International Journal of Advanced Technology in Engineering and Science www.ijates.com
Volume No 03, Special Issue No. 01, March 2015 ISSN (online): 2348 – 7550

RESOURCE MANAGEMENT IN CLOUD COMPUTING ENVIRONMENT

A.Chermaraj[1], Dr.P.Marikkannu[2]

[1]PG Scholar, [2]Assistant Professor, Department of IT,

Anna University Regional Centre Coimbatore, Tamilnadu (India)

ABSTRACT

Cloud computing is the term for delivery of hosted services over the internet. It enables companies to consume resource based on their needs scale up and down. Most of the companies move to this technology due to reduction of cost Cloud computing delivery models are categorized as Iaas, Paas, Saas. Most of the cloud service provider gives the same service with different cost and speed. So the users can select the services based on their needs and cloud providers criteria like cost. In existing system is build upon future load prediction mechanism. Based on it, virtual machine allocation was done. In this paper Cloud Booster Algorithm is used, it provides the resource management. In this system uses virtualization technology to allocate resources dynamically based on the application demands.

Keywords: Cloud Computing, Virtualization, Virtual Machine, Resource Management.

I. INTRODUCTION

Cloud computing, as the name suggests, is a style of computing where dynamically scalable and often visualized resources are provided as a service over the internet. These services can be consumed by any user over a standard HTTP medium. The user doesn't need to have the knowledge, expertise, or control over the technology infrastructure in the "cloud" that supports them. The elasticity and the low cost of capital investment is offered by cloud computing to many businesses. Cloud computing platform guarantees service, scope, quality by service level agreement (SLA) to users. It allows users can scale up and down resource usage based on their resource needs. In cloud computing environment resources are shared, if they are not efficiently distributed then the result will be resource wastage The cloud is a next generation platform that provides dynamic resource pools, virtualization, and high availability. Today, we have the ability to utilize scalable, distributed computing environments within the confines of the Internet, a practice known as cloud computing. Cloud computing is the concept implemented to decipher the daily computing problems, likes of hardware software and resource availability unhurried by computer users. The cloud computing platform guarantees subscribers that it sticks to the service level agreement (SLA) by providing resources as service and by needs. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and help the users focus on their core business instead of being impeded by IT obstacles. The main enabling technology for cloud computing is virtualization. Virtualization generalizes the physical infrastructure, which is the most rigid component, and makes it available as a soft component that is easy to use and manage. By doing so, virtualization provides the agility required to speed up IT operations, and reduces cost by increasing infrastructure utilization. On the other hand, autonomic computing automates the process through which the user can provision resources on-demand. By minimizing user involvement, automation speeds up the process and reduces the possibility of human errors.

Modern computers are sufficiently powerful to use virtualization to present the illusion of many smaller virtual machines (VMs), each running a separate operating system instance. This has led to a resurgence of interest in VM technology. "Virtualization, in computing, is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system, a storage device or network resources" classification system.

Cloud can exist without Virtualization, although it will be difficult and inefficient. Cloud makes notion of "Pay for what you use", "infinite availability- use as much you want". These notions are practical only if we have.

- Lot of Flexibility
- Efficiency in the back-end.

This efficiency is readily available in Virtualized Environments and Machines. On a cloud computing platform, dynamic resources can be effectively managed using virtualization technology. Amazon EC2 provides virtualized server instances. Whereas some resources like CPU, memory, and instance storage are dedicated to a particular instance, other resources such as the network and the disk subsystem are shared among instances. If each instance on a physical host tries to use as much of one of these shared resources as possible, each receives an equal share of that resource. However, when a resource is under-utilized, you are often able to consume a higher share of that resource while it is available. The subscribers with more demanding SLA can be guaranteed by accommodating all the required services within a virtual machine image and then mapping it on a physical server. This helps to solve problem of heterogeneity of resources and platform irrelevance. Load balancing of the entire system can be handled dynamically by using virtualization technology where it becomes possible to remap virtual machines (VMs) and physical resources according to the change in load. Due to these advantages, virtualization technology is being comprehensively implemented in cloud computing. However, in order to achieve the best performance, the virtual machines have to fully utilize its services and resources by adapting to the cloud computing environment dynamically. The load balancing and proper allocation of resources must be guaranteed in order to improve resource utility. Thus, the important objectives of this research are to determine how to improve resource utility, how to schedule the resources and how to achieve effective load balance in a cloud computing environment. Virtualization technology allows sharing of servers and storage devices and increased utilization. Applications can be easily migrated from one physical server to another.

