

Análisis de rendimiento y replicación en Bases de Datos distribuidas

Lic. Rodolfo Bertone¹
Lic. Sebastián Ruscuni²
A.C. Ivana Miatón³
Srta. Aixa Mauriello⁴

*LIDI. Laboratorio de Investigación y Desarrollo en Informática.
Facultad de Informática. UNLP.
50 y 115. 1er Piso. La Plata*

Palabras Clave

Bases de Datos Distribuidas. Sistemas Distribuidos. Replicación de Datos. Simulación.

Resumen (u Objetivos)

El objetivo del trabajo consiste en estudiar el comportamiento (en tiempo de respuesta y confiabilidad) de grandes bases de datos de imágenes sobre arquitecturas distribuidas de redes LAN y WAN. Esto significa problemas tales como: Tasa de pérdida de datos; Tiempo máximo necesario para recuperación de información; Complejidad y eficiencia de los algoritmos de recuperación; Tiempo de utilización de recursos del sistema; Incidencia del porcentaje de replicación en el tiempo de respuesta.

¹ Profesor Adjunto Dedicación Exclusiva, Facultad de Informática, UNLP

² Ayudante Diplomado Semi Dedicación, Becario de Iniciación de CIC. Director: Armando De Giusti – Rodolfo Bertone

³ Ayudante Diplomado Semi Dedicación, Facultad de Informática, UNLP

⁴ Ayudante Alumno, Becario LIDI. Director: Rodolfo Bertone

Introducción

El modelo distribuido de datos hace posible la integración de BD heterogéneas proveyendo una independencia global del administrador de bases de datos (DBMS) respecto de esquema conceptual. Además, es posible implementar una integración tal, que reúna varios modelos de datos, representado cada uno de ellos características propias de organizaciones diferentes, asociadas para un trabajo conjunto.

Este modelo de distribución normalmente tiene un soporte multiprocesador tipo MIMD sobre diferentes esquemas de comunicaciones. En tal arquitectura se reproducen los problemas de distribución óptima de datos y procesos, de migración de datos y procesos y de tolerancia a fallas que se mencionan como áreas de investigación del proyecto.

El objetivo concreto ha consistido en desarrollar un ambiente de experimentación sobre una red para estudiar grandes Bases de Datos distribuidas que deban ser accedidas en tiempo real (por ello puede requerirse tener un porcentaje de la BD on-line y otro off-line). Se tiene como modelo las bases de datos de fotografías utilizadas por los diarios en la composición de noticias.

Temas de Investigación y desarrollo

Los puntos centrales de investigación y desarrollo son los siguientes:

- Estudio de esquemas de replicación y fragmentación de datos y su incidencia en el comportamiento final del sistema de BDD.
- Estudio de protocolos de commit, para asegurar la integridad de los datos.
- Estudio de generación de casos de prueba, a fin de obtener las mejores trazas para evaluación del modelo.
- Desarrollo del modelo de simulación, el cual incluye:
 - Generador de trazas de ejecución (este trabajo se encuentra parcialmente realizado, permitiendo obtener trazas parametrizadas con un nivel de complejidad intermedio)
 - Generador de la Base de Datos Distribuida (desarrollado)
 - Generador de prueba (primer etapa finalizada)

Algunos resultados obtenidos. Líneas de Investigación en curso.

Los resultados obtenidos permiten simular la distribución y replicación de una Base de Datos con imágenes (las pruebas fueron realizadas con alrededor de 15000 fotos) sobre una red de computadoras LAN. A partir del entorno de simulación seleccionado (% de replicación, % de fragmentación y carga de cada localidad de la red) se puede obtener el esquema óptimo de trabajo para el sistema.

La ampliación que se encuentra en desarrollo consiste en simular una red de computadoras WAN, incorporando al modelo el costo de transmisión en la red.

El modelo seguirá incorporando, en ampliaciones futuras, rasgos característicos de BDD, teniendo como objetivo final poder simular el comportamiento de un esquema de replicación dinámico de datos.

Es de hacer notar que como resultado en etapas tempranas del trabajo de investigación se desarrollaron dos tesis de grado de la Licenciatura en Informática y está en desarrollo una tesis de Magíster.

Bibliografía Básica

Distributed and Multi-Database Systems Angelo Bobak ARTECH HOUSE 1996

Data Replication Marie Buretta JOHN WILEY 1997

Managing Distributed Data Bases Burleson. JOHN WILEY 1993

Experiencias en el análisis de fallas en BDD. Miaton, Ivana; Ruscuni, Sebastián; Bertone, Rodolfo; De Giusti, Armando. ANALES CACIC 98. NEUQUÉN ARGENTINA.

Análisis de Replicación en Bases de Datos Distribuidas, Marcelo Zanconi, TESIS DE MAGISTER EN CIENCIAS DE LA COMPUTACIÓN, UNIV. NAC. DEL SUR, BAHÍA BLANCA, 1996

[RUSC00] *Ambiente para la simulación de distintos ACP y recuperación de errores en Bases de Datos Distribuidas.* Ruscuni, Sebastián; Bertone, Rodolfo. Anales CACIC 2000. Ushuaia Argentina.

[WEB1] www.itlibrary.com/library/1575211971/ch45.htm

[WEB2] www.itlibrary.com/library/1575211971/ch26.htm