Proceedings of the 5th WSEAS Int. Conf. on SIMULATION, MODELING AND OPTIMIZATION, Corfu, Greece, August 17-19, 2005 (pp102-105)

University Grid: A Case study of Bangkok University Grid

PUKDESREE, SORAPAK
SUKSTRIENWONG, ANON, Asst.Prof.
LACHAROJ, VITALWONHYO
School of Science and Technology
Bangkok University
Rama IV rd., Klongtoey, Bangkok
THAILAND

Abstract: - In order to provide computer facilities for almost 20,000 students plus a thousand of faculty, university have to spend a lot of budget. The time has changed, the technology has changed either. Most of computer utilize only 5-20% of peak time, so they are almost idle. So the previous system might not have enough capacity to support the use of end-user in term of number of users and number of applications used. The new budget must be provided for the new system. Instead of spend more budget, more advance techniques such as clustering, load balancing and grid technology has been raised to improve the system performance by using current system. Our team has researched and tried to implement these techniques deployed into current system With more than 3,000 PC computers, we would have a supercomputer as powerful as 3TFlopps in performance. We have developed and implemented a small model of grid system. The process is on the progress. We would like our case to show the advantages aspect to others especially in Thailand.

Key-Words: - Grid technology, Clustering, University Grid, High Performance Computing.

Introduction

This paper presents the basic structure of Grid Technology for university environment. Most of Universities provide infrastructure of information technology such as computers, peripherals and network infrastructure to support end-user applications, e-services, security and administrative services. For end-user applications, university provides several thousands of PC computers that have adequate performance for end-user applications and activities. For many of transactions and activities, users or information such as e-services or client-server applications, most of universities utilize mid-range computer such as mini computers or high-end workstations that should have enough performance to support e-services.

Universities have spent a lot of budget to provide PC computers, and network infrastructure including the cost of maintenance. While computer technology has been changed rapidly, so the technology you have purchased today will cost less than 50-70% in the next few years. Therefore universities will loose a lot of money in the near few years. Furthermore, to provide enough capacity performance of computer technology, universities will have to spend a lot more budget to buy the new PC computers or to invest with new infrastructure such as networking and peripherals to implement new system. Also the cost of software such as operating system and application software deployed to the new system. Several thousands of PC computers in university environment, most of them are not worked together and separated on utilizing. The time's utilizing of PC computers almost idle, the utilizing ratio is quit very low approximately 5-20% of the peak time. Also the mid-range computers or high-end workstations used to be servers both Windows Server and also Linux server is utilized about 15% and mainframe computer is utilized about 15% of the peak time.

Problem Solution

If universities can collaborate the utilization of their PC computers in some of applications that would be increase the utilization ratio. Furthermore, universities will get the advantages from those applications used. Therefore several techniques such as cluster technology and grid technology have been approached. Cluster technology is a method to create a virtual super computer by grouping several PC computers connected with network connections. From the user point of view, they will use the PC cluster as a superpower computer. For the hardware and network point of view, there may have several PC computers connected together using network media. The characteristics of clustering Technology are homogenous architecture and tightly coupled. In

term of homogenous architecture means most of computers in the cluster would be the same of hardware architecture such CPU's architecture and also software architecture would the same architecture such as Linux or Windows. Hence, clustering technique has some constraints to deploy. Before utilize the technology you need to learn that technology meet your requirements. However, clustering Technology can improve system performance in term of computational and storage. Jobs, which will be processed in the clustering system, will be decomposed in to smaller pieces. Then distribute those pieces to each computational node within the cluster in parallel. That means the processing time to complete those jobs would be less than previous depending on applications used either. To improve storage of data and information, some operating systems provide tools such as software-raid with RAID-0, RAID-1, or RAID-5 that can improve the capacity of storage and disk operations to retrieve data or information with RAID system. RAID also improves fault tolerant and high availability of information.

