

A Novel Dynamic Resource Allocation Using Virtual Machines for Cloud Computing

¹D. Anil Kumar & ²P.B.V Raja Rao

¹M.Tech student, Dept. of CSE, B V C Engineering College Odalarevu, A.P, India

²Associate Professor at BVC Engineering College, Odalarevu, A.P, India

ABSTRACT:

The main potential of cloud computing is that it's capable of dealing with a tremendous quantity of developing work in a predestined manner for the utilization of the trade consumers. The foremost enabling science for cloud computing is virtualization which generalize the physical infrastructure and makes it effortless to use and manage. In this paper, we present a procedure that makes use of virtualization technology to allocate data center assets dynamically based on utility demands and support green computing by optimizing the quantity of servers in use. We introduce the proposal of "skewness" to measure the unevenness in the multi-dimensional resource utilization of a server. By way of minimizing skewness, we will combine extraordinary varieties of workloads effectively and strengthen the overall utilization of server resources. We advance a suite of heuristics that prevent overload in the method without problems while saving energy used.

Index Terms— Cloud computing; Green computing; Resource; Skewness; Virtual machine.

I. INTRODUCTION

Cloud computing is the supply of computing and storage potential as a provider to a group of finish recipients. The title comes from using a cloud formed symbol as an abstraction for the intricate infrastructure it contains in process diagrams. Cloud computing entrusts offerings with a person's information, application and computation over a community. The remote accessibility allows us to access the cloud services from at any place at any time. To achieve the highest measure of the above acknowledged advantages, the services supplied in phrases of assets should be allotted optimally to the functions going for walks within the cloud. The elasticity and the shortage of upfront capital investment furnished through cloud

computing is attractive to any firms. On this paper, we discuss how the cloud carrier provider can excellent multiplex the on hand virtual resources onto the physical hardware.

We are trying to achieve two goals in our algorithm. Overload avoidance: The capacity of a PM should be sufficient to satisfy the resource needs of all VMs working on it. Otherwise the PM is overloaded and can lead to degraded performance of its VMs. Cloud Computing become a de facto standard for computing, infrastructure as a services has been emerged as an important paradigm in IT area. By applying this paradigm we can abstract the underlying physical resource such a CPUs, Memories and Storage and offer this Virtual Resource to users in the formal Virtual Machine. Multiple Virtual Machines are able to run on a unique physical machine. Multiple VMs are able to run on a unique Physical Machine (PM).

Another important issues in Cloud computing is provisioning method for allocating resources to cloud consumers. Cloud computing environment consists of two provision. The goal is to achieve an optimal solution for provisioning resource which is the most critical part in cloud computing. To make an optimal decision the demand price and waiting-time uncertainties are taken into account to adjust the trade-offs between on-demand and oversubscribed costs.

The Bender's Decomposition is applied to divide the resource optimization problem into many sub problems to decrease the on demand cost and Reservation Cost. Scenario Reduction Technique is applied to reduce problem by reducing number of Scenarios. It will decrease Reservation cost and Expending cost. In the meantime, the advent of multi-cores has enabled the sharing of micro-architectural resources such as shared caches and memory controllers. Contention on such micro-architectural resources has emerged as a major reason for performance variance, as an application can be affected by co-running applications even though it receives the same share of

CPU, memory, and I/O. For a single system, there have been several prior studies to mitigate the impact of contention on shared caches and memory controllers by carefully scheduling threads. The prior studies rely on the heterogeneity of memory behaviors of applications within a system boundary. The techniques group applications to share a cache to minimize the overall cache misses for a system. However, if a single system runs applications with similar cache behaviors, such intra-system scheduling cannot mitigate contentions.

There is an inherent tradeoff between the two goals in the face of changing resource needs of VMs, overload avoidance. We should keep the utilization of PMs low to reduce the possibility of overload in case the resource needs of VMs increasing later, for green computing we should keep the utilization of PMs reasonably high to make efficient use of their energy.

II. RELATED WORK

In [1] “On networking and computing environments' integration: A novel mobile cloud resources provisioning approach” by Skoutas, D.N. , Skianis, C. A Mobile Cloud Resources Provisioning (MCRP) scheme, which is flexible enough to adapt to the various general MCC reference use cases being described. The main novelty feature of the employed MCC Service Admission Control algorithm lies in the fact that it jointly handles radio and computing resources rather than confronting the problem as two independent resource management sub-problems.

