Performance Enhancement of Cloud Computing using Clustering

Bhupendra Panchal and R.K.Kapoor

Dept. of Computer Technology & Application,
NITTTR, Bhopal

Summary
Cloud computing is an emerging infrastructure paradigm that allows efficient maintenance of cloud with efficient uses of servers. Virtualization is a key element in cloud environment as it provides distribution of computing resources. This distribution results in cost and energy reduction, thus making efficient utilization of physical resources. Thus resource sharing and use of virtualization allows improved performance for demanding scientific computing workloads. Number of data centers and physical servers are underutilized so they are used inefficiently. So performance evaluation and its enhancement in virtualized environment like public and private cloud are the challenging issues. Performance of cloud environment is dependent on CPU & memory utilization, Network and I/O disk operations. In order to improve the performance of the virtualization with cloud computing, one of the solutions is to allow highly available data in the cluster form. Thus replicas are available at each data centers and are highly available. In the proposed work, the I/O parameters are chosen for increasing the performance in this domain. This enhancement can be achieved through the clustering and caching technologies. The use of technology for data centers clustering is proposed in this paper. Thus performance and scalability can be improved by reducing the number of hits to the cloud database.

Key words:
Virtualization, cloud computing, clustering, performance enhancement, caching.

1. Introduction

For many computational applications, large numbers of resources are needed and this utilization of resources is for long period of time. Cloud computing, which is an emerging technology, provides the proper hosting of resources by leasing them from huge data centers only when they are needed. Cloud computing is replacing all existing technologies by offering their customer to pay only what they use. For example, an organization can buy any software or service for required period of time on the cloud rather than to purchase a machine for that purpose. It offers infrastructure, platform, software and data as services and these are subscription based services means pay-as-you-go model (C. Pelletingeas, 2010 [4]). These services are known as Infrastructure as a service (IaaS), platform as a service (PaaS), Software as a service (SaaS) and data as a service (DaaS) respectively. Infrastructure as a service ensures processing, storage, network and other fundamental computing resources to the users. Examples of IaaS based services are Amazon EC2, IBM`s Blue cloud, Eucalyptus, Rackspace Cloud etc. the platform as a service gives a high level integrated environment to build, test, deploy and host customer created applications. Examples of PaaS based services are Google App Engine, Engine Yard, Heroku etc. Software as a service is a software delivery model in that the applications are accessed by simple interface like web browser over Internet. Examples of SaaS based services are Web Mail, Google Docs, Facebook etc (G.Malathy et al [1]). Data as a service provides an infrastructure for web scale data mining and knowledge discovery in order to empower the applications and services with intelligence.

Cloud computing models such as public, private, community and hybrid models can be implemented by using virtualization. Virtualization is the virtual evaluation of computing elements like hardware, software, memory, storage, network and so on (C. Pelletingeas, 2010 [4]). It allows the sharing of physical resources and higher utilization rate with optimal storage. It also reduces the power consumption and hardware investment and improves the system management without extra cost.

Thus cloud is a package of services that offers infrastructure, platform, software and data as services. So many researches are being made for improving these flavors of services. But the dark side of using this virtualization is degradation of performance due to extra overhead. CPU usage, memory, storage and network are the performance factors for cloud computing. Since fast accessing of data and resources is highly demanded in cloud environment. Any organization adopting cloud computing certainly expect the kind of enhanced performance. But this performance is degraded due to limited bandwidth, high response time, inefficient CPU & memory utilization, scalability bottleneck and unnecessary use of data centers.

I/O virtualization poses a more difficult problem because I/O devices are shared among all virtual machines. It

Manuscript received June 5, 2014
Manuscript revised June 20, 2014

requires a privilege domain from guest VMs to access I/O. This intervention leads to longer I/O latency and higher CPU overhead due to context switches between the guest VMS and VMM (Virtual Machine Monitor). Performance of cloud computing is also dependent on the underlying cloud infrastructure. This work is aimed to address different issues that are responsible for improving the performance of cloud computing.

2. Literature Survey

In the context of cloud computing, Different researches have been made for performance enhancement. Various solutions are proposed for improving the performance of cloud computing. Thus previous works related to performance improvement are reviewed in this section:

IOSUP et al [4] has proposed performance analysis of cloud computing services for Many Task scientific Computing (MTC). MTC based computing requires resources to be leased from big data centers only when they are needed. It is obtained that existing cloud technologies are insufficient for scientific computing but they may be still good solution for those that required resources instantly and temporarily.

Donglai Zhang et al [2] proposed WSDF framework for improved data transfer performance of web service workflows. This framework allows directly data sharing between consecutive web services within a web service workflow. Thus it allows improved performance and better data transfer speed between different components of a workflow.

Wei Huang et al [3] proposed a framework for performance and management overheads with virtual machines. Two cases are considered: one is Virtual Machine Monitor (VMM) bypass I/O and another is VM image management. VMM bypass I/O extends the OS bypass feature of modern high speed interconnects such as InfiniBand and VM image management is done with three aspects: Customizing small kernels, developing fast and scalable distribution schemes for large scale clusters and VM image caching on computing nodes. Both these cases are having significant performance with VMs.

