A Survey on Enhancing Resource Allocation in Virtual Environment

Shilpa
M.Tech(SCS)
East West Institute of Technology,
Bangalore, India.

Arun Biradar, PhD
Prof & Head of the Dept,CSE
East West Institute of Technology,
Bangalore, India.

ABSTRACT
The cloud computing has become the main part networking for its vast applications all over the world. The cloud computing empowers the primary services for business point of view with the customer and also computes the variation in the resource consumption depending upon the load requirement. However, the enabling the simultaneous use of single machine over many numbers of the computer system is the most challenging thing in the cloud computing. In reality, when the workload ramp-up, techniques adopted for resource allocation cannot fulfill the speedy execution of more job, by which efficiency in the service may reduce. The efficiency can be improved by considering information on workload and analytical performance. The workload intensity in many visualized IT resource can be found to have QoS (Quality of service). Thus, the virtual machines are used to solve these issues. This paper presents a survey on resource allocation using virtual service of cloud computing in a different work situation.

Keywords
Cloud computing, Resource allocation, Virtual Machine

1. INTRODUCTION
The cloud is the latest computing model. This computing model computes as per the change in the requirement, means the manufacturers of the model supply the required hardware, software and service as per the user requirement [1]. The growth in the internet, user need is changing with the application. The cloud computation is the latest version of the parallel and distributed computing. The function is to provide quick, secure and convenient storage for data and also internet based computing service. The development and occurrence of cloud computing are actuated by some factors such as; data transportation, Web 2.0 appearance, high quality of technology, grid computation development and moreover the virtualization development. Cloud computing offers new business opportunities for service providers and their clients (e.g. organizations, enterprises, and end users), using architecture for delivering Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). A Cloud encloses the IaaS, PaaS, and SaaS inside its virtualization infrastructure, to carry out an abstraction from its underlying physical assets [2]. Typically, the virtualization of a service implies the aggregation of several proprietary processes collected in a virtual environment, called Virtual Machine (VM). The data centers are developed in such a way that they can be used comfortably to share the data with the infrastructures. These data centers have the ability to take one set of resource to another set in a non-disruptive manner. These advantageous factors have gained popularity in modern cloud computing infrastructure. The main purpose of cloud computing is to share and control the larger data centers. The virtualization is the best technology where the data control and transformation can be done efficiently [3].

For example virtual disks, virtual machines (VM) are the virtual containers and are used to host the data. The virtualization enables the application execution. The virtual containers de-coupled from underlying physical resources. The server technologies of virtualization are VMware and Xen offers the application execution which is a need to be hosted in VMs and that allows the many VMs run in a series manner to one physical machine [4]. The resource sharing in many applications takes place in this way. The growth in the new live migration of VMs allows the migration of VMs from one server to another server without downtime in the application running time [5].

The virtualization technology for storage purpose; performs the virtualization of physical storage in storage area network (SAN) as virtual so that it can be reused for application purpose. The application and the physical storage will allow consolidation over protocols and different vendors for sharing the storage resources. Also enables the live data migration, where the virtual disks can migrate from one storage sub-system to another with no downtime. EMC in Vista and IBM SAN volume controller are the virtualization products for storage, and presently these are popular in data centers. The storage and server virtualization technologies exist from past years and are entirely beneficial [6]. The combined virtualization of both storage and server with live migration will allow the applications to share the resources of storage and server, by this increasing and consolidating utilization over the data center. The combination of storage and server is used to enhance the efficiency also the combination is the best choice for every application. The combination also performs the load balancing by data optimization [7]. Section II discusses the background of the study followed by a discussion literature survey on cloud computing in Section III. Section IV problem identification and survey results are discussed in Section V. Finally Section VI gives the research gap and conclusion of this paper.

2. BACKGROUND OF THE STUDY
2.1 Cloud Computing
Cloud computing platforms have received a great deal of attention in the business world in recent years. The primary motivation for companies to consider transferring their existing systems or creating new ones on top of a cloud platform is the flexibility they promise to provide and their pricing model. The fundamental difference between a classical approach with a fixed infrastructure and cloud computing can be illustrated on an example of a startup company launching a new service on the web. Then it has to acquire the equipment (by buying or by renting), install the application and start providing the service. However, as the number of clients is often difficult to predict and changes significantly in time the company risks that it will either buy more hardware than necessary or pay unnecessary costs. On the other hand, that the hardware will not be sufficient enough, the quality of the service will be low and will discourage potential clients.