In [2] “Optimal resource allocation for multimedia cloud in priority service scheme “by He, Yifeng; Guan, Ling employ the queuing model to optimize the resource allocation for multimedia cloud in priority service scheme. Specifically, formulate and solve the resource cost minimization problem and the service response time minimization problem respectively.

In [3] “Optimization of Resource Provisioning Cost in Cloud Computing “by BuSung Lee , Niyato, D. an optimal cloud resource provisioning (OCRP) algorithm is proposed by formulating a stochastic programming model. The OCRP algorithm can provision computing resources for being used in multiple provisioning stages as well as a long-term plan, e.g., four stages in a quarter plan and twelve stages in a yearly plan. The demand and price uncertainty is considered in OCRP. In this paper, different approaches to obtain the solution of the OCRP algorithm are considered including deterministic equivalent formulation, sample-average approximation, and Benders

decomposition. Numerical studies are extensively performed in which the results clearly show that with the OCRP algorithm, cloud consumer can successfully minimize total cost of resource provisioning in cloud computing environments.

The OCRP algorithm was proposed in [4].The OCRP algorithm can find an optimal solution for resource provisioning and VM placement. It uses only two uncertainties only viz., demand and price. Here in this paper RCRP algorithm is used which is an extension of OCRP where four uncertainty factors are considered. Grid provides services which are not of desired quality. One of the major drawbacks of grid is single point of failure where one unit on the grid degrades which will cause the entire system to degrade.

Hence [5] suggests cloud which is used for adaption of various services. The benefit of cloud is that it will avoid single point of failure and also will decrease hardware cost.

III. PROPOSED SYSTEM

This proposed system consists of number of servers, predictor, hotspot and cold spot solvers and migration list. Set of servers used for running different applications. Predictor is used to execute periodically to evaluate the resource allocation status based on the predicted future demands of virtual machines.

A. System Overview

The architecture of the system is presented over in Fig. 1. Each physical machine (PM) runs the Xen hypervisor (VMM) which supports a privileged domain 0 and one or more domain U [7]. Each VM in domain U encapsulates one or more applications such as Web server, remote desktop, DNS, Mail, Map/Reduce, etc. We assume all PMs share backend storage. The multiplexing of VMs to PMs is managed using the Usher framework [8]. The main logic of our system is implemented as a set of plug-ins to Usher. Each node runs an Usher local node manager (LNM) on domain 0 which collects the usage statistics of resources for each VM on that node. The statistics collected at each PM are forwarded to the Usher central controller (Usher CTRL) where our VM scheduler runs.

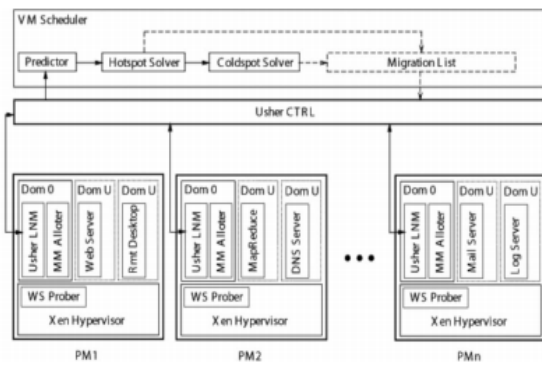


Fig. 1 System Architecture

The VM Scheduler is invoked periodically and receives from the LNM the resource demand history of VMs, the capacity and the load history of PMs, and the current layout of VMs on PMs. The scheduler has several components. The predictor predicts the future resource demands of VMs and the future load of PMs based on past statistics. We compute the load of a PM by aggregating the resource usage of its VMs. The LNM at each node first attempts to satisfy the new demands locally by adjusting the resource allocation of VMs sharing the same VMM. The MM Alloter on domain 0 of each node is responsible for adjusting the local memory allocation. The hot spot solver in our VM Scheduler detects if the resource utilization of any PM is above the hot threshold (i.e., a hot spot). The cold spot solver checks if the average utilization of actively used PMs (APMs) is below the green computing threshold.