Grid technology is another technique approached to collaborate up university's PC computers as a virtual supercomputer. But grid technology is more flexible, more reliable, and more secure. The characteristics of Grid technology are heterogeneous, lightly coupled, resource registry and security service. Computers in the grid system could be difference in term of hardware and software architecture such as Pc computers or mid-range workstation. Compute nodes in the grid system can be used in different places but they can work together collaboratively to perform virtual organization (VO). Grid technology is more flexible, since compute nodes in grid can be increased or deceased dynamically to improve performance or to downsizing the system. Grid Technology can be used to integrate multiplatform of operating system such Windows, Linux, McOS or Solaris. One of the importance things to be concerned is to manage distributed resources with performance, reliable and security. Grid technology can be used to provide difference services such as Computing Service, Application Service, Data Services, Information Service or Authentication Service. Grid technology is a distributed system that all of services is done in parallel.

From the above reasons, we have researched and working to implement grid technology in Bangkok University because grid technology can improve the current performance with low cost to deploy. Bangkok University have spent a lot of budget to provide computer infrastructure such as PC computers, peripherals and network infrastructure per year to support the use of our 17,000 – 20,000 students and a couple thousands of faculties. Bangkok University has deployed several of e-service such as online student's registration, grade report, score announcement, withdraw. E-services for faculties are individual information, time attendance, working record, announcement and etc. Bangkok University has spent a lot of budget for many of PC computers that have low capacity each year. We are a researching group to deploy technology to produce high performance computing with low cost. We have started to implement cluster computing and load balancing in our limited spaces using Linux and Windows system. Later we have researched and learned a lot more about grid technology. So, we have started our new project, Bangkok University's Grid, since 2004. We have set up our two local grid system within the same spaces. Our systems are still in our lab because we are in the progress to configure the system. We also need to find some more applications onto the system.

Hardware and Network Architecture

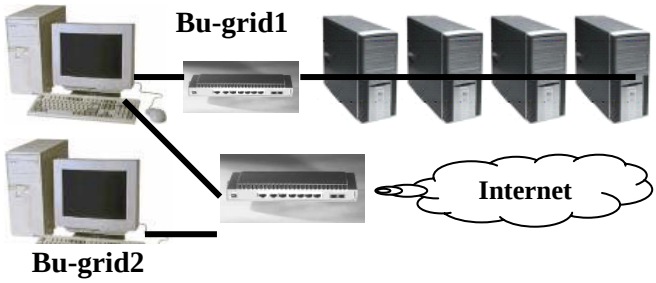


Figure1: Bu-grid1 with 4 grid nodes and Bu-grid2 without grid node.

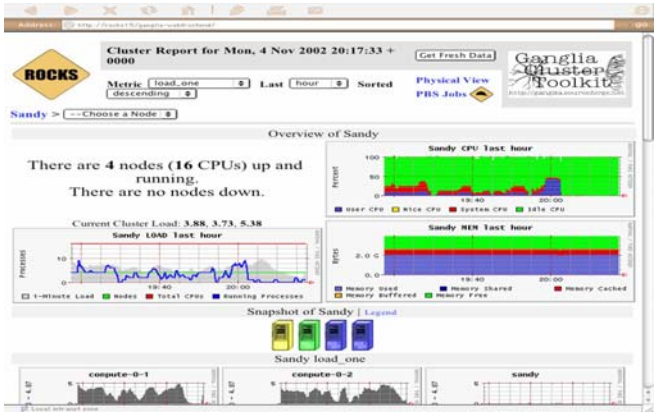
Our systems were designed to deploy grid technology in the system. We have implemented two grid systems because we would like to try to joint our grids together. We will joint our grid system to other grids outside the Bangkok University such as Thai grid, or APGrid. In Thailand there are about 12 organizations joint their together as Thai grid. Figure 1 shows a grid system with four compute nodes. (named bu-grid1). This is our main grid system. Figure 2 shows another grid system with no compute nodes. (named bu-grid2) that is our test system to joint with bu-grid1.