Cloud computing claims to provide a solution to this problem: Instead of buying hardware and building its infrastructure, the company rents capacity from a cloud provider and only pays for the time it is used [8]. According to a study of Armbrust et al. [9], the saving on costs can be of factor 5 to 7. Instead of paying a high fixed cost of installation and subsequent operating costs, the company would only pay the operating costs. The cloud computing provides the elasticity for the data centers. For new applications it allocates the resource at right time. The companies like start ups are need proper resource allocation in required time and can release the excess resource when it is of no use. This will helps in providing better QoS and saves money for unused resources. Figure. 1 and 2 gives the Comparison of classical scalability and elasticity approaches taken from the survey in internet: The figure 1. Gives the scalability plot, where x- axis indicates time in minutes and y-axis gives the load and capacity of cloud computing. During classical scalability the capacity of the cloud will be constant while the load request will be changing and it will be more than the capacity of the cloud; this means QoS will be less [9].The figure 2 gives the elasticity plot, where x- axis indicates time in minutes and y-axis gives the load and capacity of cloud computing. After adopting the elasticity, the capacity of the cloud can be varied as per the load requirement.

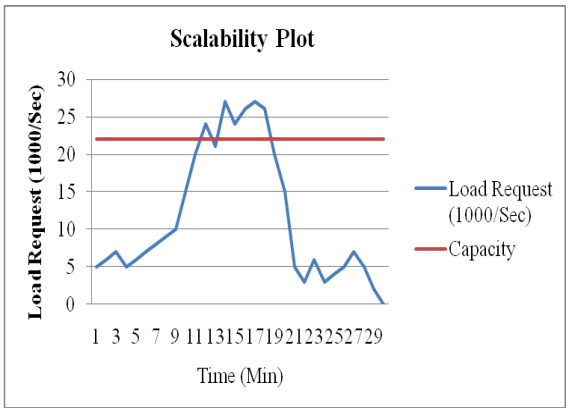

Fig 1. Shows Load Requirement (Capacity of Cloud Also) Vs Time During Cloud Computing in Scalability

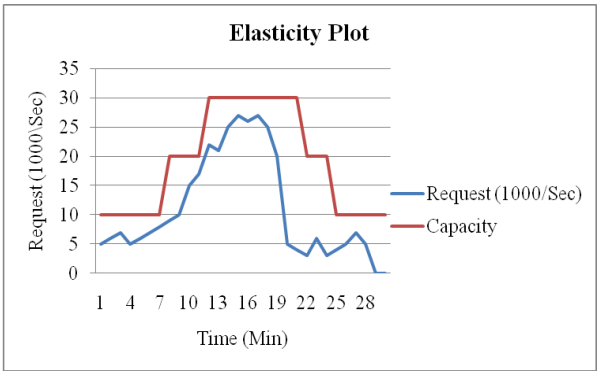

Fig.2. Shows Load Requirement (Capacity of Cloud Also) Vs Time During Cloud Computing in Elasticity.

3. CHARACTERISTICS OF CLOUD

Today, most people, each IT organizations are discussing the Clouds. Cloud computing could be a model for sanction present, simple, on-demand access to a joint pool of computing services (e.g., servers, networks, programs, storage, and so on) that may be immediately provisioned and also discharged with lowest administration effort. US government is the main customer of electronic services and, thus, one in every of the principal consumers of Cloud systems. The United States National Institute of Standards and Technology (NIST) have a collection of operating definitions that distinguish Cloud computing into service models and deployment models [10].

The characteristics of cloud can be configured based on three aspects

1. Based on Nonfunctional Aspect:-

a. Elasticity

The elasticity helps to change the cloud infrastructure as per the data size requirement of consumers.

Example: Amazon EC2

A. Reliability

This allows the continuous program process without loss.

Example: VMware.

B. Quality of Service (QoS)

This is the feature which means every consumer will be satisfied with cloud service provider.

Example: Amazon S3

C. Availability

The factor in which the errors are clearly masked during the computing.

Example: MS Azure

2. Based on Economic Aspect

A. Cost Reduction

This helps to reduce the cost of the cloud maintenance and give the consumer satisfaction for the buyers of cloud. Example: Google Apps Engine.

B. Going Green

This means making use of less power consumption.

C. Based on Technological aspect

D. Virtualization

This allows the flexibleness in the cloud computing.

Example: Virtual Box.

E. Multi-tenancy

Is the feature in which data and code are not known first and allows allocating the equal data.

Example: MS SQL.

F. compliance, Security and privacy:

In many systems data coping with critical code and information.

G. Data Management

The feature is used to store data for various sites of cloud.

Example: Web Sphere.

H. Metering

All programs and device consumption is necessary to provide pricing, charges and also billing.

I. Tools

Are required to adaptation, production, assistance and also Cloud programs use.

