IJDCST @ Nov-Dec, Issue- V-3, I-1, SW-79
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Dynamic Resource management with VM layer and Resource prediction algorithms in Cloud Architecture

[1]Shaik Fayaz, [2]Dr.V.N.Srinivasu, [3]Tata Venkateswarlu
[#1] M.Tech (CSE) from P.N.C & Vijai Institute of Engineering And Technology, Repudi, Guntur.
[#2] HOD of CSE Dept. in P.N.C & Vijai Institute of Engineering And Technology, Repudi, Guntur.
[#3] Assistant Professor in P.N.C & Vijai Institute of Engineering And Technology, Repudi, Guntur.

Abstract: Cloud computing became an emerging computation technology to offer reliable services in the form of IaaS, SaaS and PaaS. Today business entrepreneurs are updating their infrastructure to adopt for cloud environment to have the miscellaneous advantages of cloud computing architecture. Resource management is an important aspect of cloud to enhance the scalability and efficiency in terms of resource utilization of cloud. As on multiple advancements were introduced in the area of cloud resource management to achieve the efficiency. This paper introduces dynamic resource management through virtual systems is an extreme process in the area of resource allocation and management. We proposed an updated “skewness algorithm” to predicate, allocate and reclaim the resources dynamically using virtualization concept. Experiments on our approach are demonstrating high availability and efficient utilization of cloud resources.

Keywords: Cloud Computing, Resource management, virtual systems, updated skewness algorithm.

1. Introduction

Cloud computing is typically defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Based on the available features and nature of cloud environment we can assess that this is combination of distributed computing, parallel processing and grid computing. In cloud computing, the word cloud is used as a metaphor for "the Internet," so the phrase cloud computing means "a type of Internet-based computing," where different services - such as IaaS, PaaS and SaaS are delivered to an organization's computers and devices through the Internet. In Cloud computing, computing infrastructure and services should always be available on computing servers (which are distributed among all continents) such that companies are able to access their business services and applications anywhere in the world whenever they need as shown in Figure1. Cloud computing delivers the above mentioned infrastructure, platform and software (applications) as services that are made available to consumers in a pay-per-use basis.

Several providers have cloud computing solutions available, where a pool of virtualized and dynamically scalable computing power, storage, platforms, and services are delivered on demand to clients over the Internet on a pay as you go basis. This is implemented using virtualization technology where clients are just a credit card payment away from scaling their rented virtual machines (VMs) dynamically to include as many machines as they need. Network performance and resource availability is the prominent bottleneck for any cloud architecture of today.

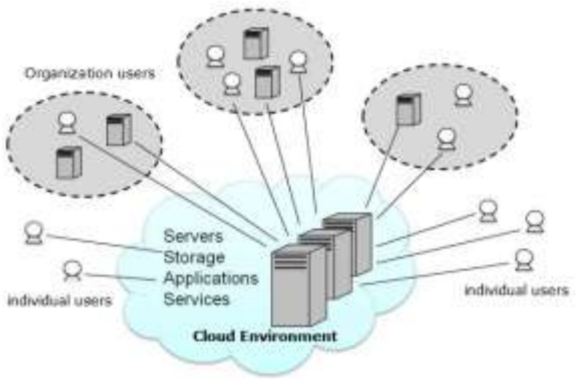

Figure.1. Basic Cloud Architecture

Cloud resource allocation and management is an evolving part of many cloud data center management issues such as virtual machine deployment in data centers, network virtualization, and multi-path network routing etc.

Majority of Business customers interested towards cloud computing and they started their app migration with cloud environment to promote their business operations to end client with low investments and high availability. Due to this increased adoption, Resource Management in Cloud (RMC) becomes an important research aspect in this area. Earlier approaches [1, 3] were used evenness procedure in resource distribution to allocate the available resources among the running applications. This approach may leads to resource over flow due to high amount of resource allocation than required and resource underflow due to less amount of resource allocation than required. Always resource needs for a running application changes from time to time depends on number of live clients.

To avoid the above problems this paper introduces dynamic resource management through virtual systems is an extreme process in the area of resource allocation and management. We proposed an updated “skewness algorithm” to allocate and reclaim the resources dynamically using virtualization concept. Experiments on our approach are demonstrating high availability and efficient utilization of cloud resources.

2. Related work

2.1) Resource allocation: However, despite the recent and notable growth of the Cloud Computing market, there are several key issues open with the process of resource allocation. In cloud computing, Resource Allocation (RA) is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation starves services if the allocation is not managed precisely. A large portion of the work in resource allocation in cloud computing mainly focused on the cost-effectiveness and easy maintenance of the systems. Increasing demand for cloud migration enforces the efficient resource management in cloud computing [1 and 2]. Static resource allocation and dynamic resource allocation were the two important methodologies in this area for resource management. In static resource allocation method [3], all systems will share the available resources (CPU time, Memory, Registers and Network bandwidth etc.) equally without having any priorities. Later dynamic resource management introduced unevenness procedure to distribute physical resources among virtual servers of cloud environment. This approach will be automatically refreshed eventually to update the allocated resources of server based on the workload. In order to predict the hot spot and cold spots [7, 9] this method will determine the threshold from periodic evaluations of workload. CPU and Network resources are monitored by scheduling algorithms and memory resources are by swap activities in this area.

