

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

Lic. Rodolfo Bertone¹
Ing. Armando De Giusti²

L.I.D.I. (Laboratorio de Investigación y Desarrollo en informática)

Resumen

En este proyecto se trata de investigar las características de replicación y fragmentación de la información, en particular la propagación y regulación de las actualizaciones de datos. El esquema de replicación planteado en este estudio se considera estático y preestablecido, interesando el comportamiento de los métodos, en cuanto al tratamiento de lecturas y modificaciones de los datos. El soporte de simulación elegido es Java, dado que permite una mayor versatilidad en las construcciones de aplicaciones.

Las métricas propuestas están relacionadas, básicamente, con el tiempo de respuesta ante operaciones de ABM-C, bajo diferentes entornos de replicación y actualización. Los resultados dependen, además, del entorno de simulación definido. Esto es, del tasa de comunicación de cada localidad, el volumen de información que contiene, el porcentaje de cada tipo de operación (alta, baja, modificación o consulta) etc. A partir de los parámetros definidos y los resultados obtenidos es posible evaluar el comportamiento del entorno distribuido según el esquema de replicación y actualización seleccionado.

Además, la investigación se combina con estudios realizados sobre protocolos de integridad para BDD (protocolos de cometido), para integrar la solución al esquema de replicación con otro esquema que asegure integridad de los datos. De esta forma se podrá incorporar a las mediciones obtenidas, la “distorsión (overhead) de tiempo de procesamiento” que produce el proceso de aseguramiento de integridad de datos, en conjunto con las operaciones que recuperan la BD ante errores de ejecución.

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 Profesor Titular Dedicación Exclusiva, Investigador Principal CONICET – Facultad de Informática - UNLP – degiusti@lidi.info.unlp.edu.ar

Introducción

El procesamiento de bases de datos distribuidas (BDD) consiste en trabajar con información en la cual la ejecución de transacciones y la recuperación y actualización de los datos se produce en dos o más computadoras independientes, por lo general separadas geográficamente.

La utilización de bases de datos distribuidas (BDD) representa una solución viable para los usuarios cuando deben optar para la generación de sus sistemas de información. La utilización de estas bases de datos para el mantenimiento de la información requiere el estudio de una gran cantidad de casos particulares, a fin de determinar las mejores condiciones de trabajo para cada problema real.

La replicación de datos es uno de los casos particulares que se presentan cuando se estudia la utilización de BDD. Esta replicación es mucho más que simplemente copiar el dato a lo largo de varios nodos de la red. Está acompañado de análisis, diseño, implementación, administración y monitoreo de un servicio que garantiza la consistencia de la información a través de múltiples localidades en el ambiente distribuido.

El propósito de distribuir datos sobre múltiples sitios es aumentar la confiabilidad, disponibilidad y eficiencia, entre otros. En estas aplicaciones, a menudo, el acceso concurrente de una gran cantidad de usuarios al sistema de base de datos, demanda alta disponibilidad y rápidos tiempos de respuesta. La replicación es el factor clave para una mayor disponibilidad (confiabilidad) y eficiencia. Los datos replicados son almacenados en múltiples sitios para que estos puedan ser accedidos por los usuarios cuando alguna de esas copias no está disponible debido a fallas en los sitios, proporcionando mayor disponibilidad. También la performance de los sistemas de base de datos distribuidas es mejorada ubicando los datos en los lugares próximos a donde se usan y poder de esta manera acceder a copias locales, reduciendo la necesidad de costosos accesos remotos.

Objetivos

El procesamiento sobre bases de datos distribuidas continúa en evolución, por lo tanto no es posible hablar de una disciplina en plena madurez, donde para todos los problemas que se presentan existe una solución óptima y efectiva; y donde existe una caracterización única que represente una solución eficiente para aplicar en todo caso. El control de concurrencia distribuido, en comparación con uno centralizado está en pleno desarrollo. Se han propuesto una serie de algoritmos para su tratamiento y algunos de ellos cuenta con implementaciones que, si bien se adaptan a problemas específicos del mundo real, distan de ser el óptimo aplicable en todas los casos.

Además, existen problemas identificados con BDD, pero con soluciones limitadas y con falta de eficiencia. Algunos resultados teóricos son difíciles de utilizar en aplicaciones reales de BDD.

El presente trabajo evoluciona en el estudio de las principales características que debe seguir un esquema de replicación de datos. A partir de cada una de las características en juego, se definió un modelo de simulación que permite evaluar el comportamiento posible del modelo de datos y su esquema de replicación. Con los resultados obtenidos es posible evaluar distintas alternativas de solución y, de esa forma, aproximarse al esquema que mejor se adecue para el problema que se está estudiando. Básicamente, se mide el tiempo de respuesta de cada traza de ejecución en función de la configuración de replicación y actualización definida. Cada traza de ejecución se define en función de parámetros como: porcentaje de replicación de datos, porcentaje de fragmentación de datos, número de localidades intervinientes, "costo" de comunicación de cada localidad, tamaño de la traza (cantidad de transacciones involucradas),

porcentaje de operaciones de la traza (altas, bajas, modificaciones y consultas), etc. Posteriormente se pretende agregar a esta definición los parámetros que definen el porcentaje de transacciones con error, cuando se integre el trabajo con los estudios realizados sobre los protocolos de cometido.

Los objetivos específicos de la primer etapa, utilizando el ambiente desarrollado en el LID [Len 2001], son:

- ✎ 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.
- ✎ Combinar los desarrollos con los protocolos de integridad
- ✎ Reevaluar los resultados obtenidos

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.