

UNIVERSIDAD NACIONAL DEL ALTIPLANO

Facultad de Ingeniería Estadística e Informática



COMPUTACIÓN PARALELA

ARTÍCULO: Aspectos de Implementación de Servidores Web con el modelo BSP de computación paralela

Presentado por:

Grylia Yaneth Chata Iscarra

Código: 190496

Semestre: VIII - Unit II

Email: gchatai@est.unap.edu.pe

Docente: Ing:

TORRES CRUZ FRED

14 de noviembre de 2022

listings

Índice

1. RESUMEN.	3
2. INTRODUCCIÓN.	3
2.1. Modelo de computación paralela BSP.	4
3. Resultados de Implementaciones con BSP.	5
4. Conclusiones.	7
5. Trabajos Futuros.	8

xcolor blindtext multicol [utf8]inputenc chicago

graphicx

1. RESUMEN.

En el marco de la Investigación sobre Paralelización Eficiente de Servidores Web, de la Universidad Nacional de la Patagonia Austral se ha abierto una línea de investigación que da continuidad al desarrollo de servidores web soportados en clusters de PC a través del modelo BSP de computación paralela y que tiene como objetivo estudiar estrategias de implementación de servidores paralelos en entornos reales. El propósito de este trabajo es presentar los resultados alcanzados en esta línea de investigación, los desarrollos en progreso y los trabajos futuros

2. INTRODUCCIÓN.

Web se ha convertido en un recurso ubicuo para la computación distribuida, haciendo relevante la investigación de nuevos caminos para proveer acceso eficiente a los servicios disponibles en los sitios dedicados[Gesto, Laguia, Trejo, Osiris, and CanumánGesto et al.2006].

El crecimiento exponencial es un modelo de datos que muestra aumentos más grandes a lo largo del tiempo que ha experimentado desde sus comienzos en cuanto al volumen de información y al número de usuarios que la utilizan hace que la búsqueda, organización, acceso y mantenimiento de sus contenidos sea cada vez más difícil.[Ismail, Malone, and Van GeestIsmail et al.2016].

En respuesta a esta expansión de las fuentes potenciales de información, los motores de búsqueda han hecho énfasis en ampliar su velocidad y cobertura, brindando poca importancia a la eficiencia.[Martínez, Palma, and VelásquezMartínez et al.2020].

—cm

Debido a esto, diversos estudios se han abocado al desarrollo de nuevas estrategias que permitan satisfacer estas demandas a través del procesamiento paralelo [18], el cual ha demostrado ser un paradigma que permite mejorar los tiempos de ejecución de los algoritmos.[Maida and PacienzaMaida and Pacienza2015]

Particularmente nuestro grupo de investigación, integrado por docentes investigadores de la Universidad Nacional de la Patagonia Austral (Argentina) y de la Universidad de Magallanes (Chile), se ha abocado al estudio y desarrollo de herramientas de programación basadas en el modelo BSP [17, 2] de computación paralela, el cual utiliza una configuración de base de datos distribuida para acelerar las consultas.[[Rivera PoloRivera Polo2021](#)].

2.1. Modelo de computación paralela BSP.

En BSP un computador paralelo es visto como un conjunto de procesadores con memoria local e interconectados a través de una red de comunicaciones de topología transparente al usuario. En este modelo, la computación es organizada como una secuencia de supersteps. Tal como lo indica la Fig. 1, un superstep está formado por una fase en la que cada procesador puede realizar operaciones sobre datos locales únicamente y depositar mensajes a ser enviados a otros procesadores. Al final del superstep, todos los mensajes son enviados a sus destinos y los procesadores son sincronizados en forma de barrera para iniciar el siguiente superstep. Es decir, los mensajes están disponibles en sus destinos al instante en que se inicia el siguiente superstep.[[Ismail, Malone, and Van GeestIsmail et al.2016](#)].