B. Skewness Algorithm

We introduce the concept of “skewness” to measure the unevenness in the multi-dimensional resource utilization of a server. By minimizing skewness, we can combine different types of workloads nicely and improve the overall utilization of server resources. Let n be the number of resources we consider and r_i be the utilization of the i-th resource. We define the resource skewness of a server p as

$$skewness(p) = \sqrt{\sum_{i=1}^N \left(\frac{r_i}{r} - 1\right)^2}$$

Where r is the ordinary utilization of all assets for server p. In apply, now not all types of resources are efficiency relevant and therefore we only have got to consider block assets in the above calculation. By way of minimizing the skewness, we can combine unique types of workloads effectively and toughen the overall utilization of server assets. The go with the flow chart represents the flow of an

algorithm in Fig 2. Our algorithm executes periodically to evaluate the useful resource allocation status headquartered on the expected future resource demands of VMs. We outline a server as a scorching spot if the utilization of any of its assets is above a sizzling threshold. We outline the temperature of a hot spot p as the square sum of its resource utilization beyond the hot threshold:

$$temperature(p) = \sum_{r \in R} (r - r_t)^2$$

where R is the set of overloaded assets in server p and r_t is the sizzling threshold for useful resource r.

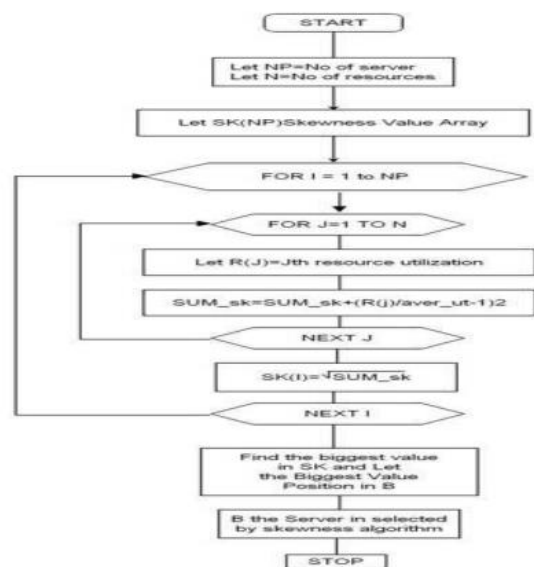


Fig. 2 Flow Chart of Skewness

We define a server as a cold spot if the utilizations of all its resources are below a cold threshold. This indicates that the server is most commonly idle and an advantage candidate to show off to save lots of energy. Ultimately, we outline the warm threshold to be a degree of useful resource utilization that is sufficiently excessive to justify having the server running however not as high as to danger becoming a hot spot in the face of transitory fluctuation of software resource needs.

C. Hotspot Mitigation

We handle the hottest one first i.e., Kind the record of hot spots in the procedure otherwise, hold their temperature as low as viable. Our purpose is emigrate the VM that can shrink the server’s temperature. In case of ties, the VM whose removing can curb the skewness of the server probably the most is chosen. We first make a decision for each and every server p which of its VMs should be migrated away. Based on the resulting temperature we sort

record the VMs of the server if that VM is migrated away. We see if we can discover a destination server to accommodate it for each and every record of within the VM. After accepting this VM the server should no longer emerge as hot spot. We choose one skewness which can also be decreased probably the most via accepting this VM among all servers. We record the migration of the VM to that server and update the envisioned load of associated servers when the destination server is located. Else we move on to the subsequent VM within the list and try to discover a vacation spot server for it.

D .Green Computing

When the resource utilization of active servers is too low, some of them can be turned off to save energy. This is handled in our green computing algorithm. Our green computing algorithm is invoked when the average utilizations of all resources on active servers are below the green computing threshold. We check if we can migrate all its VMs somewhere else for a cold spot p. For each VM on p, we try to find a destination server to accommodate it. The utilizations of resources of the server after accepting the VM must be below the warm threshold.

IV. RESULTS AND DISCUSSION

The purpose of the skewness algorithm is to mix workloads with one-of-a-kind resource requisites collectively so that the overall utilization of server potential is improved. In this scan, we see how our algorithm handles a mix of CPU, memory, and community intensive workloads. Useful resource allocation reputed of three servers A, B, C has total reminiscence allotted 500KB and resource used memory for server A 80KB, server B 170KB and server C 80K. In Fig. 4 each cloud customers provide cloud provider resource allocation in green computing. In Fig.5 show Server utilization and useful resource allocation fame for user1 utilizing Bar Chart. The cloud computing is a mannequin which permits on demand network access to a shared pool computing assets. Cloud computing environment consists of multiple consumers soliciting for assets in a dynamic atmosphere with their many possible constraints. The virtualization may also be the solution for it. It may be used to cut down energy consumption by using knowledge centers. The most important intent of the virtualization is that to take advantage of efficient use of to be had procedure assets, together with vigour. A knowledge core, putting in virtual infrastructure enables a few running techniques and purposes to run on a lesser number of servers, it might help to slash the overall energy used for the info core and the

energy consumed for its cooling. Once the quantity of servers is lowered, it additionally means that data core can diminish the building dimension as well. Some of the advantages of Virtualization which directly affects efficiency and contributes to the environment include: Workload balancing across servers, resource allocation and sharing are higher monitored and managed and the Server utilization rates will also be improved up to eighty% as compared to initial 10-15%.

