

Laboratorio 6 - Índices

Profesores: Claudio Gutiérrez
Matías Toro

Auxiliares: Benjamín Correa
Cristóbal Miranda
Raúl Silva
Florencia Yáñez

Se cuenta con varias tablas que son instancias del siguiente esquema abstracto:

- **pelicula**(nombre, año, calificación, votos)
- **actor**(nombre, genero)
- **personaje**(a_nombre, p_nombre, p_año, personaje)

En la base de datos hay dos esquemas: uno con datos indexados (**lab6i**) y otro sin índices (**lab6**). En cada esquema está la misma estructura tres veces con los datos para películas con más de 10.000 votos, más de 1.000 votos y más de 100 votos. Note que las tablas de más de 10.000 votos tienen **menos** tuplas que las 1.000 y muchas **menos** que las 100. En el laboratorio, usted medirá el efecto de utilizar índices en los tiempos de consultas, entregando un PDF con sus respuestas.

P1. 6 PUNTOS

(a) Usando el esquema **lab6** cuente el número de tuplas para cada tablas presente. Registre las consultas y sus resultados.

Puede usar el comando `\dt lab6.*` para listar todas las tablas del schema **lab6**.

(b) Cuente cuántos bloques hay en cada tabla. Para ello puede usar la siguiente consulta:

```
SELECT DISTINCT relname, relpages  
FROM pg_class WHERE relname = 'TABLA-NOMBRE';
```

Donde **relpages** corresponde al número de bloques en PostgreSQL.

(c) Calcule el número promedio de tuplas por bloque para cada tabla.

Nota para las siguientes preguntas: Observe que estos conteos no cambian en el caso de **lab6i**, así que no hay que contar las tuplas y los bloques dos veces.

P2. SIN PUNTOS Compare los índices disponibles para las tres tablas con el prefijo **10000** en ambos esquemas usando los dos comandos `\d+ lab6.TABLA10000` y `\d+ lab6i.TABLA10000`. Recuerde que Postgres agrega un índice para la llave primaria por defecto, entonces **lab6** solo tiene esos índices. (Se configuran los índices de las tablas **1000** y **100** de la misma forma.)

P3. 9 PUNTOS Use el siguiente comando para obtener los planes de consulta y tiempos de ejecución de la consulta indicada. Ejecútela en los esquemas `lab6` y `lab6i`.

```
EXPLAIN ANALYZE SELECT * FROM ESQUEMA.personaje100  
WHERE p_nombre='Up' AND p_anho=2009;
```

- (a) Registre el plan de consulta y el tiempo de ejecución.
- (b) Calcule y registre la cantidad de consultas por segundo (según la suma del tiempo de planificación y ejecución) que pueden realizarse.

P4. 9 PUNTOS De la pregunta anterior, calcule y registre una estimación de la cantidad de bloques leídos (± 1) por las dos consultas. Para esto considere el plan, el número de tuplas en el resultado, el número de tuplas en la tabla, y el número promedio de tuplas por bloque, en cada uno de los dos esquemas.

P5. 12 PUNTOS Considerando solo las tablas 100:

- (a) Consulte por las **otras** películas (nombre y año) en las cuales los actores/actrices de “Up (2009)” han participado.
Para ello, escriba dos versiones equivalentes de la misma consulta: una versión sin anidación y otra versión con anidación (usando `IN`).
- (b) Utilice `EXPLAIN ANALYZE` en ambas consultas y obtenga los planes y tiempos de ejecución para los esquemas `lab6` y `lab6i`.

P6. 24 PUNTOS Tomando la consulta de **P3**, y ambas consultas de **P5**, substituya “Up (2009)” por otra película (que aparezca en la tabla `pelicula10000`). Para las tres consultas resultantes, debe:

- (a) Ejecutar las consultas en el esquema `lab6` usando las tablas terminadas en 100, 1000 y 10000 usando `EXPLAIN ANALYZE` y registrar los tiempos.
- (b) Ejecutar las consultas en el esquema `lab6i` usando las tablas terminadas en 100, 1000 y 10000 usando `EXPLAIN ANALYZE` y registrar los tiempos.
- (c) Mostrar gráficamente (con la herramienta que estime conveniente) cómo varía el tiempo tomado (la suma del tiempo de ejecución y el tiempo de planificación) respecto al tamaño de las tablas, tanto en la versión sin índices, como en la versión indexada, como se ejemplifica en la Figura 1. Debe incluir un gráfico por consulta. Considere que el tamaño de la tabla es el número de tuplas en la tabla `personaje*`.

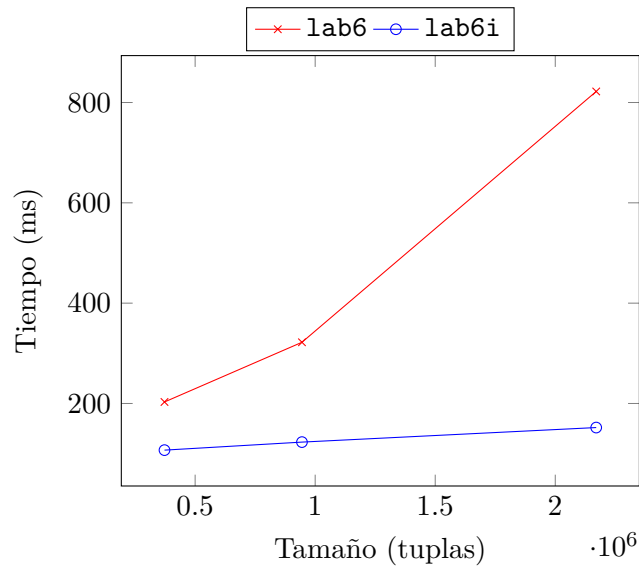


Figure 1: Gráfico de ejemplo: se muestran las curvas para la misma consulta con y sin índices y cómo varía el tiempo de ejecución respecto al tamaño (número de tuplas) de la tabla `personaje*`.

Entrega

Debe entregar un documento simple en pdf que incluya los datos recolectados en **P1**, **P3**, **P4** y **P5**. Además, debe incluir las consultas usadas en la **P5**. Finalmente, para cada consulta en la **P6**, debe incluir:

- La consulta misma (*solo* para la escala 100 en **lab6**).
- La planificación de la consulta con y sin índices (con `EXPLAIN ANALYZE`, solo para la escala 100).
- Los gráficos de comparación debidamente detallado (mirar ejemplo de Figura 1).