การเพิ่มประสิทธิภาพของเว็บเซิร์ฟเวอร์ ด้วยระบบคลัสเตอร์สองตัวกระจายงาน

High Performance Web Server by Server Clustering with Two Load Balancers

กายรัฐ เจริญราษฎร* ชาคริต มณีงาม

ภาควิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กำแพงแสน นครปฐม 73140
*Corresponding author; E-mail: kairat.j@ku.ac.th

บทคัดย่อ

ปัจจุบัน การใช้งานอินเทอร์เน็ตเพิ่มมากขึ้นอย่างต่อเนื่อง ส่งผลให้เครื่องแม่ข่าย (Server) ต้องรองรับการใช้บริการจากผู้ใช้เป็นจำนวนมากจนมักทำให้เครื่องแม่ข่ายประสบปัญหาไม่สามารถให้บริการได้อย่างมีประสิทธิภาพ ดังนั้นผู้วิจัยจึงได้ทดลองใช้ระบบคลัสเตอร์มาเพิ่มประสิทธิภาพให้กับเครื่องแม่ข่าย โดยได้สร้างเป็นระบบคลัสเตอร์ให้แบ่งภาระงานระหว่างหลายเครื่องแม่ข่ายซึ่งทำงานด้วยระบบปฏิบัติการลินุกซ์ (Linux) ที่มีตัวกระจายงาน (Load Balancer) สองตัวเพื่อเสถียรภาพของระบบที่มากขึ้น จากนั้นทำการทดสอบหาค่าประสิทธิภาพในด้านการรองรับการร้องขอ (Requests) และอัตราการโอนถ่ายข้อมูล (Transfer rate) พบว่า เครื่องแม่ข่ายที่ทำระบบคลัสเตอร์จะมีประสิทธิภาพเพิ่มขึ้น ทั้งด้านอัตราการรองรับการร้องขอ และด้านอัตราการถ่ายโอนข้อมูล โดยเครื่องแม่ข่ายเพียงเครื่องเดียวนั้นสามารถรองรับการร้องขอได้ 2,301.61 Requests/sec และมีอัตราการโอนถ่ายข้อมูลที่ 707.71 Kbytes/sec แต่ระบบคลัสเตอร์ที่ประกอบด้วยเครื่องแม่ข่าย 2 เครื่องนั้น จะสามารถรองรับการร้องขอได้ 4,617.34 Requests/sec และมีอัตราการโอนถ่ายข้อมูลได้ 1401.28 Kbytes/sec ซึ่งประมาณได้เป็นสองเท่าของการใช้เครื่องแม่ข่ายเครื่องเดียว ส่วนระบบคลัสเตอร์ที่ประกอบด้วยเครื่องแม่ข่าย 3 เครื่องนั้นจะสามารถรองรับการร้องขอได้เพิ่มขึ้นเป็น 6,204.47 Requests/sec และมีอัตราการโอนถ่ายข้อมูลได้ 1,875.66 Kbytes/sec ซึ่งประมาณได้ใกล้เคียงสามเท่า ดังนั้น จึงสรุปได้ว่าประสิทธิภาพของระบบคลัสเตอร์นี้จะเพิ่มขึ้นตามจำนวนเท่าของเครื่องแม่ข่ายภายใต้สภาพแวดล้อมที่ทำการทดสอบ

คำสำคัญ: คลัสเตอร์ ลินุกซ์ แอลวีเอส เว็บเซิร์ฟเวอร์

Abstract

Currently, the Internet is increasing constantly, then the Internet servers must support a lot of request from clients that often causes Internet servers serve ineffectively. To solve these, we have enhanced the system by using a server cluster. We have built this clustering system that can balance the workload to several Linux servers by two distribution machine (Load Balancer). We determined the efficiency of the requests received and the data transfer rate and then we found that the clustering system increase the efficiency in both the rate of support requests and the data transfer rate. The server without clustering system can accommodate 2,301.619 requests/sec and a data transfer rate of 707.71 Kbytes/sec, but the cluster system consisting two servers can accommodate 4,617.34 requests/sec. and a data transfer rate of 1401.28 Kbytes/sec, which is approximately twice that of using a single server. The clustering system consisting of three machines can accommodate 6,204.47 requests/sec and a data transfer rate of 1,875.66 Kbytes/sec, which is about as close three times of using a single server. Therefore, it is concluded that the performance of this clustering system is increased by the number of the servers in the system.

