**Índices en bases de datos**

Un **índice** es una estructura de datos que crea PostgreSQL (u otro motor de base de datos) para **acelerar** la búsqueda de registros en una tabla.

Es **similar a un índice en un libro**: en lugar de leer todo el libro (tabla) para encontrar una palabra (dato), vas al índice y directamente saltás al número de página (fila).

**¿Para qué sirven?**

* **Aceleran búsquedas** (SELECT) con WHERE, JOIN, ORDER BY, etc.
* Son especialmente útiles en **tablas grandes**, donde buscar sin índice puede ser lento.
* También mejoran el rendimiento de búsquedas semánticas si los combinas con técnicas como **embeddings**.

**¿Cómo y cuándo se usan?**

* Se **crean manualmente** con CREATE INDEX, o se crean **automáticamente** si ponés una restricción UNIQUE o PRIMARY KEY.
* Se usan cuando:
  + Hacés búsquedas frecuentes por una o varias columnas.
  + Ordenás (ORDER BY) o filtrás (WHERE) usando esas columnas.

**Ejemplo concreto para tu proyecto**:  
Si tu tabla preguntas tiene miles de registros, y buscás por texto\_embedded, fecha, o id\_autor, conviene indexar esas columnas.

***Estructura básica de un índice***

CREATE INDEX nombre\_indice

ON nombre\_tabla (columna1, columna2);

***Ejemplo sencillo***

CREATE INDEX idx\_preguntas\_texto

ON preguntas (texto);

***Después, podés hacer:***

SELECT \* FROM preguntas WHERE texto ILIKE '%discord%';

Si texto está indexada, esto va mucho más rápido.

**¿Se pueden usar desde Python?**

Sí. Y no tenés que hacer nada especial. Si un índice ya existe, **PostgreSQL lo usa automáticamente** al ejecutar un SELECT desde Python.

import psycopg2

conn = psycopg2.connect("dbname=chatbot user=lourdes password=tu\_password")

cur = conn.cursor()

cur.execute("SELECT \* FROM preguntas WHERE texto ILIKE %s", ('%discord%',))

resultados = cur.fetchall()

for r in resultados:

print(r)

cur.close()

conn.close()

**¿Cómo se diferencian de otras cosas?**

|  |  |
| --- | --- |
| Objeto | ¿Qué hace? |
| Índice | Mejora el rendimiento de consultas. |
| Procedimiento | Ejecuta tareas (ej. mantenimiento, lógica de negocio). |
| Función | Devuelve un valor o conjunto, se puede usar en queries. |
| Trigger | Reacciona automáticamente a cambios (ej. INSERT/UPDATE). |
| Vista | Resultado de una consulta que se trata como tabla virtual. |
| SELECT | Sentencia para consultar datos. Puede usar índices. |

**¿Qué más deberías saber?**

* **No todos los índices mejoran el rendimiento**. Si usás demasiados, pueden **ralentizar los INSERT y UPDATE**.
* Podés usar **índices compuestos**: ON preguntas (id\_autor, fecha).
* También existen **índices únicos** y **índices GIN** o **GiST** para búsquedas complejas (como texto o embeddings).
* Si usás ChromaDB o una búsqueda vectorial, a veces los índices son manejados por esa base vectorial, no por PostgreSQL.

**¿Cómo se llama o "usa" un índice?**

La respuesta corta es: **¡no se llama directamente!**  
El índice se **usa automáticamente** cuando hacés una consulta (por ejemplo, un SELECT) **y PostgreSQL detecta que usarlo es más eficiente** que hacer una búsqueda completa de la tabla.

**¿Cómo lo decide PostgreSQL?**

PostgreSQL tiene un **optimizador de consultas** que analiza el SELECT que estás haciendo y decide:

* ¿Conviene leer toda la tabla (table scan)?
* ¿Conviene usar un índice (index scan o index-only scan)?

Esto depende de:

* La cantidad de datos.
* Las condiciones del WHERE.
* Qué columnas pedís.
* Si hay ORDER BY, JOIN, etc.

**Ejemplo:**

Suponer que tenés esto:

CREATE INDEX idx\_texto\_pregunta ON preguntas (texto);

Y hacés este SELECT:

SELECT \* FROM preguntas WHERE texto = '¿Cómo uso índices en PostgreSQL?';

Si hay muchas filas y esa condición filtra pocas, **el índice se usa automáticamente**, sin que tengas que hacer nada especial.

**¿Cómo saber si el índice se está usando?**

Podés chequearlo con:

En postgresql: EXPLAIN ANALYZE

SELECT \* FROM preguntas WHERE texto = '¿Cómo uso índices en PostgreSQL?';

Esto te muestra si PostgreSQL está usando:

* Index Scan ✅ (sí usa el índice)
* Seq Scan ❌ (lectura secuencial, no usó índice)

**En Python:** tampoco hacés nada especial. El SELECT funciona igual:

cur.execute("SELECT \* FROM preguntas WHERE texto = %s", ('¿Cómo uso índices en PostgreSQL?',))

Si hay un índice en texto, PostgreSQL lo usa por detrás.

|  |  |
| --- | --- |
| Pregunta | Respuesta corta |
| ¿Cómo llamo a un índice? | No se llama, se usa automáticamente. |
| ¿Tengo que hacer algo especial? | No, solo asegurate de que existe y que tus consultas lo pueden aprovechar. |
| ¿Cómo veo si lo está usando? | Con EXPLAIN ANALYZE. |