International Journal of Research and Reviews in Applied Sciences And Engineering (IJRRASE) Vol 8. No.1 – 2016 Pp.34-40

available at : www.ijcns.com
ISSN: 2231-0061

ASSORTED RESOURCE ALLOCATION STRATEGIES IN CLOUD COMPUTING ENVIRONMENT: A SURVEY

J.Praveenchandar
Assistant Professor
Department of CSE
Velammal Institute of Technology
Chennai-601204, India.

Dr.A.Tamilarasi
Professor & Head
Department of Computer Applications
Kongu Engineering College
Erode-638052, India.

ABSTRACT ----In a cloud computing environment, Resource Allocation is playing a vital role. Based on the resource allocation policy used in the cloud network, the performance may differ from one cloud system to another cloud. The demand of the cloud service is increased day by day. So, it is important to enhance the performance of the cloud system .In order to enhance the efficiency of the resource allocation and to satisfy the Service Level Agreement (SLA), it is important to analyze various resource allocation policies and to identify the efficiency improvement factors in the resource allocation process. In this paper, we are going to analyze various resource allocation strategies in cloud computing environment under various situation such as normal, heavy workload and congested situation. Then various efficiency improvement factors are identified and analyzed.

KEYWORDS: Resource allocation, Dynamic Resource Allocation, QOS in cloud, Parallel Data processing

I.INTRODUCTION

In order to identify the efficiency improvement factors, it is important to analyze the various resource allocation strategies in cloud environment .In this paper, We discussed some of the real time challenges of resource allocation in cloud environment. Then, Resource allocation in normal workload is taken for consideration first. During this situation two techniques are analyzed. Using the first technique the cloud resources are allocated for the requests based on their bid value [1].Second method, based on the various time delays generated by the various data centers, the server is allocating the datacenter resources to the requests [2].It is difficult to follow the traditional systematic allocation policy in highly dynamic and distributed environment. So next, three dynamic resource allocation techniques are analyzed [3][4][5].In [3] considered the protocol, which support the dynamic resource allocation in the cloud environment In [4] discussed one of the basic framework for dynamic resource allocation for parallel data processing called Nephele architecture .Next, we analyzed, Using virtual machines, how the resources are being allocated dynamically in the cloud computing environment in[5].

Finally the resource allocation mechanism under bursty workloads and congested situation are considered. Then [6] Discussed, how resources are allocated for the requests and managed in the network, under heavy workloads. And reference [7] analyzed the optimal resource allocation mechanism for cloud system with congestion control.

II.CHALLENGES IN RESOURCE ALLOCATION:

In the virtual environment, we can list the challenges in design of the network .These are categorized as external and internal challenges. Some external challenges are geographical constrains, client demands, optimizing the service portfolio and virtual network pricing. Some internal challenges are data locality, internal datacenter network reliability and SDN design challenges. Let us address one by one. The client is unable to manage the physical location of the data. Provider will not ensure the physical location of the data. The system should take care about the client's charging mode. Data locality is also addressed as an internal issue. Our system should know how to combine the management of compute and data resources using data locality features.

Due to this, migrated data load can be minimized. Then, inside the data centers reliability of the network resources also must be consider. The Data center internal network design decisions affects performance and reliability of the data center resources. Scalability, visibility and reliability are some of the issues to be consider in this. Using centralized software defined networking controller affects reliability. Then SDN allows only a visibility of a tunnel source and an end point with the UDP traffic and hides the user identity. Software defined networking is a networking paradigm that decouples the forwarding plane form software control. It enhance the network and service adaption and improves their performance. The controller placement problem affects the performance of the control plane, its fault tolerance and the state management of distributed SDN system. Concentrating the electrical energy utilizing by the computing tasks and network resources are converted into thermal energy. Due to this, the life time of the data centers are reduced. And it affects the system availability also. To overcome this problem and to protect our devices, we need a proper cooling system for the datacenters. Even though the cost of setting up the cooling system is double the amount of setting up the datacenter, it must be established. When considering the challenges, optimizing computational resources, network resources and energy consumption are major tasks.