Joyent et al [8] has introduced Smart technologies for cloud performance. These Smart technologies offer enhanced performance with scalability, light weight virtualization, flexible resource provisioning, dynamic load balancing, storage caching and CPU bursting.

Scalability of cloud environment can be improved from its SmartOS operating system, SmartMachine virtualization technology, SmartDataCenter infrastructure management system to its smartPlatform development environment. According to Joyent, scalability is the best solution to increase the performance of cloud computing environment.

3. Problems Perceived in this Area

It is difficult to match a good definition of cloud computing due to lack of standardization and economical impact. Standardizations of reliability and security are main concern now- a- days. Because these issues are facing daily new challenges. As cloud computing is an on demand service that shares a pool of resources over the network. Thus cloud security and reliability to its users are the major issues to be researched in this area and both of them make it hard to understand.

In other hand, economical impact of cloud allows that resources could be used in more efficient and intelligent ways by reducing the cost. This can be achieved by virtualization because it requires less storage space of the servers and also reduction of power consumption.

Latency and interoperability are also major issues to be solved in cloud computing because their causes are engineered into the cloud platforms themselves.

One of the most important issues in cloud computing is the performance overhead. Since fast accessing of data and resources is highly demanded in cloud environment. Any organization adopting cloud computing certainly expect the kind of enhanced performance. But this performance is degraded due to limited bandwidth, high response time, inefficient CPU & memory utilization, scalability bottleneck and unnecessary use of data centers.

I/O virtualization poses a more difficult problem because I/O devices are shared among all virtual. It requires a privilege domain from guest VMs to access I/O. This intervention leads to longer I/O latency and higher CPU overhead due to context switches between the guest VMS and VMM (Virtual Machine Monitor). Performance of cloud computing is also dependent on the underlying cloud infrastructure.

4. Proposed Methodology

Performance is the major concern in the field of cloud computing. People are running into the scalability. But increasing number of Virtual machine and CPU is not the key solution for scalability, because cost is another issue to be researched then. So the area which is identified for improving the performance is to avoid the unnecessary use of databases. Data centers are highly loaded for accessing I/O requests. Although operations on huge amount of data in cloud computing are quite complex and lead to be performance overheads. There may be a case of server failure. So from the development, maintenance and performance perspective, all these can be serious issues to be refined. One of the solutions is to allow highly available data in the cluster form. Thus replicas are available at each data centers and are highly available. In our work, the I/O

parameters are chosen for increasing the performance in this domain. This enhancement can be achieved through the clustering and caching technologies.

Cluster based data centers are highly efficient for performing I/O operations. Clusters can be defined as collection of virtual data servers and are treated as a single machine. Clustering avoids uninterrupted access to data and also helps when network or storage connectivity is lost. Caching, in other hand, prevents over hits to the databases. Network caching and VM image caching are the two aspects that are proposed for this purpose. Thus the idea behind both these technologies is to minimize the information that move among the different cloud components.

Although clustering can be done at different level such as OS level, Application level, web server level and database level. Clustering among the data centers that are located throughout the world, allow highly available data for customers without any delay. Thus if one data center is goes down, everything in the second data center is clustered with the first, so there is no problem for the time being. And you still have database/web/app server in the second data center.

In the favour of our work several issues are identified for better performance of cloud services. In some previous studies it is established that clustering can be a key contributing factor to improve the performance in cloud computing.

G.Malathy et al [1] proposed the Reservation Cluster approach for performance enhancement in cloud computing. The concept of reservation cluster is to schedule the tasks. Unscheduled tasks are sent to the reservation cluster and in this cluster all the tasks are scheduled simultaneously without any iteration. It reduces the amount of computation time and resource usage and allows better performance.

A.Mahendiran et al [10] proposed the implementation of K-Means clustering algorithm in cloud computing for large data sets. The work was focused to implement and deploy K-Means algorithm in Google Cloud using Google App Engine with cloud SQL. Thus cloud computing allows mining of large databases and storing them with less cost. Results show that clustering works well in the cloud.

Since cloud is having the large data sets and these data sets are regularly accessed as per user requirements. Michael Shindler et al [11] have proposed fast and accurate k means algorithm as a solution where the data is too large and must be accessed sequentially. BigCross dataset and census 1990 dataset were considered as large data sets for applying the proposed clustering algorithm. Experiments show the significant results in terms of cost and response time.

P. Ashok et al [12] has proposed renovated k-means algorithm for Iris and Wine datasets. Proposed algorithm is compared with K-means, static weighted K-means and Dynamic weighted K-means on three different distance function. Less execution time and minimum iteration count of proposed k- means help to improve the performance.

Karedla, R. Et al[13] has proposed the caching strategies to increase the system response time and data throughput of the disk. Results show that caching offers the twise performance of its size.