4. ELASTICITY APPLICATIONS

The variation in the in load poses a specific challenge on applications in distributed environments. However, such changes can happen in generally any kind of application. A traditional solution to this problem has been to ensure scalability – the ability of the system to be enlarged to a size which was expected to accommodate future growth. The disadvantage of such an approach is that the size of the system has to be estimated in advance. If the growth in the load does not correspond to the estimation, the assigned capacity is not used effectively. Another disadvantage is that scaling up implies extending the physical resources, and it may be difficult to scale down on a scale up. A possible solution to the problem is to ensure elasticity. Elasticity aims to solve the problem in an opposite direction - instead of setting the physical target size of the system in advance; the system dynamically reacts to actual load by adding new virtual resources. When the load on a component increases over a given limit, new instance of the component is added to accommodate the growth. If on the other hand when the workload decreases, the additional case can be removed [11].

5. REVIEW OF LITERATURE

Many researchers have worked on cloud computing to have QOS in the virtual environment [12].

Song et al. [13] have presented a multi-tiered resource scheme to schedule the resource requirement and also resource optimization in the data center based on VM. The authors have designed an algorithm for global scheduling. They have implemented the RAINBOW computing platform (Xen based) to evaluate the multi-tiered scheduling. The experiment is outcomes with the results that the algorithm gave performance improvement of 16% with assurance in the quality of service. Kusic et al. [14] have presented power management in virtualized environment. They have used and executed the LLC framework to provide dynamic resource allocation in a virtual environment. The authors have taken care of optimization risks and concentrated on switching costs. The study outcomes with the saving of 22% in power consumption cost. Verma et al. [15] have pMapper controller to minimize the migration cost and power. The author has designed the methodology on two server platform. The experiment results in the efficient power and migration cost reduction method. Ramgovind et al. [16] have discussed the security management in cloud computing. The authors aim to give the security prospects in cloud computing, also addressed the security issues. Zhang et al. [17] have considered the trend in cloud computing. The authors have represented the cloud computing principle, applications of cloud computing, security also characteristics of cloud computing. Zhao et al. [18] have discussed the power management of VM and resource allocation and also provided a solution by the *Joulemeter*. The authors have performed an experiment on many production servers and windows live messenger. The experiment results with the VM power metering, same saving in the virtual data center as of non-virtual data center and also saves the allocation cost in virtualization. Bardasiri and Hashemi [19] have presented Metrix of Quality of Service (QoS) to perform service evaluation of Cloud computing and also security in it. Celesti et al. [20] have presented the strategy to fast access of data in virtualized cloud environment. They have overlooked on the distributed cloud service issues. The authors have an outcome with the efficient link to have large data delivery. The table gives the some more works of researchers in cloud computing with different methods and for a different purpose.

Table.1: Existing Work in Cloud Computing

Author	Method Used	Purpose of Research on Cloud
Lim et al. [21]	Replication	Performance
Roy et al. [22]	Replication	Performance
Raveendran et al. [23]	resizing	Performance
Zhang et al. [24]	replication migration	Performance energy
Vijayakumar et al. [25]	resizing	performance
Rajan et al. [26]	resizing	performance
Knauth & Fetzer [27]	migration	performance

6. PROBLEM IDENTIFICATION

The cloud computing requires the perfect allocation of resource to have better hardware and software programs configuration and also to improve the system efficiency and QoS. The QoS is required to achieve the SLAs with all consumers and for cloud investment justification. The process will have some virtualized behavior on network and resource. The smooth allocation will have following issues:

- The resource providers will face the problem in estimation, which is needful to meet the require cloud computation for the workload.
- The workload will dynamically increase as the consumers of the application increases; this is the major problem in the present generation.
- The increment will reduce the QoS and also the lack of resource allocation. The change in the workload gives challenge to have better cloud computing. Thus to overcome the issues regarding the cloud computing preferably in virtual environment better resource allocation is required.

7. RESULTS OF CLOUD SURVEY

Considering the review of a cloud of the cloud is growing shown in fig.3. Many enterprises are offering more cloud environment for both public and private sector, thus IT teams are looking forward to offer cloud services to their organizations with less cost.

