

Laboratorio 8 - Evaluación de consultas

Profesores: Claudio Gutiérrez
Matías Toro
Auxiliares: Scarlett Plaza
Daniel Radrigán
Cristian Salazar
Fran Zautzik

La entrega de este laboratorio deberá ser un informe en pdf, el cual está descrito al final de este archivo.

P1. 30 PUNTOS Al igual que el lab anterior, se cuenta con varias tablas que son instancias del siguiente esquema abstracto:

- **pelicula**(nombre, anho, calificacion, votos)
- **actor**(nombre, genero)
- **personaje**(a_nombre, p_nombre, p_anho, personaje)

En la base de datos del curso hay dos esquemas: uno con datos indexados (**opti**) y otro sin índices (**opt**). 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.

- (a) 10 PUNTOS Usando el esquema **opt** cuente las tuplas en las nueve tablas presentes. Debe notar que las tablas 10000 tienen **menos** tuplas que las 1000 y muchas **menos** que las 100. Registre las consultas y sus resultados. Usando la siguiente consulta, contar cuántos bloques (**relpages**: se llaman páginas en Postgres, cada uno de 8 kb. típicamente) hay en cada tabla:

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

Calcule el número promedio de tuplas por bloque para cada tabla. (Observe que estos conteos no cambian en el caso de **opti**, así que no hay que contar las tuplas y los bloques dos veces.)

- (b) 10 PUNTOS Use el siguiente comando para obtener los planes de consulta y tiempos de ejecución de la consulta indicada. Ejecútela en los esquemas **opt** y **opti**.

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

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

- (c) 10 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.

P2. 30 PUNTOS Considere el siguiente esquema relacional y consulta SQL. El esquema captura información sobre empleados, departamentos y finanzas de la empresa (organizadas por departamento).

- **Empleado**(eid, did, sueldo, hobby)
- **Departamento**(did, nombre, piso, telefono)

- **Finanzas**(did, presupuesto, ventas, gastos)

```
SELECT D.nombre, F.presupuesto
FROM Empleado E, Departamento D, Finanzas F
WHERE E.did = D.did AND D.did = F.did AND D.piso = 1
AND E.sueldo >= 59000 AND E.hobby = 'magic'
```

- (a) 10 PUNTOS Identifique los posibles árboles de álgebra relacional que reflejen el orden de las operaciones que un optimizador podría considerar.
- (b) Suponga que la siguiente información adicional está disponible: existen índices de árbol B+ no agrupados en **Empleado.did**, **Departamento.piso**, **Departamento.did** y **Finanzas.did**. Las estadísticas/catálogo del sistema indican que los salarios de los empleados oscilan entre 10.000 y 60.000, los empleados disfrutan de 200 hobbies diferentes y la empresa posee 2 pisos en el edificio. Hay un total de 50.000 empleados y 5.000 departamentos (cada uno con información financiera correspondiente) en la base de datos. Suponga una distribución uniforme. El sistema de gestión de bases de datos utilizado por la empresa tiene solo un método de join disponible, que es el loop anidado con índice.
- 10 PUNTOS Para cada una de las relaciones base de la consulta (Empleado, Departamento y Finanzas), estime la cantidad de tuplas que se seleccionarían, si se aplican las selecciones correspondientes (filtros), antes de hacer los joins.
 - 10 PUNTOS Construya dos árboles distintos, y argumente cuál cree usted que será más eficiente de evaluar.