Keyword: Cluster, Linux, LVS, Web server

บทนำ

ด้วยความเจริญทางเทคโนโลยีอินเทอร์เน็ตในปัจจุบันที่ก้าวหน้าอย่างรวดเร็ว ส่งผลให้มีผู้ใช้งานเครือข่ายอินเทอร์เน็ตเพิ่มขึ้นเป็นจำนวนมาก และมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง โดยการใช้บริการส่วนใหญ่จะเป็นบริการในรูปแบบ World Wide Web (WWW) [1] ที่จะให้บริการโดยเครื่อง Web server [2] ซึ่ง Web Server คือเครื่องคอมพิวเตอร์ที่ติดตั้งโปรแกรมให้ให้ทำหน้าที่ให้บริการข้อมูล แก่ Client หรือก็คือเครื่องคอมพิวเตอร์ลูกข่ายที่ขอรับบริการในรูปแบบสื่อผสมผ่านระบบเครือข่าย โดยสามารถแสดงผลผ่านเว็บเบราว์เซอร์ ซึ่งการให้บริการของ Web server เพียงเครื่องเดียวนั้น เมื่อมีผู้ใช้บริการอินเทอร์เน็ตจำนวนมากเกินไปจนทำให้ระบบไม่ตอบสนองหรือ ระบบไม่สามารถใช้งานได้ สำหรับเว็บไซต์ที่ต้องการให้บริการตลอด 24 ชั่วโมง และต้องการอัตราการใช้งานไม่ได้ ของระบบ (Down Time) ต่ำที่สุดนั้นจะต้องดูในองค์ประกอบหลายส่วนด้วยกัน [3] เช่น ระบบไฟฟ้าที่ต้องมั่นใจว่ามีระบบสำรองเพียงพอ ระบบเครือข่ายที่มีลิงค์สำรองในกรณีลิงค์หลักใช้งานไม่ได้ และมีเครื่องสำรองที่สามารถทำงานได้ทันทีเมื่อเครื่องเว็บเซิร์ฟเวอร์หลักให้บริการไม่ได้ เป็นต้น ดังนั้นจึงได้เกิดการทำระบบคลัสเตอร์เซิร์ฟเวอร์เพื่อแก้ปัญหาต่าง ๆ ดังที่ได้กล่าวมา

คลัสเตอร์ [4] คือ กลุ่มของคอมพิวเตอร์ที่ถูกกำหนดให้ทำงานร่วมกันเพื่อให้ได้ประสิทธิภาพในการทำงานสูงขึ้น ระบบคลัสเตอร์ในปัจจุบันสามารถจำแนกได้ 3 รูปแบบคือ 1) คลัสเตอร์ประสิทธิภาพสูง (High Performance Cluster: HPC) ใช้สำหรับประมวลผลกับข้อมูลขนาดใหญ่ที่ต้องการกำลังประมวลผลสูงโดยให้ได้ผลลัพธ์ในเวลาอันรวดเร็ว ซึ่งส่วนมากเป็นงานทางด้านวิทยาศาสตร์และวิศวกรรมศาสตร์ 2) คลัสเตอร์ประมวลผลปริมาณงานสูง (High Throughput Cluster: HTC) ใช้สำหรับประมวลผลกับงานที่มีจำนวนมาก โดยที่

