
Métricas de Performance en Administración de BDD en redes LAN y WAN.

*Bertone R.*¹

*LIDI – Laboratorio de Investigación y Desarrollo en Informática*²
Facultad de Informática UNLP

Resumen

El proyecto, continuidad del presentado para WICC 2002, plantea el estudio, modelación y simulación de entornos de BDD. En particular las investigaciones que se están realizando se centran en los aspectos concernientes a la replicación y fragmentación de la información, y como afectan las decisiones tomadas respecto de estos aspectos en la performance final del sistema de datos distribuidos.

Los factores relacionados con la replicación y fragmentación de los datos de un sistema de información están altamente relacionados con la forma de propagación y la regulación de las actualizaciones sobre la información. Los estudios realizados se plantearon sobre un modelo de datos preestablecido y con el esquema de ubicación de datos estático, y los planes de trabajo actuales apuntan a estudiar las características de entornos con replicación dinámica de información. El soporte de simulación elegido es Java, dado que permite una mayor versatilidad en las construcciones de aplicaciones.

La investigación combina los estudio de replicación junto con protocolos de cometido de transacciones que preservan la integridad de la información. De esta forma se obtienen resultados que integran resultados de actualización de réplicas y, al mismo tiempo, aseguran la integridad de la información.

Palabras Clave

Bases de Datos Distribuidas – Replicación de Información – Métricas de procesos

¹ Profesor Adjunto Dedicación Exclusiva – Facultad de Informática – UNLP – pbertone@lidi.info.unlp.edu.ar

² LIDI- Facultad de Informática – UNLP – Calles 50 y 115 Primer Piso – (1900) La Plata, Argentina. TE/FAX +54 221 4227707. <http://www-lidi.info.unlp.edu.ar>

Introducción

Una BDD es una colección de datos que están pertenecen lógicamente al mismo sistema pero que físicamente se encuentra divididos en múltiples sitios de una red de computadoras. Son varios los factores que hacen al desarrollo de sistemas distribuidos de BD, entre ellos:

- La naturaleza distribuida de algunas aplicaciones (por ejemplo bancos)
- Incrementar la disponibilidad y confiabilidad de la información
- Permitir compartir datos mientras se permite el control local de la información
- Mejorar la performance.

La distribución incrementa la complejidad del diseño e implementación del sistema. Para conseguir las ventajas anteriormente listadas, los DBMS distribuidos deben proveer características adicionales sobre los sistemas centralizados:

- La habilidad de acceder a sitios remotos y transmitir consultas y datos entre varios sitios, vía la red de computadoras
- La habilidad de mantener la información sobre la distribución y replicación de la información.
- La habilidad de poder determinar estrategias para consultas y transacciones a lo largo de múltiples sitios.
- La habilidad de decidir que copia de datos replicada conviene acceder en cada momento.
- La habilidad de mantener consistencias entre copias (réplicas) de datos
- La habilidad de recuperarse ante errores del sistema y actualizar la información.

Estas funciones, en si mismas, incrementan la complejidad de un sistema de BDD. Es importante estudiar los mejores algoritmos para asegurar estas habilidades del las BDD, lo que representa la base del trabajo de investigación que se realiza.

La replicación es útil para mejorar la disponibilidad de la información. El caso más extremo de replicación es generar copia de toda la base de datos, en cada sitio del sistema distribuido, creando lo que se denomina BD *completamente replicada*. El otro extremo es el caso de *no replicación*, cada fragmento de la BD se almacena exactamente en un sitio. Entre ambos extremos, se tiene el espectro de la *replicación parcial* de datos, esto es algunos fragmentos de la BD pueden estar replicados mientras otros no lo están. Una descripción de la replicación de los fragmentos se denomina *esquema de replicación*.

La habilidad de encontrar el esquema de replicación que mejor se adapte a un problema particular es el objetivo de este estudio.

Objetivos

Los objetivos específicos de la primer etapa del trabajo (presentada para WICC 2002) fueron:

- Implementar técnicas de replicación basadas en diferentes esquemas de regulación y propagación de actualizaciones sobre las réplicas.
- Medir las características de performance de dichas técnicas.
- Comparar las técnicas evaluando las mediciones de los resultados obtenidos.
- Evaluar los beneficios e inconvenientes de la replicación de datos en general.
- Reevaluar los resultados obtenidos

Los objetivos de esta segunda etapa apuntan a los siguientes aspectos:

- Profundizar las técnicas de regulación de actualización de réplicas: Master–Slave vs. Group, y Eager vs. Lazy.
- Incorporar el estudio realizado sobre mantenimiento de integridad de la información (protocolos de cometido)
- Incorporar el manejo de esquemas de replicación dinámica de datos. De esta forma, la ubicación de cada fragmento puede migrar entre los nodos de la red, permitiendo poner dinámicamente los datos cerca de los sitios donde más se utiliza.

