Volume 3, Issue 10, October 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Survey of Resource Allocation Policies in Cloud Computing

Srushti Patel
Computer Science and Technology,
Rajiv Gandhi University, Bhopal, India

Krunal Suthar
Department of Computer Engineering,
S. P. College of Engineering, Visnagar, Gujarat, India

Abstract— ***Cloud computing is the use of the Internet for the tasks you perform on your computer. Recently, there has been a dramatic increase in the popularity of cloud computing systems that rent computing resources on-demand, bill on a pay-as-you-go basis, and multiplex many users on the same physical infrastructure. These cloud computing environments provide an illusion of infinite computing resources to cloud users so that they can increase or decrease their resource consumption rate according to the demands. To achieve good resource allocation in Web applications are difficult due to unpredictable traffic patterns. We address "the cloud resource management problem" which aim is to allocate and schedule computing resources in such a way that its provide high resource utilization and users meet their applications' performance requirements with minimum expenditure. Lots of researcher working with this problem in different directions. Our main purpose is to identify problems in existing scheduling algorithms and design optimized scheduling algorithm which can provides more better and powerful environment to its users.***

Keywords— ***Cloud computing, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Resource management, Load balancing.***

I. INTRODUCTION

Cloud computing is model for enabling ubiquitous, convenient, on demand network access to a share pool of configurable compute deploying resources(e.g. network, server, storage, applications and services).[6] In cloud computing there are mainly two broad classifications of models viz. service model and deployment model. Software as a service (SaaS), Platform as a service (PaaS), Infrastructure as a service (IaaS) are the three commonly known service models. Deployment models include public, private, community and hybrid types. The accessibility of Public cloud is almost anywhere. Anyone having Internet connectivity and required credentials can use the Public cloud from anywhere, anytime. Private cloud is use and managed by the single organization or third party and it is located on-premise or off-premise. Private clouds are used for safety reasons where organization does not want its resourced to be accessed by anyone except its own employees within the four walls. For the sharing purpose in different organization or specific sharing community we can use Community cloud. Hybrid cloud is a combination of a one or more public, private and community cloud. Irrespective of the service model and Deployment type, a common problem in cloud adaptation is security, as the data owner loses her control over data. There has not been any universally adopted security model for cloud computing which is trusted by cloud users. Cloud has centralized management of resources so it can reduce the unnecessary cost to system which is operating by the users. Cloud storage is built on the network computing environment. There are many benefits to move data into the cloud. For example, users do not have to worry about the complexities of direct hardware management. But since users store their data in the cloud, they lose its control data. Security becomes a significant issue to be addressed. Data security is always an important aspect of quality of service and it is also a key issue in cloud computing.

Here I have reviewed various papers on Resource allocation in cloud computing. The various research papers have shown what is cloud computing and how resource allocation take place in cloud computing.

II. REVIEW OF RELATED WORK

In [1], author presents the resource allocation at the application level, instead of studying to map the physical resources to virtual resources for better resource utilization in cloud computing environment. Paper propose a multi-dimensional resource allocation scheme for cloud computing that dynamically allocates the virtual resources among the cloud computing applications to reduce cost by using fewer nodes to process applications. In this model author adopt a two-stage algorithm to solve this multi-constraint integer programming problem. Most of recently research merely considered resources and workflow in one-dimension which cannot actually metric resources in cloud. Aiming at improving resource utility in data centre, authors propose MDRA, which can solve multi-dimensional resource allocation problem. Experiment results show that our algorithm can save resources and increase resource utilization as well as centralize working nodes. Moreover, the proposed algorithm centralized working nodes and in the long run, it can save power efficiently when the demand of user is decreasing. In [2], Adaptive Management of virtualized resources in cloud computing is based on VM Architecture which is proposed on adaptive management of virtualized resources in cloud computing environments. VM based architecture provides sound concealment between different applications which concurrently run in the virtual resource pool and concede the dynamic allocation of resources to application. All applications to swallow and shrinkage based on resource requirements to achieve SLOs .A critical problem under the VM-based architecture is how to allocate resources to every application on resource and in response of time management based on workloads.

We have designed an adaptive manager which includes three controllers,

1) CPU controller, 2) Memory controller, 3) I/O controller. The adaptive manager is modeled by the dynamic state-space feedback control method. They adopt KVM as the infrastructure of virtual system instead of Xen for analysing the virtualization of architecture .At the moment, they are authenticating the validate and executing the function of our proposed architecture and model strategy. Our assessment exhibit that there is hope for further improvement in network I/O performance. Moreover we are working on direct I/O assignment for the issues and better modelling of network sharing.

In [3], Current cloud providers don't guarantees about satisfactory response time of web applications hosted on cloud. In current scenario it's hard to get maximum average response time guarantee for web applications because of unpredictable traffic patterns. A multi-tier web applications structure is very complex because of that it create complex bottlenecks conditions and resolving then automatically is non achievable. It can be minimize the bottlenecks conditions by adding virtual machine on cloud. This research mainly on focuses on enabling clouds to offer multi-tier web application owners to maximum response time guarantees with minimizing resource utilization. At the moment, experiments on EUCAYPTS based cloud proved that dynamic bottleneck detection and resolution for multi-tier web application hosted on the cloud will help to offer SLAs can offer SLAs that can offer response time guarantees.