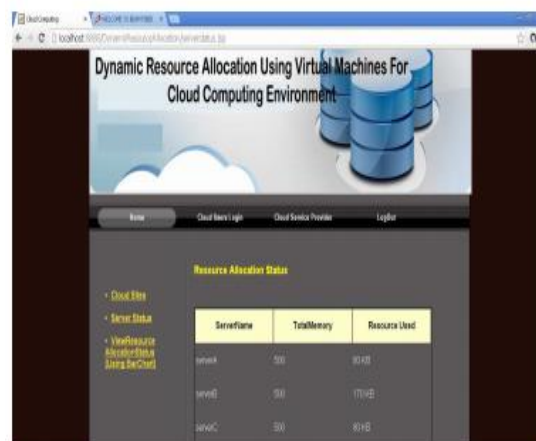


Fig. 3 Resource Allocation Status

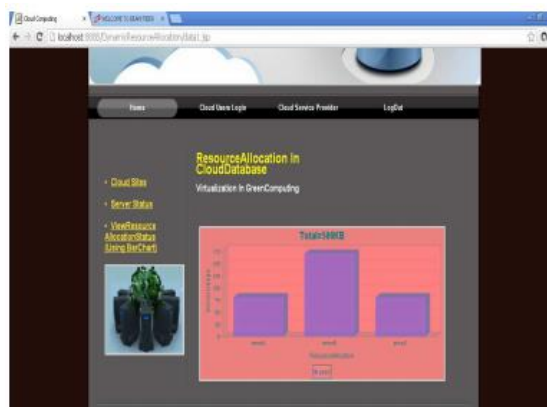


Fig. 4 View Resource Allocation Status using Bar Chart

The results are clear and having good contribution:

- 1) Allocation of resource is done dynamically.
- 2) Saves the energy using the green computing concept
- 3) Proper utilization of servers and memory utilization is taken care using skewness.
- 4) Minimize the total cost of both the cloud computing infrastructure and running application.

V. CONCLUSION

We have now presented the design, implementation, and evaluation of a useful resource administration procedure for cloud computing services. Our process multiplexes virtual to bodily assets adaptively based on the altering demand. We use the skewness metric to mingling VMs with specific useful resource characteristics competently so that the capacities of servers are well utilized. Our algorithm achieves both overload avoidance and green computing for methods with multi useful resource constraints.

REFERENCES

- [1] M. Armbrust et al., "Above the Clouds: A Berkeley View of Cloud Computing," technical report, Univ. o California, Berkeley, Feb. 2009.
- [2] L. Siegele, "Let It Rise: A Special Report on Corporate IT," *The Economist*, vol. 389, pp. 3-16, Oct. 2008.
- [3] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the Art of Virtualization," *Proc. ACM Symp. Operating Systems Principles (SOSP '03)*, Oct. 2003E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagate.*, to be published.
- [4] "Amazon elastic compute cloud (Amazon EC2)," <http://aws.amazon.com/ec2/>, 2012.
- [5] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live Migration of Virtual Machines," *Proc. Symp. Networked Systems Design and Implementation (NSDI '05)*, May 2005.
- [6] M. Nelson, B.-H. Lim, and G. Hutchins, "Fast Transparent Migration for Virtual Machines," *Proc. USENIX Ann. Technical Conf.*, 2005.M. Young, *The Technical Writers Handbook*. Mill Valley, CA: University Science, 1989.
- [7] N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Virtual Machines for Managing SLA Violations," *Proc. IFIP/IEEE Int'l Symp. Integrated Network Management (IM '07)*, 2007.
- [8] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-Box and Gray-Box Strategies for Virtual Machine Migration," *Proc. Symp. Networked Systems Design and Implementation (NSDI '07)*, Apr. 2007.

BIODATA

Author



D. Anil Kumar has received his B.Tech (I.T), from Bonam Venkata Chalamayya Institute of Technology, East Godavari in 2013. Presently he is pursuing his M.Tech in Software Engineering from BVC Engineering College Odalarevu, A.P, India. Her research interests include Software testing Data structures and web services.

Co-Author

P.B.V Raja Rao has received his M.Tech. Presently working as Associate Professor in BVC Engineering College Odalarevu, A.P, India.