แต่ละงานเป็นอิสระต่อกัน ใช้เวลาประมวลผลนานเป็นเดือนหรือปี โดยเป็นงานทางด้านวิทยาศาสตร์และวิศวกรรมศาสตร์ แต่ส่วนใหญ่จะเป็นงานวิจัยทางด้านวิเคราะห์ทางเคมี ชีววิทยาและฟิสิกส์ 3) คลัสเตอร์ที่มีสภาพพร้อมใช้งานสูง (High Availability Cluster : HA) เพื่อใช้เพิ่มเสถียรภาพของการประมวลผลใช้สำหรับงานที่สำคัญและหยุดการประมวลผลไม่ได้ โดยต้องการประมวลผลคู่ขนานจำนวน 2 หน่วยประมวลผล เป็นอย่างน้อยซึ่งส่วนใหญ่เป็นงานทางด้านธุรกิจการเงินและธนาคาร ดังตัวอย่างโครงงานของประกายนาดี [5] ได้นำเสนอการออกแบบสถาปัตยกรรมระบบเว็บคลัสเตอร์ที่สามารถคงทนต่อความผิดพลาดและขยายระบบได้และสามารถรองรับการเพิ่มประสิทธิภาพให้ระบบได้ด้วยการเพิ่มเติมโปรแกรมเข้าไปในระบบในลักษณะโมดูลนอกจากนั้นโครงงานนี้ได้เสนอกระบวนการกระจาย ภาระงาน 2 วิธี เพื่อเพิ่มประสิทธิภาพในการให้บริการของเว็บคลัสเตอร์ และโครงงานของ ธนาวุฒิ ธนวาณิชย์ ภูมิพงษ์ ดวงตั้ง และ อธิคม ศิริ [4] ซึ่งเป็นระบบประมวลผลแบบขนานโดยวิธีการส่งผ่านข้อความ (Message Passing) ผ่านเครือข่ายภายในกลุ่มของคอมพิวเตอร์ที่เรียกว่า ระบบประมวลผลแบบคลัสเตอร์ทำให้กลุ่มคอมพิวเตอร์นั้น ๆ มีประสิทธิภาพสูงขึ้นมาก ซึ่งจากการศึกษาทั้ง 2 โครงงานนี้ผู้วิจัยพบว่า ถ้าเครื่องกระจายงาน (Load Balancer) เกิดความผิดพลาด ไม่สามารถใช้งานได้ ระบบทั้งหมดก็จะไม่สามารถใช้งานได้ตามไปด้วย

ดังนั้นงานวิจัยนี้จึงมุ่งเน้นแนวทางของคลัสเตอร์ โดยจะนำทั้งระบบคลัสเตอร์ประมวลผลปริมาณงานสูง และระบบคลัสเตอร์ที่มีสภาพพร้อมใช้งานสูง มาเพื่อเพิ่มประสิทธิภาพของเว็บเซิร์ฟเวอร์ให้รองรับผู้ใช้งานจำนวนมากได้ อีกทั้งยังมีระบบสำรองการกระจายงาน กรณีที่เครื่องกระจายงานตัวหลักใช้งานไม่ได้อีกด้วย

ภาพรวมของระบบ

การติดตั้งระบบได้ใช้ระบบ LVS (Linux Virtual Server) [6] ซึ่งประกอบด้วยโครงสร้าง 4 ลำดับชั้น ดังแสดงในรูปที่ 1

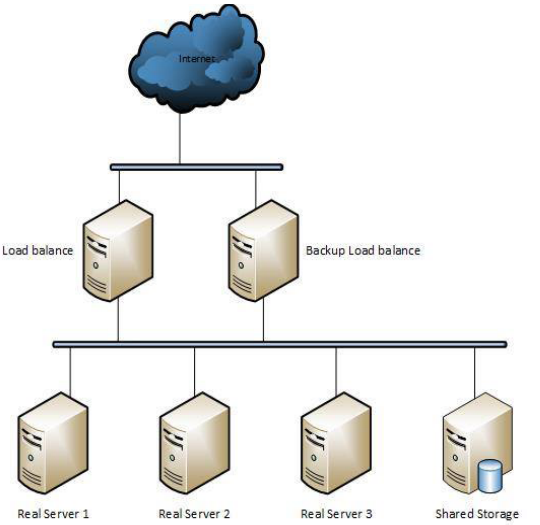

Figure 1. System Structure

Load Balancer ทำหน้าที่รองรับการเชื่อมต่อจากลูกข่ายในการให้บริการของระบบ LVS ของระบบทั้งหมด โดยที่ลูกข่ายจะเห็นว่าการให้บริการนั้นมาจาก Load balancer เพียงตัวเดียวเท่านั้น โดย Backup Load Balancer จะทำหน้าที่ทำงานแทน Load Balancer ในกรณีที่ Load Balancer ตัวหลักไม่สามารถทำงานได้

