

# Entrega 4

Mauricio Monsalve

June 19, 2007

## 1 Objetivos

Evaluar la eficiencia lograda en la base de datos al usar diferentes planes de índices para un conjunto de consultas críticas.

## 2 Requerimientos

Para esta entrega, es necesario:

1. Definir tres consultas complicadas: que incluyan al menos tres tablas en la evaluación y tengan agrupación y/o consultas anidadas. Las consultas deben ser costosas de evaluar (ojo con las condiciones WHERE).
2. Ser capaces de llenar tablas hasta con 10.000 elementos (cada una de las tablas importantes de las consultas) dinámicamente. O sea, variar el contenido de las tablas. Esto se puede hacer con un script externo o con copias de las tablas con una cierta cantidad de datos en ellas. Por ejemplo, `SELECT ... FROM A,B,C WHERE...` requiere A, B, C, por lo que se podrían tener A1000, B1000, C1000 como tablas con 1000 registros, A5000, B5000, C5000 con 5000 registros, etc.
3. Definir planes de índice con sólo un tipo de índice (variable). Deberán usar planes sin índices, con índices B+ y con índices de hashing.

## 3 Experimentos

En el trabajo será necesario realizar experimentos del rendimiento de la base de datos ante distintos planes de índices. En particular,

- Para cada consulta:
  - Por cada plan (sin índices, B+ y hashing):
    - \* Obtener los tiempos de consulta con 100, 1000, 5000 y 10000 registros por tabla.

Nótese que cada experimento debe realizarse varias veces para absorber los rendimientos variables del sistema. Así, el tiempo promedio por experimento debiera suavizar rendimientos extraños de sistema.

El número de veces que se repite un experimento se deja a criterio, pero debiera ser superior a las 10 veces.

Nótese que, sin contar las veces que se repite el experimento, hay que realizar  $3 \times 3 \times 4 = 36$  experimentos distintos. Se recomienda usar un script para automatizar el experimento.

## 4 Estructura del informe

En el informe se deben presentar las consultas importantes, su significado y la justificación de su costo (que sean algo más de un simple join natural entre tres tablas). Luego se debe indicar qué atributos tendrán índices y por qué. Después de esto se detallarán los resultados experimentales.

Los resultados experimentales se presentarán como gráficos de N v/s tiempo, donde N es el tamaño de cada tabla (100, 1000, 5000 y 10000). Esto significa que habrán tres gráficos por consulta, significando un total de nueve gráficos.

La presentación del informe será considerada. Nótese que deberá haber coherencia entre el plan de índices y los resultados experimentales; un buen plan de índices logrará resultados coherentes con la teoría en cuanto al costo de las consultas.

Una vez terminados los gráficos, se deberán presentar las conclusiones. Contraste la teoría con la práctica en términos de aproximación de costos y vea cuál es el beneficio real (cuantitativo) del uso de índices. Vea cómo un índice supera a otro en determinados casos.