

Optimización Física de Consultas

Profesor: Teófilo Chambilla
ACL: Alexandra Shulca Romero

Se cuenta con la siguiente estructura:

- **Pelicula**(nombre:string, anho:int, calificacion:float, votos:int)
- **Actor**(nombre:string, genero:char)
- **Personaje**(a.nombre:string, p.nombre:string, p.anho:int, personaje:string)

Construya la base de datos con los scripts y archivos adjuntos, en dicha BD podemos encontrar dos esquemas, uno con datos indexados y otro sin índices. Cada esquema tiene las mismas tablas replicadas tres veces: en la primera se encuentran los datos para películas con más de 10.000 votos, en la segunda los datos de películas con más de 1.000 votos y en la tercera para las películas con más de 100 votos. En el laboratorio, usted deberá medir el efecto de utilizar índices en consultas complejas y deberá contrastar los resultados prácticos con los teóricos. Ud. debe entregar un breve reporte con sus respuestas. Tenga en cuenta que éste mismo procedimiento deben realizar en último capítulo del informe de proyecto.

P1: Usando el esquema **lab11** cuente las tuplas de las tablas presentes. Debe notar que las tablas 10k tienen menos tuplas que las **1k** y muchas menos que las **100**. Registre sus resultados. En el esquema **lab11i** explore los índices que están a su disposición usando `\d + tabla` o haciendo una consulta sobre la vista de catalogo **pg_catalog.pg_indexes**. Recuerde que postgres agrega un índice para la llave primaria por defecto, entonces **lab11** sólo tiene esos índices.

P2: Use la consulta detallada más adelante. Ejecútela en ambos esquemas **lab11** y **lab11i**. Utilizando **EXPLAIN ANALYSE** obtenga los planes de consulta y tiempos de ejecución. Registre estos datos e indique cantidad de queries por segundo que pueden realizarse y la cantidad de registros accedidos en cada uno de los esquemas.

```
SELECT * FROM personaje100 WHERE p_nombre='Fight Club'
```

P3: Seleccione tres consultas complejas: una que use una consulta por rango (mientras más pequeño el rango, más se beneficia la consulta del indexamiento) una que requiera joins y una que utilice consultas anidadas. Para cada consulta ud. debe:

- Ejecutar las consultas en el esquema lab11 usando las tablas terminadas en 10k, 1k y 100 usando EXPLAIN ANALYSE y registre los tiempos.

- Ejecutar las consultas en el esquema lab11i usando las tablas terminadas en 10k, 1k y 100 usando EXPLAIN ANALYSE y registre los tiempos. Note que es imprescindible que su consulta **utilice** alguno de los índices **proporcionados** (es decir, no los índices por defecto de las llaves primarias), sino deberá seleccionar otra consulta

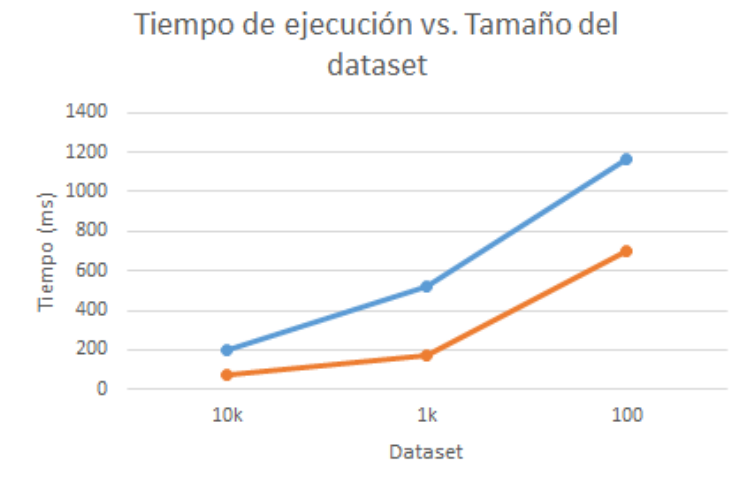


Figure 1: Gráfico de ejemplo: se muestran las curvas para las consultas con y sin índices y cómo varía el tiempo de ejecución respecto al tamaño de las tablas.

- Muestre gráficamente (usando la herramienta que estime conveniente) cómo varía el tiempo de ejecución respecto al tamaño de las tablas, tanto en la versión sin índices, como en la versión indexada. Puede ver un ejemplo del gráfico esperado en la Fig. 1.

En su reporte entonces debe mostrar los datos recolectados en **P1** y **P2** y además, para cada consulta:

- El SQL de la consulta
- La planificación de la consulta con y sin índices
- El gráfico de comparación
- La cantidad (aprox.) de registros visitados en la base de datos más grande para ambos casos
- Opcional. Agregue el resultado teórico que debiera darse al aplicar los planes. ¿Tienen sentido al mirar las curvas?