El modelo práctico de programación paralela en BSP es el conocido SPMD (Simple Program Multiple Data), el cual es realizado mediante P copias del mismo programa corriendo en un cluster de P procesadores, cada una actuando sobre un subconjunto de los datos, donde la comunicación y sincronización de las copias es realizada mediante librerías tales como BSPLib o BSP-PUB.[[Gesto, Laguía, Sofia, Marín, and CanumánGesto et al.2007](#)]

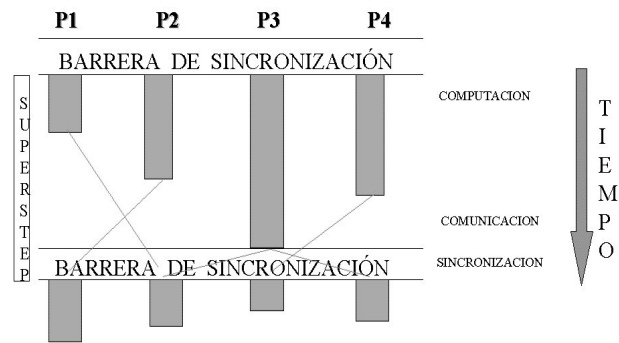


Figura 1: Figura 1: Modelo BSP y supersteps.

—cm

En el marco del Proyecto de Investigación Paralelización Eficiente de Servidores Web, de la Universidad Nacional de la Patagonia Austral se ha abierto una línea de investigación que da continuidad al desarrollo de servidores web soportados en clusters de PC a través del modelo BSP de computación paralela y que tiene como objetivo estudiar estrategias de implementación de servidores paralelos en entornos reales.

El propósito de este trabajo es presentar los resultados alcanzados en esta línea de investigación, los desarrollos en progreso y los trabajos futuros.

3. Resultados de Implementaciones con BSP.

Para la implementación de una aplicación que permitiera una aproximación a la solución para el problema planteado del acceso a grandes volúmenes de información, se realizaron varios estudios, tendientes inicialmente a la implementación de clúster y la distribución de los registros de la base de datos [9, 7, 8]. La solución adoptada implementa una base de datos distribuida del tipo relacional, sobre una plataforma computacional formada por un grupo de PC con sistema operativo Linux, los cuales están conectados en red mediante un switch de alto desempeño y en estas máquinas se ejecutan los programas con supersteps de BSP. En cada PC se encuentra instalado un administrador de bases de datos relacional llamado MySQL [15] y la librería ApiMySQL que proporciona clases que permiten enviar string con sentencias SQL desde un

programa C++ al servidor
MySQL.[[Maida and PacienzaMaida and Pacienza2015](#)]

Esta implementación es de bajo costo, ya que el software involucrado es de dominio público y el equipamiento corresponde a computadoras de escritorio de bajo costo, en contraposición a los servidores secuenciales de alto desempeño y costo. La base de datos se distribuye uniformemente para permitir el balance de carga en los procesadores involucrados. Cada superstep envía la consulta a todos los procesadores, y los servidores de bases de datos locales realizan la consulta sobre su porción de la base de datos de manera secuencial. El resultado de la consulta SQL corresponde a la unión de todas las respuestas parciales de los procesadores del cluster. La estructura del modelo BSP facilita la predicción del desempeño de programas y algoritmos. El costo de un programa está dado por la suma del costo de todos sus supersteps, donde el costo temporal de cada uno de ellos está dado por la suma del tiempo de computación sobre datos locales, el tiempo de comunicación entre procesadores y el tiempo de sincronización. Aprovechando

esta facilidad del modelo BSP se ha desarrollado una herramienta gráfica [11, 10, 14, 12, 13] que a través de metáforas visuales permite observar y administrar la cola de consultas y los supersteps ejecutados con distintos niveles de granularidad, evaluando el desempeño de cada uno de los parámetros que los integran. En los objetivos del proyecto se encuentra el desarrollo de una aplicación funcional, para lo cual se ha elegido la implementación de un Digesto Digital Institucional [5, 6] que realiza consultas complejas del tipo Join [13] y la misma posee características de Base de Datos Textual.[[Martínez, Palma, and VelásquezMartínez et al.2020](#)] Esta

herramienta de software se compone de los siguientes elementos: una aplicación web desarrollada fundamentalmente con PHP, instalada en un servidor Web con infraestructura LAMP (Linux, Apache, MySQL, PHP), un servidor de base de datos distribuida implementado con MySQL, un programa ejecutable desarrollado bajo el modelo de computación paralela BSP y un broker realizado en base a la modificación del daemon del BSP-PUB, pubd. Estos componentes se interrelacionan de la siguiente manera: en primer

