

# Proyecto de CC42A

## **Etapa Final**

Profesor: Claudio Gutiérrez  
Auxiliar: Mauricio Monsalve

Otoño 2005

### **1 Introducción**

La etapa final del proyecto de curso consiste en utilizar un plan de índices para la base de datos y estimar su eficiencia comparativa respecto a la situación sin índices.

### **2 Objetivos**

Objetivos específicos de esta etapa son:

- La identificación de las consultas más relevantes y su escritura en SQL.
- La generación un plan de índices adecuado para optimizar el tiempo de respuesta de la base de datos ante las consultas presentadas.
- La realización de experimentos que demuestren la eficiencia comparativa entre el caso con índices y sin índices.
- La representación visual (tablas y gráficos) de la evolución del costo asintótico de las consultas.
- La justificación de los resultados anteriores según la teoría.

### 3 Estructura del informe

A continuación se presenta una estructura tentativa<sup>1</sup> del informe:

1. **Portada**

Debe contener el título de la entrega y el nombre de cada integrante del grupo, ordenado alfabéticamente.

2. **Situación**

Esta sección debe indicar, de forma precisa y concisa, cuál es el tema de la base de datos. No debe superar las 15 líneas.

3. **Consultas más típicas**

Se debe indicar la consulta y su equivalente en SQL. Cada consulta debe ser acorde a lo presentado en la sección anterior<sup>2</sup>.

4. **Plan de índices**

Se debe especificar un plan de índices para la base de datos, aplicada para las consultas más típicas<sup>3</sup>. Esta sección no debiera superar las 10 líneas.

5. **Pruebas de rendimiento**

En esta sección se deben presentar gráficos mostrando el resultado de los experimentos.

6. **Conclusión**

Finalmente una breve conclusión en la que expliquen lo aprendido con los experimentos.

En general se espera un informe de tamaño reducido (no debería superar las 15 páginas, siendo ideal sólo 10) y autoexplicativo. Es muy importante que el informe esté bien redactado, de manera que los ayudantes corrijan en el menor tiempo posible.

---

<sup>1</sup>Queda a discreción de cada grupo la estructura que ha de seguir. Una estructura es aceptable si presenta toda la información solicitada y es de tamaño reducido.

<sup>2</sup>No tiene sentido que una consulta sea típica si es que no tiene relación con el tema.

<sup>3</sup>Optimizar la acción más recurrente lleva a optimizar el rendimiento general.

## 4 Consultas más típicas

Se espera que las consultas más típicas tengan una complejidad razonable. Esto es, que hagan producto cartesiano sobre 4 ó más tablas. Por razones de tiempo, es requisito que, de las consultas más típicas, al menos una contenga una consulta anidada y al menos una sea agrupada (pueden ser la misma). Respecto al número de consultas, se espera que sean tres al menos.

## 5 Pruebas de rendimiento

Se espera que se hagan dos tipos de experimento:

- Consultas en tablas con índices.
- Consultas en tablas sin índices.

Se deben probar todas las consultas. Además, se deberán realizar distintas pruebas con tablas pobladas a distinto tamaño<sup>4</sup>. Esto para ilustrar la evolución del tiempo de respuesta de cada consulta en función del tamaño de las tablas.

Para reducir errores experimentales, cada prueba se deberá realizar varias veces<sup>5</sup> y luego se deberá obtener un valor promedio del tiempo de respuesta de la consulta. No interesa la desviación estándar ni nada de eso, solamente el promedio. Es recomendable que los registros se generen de manera aleatoria y que cambien sus valores entre distintos experimentos. Pero ojo, **sólo se debe medir el tiempo de respuesta de las consultas, no se deben medir tiempos de inserción o de borrado**, así que deberán filtrar, de alguna manera, todo lo que no sea tiempo de consulta.

Todas las combinaciones deberán ser probadas. Para presentar sus resultados deberá ocupar gráficos. Los gráficos serán de la forma  $X=N\_Registros$ ,  $Y=Milisegundos$ . Si las consultas tardan mucho tiempo, puede ocupar escalas en segundos o minutos. En lo que concierne al rendimiento comparativo, los gráficos deberán ser de la forma  $X=N\_Registros$ ,  $Y=Mejora\%$ .

---

<sup>4</sup>Tamaño mediano. Valores como 50, 100, 300 registros no son de interés. Valores como 1000, 5000, 10000 son valores de interés.

<sup>5</sup>A criterio. Si la consulta es resuelta muy rápido, puede repetir cientos de veces (o miles) y luego dividir para evitar errores de precisión en milisegundos. Si la consulta resulta ser muy lenta de responder, repetir 10 veces puede ser suficiente.

## 6 Ejemplo de pruebas de rendimiento

Este ejemplo muestra el proceder necesario para realizar los experimentos.

A continuación se presentan los resultados experimentales:

1. Consulta: “Los átomos con electronegatividad mayor a  $\epsilon$ ”

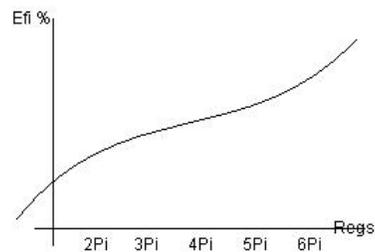
(a) Experimento con tablas sin índices.



(b) Experimento con tablas con índices.



(c) Rendimiento comparativo.



2. Consulta: “Los líquidos con temperatura de ebullición mayor a 100 Celcius”

(a) Esto es sólo un ejemplo...

## 7 Recomendaciones generales

No se espera un informe de detalle técnico con letra de 9pt, precisión de 9 decimales en notación científica, etc. **Sólo se pide un informe breve y conciso, bien ilustrado en los experimentos.**

No se debe perder tiempo dilatando el texto del informe. Oraciones cortas, simples y precisas serán un premio. Su informe tendrá mejor presentación y ustedes ahorrarán tiempo en confeccionarlo. Tiempo que, por cierto, es valioso por estos días.

## 8 Condiciones de entrega

El informe debe ser entregado **en papel** en Secretaría Docente.

El plazo de entrega acaba el día **viernes 1 de Julio a las 16:00 horas**<sup>6</sup>.

No esperen la última hora para entregar.

**Recuerden: Tener un informe con nota inferior a 4,0 significa reprobar el curso.**

---

<sup>6</sup>Es necesario darse una holgura para recoger los informes antes que Magaly cierre la Secretaría Docente. Se debe corregir durante el fin de semana. Por lo tanto no se puede colocar plazo hasta las 17:00 horas. Menos se puede postergar el día.