International Journal of Research Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment

[1]M.Pavani, [2]Dr. Ravindar Reddy Thokala

[1] M.Tech Student, Dept of CSE, Brilliant grammar school educational institutions group of institutions integrated campus, T.S, India

[2]Associate Professor, Dept of CSE, Brilliant grammar school educational institutions group of institutions integrated campus, T.S, India

Abstract

Cloud computing allows business customers to scale up and down their resource usage based on needs. Many of the touted gains in the cloud model come from resource multiplexing through virtualization technology. In this paper, we present a system that uses virtualization technology to allocate data center resources dynamically based on application demands and support green computing by optimizing the number of servers in use. 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. We develop a set of heuristics that prevent overload in the system effectively while saving energy used. Trace driven simulation and experiment results demonstrate that our algorithm achieves good performance

INTRODUCTION

Cloud computing is the delivery of computing and storage capacity as a service to a community of end recipients. The name comes from the use of a cloud shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts services with a user's data, software and computation over a network. The remote accessibility enables us to access the cloud services from anywhere at any time. To gain the maximum degree of the above mentioned benefits, the services offered in terms of resources should be allocated optimally to the applications running in the cloud.

The elasticity and the lack of upfront capital investment offered by cloud computing is appealing to any businesses. In this paper, we discuss how the cloud service provider can best multiplex the available virtual resources onto the physical hardware. This is important because much of the touted gains in the cloud model come from such multiplexing. Virtual Machine Monitors (VMMs) like Xen provide a mechanism for mapping Virtual Machines (VMs) to Physical Resources This mapping is hidden from the cloud users. Users with the Amazon EC2 service for example, do not know where their VM instances run. It is up to the Cloud Service Provider to make sure the underlying Physical Machines (PMs) has sufficient resources to meet their needs VM live migration technology makes it possible to change the mapping between VMs and PMs.

This is challenging when the resource needs of VMs are heterogeneous due to the diverse set of applications they run and vary with time as the workloads grow and shrink. The capacity of PMs can also be heterogeneous because multiple generations of hardware co-exist in a data center. To achieve the overload avoidance that is the capacity of a PM should be sufficient to satisfy the resource needs of all VMs running on it. Otherwise, the PM is overloaded and can lead to degraded performance of its VMs. And also the number of PMs used should be minimized as long as they can still satisfy the needs of all VMs. Idle PMs can be turned off to save energy. In this paper, we presented the design and implementation of dynamic resource allocation in the Virtualized Cloud Environment which maintains the balance between the following two goals.

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.

We propose contention-aware cloud scheduling techniques for cache sharing and NUMA affinity. The techniques identify the cache behaviors of VMs on-line, and dynamically migrate VMs, if the current placements of VMs are causing excessive shared cache conflicts or wrong NUMA affinity. Since the techniques identify the VM behaviors dynamically and resolve conflicts with live migration it will not required any prior knowledge on the behaviors of VMs. The first technique, cache-aware cloud scheduling minimizes the overall last-level cache (LLC) misses in a cloud system.

The second technique, NUMAaware cloud scheduling extends the first technique by considering NUMA affinity. We evaluate our proposed schedulers using selected Speculum 2006 applications in various combinations. The experimental results show that the cache-aware scheduler can significantly improve the performance compared with the worst case. With our preliminary NUMA optimization, the performance is slightly improved for our benchmark applications, compared with that of the cacheaware scheduler.

RELATED WORK

In author proposed architecture, using feedback control theory, for adaptive management of virtualized resources, which is based on VM. In this VM-based architecture all hardware resources are pooled into common shared space in cloud computing infrastructure so that hosted application can access the required resources as per there need to meet Service Level Objective (SLOs) of application. The adaptive manager use in this architecture is multi-input multi-output (MIMO) resource manager, which includes 3 controllers:

CPU controller, memory controller and I/O controller, its goal is regulate multiple virtualized resources utilization to achieve SLOs of application by using control inputs per-VM CPU, memory and I/O allocation. The seminal work of Walsh et al. proposed a general two-layer architecture that uses utility functions, adopted in the context of dynamic and autonomous resource allocation, which consists of local agents and global arbiter. The responsibility of local agents is to calculate utilities, for given current or forecasted workload and range of resources, for each AE and results are transfer to global arbiter.

Where, global arbiter computes near-optimal configuration of resources based on the results provided by the local agents. In authors propose an adaptive resource allocation algorithm for the cloud system with preempt able tasks in which algorithms adjust the resource allocation adaptively based on the updated of the actual task executions. Adaptive list scheduling (ALS) and adaptive min-min scheduling (AMMS) algorithms are use for task scheduling which includes static task scheduling, for static resource allocation, is generated offline. The online adaptive procedure is use for re-evaluating the remaining static resource allocation repeatedly with predefined frequency.