II. RELATED WORKS

In this section, I present a general review of project work on resource allocation and management on the cloud computing environment with respective algorithm. This paper completely deals with virtualization [1]. The key concept of cloud based solution is virtualization. The active resource allocation is performed by the virtual machine with the aid of virtual machine related features like flexible resource provisioning.In this report, we introduce a scheme 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 present the concept of "skewness" to measure the variability in the multidimensional resource utilization of a host. By minimizing skewness, we can mix different types of workloads nicely and better the overall utilization of host resources. In this paper, VM resource scheduling only considers the future resource need and ignores the current allocation & performance of the nodes. Grounded on this factor VM resource allocation is performed. During this allocation some nodes get more than it deserves and some other not get. During VM migration, there is no suitable criteria for unique identification and location of the VM, that means which VM is migrated and where to be migrated. The total migration cost becomes a problem when all the VM resources are migrated.

This report offers a parallel data processing framework exploiting dynamic resource allocation in Infrastructure-as-a-Service (Iaas) clouds [2]. This scheme fits well in a cloud for efficiently parallelizing incoming set of projects using large data. In this, a job initiates one virtual machine (VM) and on the go, based on this and complexity of subtasks, further VMs are allocated and de-allocated. It shows the architecture of this fabric. In parliamentary law to perform scheduling of tasks to VMs in a cost-effective way, they used a load balancing algorithm named Join-Idle-Queue algorithm. It brings in an idle queue between the Job Manager and the Task manager (VM). Whenever a VM becomes idle, it joins the queue. When a job approaches the Job manager, it fetches the first VM in the queue and allocates the task to it. This algorithm divides avoids the Job manager from inquiring every VM for its availability. Hence this method reduces communication overhead and thereby improves throughput of the information processing system. It comes after a centralized architecture and also ensures automatic adaptation to under and over-use of resources. Only this arrangement does not consider the heterogeneity aspects of resources in the swarm.

This report addresses a novel access for dynamic resource management in clouds [3]. Virtualization is a central concept in enabling the "computing-as-a-service" vision of cloud-based resolutions. Virtual machine related features such as flexible resource provisioning, and isolation and migration of machine state have improved efficiency of resource use and dynamic resource provisioning capabilities. Live virtual machine migration transfers "state" of a virtual car from one physical machine to another, and can mitigate overload conditions and enables uninterrupted maintenance activities. The focal point of this article is to give the details of virtual machine migration techniques and their usage toward dynamic resource management in virtualized environments. The migrations in a data center can be triggered periodically. For example, data centers in one part of the world may be heavily used in daytime (9 a.m. to 9 p.m.), whereas they may be under load during the night. Such "time of day" based migration of VMs ensures that VMs are "near" clients, and the communication delays and operating costs are minimized. Migrations can also be executed periodically to consolidate the reduced loads.

Virtual machines change their resource requirements dynamically. This dynamism leads to asymmetries in the resource usage levels of different PMs. Some PMs can get heavily loaded while others may be lightly loaded. In a data center, resource utilization levels of PMs are monitored continuously. If there is large discrepancy in the utilization levels of different PMs, load balancing is triggered. Load balancing involves migration of VMs from highly loaded PMs to low loaded ones. An overloaded PM is undesirable as it causes delays in service of user requests. Likewise, the PMs that are lightly loaded cause inefficient resource usage.