Bibliografía

- [Bell et al, 1992] *Distributed Database Systems*. Davil Bell, Jane Grimson. Editorial: Addison Wesley. 1992.
- [Bhargava, 1987] *Concurrency and Reliability in Distributed Database Systems*. B. Bhargava (editor). Van Nostrand Reinhold, 1987.
- [Buretta 1997] *Data Replication. Tools and techniques for managing distributed information*. Marie Buretta. Editorial: John Wiley & Sons, Inc. 1997.
- [Carey et al, 1991] *Conflict Detection Tradeoff for Replicated Data*. Michael Carey, Miron Livny. ACM Transactions on Database Systems, Vol 16, 4 December 1991, pp 703-746
- [Di Paolo et al., 1999] *Ambiente Experimental para Evaluación de Bases de Datos Distribuidas*. Monica Di Paolo, Rodolfo Bertone, Armando De Giusti, Anales ICIE 99. International Congress of Information Engineering. Buenos Aires 18-20/08/99. pp 729-743
- [Di Paolo 1999] *BDD. Estudio de consistencia de transacciones con dos modelos de subsistemas de comunicaciones*. Mónica Di Paolo. Tesina de Grado. Facultad de Ciencias Exactas. 1999. Director: Rodolfo Bertone
- [Hoffer 2002] *Modern Database Management*. Jeffrey Hoffer, Mary Prescott, Fred McFadden. Editorial: Prentice Hall. 2002
- [Gray et al., 1996] *The dangers of replication and a solution*. Jim Gray, Pat Helland, Patrick O’Neil, Dennis Shasha. SIGMOD ’96. 6/96. Montreal, Canada., pp 173-182.
- [Len et al., 2001] *Estudio de actualización de Réplicas de datos en entornos de BDD*. Sergio Len. Rodolfo Bertone. Sebastián Ruscuni. Análes: Cacic 2001. Congreso Argentino de Ciencias de la Computación. El Calafate. Octubre 2001. pp 695-706.
- [Len 2001] *Actualización de Réplicas de datos en entornos de BDD*. Sergio Len. Tesina de Grado. Facultad de Informática. 2000. Director: Rodolfo Bertone.
- [Miaton et al., 1998] *Expediencias en el Análisis de Fallas en Bases de Datos Distribuidas*. Ivana Miatón, Sebastián Ruscuni, Rodolfo Bertone, Armando De Giusti.

-
- Análes: Cacic 98. Congreso Argentino de Ciencias de la Computación. Neuquen. Octubre 1998. pp 265-276
- [Miaton et al., 1999] *Ambiente de simulación para la Recuperación en un Entorno con BDD*. Ivana Miatón, Sebastián Ruscuni, Rodolfo Bertone, Armando De Giusti. Análes: Cacic 99. Congreso Argentino de Ciencias de la Computación. Tandil. Octubre 1998.
- [Özsu et al., 1991] *Principles of Distributed Database Systems*. Second Edition. M.Tamer Özsu, Patrick Valduriez. Prentice Hall. 1991.
- [Ruscuni et al., 2000] *Evaluación de Replicación y Consistencia en Bases de Datos Distribuidas*. Sebastián Ruscuni, Rodolfo Bertone. Cacic 2000. Congreso Argentino de Ciencias de la Computación. Ushuaia, Octubre 2000. pp 145-158
- [Ruscuni 2000] *Estudio de Recuperación de errores en BDD*. Sebastián Ruscuni. Tesina de Grado. Facultad de Informática. 2000. Director: Rodolfo Bertone.
- [Silberschatz et al, 1998] *Fundamentos de Bases de Datos*. Abraham Silberschatz, Henry F. Korth, S. Sudarshan. Editorial: Mc Graw Hill. 1998. Tercera Edición
- [Simon 96] *Distributed Information Systems. From C/S to distributed multimedia*. Errol Simon. Editorial: Mc Graw Hill. 1996.
- [Stamos et al., 1993] *Coordinator log transation execution Protocol*. J. Stamos, F. Cristian. Journal of Distributed and Parallel Databases. 1993.
- [Triantafillou et al., 1995] *The location-Based Paradigm for replication: Achieving Efficiency and Availability in Distributed Systems*. Peter Triantafillou, David Taylor. IEEE Transaction on Software Engieneering, Vol 21, No. 1, January 1995.