

Esquema Presentación Tarea3 – Base de datos

Profesor: Claudio Gutiérrez
Ayudante: Luis G. Silvestre Quiroga
V 2.0 (20-05-2011)

A continuación se describe la estructura que debe tener el **Informe Final de la Tarea 3**, considerar que la “experimentación y optimización” debe realizarse mediante 2 o 3 consultas “genéricas”, en 1.000(mil) datos, 10.000(diez mil) datos y 100.000(cien mil) datos almacenados en la base de datos (deberían tener 3 dump de su proyecto).

1. PARTE I. Requisitos

- 1.1. Introducción.
- 1.2. Descripción general del problema/organización/empresa.
 - 1.2.1. Necesidad/usos de la base de datos.
- 1.3. Cómo resuelven el problema hoy?
 - 1.3.1. Cómo almacenan/procesan los datos hoy?
 - 1.3.2. Flujo de datos (como/para-que se usan).
- 1.4. Descripción detallada del sistema.
 - 1.4.1. Objetos de información actuales.
 - 1.4.2. Características y funcionalidades esperadas.
 - 1.4.3. Tipos de usuarios existentes/necesarios.
 - 1.4.4. Tipos de consulta, actualizaciones, etc.
- 1.5. Eventualidades.
 - 1.5.1. Problemas que pudieran encontrarse en el proyecto.

2. PARTE II. Modelo Entidad Relación

- 2.1. Modelo Entidad Relación
- 2.2. Especificaciones y consideraciones sobre el Modelo Entidad Relación

3. PARTE III. Modelo Relacional

- 3.1. Modelo Relacional.
- 3.2. Especificaciones del Modelo Relacional
- 3.2. Diccionario de Datos Semántico.

4. PARTE IV. Optimización y Experimentación

- 3.1. Implementación de la Base de Datos.
 - 3.1.1. Carga de Datos Desde Archivos de Datos Originales (si existieran).
 - 3.1.2. Simulación de Datos Faltantes.
- 3.2. Consultas SQL para el experimento.
 - 3.2.1. Consultas Seleccionadas.
 - 3.2.2. Implementación de Consultas en SQL.
- 3.3. Metodología del experimento.
- 3.4. Optimización de Consultas.
 - 3.4.1. Planes de Índices Para Consulta 1.
 - 3.4.2. Planes de Índices Para Consulta 2.
- 3.5. Plataforma de Pruebas.
- 3.6. Medición de tiempos.
- 3.7. Resultados.
 - 3.7.1. Consulta 1
 - 3.7.1.1 Resultados y gráficos en tiempo de ejecución con 1 000 datos
 - 3.7.1.2 Resultados y gráficos en tiempo de ejecución con 10.000 datos
 - 3.7.1.3 Resultados y gráficos en tiempo de ejecución con 100.000 datos
 - 3.7.2. Consulta 2
 - 3.7.2.1 Resultados y gráficos en tiempo de ejecución con 1 000 datos
 - 3.7.2.2 Resultados y gráficos en tiempo de ejecución con 10.000 datos
 - 3.7.2.3 Resultados y gráficos en tiempo de ejecución con 100.000 datos
- 3.8. Análisis y Discusión.

5. PARTE V Conclusiones

- 5.1. Conclusiones y Recomendaciones

6. PARTE VII Anexos:

- 6.1 Algunas salidas de comando "EXPLAIN ANALYZE" del SGBD para consultas y planes de índices.
- 6.2. Consultas SQL e índices.
 - 6.2.1. Alternativa 0 (*).

6.2.2. Alternativa 1 (*).

6.2.3. Alternativa 2 (*).

(*) Las alternativas se refieren a tratamiento de las consultas mediante otros índices combinados como ser: Hash, B+, etc.

NOTA: Tres “dump” y el informe en PDF debe ser entregado por **U-Cursos** en la fecha agendada, *además del **informe impreso** en la Secretaria Docente (Piso 4) con **Sandra**.*

Pauta Presentación Tarea3 – Base de datos

Profesor: Claudio Gutiérrez
Ayudante: Luis G. Silvestre Quiroga
V 2.0 (20-05-2011)

Objetivo

El objetivo principal es aplicar los conocimientos de índices explicados en el curso, mediante la experimentación sobre la base de datos (o un sub-conjunto representativo de la misma) de los proyectos grupales del semestre.

Metodología

Se trabaja experimentalmente sobre la base de datos del proyecto, utilizando datos ficticios (o reales si se poseen), midiendo valores de tiempo sobre consultas que representen una carga importante para la base de datos. Se entregara un informe final, que contiene todo lo relacionado con el informe 1 y 2, más los cambios que se sugirieron en el informe 2, más lo pedido para esta entrega. La metodología final consiste en comparar los tiempos que tarda Postgres, en consultas con y sin índices.

Pasos a seguir:

- 1) Estudiar la materia de índices del ramo. Estudiar la materia de plan de índices, arboles de consulta, tipos de índices mas comunes, estructuras, etc.
- 2) Implementar la estructura de tablas en Postgres.
- 3) Explicar brevemente en el informe porque se recomienda utilizar Postgres.
- 4) Poblar la base de datos con 1.000, 10.000, 100.000 datos por tabla (explicar de dónde se obtuvieron los datos, y en caso de ser ficticios como se generaron). Deberían generar 3 “dump” de la base de datos.
- 5) Construir al menos 2 consultas sobre la base de datos, exponer el código SQL que las representa y explicar que se espera de la consulta.
- 6) Investigar en internet el uso del comando ANALYZE y EXPLAIN. Entender y explicar el output de ambos, además de entender el funcionamiento de los dos.
- 7) Investigar en internet el cache de Postgres (que es, para que sirve y como se limpia). Explicar brevemente en el informe.
- 8) Eliminar cualquier índice de las tablas de la base de datos.
- 9) Utilizando los comandos ANALYZE y EXPLAIN, ejecutar al menos 12 veces cada una de las 3 consultas, anotando en tablas los resultados obtenidos para los tiempos (cada valor con su

respectivo promedio y desviación estándar). Cuando ejecute este experimento, recuerde que las tablas NO deben tener asignados índices.

- 10) Interpretar el árbol de consulta generado por el comando EXPLAIN y exponerlo de forma clara en el informe.
- 11) Analizar el árbol de consulta generado por el comando EXPLAIN, y proponer en el informe un nuevo plan de índices, de modo de minimizar los tiempos de consulta.
- 12) Agregar el plan de índices a la base de datos.
- 13) Ejecutar nuevamente las 12 repeticiones de cada una de las 3 consultas, anotando nuevamente los tiempos promedios y desviaciones estándar (tenga en cuenta y considere la limpieza del cache de postgres).
- 14) Analizar las tablas obtenidas con índice y sin índices. Discutir los resultados, explicar porque se obtuvo dicho comportamiento apoyándose en la herramienta EXPLAIN.
- 15) Proponer un nuevo plan de índices (de no ser necesario, explicar porque).
- 16) De haber propuesto un nuevo plan, expóngalo y arguméntelo, repitiendo el experimento opcionalmente.
- 17) Concluya respecto a los experimentos y concluya respecto a todo el trabajo final considerando las entregas anteriores.

NOTA: Tres “dump” y el informe en PDF debe ser entregado por **U-Cursos** en la fecha agendada, *además del informe impreso en la Secretaria Docente (Piso 4) con Sandra.*