This report addresses the problem of resource management for large-scale cloud environments that hosted sites by contributing a key element called Gossip protocol [4]. Generally large-scale cloud environment includes the physical base and associated control functionality that enables the provisioning and management of cloud services. In addressing the problem of resource management, specific use case of a cloud service provider which hosts sites in a cloud environment is deployed. The cloud infrastructure includes many computational resources and storage resources in the machines within the swarm. The cloud service provider controls the cloud infrastructure which host sites in the cloud in the platform as a service concept. The required services will be offered by the site owners through the websites hosted on the swarm. The services can be made available to the end users through internet using a cloud tenant with host virtual machines in the cloud. The key contribution of this paper is a generic gossip protocol, which executes in the middleware and performs the resource allocation in the large cloud environment and achieves the objective fairness, adaptability and scalability. It owns the structure of a troll-based broadcast algorithm. Two nodes exchange state information, process this information and update their local status during a round. This algorithm is also subjected to constraints that allocated CPU and memory resources cannot larger than available resources. Each machine interacts with others by maintaining an Overlay network. A node which initiates protocol is called active server and the node which communicates with an active node for load distribution is called passive server. Active server compares the load with the passive server and if its load is higher than that of passive load, then the particular site request is handled by passive server otherwise that request is handled by active server itself.

This paper proposes to provide better and efficient VM scheduling for Cloud Data Centers with Minimum energy and VM migration [5]. This report is a contribution to the diminution of such excessive energy consumption using energy aware allocation. Our study offers an exact energy aware allocation algorithm. Modeling of energy aware allocation and consolidation to minimize overall energy consumption leads us to the compounding of an optimal allocation algorithm with a consolidation algorithm relying on migration of VMs at service departures. This paper aimed to achieve energy efficient VM placement consists of two steps. 1. Screening out the requested VMs in decreasing order of power use of goods and services. This builds somehow an ordered stack that is used in the second step for packing VMs in available servers. 2. The sorted VMs are handled starting from the top of the stack and attempting to place the most power consuming VMs on the server with the smallest remaining power consumption budget until a VM down the stack fits in this target server. The process repeats or continues until all VMs in the stack are set and carried as much every bit possible in the most occupied servers. This will tend to free servers for sleep mode or switching off. Merely it is ineffective to oversee energy efficiency under high load.

This report presents a precise VM placement algorithm that ensures energy efficiency and also prevents Service Level Agreement (SLA) violation [6]. The basic principle of VM placement is to allocate as many VMs on a physical server as possible, while satisfying various constraints specified as percentage of the scheme demands. In this newspaper, we design and deploy a distributed decision support system, which is deployed as a cloud-based solution. The road system is a cloud-based solution. Resource usage information from each VM and host is logged to a centralized cloud location. The information aggregation process can be performed by a hypervisor or third party software. Data analytic servers that employ the R framework as an engineer are pre-packaged Linux images that can run on VMs. The number of data analytic servers can be scaled up on demand since its cloud-based. A group of host machines is controlled by a Controller. The Controller has three duties. Firstly, it segments historical data to a specified distance for each VM or host; this segmented data will be utilized for building forecast models. Secondly, the Controller passes the address of data and the specific modeling algorithm to the roads. But in this scheme, there is no load balancing scheme.

In this report, we studied the problem of active resource allocation and power management in virtualized data centers [7]. Prior study in this area uses prediction based approaches for resource provisioning. In this study, we have utilized an alternate approach that makes usage of the queuing information available in the system to make online control decisions. This approach is adaptable to unpredictable changes in workload and does not require estimation and prediction of its statistics. We have been using the recently developed technique of Lyapunov Optimization to design an online admission control, routing, and resource allocation algorithm for a virtualized data center. This algorithm maximizes a joint service program of the average application throughput and energy costs of the information center.