xlugar, los clientes acceden a la aplicación a través del servidor web, donde elaboran la consulta a realizar. El servidor web envía la consulta al broker a través de sockets, ejecutando el programa BSP en cada procesador que forma parte del cluster. En cada procesador se ejecuta la consulta en el servidor local de base de datos MySQL y, cuando se produce la etapa de sincronización del superstep BSP, los resultados son enviados a la máquina broker, quien los agrupa y envía al servidor Web a través de los sockets definidos para que de esta manera el cliente pueda obtener los resultados requeridos.[[Maida and PacienzaMaida and Pacienza2015](#)]

4. Conclusiones.

En este trabajo se ha presentado una de las líneas de investigación del Proyecto de Investigación Paralelización Eficiente de Servidores Web, de la Universidad Nacional del Altiplano que da continuidad al desarrollo de servidores web soportados en clusters de PC a través del modelo BSP de computación paralela y que tiene como objetivo estudiar estrategias de implementación de servidores paralelos en entornos y aplicaciones reales. Los trabajos realizados presentan una solución concreta al problema de acceso a grandes volúmenes de datos a través de la Web, mediante el desarrollo de varios componentes que conforman un motor de búsquedas paralelo, con acceso a una base de datos distribuida, implementando el modelo de computación paralela BSP, en particular a través de la librería BSP-PUB. Este modelo soporta una metodología estructurada de diseño de software que es simple de utilizar y permite el uso de tecnología existente y gratuita.

La distribución de los registros influye significativamente en la velocidad de respuesta de las consultas, obteniendo hasta el momento los mejores resultados con una distribución uniforme. Se han desarrollado herramientas gráficas para visualizar y administrar el desempeño del servidor de base de datos, mediante distintas metáforas visuales, que permiten observar con distintos niveles de granularidad las consultas realizadas y los supersteps generados al largo de la ejecución de la aplicación. Actualmente nos encontramos estudiando distintas alternativas de implementación de la cola de consultas [16, 3, 4] para mejorar el rendimiento ajustando parámetros y características de la cola de consultas

SQL, de manera de tomar en cuenta el tiempo de espera de los clientes que han generado las consultas para maximizar su satisfacción al realizar un acceso a la aplicación.[Maida and PacienzaMaida and Pacienza2015]

5. Trabajos Futuros.

Entre nuestros trabajos futuros se encuentran el estudio de la recuperación de la base de datos distribuida a raíz de la caída de una máquina del cluster, mediante la replicación de registros u otra estrategia. También es posible su aplicación a redes de Datos del tipo Grid, en el supuesto de la necesidad de contar con una base de datos ditribuida geográficamente para respetar la autonomía de las organismos intervinientes para una aplicación dada, tal como el Digesto Digital Institucional.[Rivera PoloRivera Polo2021]

Referencias

- [Gesto, Laguía, Sofia, Marín, and CanumánGesto et al.2007] Gesto, E., D. Laguía, O. Sofia, M. Marín, and J. Canumán (2007). Aspectos de implementación de servidores web con el modelo bsp de computación paralela. In *IX Workshop de Investigadores en Ciencias de la Computación*.
- [Gesto, Laguia, Trejo, Osiris, and CanumánGesto et al.2006] Gesto, E., D. Laguia, N. B. Trejo, S. Osiris, and J. Canumán (2006). Implementación de un digesto digital paralelo. In *VIII Workshop de Investigadores en Ciencias de la Computación*.
- [Ismail, Malone, and Van GeestIsmail et al.2016] Ismail, S., M. S. Malone, and Y. Van Geest (2016). *Organizaciones exponenciales*. Bubok.
- [Maida and PacienzaMaida and Pacienza2015] Maida, E. G. and J. Pacienza (2015). Metodologías de desarrollo de software.
- [Martínez, Palma, and VelásquezMartínez et al.2020] Martínez, R., A. Palma, and A. M. Velásquez (2020). Revolución tecnológica e inclusión social:

reflexiones sobre desafíos y oportunidades para la política social en américa latina.

[Rivera PoloRivera Polo2021] Rivera Polo, F. (2021). Creación de la universidad de aysén en la patagonia chilena y su impacto en el desarrollo regional. estudio de caso.