION BY nombre, correo ORDER BY id) AS fila  
  FROM clientes  
)  
DELETE FROM clientes WHERE id IN (SELECT id FROM Duplicados WHERE fila > 1);
```

◆ **Notas:**

- **Si no tienes un id único**, debes definir reglas claras para conservar solo una fila (por ejemplo, la más reciente usando ORDER BY fecha_creacion DESC).
- Antes de eliminar, revisa los duplicados con SELECT * FROM Duplicados WHERE fila > 1;.

¿Necesitas adaptar estas consultas a un caso específico en tu base OLTP?