III. SYSTEM ARCHITECTURE

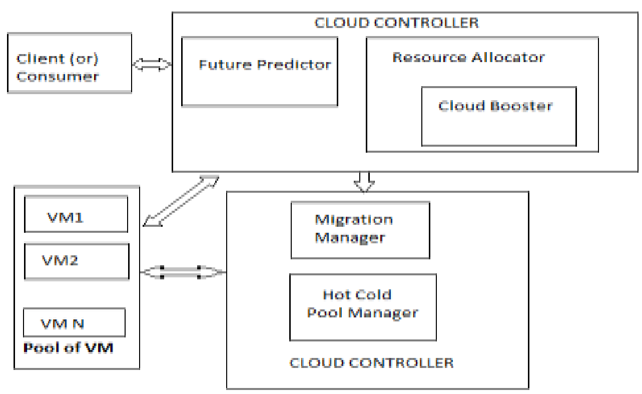

Fig No 3.1: System Architecture

Fig no3.1.shows the architecture of the systemCloud is a model for enabling convenient; on demand network access to shared pool of configurable computing resources. It allows business clients to scale upwards and down their resource use based on demands. The objective of this project is to create and implement Cloud booster algorithms for VM allocation which reduce hot spot nodes in the cloud and to design and implement the managing of the overloading process to Migrate VM from overloaded PM to overcome overloading.

- Cloud Node (Requestor) gets the job request from user to cloud controller.
- The cloud booster algorithm will calculate current allocation & Node capabilities.
- Future predictor will predict the future, loads based on the allocation details from the job allocator.
- Job allocator allocates the jobs based on the allocation sequence from the cloud buster.

- Hot cold pool manager gets the allocation details from the job allocator, and discoveries which all nodes are in hot and cold threshold.
- Based on the node load details of the Hot cold pool manager, the Migration manager overload detector will finds the victim job and remote node, where it can be able to migrate the victim job.
- PMs give the results back to user through the cloud controller.

IV. ALGORITHM

4.1Cloud Booster Algorithm

Resource Weight Calculation

- Collect resource information from various parallel nodes
- Calculate weight for each resource separately.
- Take the sum of all the resource's weights to determine the weight of a node.

Resource information

Here CPU Utilization and Memory usage are negative factors and the Available CPU and Memory are the positive elements. The CPU has greater impact on the performance of a machine comparing in memory or other factors. So it has a maximum weight constant. Weight for the above machine configuration is computed as follows:

Memory Utilization = (Total Memory – Used Memory) / Total Memory

CPU Utilization = Total CPU Usage of Processes / No. of Processes

Available CPU = Clock Speed * (1 – CPU Utilization) **37**

Resource	*Value (Example)*	*Weight constant*	*Weight of Each resource*
CPU Speed(MHZ)	2000	0.2	W_{cpu}= 2000*0.2=400
Cache memory(kb)	2500	0.2	W_{cache}=500
Main memory(kb)	1000000	0.15	W_{ram}=150000
CPU Usage (%)	20%	0.2	$W_{cpu\ usage}$=0.04
Memory usage (kb)	750000	0.15	W_{mem}=112500

Table No 4.1: Resource Table

Available Memory = Main memory * (1 – Memory Utilization)

For converting each parameter into same scale between 0 and 1, divide each value with the maximum value. Then weight for each resource is separately determined.

CPU Weight (WCPU) = Available CPU * Weight constant for CPU

Memory Weight (WMem) = Available Memory * Weight constant for Memory.

Weight for current load (WLoad) = Current load * Weight constant for load

Weight of Machine = WCPU + WMem – WLoad

In Load distribution phase selects the best parallel machine for job allocation. Here job Size is considered as load. Before job allocation, see the job percentage required by each machine. This is calculated by using the formula.

Required job percentage for a machine say X, PX = Wx / Σ WI*, here Σ WI *is the total weight of all the nodes.