2.2) Virtual Machine (VM): A system which can automatically scale its infrastructure resources is designed in [2 and 4]. The system composed of a virtual network of virtual machines capable of live migration across multi- domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources. But the above work considers only the non-preempt able scheduling policy [5]. Virtualization allows on-demand remapping of virtual resources over physical resources and thus enables the adaptation of systems to dynamic workloads. This remapping primitive is defined as migration and allows workload transfers among different physical machines without interrupting their execution. Currently, migration is manually triggered by network managers to load balance data centers [6]. This reallocation scheme is inefficient because of its high reaction time, which is inadequate to dynamic workload environments like clouds.

3. Dynamic resource allocation with updated skewness algorithm

Recent cloud architectures are facing resource management problem, due to unexpected requirement of huge resources (CPU time, memory, networks etc.) in an asynchronous manner. Current dynamic resource allocation methodologies [5, 8] are having the capability to map virtual system resources with physical systems dynamically depends on work load. This process will adjust the available resources with the help of hotspot and cold spot migration [8] among the physical machines. Dynamic resource allocation may fail under some circumstances, when suddenly there is an unexpected huge resource requirement for a physical machine. To address the above problem, in this paper we introduced dynamic resource management through virtual systems is an extreme process in the area of resource allocation and management. We proposed an updated "skewness algorithm" to allocate and reclaim the resources dynamically using virtualization concept. This algorithm is an extension to dynamic resource allocation and management architecture as shown below figure 2.

3.1) Management layer: This layer is the conjunction between cloud layer and the VM layer and having collection of modules like VM Manager, Resource prediction unit, load management resource grid and job scheduler module etc. Secured communication relevant to all aspects is implemented in this layer along with https like secured protocols [9]. This layer contains the high speed wired and wireless network components to support high speed commutation among clouds. This layer is the backbone for this architecture because intra cloud communication is highly demanded in this implementation for resource, process sharing and management.

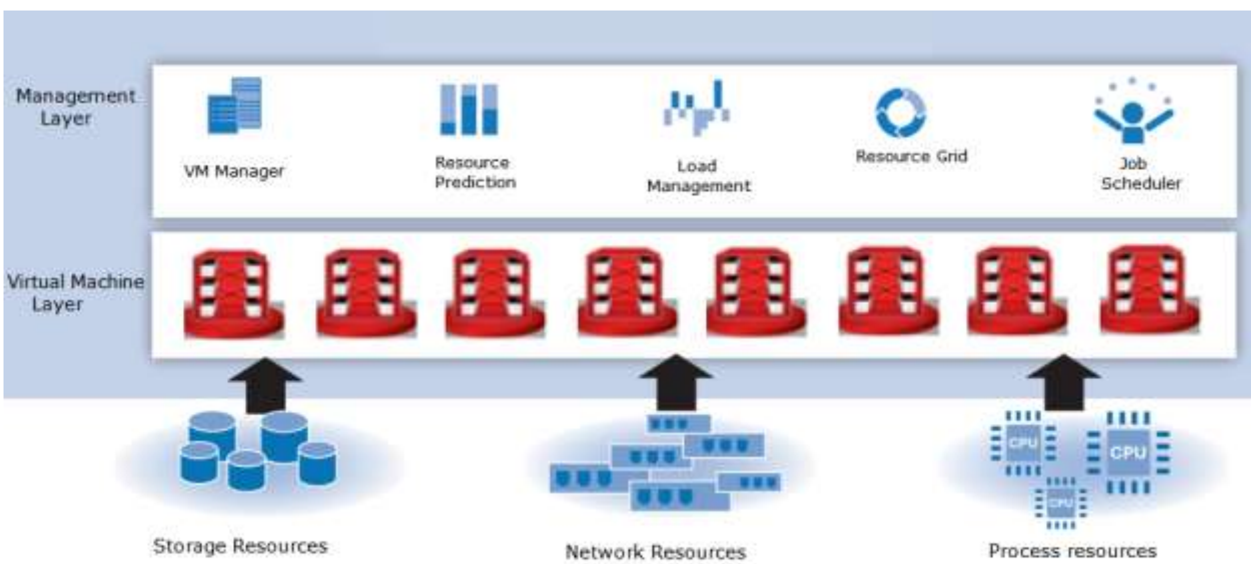

Figure 2. Dynamic resource allocation strategy through VM layer

3.2) VM layer: This layer is a list of virtual machine groups where each group is a set of virtual machines for physical system mapping. In this layer we proposed a separate virtual machine group for single cloud architecture to achieve the operational feasibility. Each VM is a virtual representation of a cloud physical machine at VM Layer level. This is the prominent way to perform mapping between logical resources usage with physical requirements. Every VM is having the OS part for virtual process management and the data part for data manipulations. Virtual machines are always available to resource grid module for monitoring and resource mapping.