III.RESOURCE ALLOCATION BASED ON THE BID PROPORTION:

In a cloud computing environment based on the utilization of resources, the users will pay the money to service providers. Basically the cloud resources are shared among various customers in the network. In this virtual environment, customers are not aware of all information's about the resources being used by them. But the allocation of resource and Quality of Service will be consistent for all customers. But it is considerable, to provide the best QoS to the customers who is ready to pay more, for the better services. FeiTeng[1] has used the game theory solve this resource allocation problem in cloud environment .In this approach, resource allocation depends on their bid value, is proposed. It deals, how to allocate the resource dynamically, when 'k' no of users are seeking the same resource with different financial capacities. In this

approach, When users are proposing their request for the cloud resources, all the users are asked to offer their bids at the same time. Each user only know their own bids and they will have their own bidding function, which calculates the suitable estimated cost to purchase a resource depending on the task size, priority, QoS requirements, budget and deadline. Then the resources will be allocated based on their bid proportions. This framework can support the resource allocation for both Grid and cloud environments. It integrates some challenging issues. In auctions, one user can not know how much others would like to pay for bid item, because they are at widely scattered locations without communication. It turn to be dynamic repeated gambling process. Each users can adjust their bid price in the next stage in terms of other prior behaviors. Finally a cloud resource management policy is achieved by sequent gambling auction, which will realize Nash equilibrium allocation among users under budget and time constraints. This strategy satisfy the heterogeneous demand of cloud users.

IV.OPTIMAL RESOURCE ALLOCATION IN DATACENTERS

Inall resource allocation techniques, the resource type is considered individually. The system assumes that many resources are allocated to the requests simultaneously. Then all allocated data centers are providing the different network delays to the users at various locations. In this approach, the generalized model for resource allocation in cloud computing is analyzed. From a common resource pool, various resources with discrepant resource attributes are taken. Then they are allocated for each requests with in a period of time. Processing ability and bandwidth are taken as the resources to be distributed for consideration. Each data centers has got the server to provide the processing ability and physical network to provide the bandwidth.

When we get a request, any one optimal data center will be chosen from n datacenters. Processing ability and bandwidth

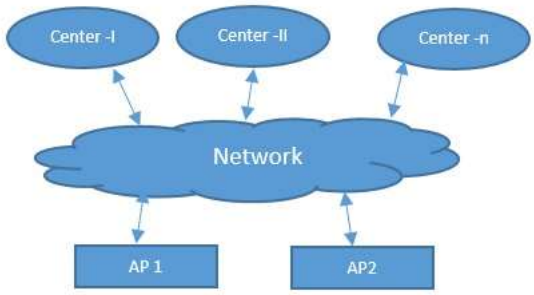


Figure 1:Optimal Allocation in Data centers

of that center will be allocated to the request with in a period of time. If the data centers has no sufficient resource, the request will be rejected. But in this allocation, resources are allocated without considering the network delay. Because all access points having different network delay. In order to reduce the average request loss probability for all access points and to increase the efficiency of the scheduling process,two solutions has been discussed.

First one is, the data center only accepts the requests that has got the maximum quality and the threshold value is calculated for all requests from various access points. The requests above threshold level will be accepted and the resources will be allocated. Then the requests bellow threshold level will be dropped. Then the next solution is, each data center reserves some amount of resources for all access points then remaining resources will be shared. Figure 1 represents first solution, it shows how the data centers has selected the requests based on their threshold value. Center 1 has got short delay for AP1 and long delay for AP2.center n has

got long delay for AP1 and short delay for AP2.Based on the threshold value of the request the resources are allocated.