In [4], A sound research on Iaas (Infrastructure as a service) based cloud systems of open source, we suggest an optimized scheduling algorithm to accomplish the optimization or sub-optimization for cloud scheduling problems. We look into the possibility to assign the virtual machines in a flexible way to permit the maximum usage of physical resources. To achieve this we used an improved genetic algorithm for the automated scheduling policy. The IGA uses the shortest genes and present the idea of dividend policy in economics to select an optimal or suboptimal allocation for the VMs requests. The simulation research conclude that dynamic scheduling policy performs far better than of Eucalyptus, open nebula, nimbus Iaas cloud etc also the speed of IGA is approximately twice the conventional GA scheduling method in grid environment and the performance rate of resources always higher than the open-source IaaS cloud systems.

In [10], cloud computing concede business customers to speed up and down on their resource allocation based on requirements. Couple of get a boost gains in the cloud model from resource multiplexing through visualization technology. In Existing system, they consider a structure that uses visualization technology to assign datacenter resources dynamically based on application demands and support green computing by the allocated number of server in use. For better utilization they present concept of "skewness" to measure up down utilization multi-dimensional resource on a server. By reducing skewness, they can associate different types of workloads nicely and improve the overall performance of server resources. They develop a system which prevent overload in the workflow effectively as well as used saving energy. Trace driven simulation and experiment conclude that demonstrated our algorithm improve significant performance.

In [11], In Priority driven auction strategy (PDAS) for resource allocation based on the priority of resources and auction between cloud user and datacenters through the resource manager, in which model proposed Load-balancing policy assigning the priorities wise resources. In existing system resources being core of datacenters and managing resources and allocating them in efficient manner through PDAS. In PDAS author use dynamic priority assignment to resources facilitates load-balancing of datacenters.

III. CONCLUSION

Cloud computing technology virtualizes and offers many services across the network. It mainly aims at scalability, availability, throughput, and resource utilization. Emerging techniques focus on scalability and availability. However, cloud computing must be advanced to focus on resource utilization and resource management. The resource allocation in clouds, as an automatic control problem, is much harder than can be thought at first sight. Provisioning customer applications in the Cloud while maintaining the application’s required quality of service and achieving resource efficiency are still open research challenges in Cloud computing. Scheduling and deployment strategies are means of achieving resource provisioning in Cloud environments.

IV. FUTURE WORK

Resource allocation, job scheduling at large scale allocation resource significantly reduce cost in cloud. In dynamic allocation process during virtual resources the cloud computing application dramatically reduces cost using fewer nodes. Strengthen this document proposed scheme can improve resource utilization and reduce cost of data centre. Which logically make smooth process and reduce user usage cost. we defined capacity of each node then when it reach its own capacity all remaining jobs pass to relevant or less utilized node and we specify requirements of each node and if node fulfil requirements of job then it accept otherwise it pass to relevant nodes or fail during process. In conclusion, I believe that if we address mentioned solutions then it improves resource utilization, reduce the user usage cost and save power efficiently on central system.

ACKNOWLEDGMENT

I would like to sincerely thank to my co-author for his valuable advice and guidance at the start of this article. He is the real dynamic motivator for whole this article. My sincerely gratitude goes to my guide who helps a lot for this research work.

REFERENCES

[1] "A Multi-dimensional Resource Allocation Algorithm in Cloud Computing", Journal of Information & Computational Science 9: 11 (2012) 3021–3028. [Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu, State Key Laboratory of Networking and zSwitching Technology, Beijing University of Posts and Telecommunications, Beijing]

[2] "Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control," in First International Conference on Information Science and Engineering, April 2010, pp. 99-102.

[3] "SLA-Driven Dynamic Resource Management for Multi-tier Web Applications in a Cloud", 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.

[4] "An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems" , by Hai zhong,Kun Tao, Xuejie Zhang, The Fifth Annual ChinaGrid Conference.

[5] "Optimal Provisioning of Resource in a Cloud Service". [Ctrls Yee Ming Chen1 Shin-Ying Tsai Department of Industrial Engineering and Management, Yuan Ze University 135 Yuan-Tung Rd., Chung-Li, Tao-Yuan, Taiwan, ROC.] IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010.

[6] "Resource Allocation for Distributed Cloud: Concepts and Research Challenges", IEEE Network, july/august 2011.

[7] "Optimal Resource Allocation in Clouds" by Fang he Chang,Jenniffer Ren,Ramesh Viswanathan. *2010 IEEE 3rd International conference on* cloud computing FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

[8] " Design and Implementation of an efficient Two-level Scheduler for Cloud Computing Environment", *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[9] " Exploiting Dynamic Resource Allocation For Efficient Parallel Data Processing In The Cloud" by Daniel Warneke, Member, Ieee, And Odej Kao. Eee Transactions On Parallel And Distributed Systems, Vol. 22, No. 6, June 2011,Page-985 To 997.

[10] " Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment" by Zhen Xiao, *Senior* Member, IEEE, Weijia Song, and Qi Chen. Ieee transaction on parallel and distributed systems (tpds), vol. N, no. N, month year, IEEE-2012.

[11] *"*Priority Driven Auction Strategy for Resource Allocation in Cloud Computing*", by Ronak Patel, International journal of innovative research* & development – may-2013.vol-2 Issue -5.