Bu-grid1 system composes of 5 PC computers, one used for front-end server, the controller of system,

acts as gateway to connect system to internet, also waits for request from user, and then distributes user's requests to grid nodes by decompose the jobs request into smaller pieces. When each compute nodes finished their jobs, they will send result back to front-end server. Then front-end composes those pieces of result and sends them back to user. The system has four grid nodes. Bu-grid2 has only one PC computer which is used to test to joint with bu-grid1. Under constraints in term of budget, resources and manpower, our PC computer's specifications are vary such as [AMD Sempron +2200, 1 GB RAM, 80GB], [AMD Athlon XP 900, 512 MB RAM, 40GB], and [Pentium II 350, 256 MB RAM, 20GB].

Each grid nodes are connected together with 24 ports 10 Mbit hub, even though each NIC card has maximum capacity of 100 Mbit. Both bu-grid1 and bu-grid2 are connected to other systems using 8-ports fast Ethernet 10/100 Mbit hub. Bu-grid1 has two network interface cards. One NIC is used to connect other systems and the other NIC is used to connect local grid nodes. That can reduce network traffic handled only one NIC card.

Software Architecture



Linux Operating System - ROCKS Cluster Distribution

Figure2: Grid Monitoring system (ROCKS Cluster)

The system has been implemented on a PC computer based on ROCKS Cluster Distribution. ROCKS The Rocks development community includes the Cluster Development Group at San Diego Supercomputer Center, Scalable Systems in Singapore, the HPC Group at University of Tromso in Norway, the SCE Group at Kasetsart University in Thailand, and the cluster development group at KISTI in Korea. Development of Rocks is funded from NSF, and aided by generous equipment donations from Sun Microsystems, Dell, AMD, Infinicon Systems, and Intel. The Rocks project was started at SDSC in early 2000, with the goal of "making clusters easy". Our current system is used ROCKS 4.0.0 as our operating system and tools. ROCKS 4.0.0 is based on Cent OS which is fundamentally as Redhat enterprise Linux 4 (RHEL 4). ROCKS 4.0.0 provides security, server services, and stability as RHEL 4. The grid configuration of ROCKS 4.0.0 is not too complicate. We are also familiar with RHEL 4 because in March 2005, Bangkok University had signed agreement with Redhat Inc., to set up Redhat Academy Program. This will be the first University in South East Asia. Redhat will provide materials and instructions for faculty of Bangkok University. The University will use all materials to teach students in South East Asia who are interested in Redhat or Linux operating. Redhat is one of the most growing up rate of commercial Linux operating system. Redhat is based on Fedora Project, community of Linux. But Fedora is free without supporting.

Conclusion

We have finished up the process of grid configuration of both bu-grid1 and bu-grid2. Our two systems can now be jointed together. We can submit jobs and distribute them to computational nodes in the system successfully. Currently we use our grid only for our web server and to test the grid configurations. We also use the grid system as our web server. We plan to do some more application such as grid database applications by using Oracle 10g or may be MySql clustering. Both are distributed database system and would be compared to each other to discover their performance. From the research, we have found that there are many applications on grid technology that we would like to discover such as drug design, access grid or rendering on grid. If Bangkok University agrees to deploy grid technology, with more than 3,000 PC computers, we would have a supercomputer as powerful as more than 3TFlops in performance.

We are amateur researchers in grid technology. All information we have searched from the internet. Grid technology in Thailand is also new and few researchers know and would deploy the technology. Even though there are many success cases of universities utilizing grid in US or EU. In our cases we also want the case to be a good case of university to deploy grid technology. We have planed to joint our grid to Thai grid, the largest grid community in Thailand. We have developed our

system with many constraints such as time, people, resources, spaces, and budget. Our team has only 2-3 men. Most of computers and resources are provided by ourselves. But we try to present that grid technology can be deployed in many applications and shows them to understand grid technology. Many commercial IT companies have announced to support and move their applications to grid technology. So whatever we will see and use grid technology in the next few years.