V.EFFICIENT UTILIZATION OF VIRTUAL MACHINES FOR DYNAMIC RESOURCES ALLOCATION

Mostof the cloud environments have implemented the resource multiplexing through virtualization technique. In this approach, based on the application demand, the data center resources are allocated using virtualization technology .Then, by optimizing the number of servers, it supports green computing. Unevenness of the multidirectional resource utilization of a server is measured. By minimizing that, different types of workloads are combined. This process improves the over all utilization of the server resources. In order optimize the unevenness in the resource utilization of the server, the capacity of a Physical machines is taken for consideration. It must satisfy the resource needs of all virtual machines running on it. Otherwise the physical machine will be overloaded.

Then it will affect the performance of the virtual machine. To support green computing, number of physical machines used must be minimized as long as they can satisfy the needs of all virtual machines. Then resource allocation model has been developed to avoid overloading while minimizing the number of servers. Then an algorithm has been proposed to predict the future resource use of an application to avoid overloading. We use following formula to predict the CPU load.

$E(t) = \alpha * E(t-1) + (1- \alpha) * O(t)$, $0 \leq \alpha \leq 1$

Where E*(t)-Estimated load* and *O(t)-Observed* load at time *t* , α – *tradeoff* between the stability and responsiveness.

The formula to minimize the unevenness of the multidimensional resource utilization of the server (also called as "*skewness*")

Skewness (p) = $\sqrt{\sum_{i=1}^{n} [(r_i / r)-1]^2}$

n -number of resources, r_i- utilization of i^{th} resource, r – average utilization of all resources for server p.

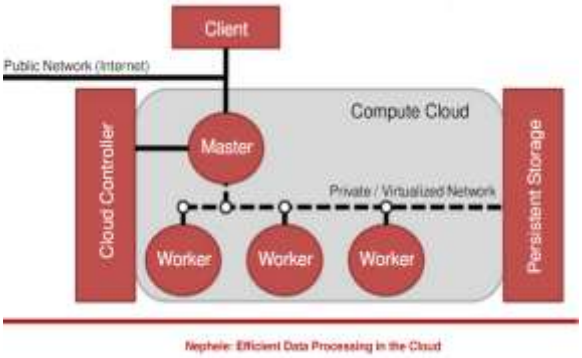
When the resource utilization of server is very low, some servers can be turned off to save energy. Here the challenge is, during the low load, when we reduce the number of active servers, the performance should not be sacrificed. Then that should be maintained in the future. Then, green computing is achieved by identifying the servers which are running in below the green computing threshold level. Based on the ascending order of their memory size cold spots are shortlisted. Then we migrate all its virtual machines before we shut down the underutilized server. The energy consumption in data centers is achieved. Finally the capacities of the server are utilized effectively and the green computing and the overload avoidance is achieved.

VI.PARALLEL DATA PROCESSING AND DYNAMIC RESOURCE ALLOCATION IN CLOUD ENVIRONMENT

Parallel data processing is most important application in the Infrastructureas a service clouds. In this approach, threats in efficient parallel data processing and dynamic resource allocationsare analyzed. Nephele approach is discussed to schedule and execute the task. Here, a Simple scheduling algorithm

is followed. Particular load of a processing job is assigned to a various types of virtual machines. Those are automatically instantiated and terminated during the job execution. Then extended evaluations of Map Reduce–inspired processing jobs on an IaaS cloud system is also considered. It follow the classic master-slave pattern. Initially the job manager (master) start receiving the client's job. Then it schedule and coordinates its execution. The cloud controller used to control the instantiation of virtual machines. Task manager (worker) receives one or more tasks form job manager at a time, executes them. Then persistence storage is supported to store the job's input data and receive its output data.

It is accessible for both Job manager and the set of Task manager. When we get a job from user, job manager transforms that into a primary data structure called *execution graph* to schedule and monitor that. Then that models task parallelization and mapping of tasks to instances. It means one job is split in to two parallel subtasks. This approach allows the task to be executed on its own instance type, so the characteristics of the requested virtual machines can be adapted to the demands of the current processing phase. Here, we use three channels for scheduling, first one is the network channel, to exchange the data between two subtasks. Second, In-memory channel to enable pipeline processing. Then, file channel is to exchange the data via local file system.