Server Cluster เป็นกลุ่มของเครื่องแม่ข่ายที่ให้บริการจริงแก่เครื่องลูกข่าย เราเรียกว่า Real Server ซึ่งเป็นเครื่องที่ให้บริการทางอินเทอร์เน็ตต่าง ๆ เช่น Web Mail FTP และ DNS เป็นต้น โดยมี Shared Storage เป็นส่วนที่ให้บริการข้อมูลระหว่าง Real server ด้วยกันเอง เพื่อรับประกันว่า Real server ทุกตัวที่ให้บริการนั้น จะมีข้อมูลเหมือนกัน ซึ่งอาจเป็นการใช้เครือข่ายฐานข้อมูล (Database server) NFS (Network File System) และ Distributed File System โดยที่เมื่อมีการใช้แบบ Database System นั้น ข้อมูลจะมีการเปลี่ยนแปลงโดยอัตโนมัติ เมื่อมีการเขียนหรืออ่านข้อมูลจากระบบ จึงรับประกันได้ว่า Real Server นั้น จะมีข้อมูลที่เหมือนกันทุก ๆ ตัว

สำหรับข้อมูลที่เป็นแบบ Static Data หรือเป็นการอ่านเขียนแบบไฟล์ปกติทั่วไป เช่น ไฟล์บนบริการเว็บต่างๆ ได้แก่ PHP Script หรือ HTML Script อาจจะมีการใช้ NFS (Network File System) [7] ในการแชร์ข้อมูลระหว่าง Real Server ซึ่งจะมีข้อจากัดในด้านประสิทธิภาพของการแชร์ข้อมูลไม่เกินประมาณ 8 ตัว ในกรณีที่ข้อมูลใหญ่กว่านี้อาจจะมีการใช้ Distributed/Cluster File System แทน

วิธีการวิจัย

ฮาร์ดแวร์ที่ใช้ในการทดลอง ประกอบไปด้วยคอมพิวเตอร์ทั้งหมด 7 เครื่อง ได้แก่ Load Balancer 1 เครื่อง, Backup Load Balancer 1 เครื่อง, Real Server 3 เครื่อง, Shared Storage Server 1 เครื่อง, เครื่องลูกข่าย 1 เครื่อง โดยเชื่อมต่อกันเป็นเครือข่ายแลนด้วย LAN Switch

ส่วนซอฟต์แวร์ที่ใช้ในการทดลอง ได้แก่ ระบบปฏิบัติการ CentOS Linux, Linux Virtual Server สำหรับทำระบบคลัสเตอร์, และ Apache Web Server, Heartbeat ใช้ให้สามารถมีตัวกระจายงานได้ 2 ตัว, ระบบเรียกใช้ข้อมูลผ่านเครือข่าย Network File System และโปรแกรมทดสอบประสิทธิภาพ Apache Bench

ในงานวิจัยนี้ได้ทำการเปรียบเทียบและพิสูจน์ประสิทธิภาพความสามารถในการรองรับการร้องขอจากเครื่องลูกข่าย และอัตราการถ่ายโอนข้อมูลของเครื่องแม่ข่ายในระบบระบบคลัสเตอร์ว่าดีขึ้นกว่าระบบเครื่องแม่ข่ายธรรมดา โดยในการทดลอง ผู้วิจัยใช้เครื่องพีซีเก่ามาเป็นเครื่องแม่ข่าย และเครื่องกระจายงาน ด้วยซีพียู Intel Pentium 1.7 GHz, หน่วยความจำหลัก 512 MByte และเชื่อมต่อเครือข่ายแลนด้วยความเร็ว 100 Mbps ส่วนเครื่องมือที่ใช้วัดประสิทธิภาพ คือ Apache Bench เวอร์ชั่น 2.0.40 ในการสร้างภาระงานให้แก่ระบบ ค่าใช้วัดประสิทธิภาพของระบบเครื่องแม่ข่าย คือ การรองรับการร้องขอ (Requests) และอัตราการโอนถ่ายข้อมูล (Transfer rate) โดยหลังจากเตรียมระบบที่ใช้ทดลองตามโครงสร้างของระบบแล้วจึงสร้างภาระงานด้วยคำสั่ง ab –n 100 –c 20 http://192.168.1.100/