References:

[1] Chien, Pakin, Lauria, Buchanan, Hane, Giannini. *The High Performance Virtual Machine (HPVM): Clusters with Supercomputing API's and Performance*. Eighth SIAM Conference on Parallel Processing for Scientific Computing (PP97); March 1997.

[2] Dan Kusnetzky Carl W. Olofson, *Oracl e 1 0 g : Putting Grids t o Work*, April 2004

[3] D.E. Culler, D.A. Patterson, T.E. Anderson. *The Berkeley Network of Workstation (NOW) Project*. 40th IEEE Computer Society International Conference (COMPCON'95) San Francisco California USA, March 05 - 09, 1995

[4] G.C. Fox. *What have we learnt from using real parallel machines to solve real problem?* In Proc. Of the 3rd conference on Hypercube concurrent computers and applications, pages Vol.2 pp 897-955, Pasadena, CA, January 1998.

[5] Global Grid Forum website, http://www.ggf.org

[6] Globus website, http://www.ggf.org

[7] Ian Foster, Carl Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*, second edition, Morgan Kaufmann Publishers, 2004

[8] Ian Foster, Carl Kesselman, and Steven Tuecke, *"The Anatomy of the Grid: Enabling Scalable Virtual Organizations"*

[9] Ian Foster, Carl Kesselman, Jeffrey M. Nick, and Steven Tuecke, *"The Phisiology of the Grid: An Open GridService Architecture for Distributed Systems Integration"*

[10] Ishikawa, Tezuka, Hori, Sumimoto, Takahashi, O'Carroll, Harada. *RWC PC Cluster II and Score Cluster System Software High Performance Linux Cluster.* The Fourth International Conference/Exhibition on High Performance Computing in Asia-Pacific Region (HPCAsia2000), Beijing, China, May 14-17, 2000

[11] Luis Ferreira, Bart Jacob, Sean Slevin, Michael Brown, Srikrishnan Sundararajan, Jean Lepesant, and Judi Bank, *"Globus Toolkit 3.0 Quick Start,"* IBM RedPaper, September 2003

[12] Napat Chalakornkosol and Putchong Uthayopas, *"Monitoring the Dynamics of Grid Environment using Grid Observer "*, Poster Presentation in IEEE CCGRID2003, Toshi Center, Tokyo, May 12-15, 2003.

[13] *ORACLE DATABASE 10G ENTERPRISE EDITION*, www.oracle.com

[14] *Oracle Grid Index Report: Mapping The Global Journey To Grid Computing*, http://www.oracle.com/global/eu/pressroom/nagridreport.pdf, April 2005

[15] NPACI Rocks Cluster Distribution, www.rocksclusters.org

[16] R. Buyya. *"High Performance Cluster Computing"* Vol 1. And Vol II, Prentice Hall, 1999

[17] Red hat Inc website, www.redhat.com

[18] Robert G. Shimp, *Oracle Grid Computing*, February 2005

[19] Sterling, J. Becker, Savarese, E. Dorband, A. Ranawak, V. Packer. *BEOWULF: A PARALLEL WORKSTATION FOR SCIENTIFIC COMPUTATION*. Proceedings, International Conference on Parallel Processing, 1995.

[20] T. Grabs, K. Bohm, and H.-J.Schek. *High-level Parallelization in a Database Cluster: a Feasibility Study Using Document Services.* In Proc. of the International Conference on Data Engineering (ICDE), Heidelberg, Germany, April 2001. IEEE Computer Society.

[21] ThaiGrid community website, www.thaigrid.net

[22] The Globus Alliance website, http://www.globus.org

[23] The Grid Computing Information Centre (GRID Infoware: http://www.gridcomputing.com)

[24] The Java Technology website, http://java.sun.com/j2se/index.jsp

[25] The Junit website, http://www.junit.org

[26] YOSHIO , MOTOHIKO , KAZUTO, MITSUHISA. *COMPaS : A Pentium Pro PC-based SMP Cluster and Its Experience (Special Issue on Parallel Processing).* In 1PPS Workshop on PC NOW, pages 486497. LNCS, 1998