Nephele: Efficient Data Processing in the Cloud

Diagrammatic representation of the system is given bellow,

VII.AN ALTERNATE RESOURCE ALLOCATION FOR CLOUD UNDER HEAVY WORKLOADS

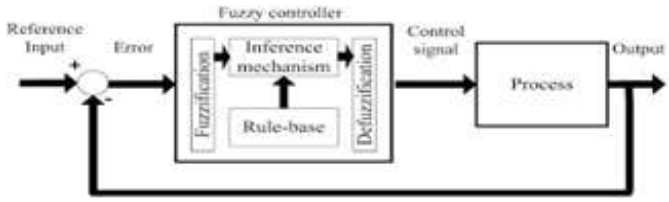
In this session, there is an alternate resource allocation mechanism has been analyzed under heavy workloads .Now a daysAs we know, Cloud computing become very popular network by offering the variety of resources. Heavy load is one of the problem to degrade the performance of an application. Due to this, application is unable to meet the service level agreements and satisfy peak user demands. So some efficient resource allocation is needed with efficient load balancers .Here burstiness-awear algorithm is taken for consideration to balance heavy workloads. It improves overall system performance of the system. The smart load balancer leverages the knowledge of burstiness to predict the changes in the user demands.in this approach the resource allocation has done at two levels. once the application is uploaded to the cloud ,the load balancers assigns the requested instance to the physical computers. Then an application is receiving multiple requests ,these requests should be each assigned to a specific application to balance the computational load across the set of instance of same application. In this approach on-off prediction method is used to forecasts changes in user demands accurately by leveraging the knowledge of burstiess in workload. And smart load balancer which on-the-fly shifts between the schemes that are greedy and random based on the predicted information.

VIII.IMPROVING QOS IN DYNAMIC RESOURCE ALLOCATION

Application level performance is highly dependent on characteristics of workload and sensitive to cloud dynamics. Self-tuning fuzzy control approach is available for response time assurance in web servers to resource allocation. Then for better adaptability and stability one more technique is used. That is adaptive output amplification and flexible rule selection**.** Because of the workload and cloud dynamics the relationships between allocated capacity and received service quality exhibits considerable nonlinearities. The relationship can often be linearized at fixed operating points. It is well known that the linear approximation of the nonlinear system is accurate only with in the neighborhood of the operating point. Abrupt changes in workload traffics and non-determinism in virtual machine capacity possibly make the simple linearization inappropriate. Instead of modeling the system in mathematical equations, fuzzy control employs the control rules of conditional linguistic statements on the relationship of allocated resources and high level objectives. Self tuning fuzzy controller consists of three components namely the fuzzy logic controller, the scaling factor controller and the output amplifier. The fuzzy logic controller implements the static fuzzy control logic. Scaling factor controller together with the output amplifier makes the basic controller adaptive to dynamic server capacity. let us take control interval k+1that is u(k+1) and change of error ⌂e(k) using the reference value r(k),error e(k) is calculate as bellows

$$E(k)=\left(\frac{r(k)-y(k)}{r(k)}\right) 0 \le y(k) \le 2r(k) \quad -1$$

y(k)>2r(k).



Based on this fuzzy controller we can extend two layer provisioning framework, DynaQos that supports adaptive multi objective resource allocation and differentiation**.**

IX.CONCLUSION

In this research paper, both internal and external challenges of resources allocation in cloud computing are analyzed. And some resource allocation polices are discussed .Such as Based on bid proposition how, the resources are allocated .Then ,how optimal resource allocation is achieved in the data centers .Three dynamic resource allocation techniques have identified, such as minimizing the unevenness of multidimensional resource utilization of server, and parallel data processing called as nephele approach. Then dynamic resource allocation with efficient load balancing. Finally, dynamic resource allocation with improved QoS. from the above discussion various factors affecting the resource allocations in cloud infrastructure are analyzed and identified. As a future work ,we are going to apply those identified factors on one another. And simulation will be done on each possible iteration then results are observed. Then improved efficient dynamic resource allocation will be obtained.