จากคำสั่งดังกล่าวสามารถอธิบายได้ว่า ต้องการทดสอบเครื่องแม่ข่าย http://192.168.1.100/ จำนวนทั้งหมด 100 ครั้ง แต่ให้ส่งการทดสอบพร้อม ๆ กัน ครั้งละ 20 การร้องขอ หลังจากทดสอบแล้วจะได้ค่าการรองรับการร้องขอ และอัตราการโอนถ่ายข้อมูล เพื่อนำมาวิเคราะห์ต่อไป

ผลการวิจัย

จากการทดลองสามารถนำค่าที่ได้มาเปรียบเทียบประสิทธิภาพของเครื่องแม่ข่ายธรรมดา กับเครื่องแม่ข่ายที่ใช้ระบบคลัสเตอร์ โดยนำค่าเฉลี่ยการรองรับการร้องขอ (Requests) และอัตราการโอนถ่ายข้อมูล (Transfer rate) มาเปรียบเทียบกันได้ดังตารางที่ 1 และตารางที่ 2 ตามลำดับ

Table 1. Comparison of efficiency received data rate from cilent

ครั้งที่	Server					
	Real server 1 (Request/sec)	Real server 2 (Request/sec)	Real server 3 (Request/sec)	Average of 3 Real server (Request/sec)	Cluster servers	
					2 servers (Request/sec)	3 servers (Request/sec)
1	2162.02	2058.42	2173.87	2131.44	4607.35	6256.26
2	2472.55	2339.73	2481.02	2431.10	4701.36	6145.23
3	2442.84	2632.20	1951.83	2342.29	4543.32	6211.92
average	2359.13	2343.45	2202.24	2301.61	4617.34	6204.47

Table 2. Comparison of efficiency received data rate from servers

ครั้งที่	Server					
	Real server 1 (Kbytes/sec)	Real server 2 (Kbytes/sec)	Real server 3 (Kbytes/sec)	Average of 3 Real server (Kbytes/sec)	Cluster servers	
					2 servers (Kbytes/sec)	3 servers (Kbytes/sec)
1	611.61	744.48	763.92	706.67	1424.29	1873.54
2	611.40	743.71	755.10	703.40	1374.99	1855.41
3	602.77	771.37	765.05	713.06	1404.56	1898.05
average	608.59	753.18	761.35	707.71	1401.28	1875.66

การอภิปรายผล

จากค่าเฉลี่ยของตารางที่ 1 และ 2 จะเห็นได้ว่าเครื่องแม่ข่ายเพียงเครื่องเดียวนั้นสามารถรองรับการร้องขอได้ 2,301.61 Requests/sec และมีอัตราการโอนถ่ายข้อมูลจากเครื่องแม่ข่ายไปยังลูกข่าย ที่ 707.71 Kbytes/sec แต่ระบบคลัสเตอร์ที่ประกอบด้วยเครื่องแม่ข่าย 2 เครื่องนั้น จะสามารถรองรับการร้องขอได้ 4,617.34 Requests/sec และมีอัตราการโอนถ่ายข้อมูล 1401.28 Kbytes/sec ส่วนระบบคลัสเตอร์ที่ประกอบด้วยเครื่องแม่ข่าย 3 เครื่องนั้น จะสามารถรองรับการร้องขอได้เพิ่มขึ้นเป็น 6,204.47 Requests/sec และมีอัตราการโอนถ่ายข้อมูลได้ 1,875.66 Kbytes/sec