The dynamic resource allocation based on distributed multiple criteria decisions in computing cloud explain in it author contribution is tow-fold, first distributed architecture is adopted, in which resource management is divided into independent tasks, each of which is performed by Autonomous Node Agents (NA) in ac cycle of three activities:

(1) VM Placement, in it suitable physical machine (PM) is found which is capable of running given VM and then assigned VM to that PM,

(2) Monitoring, in it total resources use by hosted VM are monitored by NA,

(3) In VM selection, if local accommodation is not possible, a VM need

to migrate at another PM and process loops back to into placement and second, using PROMETHEE method, NA carry out configuration in parallel through multiple criteria decision analysis. This approach is potentially more feasible in large data centers than centralized approaches.

Execution Time Different kinds of resource allocation mechanisms are proposed in cloud. The actual task execution time and pre-emptible scheduling is considered for various resource allocations. It overviews the problem of resource contention and increases resource utilization by using different modes of renting computing capacities. But estimating the execution time for a job is a hard task for a user and errors are made very often. But the VM model considered in is heterogeneous and proposed for IaaS.

Policy Since centralized user and resource management lacks in scalable management of users, resources and organization-level security policy, we proposed a decentralized user and virtualized resource management for IaaS by adding a new layer called domain in between the user and the virtualized resources. Based on role based access control (RBAC), virtualized resources are allocated to users through domain layer.

Virtual Machine (VM) A system which can automatically scale it's infrastructure resources is designed the system composed of a virtual network of virtual machines capable of live migration across multi- domain physical infrastructure Cloud computing services providers deliver their resources based on virtualization to satisfy the need of users. In cloud computing, the amount of resources required can vary preserve request. Therefore the providers have to offer Different amounts of virtualized resources per request.

Overload avoidance:

The capacity of a PM should be sufficient to satisfy the resource needs of all VMs running on it. Secondly, the PM is overloaded and can lead to degraded performance of its VMs. Green computing: The number of PMs should be minimized as long as they can satisfy the needs of all VMs. Idle PMs can be switch off to save energy. We develop a resource allocation system that can avoid overload in the system effectively while minimizing the number of servers used. We introduce the concept of “skewness” to measure the uneven utilization of a servers By minimized skewness we can increase the overall utilization of servers interface of multidimensional resource constraints.

We propose a load prediction algorithm that can capture the future resource usages of applications accurately without looking the VMs. The algorithm captures the rising trends of resource usage patterns and help reduce the placement churn significantly. The cloud computing is a model which enables on demand network access to a shared pool computing resources. Cloud computing environment consists of multiple customers requesting for resources in a dynamic environment with their many possible constraints. In existing system cloud computing allocating the resource efficient is a challenging task.

In this paper we proposed allocates resource with less wastage and provides much profit. The developed resource allocation algorithm is based on different parameters as: time, cost, No of processor, request etc. Priority model that mainly decides priority among different user request based on many parameters like cost of resource, time needed to access, task type, number of processors needed to run the job or task. In this model client send the request to the cloud server. The cloud service provider runs the task submitted by the client. The cloud admin decides the priority among the different users request. Each request consists of different tasks. It have the different parameters such as Time-computation, timeneeded to complete the particular task, Processor requestrefers to number of processors needed to run the task.

The more number of processor faster the completion of task importance-refers to how important the user to a cloud administrator (admin) that is whether the user is old customer to cloud or new customer. Price-refers to cost charged by cloud admin to cloud users. Cloud computing is a model which enables on demand network access to a shared pool computing resources. A cloud environment consists of multiple customers requesting for resources in a dynamic environment with possible constraints. In existing system cloud computing, allocating the resource usually is a challenging job. The cloud does not show the quality of services.

Scenario Tree Generation

One of the principal challenges in the field of stochastic programming handles with finding effective ways to assess the significance of atoms, and to make use of that data to reduce the tree of scenario's in such a way that the solution to the smaller best possible solution difficulty is not much dissimilar than the difficulty stated with the original tree. The Generation of Scenario Tree algorithm is a finite element technique that deals with this difficulty for the class of LSMP with random variables Scenario Tree Reduction The Scenario Tree Reduction (STR) algorithm is a finite element technique that deals with this problem for the class LSMP with random variables.

Implementation Details are performance of the proposed resource optimization framework is implemented using eclipse based java platform. Both Under provisioning and Over provisioning is solved by using different optimization problems under uncertainty. Resource Utilization and Cost Calculation is implemented in the Proposed system. In Fig 5. It displays the Customer details, resources utilization and cost calculation.

CONCLUSION

We have presented the design, implementation, and evaluation of a resource management system for cloud computing services. Our system multiplexes virtual to physical resources adaptively based on the changing demand. We use the skewness metric to combine VMs with different resource characteristics appropriately so that the capacities of servers are well utilized. Our algorithm achieves both overload avoidance and green computing for systems with multi resource constraints. we propose a system that uses virtualization technology to allocate data center resources dynamically based on application needs and support green computing by optimizing the number of servers in use. We proposed the concept of "skewness" to measure the un-evenness in the multidimensional resource utilization of a server. By minimized skewness, we can combining different of workloads and improve the over-all utilization of server resources. We develop a set of heuristics that prevent overload in the system effectively while saving energy used. Trace driven simulations and experimental results demonstrate that ours algorithm achieves good performance.

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 Propagat., 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 Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.