

CC42A – BASES DE DATOS
Profesores: Claudio Gutiérrez, Gonzalo Navarro
Auxiliar: Mauricio Monsalve

ENUNCIADO ETAPAS 4 Y 5

ETAPA 4

*Esta etapa **no será evaluada**, sin embargo debe ser realizada.*

Esta etapa consiste en la creación de 2 ó 3 consultas de alta complejidad para la base de datos. Debe constar de varios **joins** entre tablas (4 ó 5), además de tener **sentido** en el contexto de su base de datos (debe ser utilizable por el cliente de la base de datos). Se deben realizar operaciones de selección y proyección también. No es la idea que la base de datos lea al final todas las filas de las tablas bajo consulta, sólo las necesarias según la selección.

Forma de entrega: La entrega se puede realizar por UCursos o por mail al auxiliar (mnmonsar-aroba-dcc). No es la idea que se entregue texto plano. Se debe entregar el informe con una portada como de costumbre y las consultas. Las consultas deben estar detalladas y explicadas en su funcionamiento (qué hace cada parte del SELECT) y su importancia (qué hace la consulta, para qué sirve y por qué la usaría un cliente).

Formatos de archivo aceptados: Se aceptan archivos **Postscript** (.ps que se puedan leer en anakena o con un visor Ghostscript), **Acrobat Reader** (.pdf que se pueda ver con Acrobat Reader o XPdf), **Hipertexto web** (.htm o .html formato plano estándar) o, adicionalmente, **Word** (.doc un pseudoestándar).

Fecha de entrega: Sábado 30 de Octubre a las 24:00 por UCursos o por mail.

Recomendaciones: Prueben intensamente las consultas en sus bases de datos Postgres (en anakena o en su casa) para evitar entregar resultados no coherentes. Cualquier duda al auxiliar a través de UCursos o mail.

Revisión: La revisión a la etapa se hará de la forma más rápida posible y los resultados serán comunicados por UCursos. Se harán algunos alcances y correcciones a las consultas entregadas.

ETAPA 5

Etapa final del proyecto del curso. Toma varios días y es recomendable comenzar a hacerla apenas tengan los resultados de las consultas.

Esta etapa consiste en la creación de estadísticas de rendimiento de la base de datos en distintos escenarios planteados. Los distintos escenarios serán tablas sin índices, con índices árbol-B y con índices hashing.

Las actividades a realizar son las siguientes:

1. Cálculo de costos de inserción y consulta en las tablas sin índices

a) Para cada tabla de la base de datos se debe calcular el costo promedio de inserción de la siguiente manera:

- Se debe calcular cuanto demora insertar N registros en la tabla T.
- El paso anterior se debe repetir 20 veces para finalmente obtener un promedio final con los 20 resultados (costo promedio de inserción de N registros en la tabla T).
- Los valores para N son 500, 1000, 2500, 5000, 7500 y 10000.

El proceso anterior les permitirá generar el gráfico “Costo de inserción sin índices” donde se deben visualizar los costos de inserción en cada tabla con cada N cantidad de registros (Tiempo v/s Número de Registros).

Nota: A partir de este punto cuando se indique un costo promedio implica la repetición de un proceso 20 veces y así poder obtener el promedio. Lo mismo se da para los valores de N. Todo gráfico que se mencione a continuación debe ser del tipo “Tiempo v/s Número de Registros.”

b) Para cada tabla se debe calcular un costo promedio de consulta. Esto deberá hacerse de preferencia sobre el campo clave o el que según ustedes consideren sea el mas utilizado. Este paso genera el gráfico “Costo de consulta sin índices.”

c) Finalmente se debe calcular el costo promedio de ejecución de la consulta que se elaboró para la etapa 4 (desde ahora llamada consulta principal). Este paso genera el gráfico “costo de la consulta principal sin índices.”

2. Calculo de costos de inserción y consulta en las tablas con índices

Basándose en los resultados obtenidos en el paso 1, se deben analizar los resultados y junto a los atributos que participen en la consulta principal, se deberán escoger índices para cada una de las tablas de la base de datos. Este paso se debe aplicar tanto para el tipo de índice B-Tree como para el tipo Hash. En este paso se deben hacer las siguientes pruebas.

a) Para cada tabla se deberá calcular el costo promedio de inserción. Este paso genera el gráfico “Costo de inserción con índices (Tipo de índice).”

b) Para cada tabla se deberá calcular el costo promedio de consulta. La consulta deberá ser sobre el campo índice y así comprobar el comportamiento del motor de base de datos. Este paso genera el gráfico “Costo de consulta con índices (Tipo de índice).”

c) Tomando en cuenta los índices que participan en la consulta principal, se deberá calcular el costo promedio de la consulta principal para cada combinación de estos índices. Por ejemplo:

Si se tienen 3 tablas (A,B,C), y un índice en cada tabla, se deberán generar los siguientes gráficos:

- Costo de la consulta principal con índice en A.
- Costo de la consulta principal con índice en B.
- Costo de la consulta principal con índice en C.
- Costo de la consulta principal con índice en AB.
- Costo de la consulta principal con índice en AC.
- Costo de la consulta principal con índice en BC.
- Costo de la consulta principal con índice en ABC.

Nota. El paso 2 deberá hacerse para índices B-Tree y Hash por separado.

d) Un último gráfico “Costo de la consulta principal sin índices Vs. con índices”, deberá contener los resultados de la prueba 1c y los resultados de la combinación de índices que presento mejores resultados en el paso 2c (tanto para índices B-Tree como Hash).

3. Conclusiones

Presente las apreciaciones y conclusiones sobre el uso o no de distintos tipos de índices para las bases de datos (recuerde que puede construir índices sobre cualquier atributo de una tabla), y mostrar como afectan (mejorando o empeorando el desempeño) tanto en el tiempo para las inserciones y para las consultas.

Nota. El tiempo puede ser medido con el comando de postgres:

SELECT CURRENT_TIME

Para mayor referencia bajar los manuales de postgres de la página web del curso o visitar <http://www.commandprompt.com/ppbook>.

Forma de entrega: Se debe entregar un informe escrito en secretaría docente hasta las 17:00 hrs. del día Martes 16 de Noviembre.

Revisión: Como en el resto de los informes evaluados, la nota aparecerá en UCursos.