Q. Luo et al [14] has proposed the Active Query Caching for improving scalability of database web servers. It allows the load sharing of database in order to reduce the network traffic. This caching is applied at query level for simplifying query containment checking and query evaluation at the proxy. This query level caching allows fast response to the user queries and hence improve the web server performance. Proposed architecture stack using clustering is shown in figure:

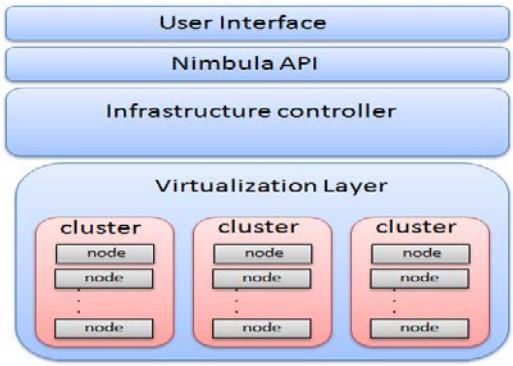

Fig1: Architecture for cloud computing using Clustering

Proposed clustering and caching technologies are integrated at the physical level where the data servers are located. Thus clusters of data centers are formed at this level. And abstract level (i.e. virtualization layer) is responsible for load sharing among the datacenters. In order to scale the performance of data centers, a pool of data centers tied together to act as a single unit called cluster. These clustering technologies are transparent to client applications. Clustering and caching techniques can be seemed a better solutions for fast accessing of data and I/O operations. In this way, proposed methods for reducing data traffic and also minimizing the database hits can have better performance than cloud environments that are not using clustering of data centres.

5. Conclusion

Performance of virtualization with cloud computing is a major issue to be researched. Poor performance can lack the interest of customers. Clustering and caching are the proposed methodologies for improving the performance in

this work. Both these technologies have their own significant. Clustering of data centers, network caching and VM image caching are the key points that are used as performance parameters. Expected results can have better performance than existing one.

Acknowledgment

I would like to express my special thanks of gratitude to my guide Dr. R.K.Kapoor for his support, his supervision and his time. He provides me better opportunity to understand the concepts of cloud computing.

REFERENCES

[1] G. Malathy, Rm. Somasundaram, "Performance Enhancement in Cloud Computing using Reservation Cluster", European Journal of Scientific Research, ISSN 1450-216X Vol. 86 No 3 September, 2012, pp.394-401.
[2] Donglai Zhang, Paul Coddington and Andrew Wendelborn," Improving Data Transfer Performance of Web Service Workflows in the Cloud Environment", Int. J. Computational Science and Engineering, Vol. 1, No. 1/1, 2012.
[3] Wei Huang , Jiuxing Liu, Bulent Abali and Dhabaleswar K. Panda, "A Case for High Performance Computing with Virtual Machines", ICS '06 Proceedings of the 20th annual international conference on supercomputing, pages 125-134.
[4] C. Pelletingeas, "Performance Evaluation of Virtualization with Cloud Computing", MSc Advanced Networking, 2010,
[5] Alexandru Iosup, Simon Ostermann, M. Nezih Yigitbasi, "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE Transaction on parallel and distributed system, VOL. 22, NO. 6, JUNE 2011.
[6] Mattias Sunding, Maximizing Virtual Machine Performance, vkernel corporation, A Quest Software Company, http://www.vkernel.com
[7] Devarshi Ghosal, R.Shane Canon and Lavanya Ramakrishnan,"I/O Performance of Virtualized Cloud Enviornment",
[8] Joyent White paper, Performance and scale in cloud computing,
[9] Nikolaus Huber, Marcel von Quast, Micahal Hauck and Samuel Konev, "Evaluating and Modeling Virtualization Performance Overhead for Cloud Enviornments"
[10] A. Mahendiran, N. Saravanan, N. Venkata Subramanian and N. Sairam, "Implementation of K-Means Clustering in Cloud Computing Environment" , Research Journal of Applied Sciences, Engineering and Technology 4(10): 1391-1394, 2012 ISSN: 2040-7467.
[11] Michael Shindler, Alex Wong and , and Adam Meyerson "Fast and Accurate k-means For Large Datasets" ,
[12] P. Ashok, Dr. G.M Kadhar Nawaz, E.Elayaraja and V. Vadivel, "Improved Performance of Unsupervised Method by Renovated K-Means"
[13] Karedla, R., Love, J.S. and Wherry,B.G.,"Caching strategies to improve disk system performance",IEEE computer Society, VOL. 27, March 1994.
[14] Q. Luo, J.F.Naughton, R.Krishnamurthy, Pei Cao and Yunrui Li,"Active Query Caching for Database Web Servers", WebDB 2000, LNCS 1997, pp. 92-104, 2001.

Bhupendra Panchal received the Bachelor of Engineering from Mahakal Institute of Technology, Ujjain(M.P.) in 2011 and pursuing M.Tech in Computer Technology & Application from National Institute of Technical Teachers` Training & Reserch ,Bhopal(.P.)

Prof. R. K. Kapoor received B.Sc. ,M.Sc. and Post Graduation diploma in Computer Science and Apllication from Barkatullah University, Bhopal in 1985, 1988 and 1990 respectively. He held MBA in 1998 from Devi Ahilya Vishvavidyalaya, Indore and also awarded with Ph.D. in 2012 from MANIT, bhopal. He is working as associate professor in National Institute of Technical Teachers` Training & Research ,Bhopal. (from 30/12/10)