REFERENCES

[1] FeiTeng and Fr´ed´ericMagoul`es "Resource Pricing and Equilibrium Allocation Policy in Cloud Computing" 10th IEEE International

Conference on Computer and Information Technology, 2010

[2] Yuuki AWANO and Shin-ichiKuribayashi , "A joint multiple resource allocation method for cloud computing environmentswith different QoS to users at multiple locations" IEEE,2013

[3] Zhen Xiao , weijia song and Qi chen "Dynamic resource Allocation using virtual machines for cloud computing" IEEEParallel and Distributed systems,2012

[4] Daniel warneke and odejkao " Exploiting Dynamic resource allocation for efficient parallel data processing in the cloud"IEEE Parallel and Distributed systems,2011

[5] Jianzhe Tai, Juemin Zhang, Jun Li, WaleedMeleis and NingfangMi "ARA: Adaptive Resource Allocation for Cloud Computing Environments under Bursty Workloads"IEEE,2011

[6] Jia Rao ,Yudiwei,jiayu Gong and cheng-zhongXu"Qos Guarantees and service differentiation for dynamic cloud applications",IEEE 2013

[7] Mohamed Abu Sharkh, ManarJammal, AbdallahShami, and AbdelkaderOuda," Resource Allocation in a Network-BasedCloud Computing Environment: Design Challenges" IEEE Communications Magazine , November 2013

[8] F. Teng and F. Magoul`es. *Fundamentals of Grid Computing:Theory, Algorithms and Technologies*, chapter Futureof Grids Resources Management, pages 133–153. Chapman & Hall/CRC, 2009.

[9] Z.-j. Li and C.-t. Cheng. Utility-driven solution for optimal resource allocation in computational grid. *Computer Languages,Systems and Structures*, 35(4):406–421, 2009

[10] J. Zhang, N. Mi, J. Tai, and W. Meleis, "Decentralized scheduling of bursty workload on computing grids," in *IEEEInternational Conference on Communications (ICC)*, 2011.

[11] A. Caniff, L. Lu, N. Mi, L. Cherkasova, and E. Smirni, "Fastrackfor taming burstiness and saving power in multi-tiered systems," in *Proceedings of the 22nd International Teletraffic Congress (ITC'10)*, Amsterdam, The Netherlands, 2010.

[12] M. de Assuncao, A. di Costanzo, and R. Buyya, "Evaluating the Cost benefit of using cloud computing to extend the capacity ofClusters," in *HPDC '09: Proceedings of the 18th ACMInternational symposium on High performance distributed Computing*, 2009, pp. 141–150

[13] M. Armbrust*et al.*, "Above the Clouds: A Berkeley View of Cloud Computing," tech. rep. UCB/EECS-2009-28, EECS Dept., UC Berkeley, Feb 2009.

[14] L. Minas and B. Ellison, "The Problem of Power Consumptionin Servers," prepared in Intel Lab, *Dr. Dobbs J.*, Mar 2009

[15] B.G. Chun *et al.*, "An Energy Case for Hybrid Datacenters,"*ACM SIGOPS Op. Sys. Rev.*, vol. 44, no. 1, Jan. 2010.

[16] M. Jammal*et al.*, "Software Defined Networking: A Survey,"submitted to *IEEE Communication. Surveys and Tutorials*,July 2013.

[17] I. Fajjari, N. Aitsaadi, G. Pujolle, and H. Zimmermann, "VNE-AC:Virtual Network Embedding Algorithm Based on AntColonMetaheuristic," Proc. IEEE Int'l Conference Communication. (ICC),1-6, June 2011, doi: 10.1109/icc.2011.5963442.