3.3) Job Scheduler: This module has to schedule the programs as per workload manager instructions in a concurrent way. Job scheduler will make the CPU time utilization efficiently to overcome bottleneck problems due to network congestion, I/O delays from user. This job scheduler uses the dynamic task scheduling algorithms [8] for job scheduling and implements the multi-tasking with parallel programming.

3.4) Resource grid: Resource grid is a logical module to manage the resource pool and resource allocation strategies. Resource pool is a collection of Databases, server resources, processing resources, main memory resources and Storage areas of IaaS. As per the job scheduler requirements VM manager make available the resources to allocate for process execution. Resource grid follows the dynamic resource allocation procedure to manage the resources of multiple clouds among distributed processing environments of cloud architecture.

3.5) Resource prediction: we are introducing the naïve concept is time bounded resource prediction for future. After analyzing many existing resource allocation methodologies we observed that resource allocation requirements are changing from time to time, here the time stands for hour to hour and day to day. Henceforth predicting resource requirements for future will be a time bound operation to support the green computing policy specifications. To avoid the complexity of time bounded prediction we are analyzing from hour to day. While considering accuracy and effectiveness in resource prediction the time bounded prediction is having the high scalability.

4. Updated Skewness Algorithm

This updated skewness algorithm outputs the allocation strategy for dynamic resource management. This algorithm identifies the resource overflow and underflow problems at every resource pool level and plays a vital role in the process of allocation, de allocation and transformation of resources in resource grid. This algorithm takes the present load, allocated resource, growth rate and sentiment analysis report as input to suggest either resource allocation or de allocation as shown below:

Updated Skewness Algorithm:

Begin

Input: present load, allocated resources, growth rate

Output: resource overflow or underflow indicator

step 1: Map the process execution with VM pool & consider a VM for analysis.

step 2: Examine the current resource load (C) vs allocated resources (A)

step 3: Extract the recent past growth rate of the process from VM manager.

step 4: cross check sentiment analysis against present executing process and save the results (R).

step 5: assess the future requirements (F) based on growth rate.

step 6: if ((F&&R) > (A&&C))

Indicate Underflow;

Else indicate Overflow;

End

5. Conclusion

In this paper we concentrated on the cloud issues like elasticity, availability, resource management and resource distribution of cloud architecture. To overcome the above issues, we presented dynamic resource allocation with updated skewness algorithm for efficient resource management in cloud environment. This approach is an extension to dynamic resource allocation and management architecture. This logical extension managed the resources efficiently by managing the resource pool, achieved the security by distributing the data among multiple clouds and elasticity by sharing the resources depends on workload among all cloud environments.

6. References

1) Atsuo Inomata, TaikiMorikawa, Minoru Ikebe, Sk.Md. Mizanur Rahman: Proposal and Evaluation of Dynamin Resource Allocation Method Based on the Load Of VMs on IaaS (IEEE,2010),978-1-4244-8704-2/11.

2) C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, “A scalable application placement controller for enterprise data centers,” in Proc. Of the International World Wide Web Conference (WWW’07), May 2007.

3) A. Celesti, F. Tusa, M. Villari, and A. Puliafito, “How to enhance cloud architectures to enable cross-federation,” in 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD). Los Alamitos, CA, USA: IEEE Computer Society, 2010, pp. 337–345.

4) N. Santos, K. P. Gummadi, and R. Rodrigues. Towards trusted cloud computing. In Hot Topics in Cloud Computing (HotCloud’09). USENIX Association, 2009.

5) N. Bobroff, A. Kochut, and K. Beaty, “Dynamic placement of virtual machines for managing sla violations,” in Proc. of the IFIP/IEEE International Symposium on Integrated Network Management (IM’07), 2007.

6) 2. J. S. Chase, D. C. Anderson, P. N. Thakar, A. M. Vahdat, and R. P. Doyle, “Managing energy and server resources in hosting centers,” in Proc. Of the ACM Symposium on Operating System Principles (SOSP’01), Oct. 2001.

7) 3. M. Jensen, J. Schwenk, N. Gruschka, and L. Lo Iacono, “On Technical Security Issues in Cloud Computing,”Proc. IEEE Int’l Conf. Cloud Computing (CLOUD-II),2009.

8) M. Nelson, B.-H. Lim, and G. Hutchins, “Fast transparent migration for virtual machines,” in Proc. of the USENIX Annual Technical Conference,2005.

9) A.Singh,M.Korupolu and D. Mohapatra “Server-storage virtualization inte-gration and load balancing in data centers,” in Proc. of the ACM/IEEE conference on Supercomputing, 2008