After estimating Job volume for each machine PXsed on theWI node weiWItagethe total weightecution of task is started in parallel processing system and output is generated.

Weight of Each node = $W_{cpu} + W_{cache} + W_{ram} - W_{cpu\ usage} - W_{mem}$

b) Resource Allocation

- Represent all parallel nodes as a Connected Graph.
- Calculate mean deviation for each node in the path by taking the difference between current Job Weight and the weight of the node.
- Repeat this process for all nodes in the set.

4.2 Load Balancing Algorithm

This Algorithm mainly focused on dynamic load balancing or management. This dynamic load balancing algorithm based upon virtual machine migration in cloud computing environment. It proposes triggering strategy that was based on fractal method. It provide better performance, better resource usage and more balanced. The triggering strategy was based on the specific threshold that resulted into instantaneous peak load triggered then the virtual machine was migrated in the cloud computing environment. Here the migration decision was made on the basis of the history of the load indicators load information to get n load value. When k load values can be exceed the specified certain value, migration was triggered. Thus this strategy avoids the problem of instantaneous peak load triggering once the virtual machine is migrated.

- CPU load indicator is measured as

$$CPU = \sum_{K=1}^{n} Ak/n$$

Here n is the number of nodes, CPU utilization represented as Ak.

- Memory load indicator represented as

$$Memory=\sum_{K=1}^{n}(Memuse+MemSwap)/Tmem$$

Here memory load consists of memory usage and the swapped memory. Total memory is Tmem.

- Bandwidth load indicator represented as

$$Band=\sum_{K=1}^{n}BandK/TBand$$

V. CONCLUSION

Cloud is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources. It allows business clients to scale upwards and down their resource use based on demands. Different organization provides the same service with different service charges and waiting time. Then clients can select services from these cloud providers according to their criteria like price and waiting time.

We have submitted the conception, execution, and evaluation of an efficient resource management scheme for cloud computing services. I use Cloud Booster Algorithm for determining the node's capabilities, Resource allocation and Load Balancing Algorithm for VM migration. My algorithm achieves both overload avoidance and energy saving for systems with multi-resource constraints. In a future enhancement will propose a new algorithm for resource scheduling and comparative with existing algorithms for reducing the number of migrations.

REFERENCES

[1]. Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment" IEEE - 2013.

[2]. A Ajitha, D Ramesh, member ELSEVIER, "Improved Task Graph-Based Parallel Data Processing for Dynamic Resource Allocation in Cloud" in 2012.

[3]. AnirudhaSahoo, Mayank Mishra, AnweshaDasandPurushottamKulkarni "Dynamic Resource Management Using Virtual Machine Migrations", IEEE- 2012.

[4]. A.Padmashree, M.Sasitharagai1, T.Dhanalakshmi&S.Gowri, ISSN proposed "Dynamic Resource Management in Large Cloud Environments using Distributed Gossip Protocol" in 2012.

[5]. ChaimaGhribi, Makhlouf and DjamalZeghlache, "Energy Efficient VM Scheduling for Cloud Data Centers: Exact allocation and migration algorithms" IEEE-2013.

[6]. Dapeng Dong and John Herbert, "Energy Efficient VM Placement Supported by Data Analytic Service", IEEE/ACM International Symposium-2013

[7]. Rahul Urgaonkar, Ulas C. Kozat, Ken Igarashi, Michael J. Neely "Dynamic Resource Allocation and Power Management in Virtualized Data Centers"

[8.]HaozhengRen, YihuaLan, Chao Yin "The Load Balancing Algorithm in Cloud Computing Environment"Proceedings of 2012 IEEE and 2012 2nd International Conference on Computer Science and Network Technology.

[9]. S.R.Suraj, R.Natchadalingam "Adaptive Genetic Algorithm for Efficient Resource Management in Cloud Computing" International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 2, February 2014