จากผลการทดลองสามารถสรุปได้ว่า การใช้ระบบคลัสเตอร์ที่ประกอบด้วยเครื่องแม่ข่าย 2 เครื่องนั้นมีประสิทธิภาพการรองรับการร้องขอเพิ่มขึ้นเป็น 200.61% และอัตราการโอนถ่ายข้อมูลเพิ่มขึ้นเป็น 198.03% จากการใช้เครื่องแม่ข่ายเครื่องเดียวที่ไม่ได้ทำระบบคลัสเตอร์ และเมื่อเพิ่มเครื่องแม่ข่ายเป็น 3 เครื่อง จะทำให้มีประสิทธิภาพการรองรับการร้องขอเพิ่มขึ้นเป็น 269.57% และได้อัตราการโอนถ่ายข้อมูลเพิ่มขึ้นเป็น 265.03% จากการใช้เครื่องแม่ข่ายเครื่องเดียวที่ไม่ได้ทำระบบคลัสเตอร์ นอกจากนั้นระบบคลัสเตอร์นี้ยังมีเสถียรภาพสูง เพราะเมื่อทดลองปิดเครื่องแม่ข่ายบางตัว หรือแม้กระทั่งปิดการทำงานของตัวกระจายงานหลัก ระบบก็ยังสามารถทำงานต่อได้ด้วยเครื่องแม่ข่ายที่เหลือ และตัวกระจายงานสำรองตามลำดับ

สรุป

จากการทดลองระบบคลัสเตอร์ สามารถสรุปได้ว่า การทำระบบคลัสเตอร์ให้กับเครื่องแม่ข่ายมีประสิทธิภาพเพิ่มขึ้นกว่าการไม่ทำระบบคลัสเตอร์เกือบเท่ากับจำนวนเท่าของเครื่องแม่ข่ายที่นำมาทำระบบคลัสเตอร์ แต่ประสิทธิภาพที่ควรจะได้เพิ่มขึ้นนี้มีขีดจำกัด คือเมื่อเพิ่มจำนวนเครื่องแม่ข่ายมากขึ้นจนเกินขีดจำกัดของเครื่องกระจายงาน ดังในการทดลองของงานวิจัยนี้ ที่ใช้เครื่องแม่ข่าย 3 เครื่อง จะพบว่าประสิทธิภาพการรองรับร้องขอน้อยลงไปกว่าที่ควรจะเป็นคือ 300% ถึง 31.43% และอัตราการโอนถ่ายข้อมูลน้อยกว่า 300% อยู่ 34.97% ซึ่งเป็นเพราะประสิทธิภาพที่ไม่เพียงพอของเครื่องกระจายงานที่ใช้ แม้กระนั้นก็ยังถือว่าการทำระบบคลัสเตอร์ทำให้ประสิทธิภาพการให้บริการที่สูงขึ้นทั้งในด้านปริมาณการรับการร้องขอจากเครื่องลูกข่ายและอัตราการถ่ายโอนข้อมูล ทั้งยังสามารถเพิ่มเครื่องแม่ข่ายให้กับระบบคลัสเตอร์เมื่อต้องการขยายระบบได้ง่าย และมีเสถียรภาพสูงเพราะมีเครื่องแม่ข่ายและเครื่องกระจายงานหลายเครื่อง

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนการวิจัยจากคณะวิศวกรรมศาสตร์ กำแพงแสน มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน

เอกสารอ้างอิง

1. Saiseesod S. (2006). The role of the Internet on public relations. All book Publishing. Bangkok, (In Thai).
2. Sanguanpong S, Numprasertchai S. (2000). Scalable Cluster Web Server with Load-Balancing. NRCT Report. Kasetsart University, (In Thai).
3. Rodkhongtee S. (2009). Installation and usage of CentOS Linux server. [E-book Online], [cited 2013 Nov 5]. Available from: URL: http://linux.sothorn.org/node/558. (In Thai).
4. Thanavanich T, Duongtang P, Siri A. (2009). Computing System with Rocks Clusters. Research Article. Chiangrai Rajabhat University, (In Thai).
5. Nadee P. (2003). Effective Load Balancing Algorithm for Scalable Web Cluster [MEng thesis]. Bangkok: Kasetsart University, (In Thai).
6. Amandossov, A. (2011). Building LINUX Virtual Server by Network Address Translation Technology. *Computer Modelling and New Technologies.* 15: 58-65.
7. Nelson, M. N., Welch, B. B. and Ousterhout, J. K. (1988). Caching in the Sprite network file system. *ACM Transactions on Computer*, 6: 134-154.