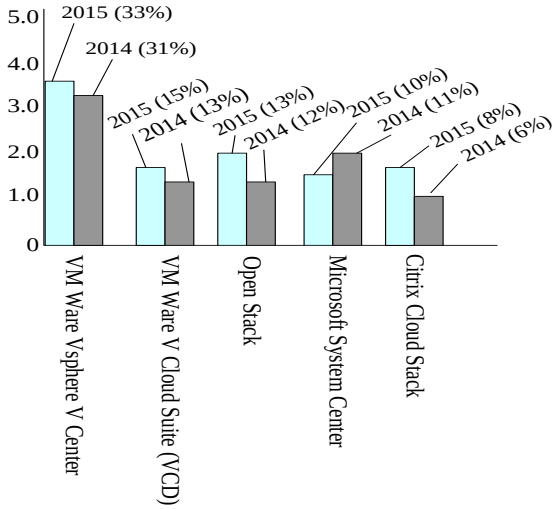

Fig.3. shows the plot of cloud adoption in many companies in 2014 and 2015.

From the figure.4 Shown, an increase in workload in data centers will affect the applications based on cloud.. Between 2000 and 2008, it states, peer-to-peer traffic dominated consumer Internet traffic, which didn't put a strain on data centers. In 2015, 37 % of consumer traffic is from cloud applications, while in 14 % in 2010.

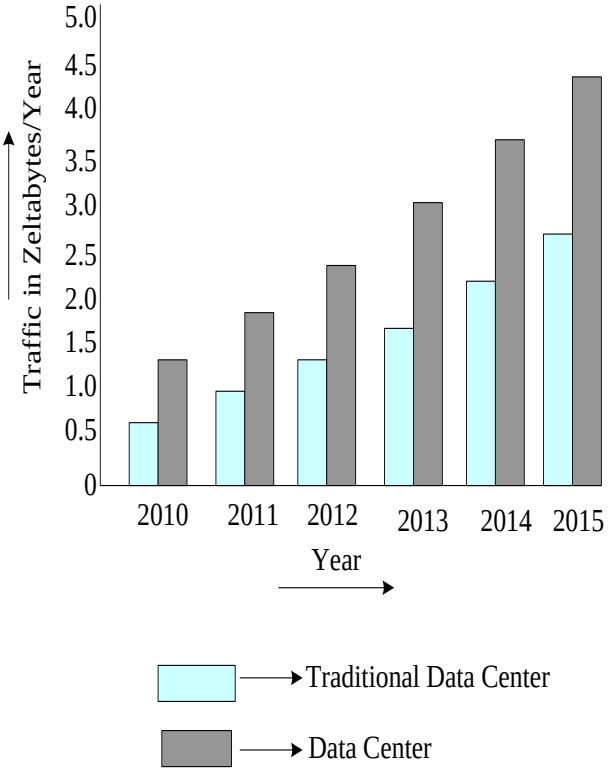

Fig.4. shows the plot of cloud adoption in many companies in 2010 and 2015.

8. CONCLUSION AND RESEARCH GAP

The resource allocation in cloud computing is selection, arrangement and also the control of software runtimes such as workload balancer and database servers. The hardware resources such as storage, network, and CPU, etc., are can be used to provide application performance. The above functions are used to enhance the performance, reduce the energy usage, improve the time response, and provide quality of service (QoS) and SLA. The resource allocation primary function is to enhance the profit rate for cloud service provider and cloud service user aims to reduce the cost. The existing resource allocation has many issues in the virtual environment; they cannot meet the target in workload variation. The new techniques are needed to be used based on virtual machines so that it can overcome the above issues on cloud computing. Architecture needs to be adopted by which it should work for real-time workload and Data intensive-HPC applications. An efficient technique has to adopt to get Quality of Service (QoS), and Mechanisms have to be proposed to make efficiently of cloud resources so that QoS and also during dynamic resource allocation SLA violation to be achieved in minimized in hybrid clouds. These allocation techniques are also needed to use in both SaaS and IaaS users.

9. REFERENCES

[1] Avram, Maricela-Georgiana. "Advantages and challenges of adopting cloud computing from an enterprise perspective." Procedia Technology 12 (2014): 529-534.

[2] Abolfazli, Saeid, et al. "Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges." Communications Surveys & Tutorials, IEEE 16.1 (2014): 337-368.

[3] Beloglazov, Anton, Jemal Abawajy, and Rajkumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." Future generation computer systems 28.5 (2012): 755-768.

[4] Barroso, Luiz André, Jimmy Clidaras, and Urs Hölzle. "The datacenter as a computer: An introduction to the design of warehouse-scale machines."Synthesis lectures on computer architecture 8.3 (2013): 1-154.

[5] Soni, Gulshan, and Mala Kalra. "Comparative Study of Live Virtual Machine Migration Techniques in Cloud." International Journal of Computer Applications, Published by Foundation of Computer Science, New York, USA, 84 (14): 19-25, December 2013, ISBN: 973-93-80879-34 (2013).

[6] Adhikari, Sameer, et al. "Best Practices for Building an Enterprise Private Cloud." Intel IT Centre (2011).

[7] Wang, Lizhe, and Samee U. Khan. "Review of performance metrics for green data centers: a taxonomy study." The Journal of Supercomputing 63.3 (2013): 639-656.

[8] Fernando, Niroshinie, Seng W. Loke, and Wenny Rahayu. "Mobile cloud computing: A survey." Future Generation Computer Systems 29.1 (2013): 84-106.

[9] Sobeslavsky, Petr, et al. "Elasticity in cloud computing." Master's thesis, Joseph Fourier University, ENSIMAG, Grenoble, France (2011).

[10] Bardsiri, Amid Khatibi, and Seyyed Mohsen Hashemi. "QoS Metrics for Cloud Computing Services Evaluation." International Journal of Intelligent Systems and Applications (IJISA) 6.12 (2014): 27.

[11] Herbst, Nikolas Roman, Samuel Kounev, and Ralf Reussner. "Elasticity in Cloud Computing: What It Is, and What It Is Not." ICAC. 2013.

[12] Aguiar, Alexandra, and Fabiano Hessel. "Embedded systems' virtualization: The next challenge." Rapid System prototyping (RSP), 2010 21st IEEE International symposium on. IEEE, 2010.

[13] Song, Ying, et al. "Multi-tiered on-demand resource scheduling for VM-based data center." Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid. IEEE Computer Society, 2009.

[14] Kusic, Dara, et al. "Power and performance management of virtualized computing environments via lookahead control." Cluster computing 12.1 (2009): 1-15.

[15] Verma, Akshat, Puneet Ahuja, and Anindya Neogi. "pMapper: power and migration cost aware application placement in virtualized systems."Middleware 2008. Springer Berlin Heidelberg, 2008. 243-264.

[16] Ramgovind, Sumant, Mariki M. Eloff, and Elme Smith. "The management of security in cloud computing." Information Security for South Africa (ISSA), 2010. IEEE, 2010.

[17] Zhang, Shuai, et al. "Cloud computing research and development trend."Future Networks, 2010. ICFN'10. Second International Conference on. Ieee, 2010.

[18] Zhao, Kansal, A., F., Liu, J., Kothari, N., & Bhattacharya, A. A. (2010, June). Virtual machine power metering and provisioning. In Proceedings of the 1st ACM symposium on Cloud computing (pp. 39-50). ACM.

[19] Bardsiri, Amid Khatibi, and Seyyed Mohsen Hashemi. "QoS Metrics for Cloud Computing Services Evaluation." International Journal of Intelligent Systems and Applications (IJISA) 6.12 (2014): 27.

[20] Celesti, Antonio, et al. "Virtual machine provisioning through satellite communications in federated cloud environments." Future Generation Computer Systems 28.1 (2012): 85-93.

[21] H. C. Lim, S. Babu, J. S. Chase, and S. S. Parekh, "Automated control in cloud computing: challenges and opportunities," in Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds, ser. ACDC 2009. ACM, 2009, pp. 13–18.

[22] N. Roy, A. Dubey, and A. Gokhale, "Efficient autoscaling in the cloud using predictive models for workload forecasting," in Proceedings of the 4th Intl. Conference on Cloud Computing, ser. CLOUD 2011. IEEE, 2011, pp. 500–507.

[23] Raveendran, T. Bicer, and G. Agrawal, "A framework for elastic execution of existing mpi programs," in Proceedings of the Intl. Symposium on Parallel and Distributed Processing Workshops and PhD Forum, ser. IPDPSW 2011. IEEE, 2011, pp. 940–947.

[24] X. Zhang, A. Kunjithapatham, S. Jeong, and S. Gibbs, "Towards an elastic application model for augmenting the computing capabilities of mobile devices with cloud computing," Mob. Netw. Appl., vol. 16, no. 3, pp. 270–284, Jun. 2011.

[25] S. Vijayakumar, Q. Zhu, and G. Agrawal, "Dynamic resource provisioning for data streaming applications in a cloud environment," in Proceedings of the 2nd Intl. Conference on Cloud Computing Technology and Science, ser. CLOUDCOM 2010. IEEE, 2010, pp. 441–448.

[26] D. Rajan, A. Canino, J. A. Izaguirre, and D. Thain, "Converting a high performance application to an elastic cloud application," in Proceedings of the 3rd Intl. Conference on Cloud Computing Technology and Science, ser. CLOUDCOM '11. IEEE, 2011, pp. 383–390.

[27] T. Knauth and C. Fetzer, "Scaling non elastic applications using virtual machines," in Proceedings of the 4th Intl. Conference on Cloud Computing, ser. CLOUD 2011. IEEE, 2011, pp. 468–475.