ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

A Survey on Dynamic Resource Management Technologies in Cloud Datacenter

Ashwini.K.L[1], Dr Suresh.S[2]

P.G.Scholar, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[1]

Associate Professor, Department of CSE, Adhiyamaan College of Engineering, Hosur, India[2].

Abstract: Cloud computing offers infinite amount of resources which is available anywhere anytime based on the demand. Resource management is the attractive technology and it is based on pay-as-you-go model and the reservation model. Resource management has become one of the major issues in the cloud computing environment due to scale of modern data centers, heterogeneity of resource types and interdependencies. There are many resource management techniques used that must satisfy SLA. The work is about to survey various resource management strategies in cloud computing.

Keywords: Cloud computing, Resource management, Resource monitoring, SLA.

I. INTRODUCTION

Cloud computing emerged a paradigm for managing and delivering the services over the internet. Cloud computing provides resources to the customers which is available anywhere, anytime, on demand and reservation. Cloud computing has some challenges like service delivery, cost, portability, reliability. Service delivery is one of the important challenges because it is very reluctant to switch to cloud without a strong service Quality guarantee.

Todays IT infrastructure has a great intensity and also we find difficult to keep up. It is a natural evolution of the widespread adoption of multiple technical advances in the distributed computing area including virtualization, grid computing, autonomic computing, utility computing and software as a service [1].

Cloud computing is much more elastic and scalable for acquiring the resources. Resource management is used for monitoring the availability of resources, allocating the resources and provisioning of the resources. The resource monitoring process is used for controlling and managing the hardware and the software services. The allocating process is used for assigning the available source over the internet. The allocation is based on pay-per usage and reservation method. The provisioning process is the combination of allocation and monitoring of resources. This paper highlights the need of resources in datacenters and surveys various strategies for allocating the resources.

II. NEED FOR RESOURCE MANAGEMENT

Resource management has become major challenge in the cloud computing. For minimizing the operating cost Amazon uses work consolidation policy where as private enterprise uses load balancing policy for ensuring Quality of service. Resource management must be more efficient for handling large scale problems. In Anchors clients and operators are able to express a variety of distinct resource management policies in the stable management framework [2].

In the large scale distributed computing resource management is one of the issue in the cloud computing. Mapping the set of applications to a system of machines, for each such machine we need to assign local resources for those applications which is capable of executing on it. The resource demand of an application can change over time. In response to such changes, the resource allocation process needs to be repeated many times over; in other words, it has to be dynamic, in order to ensure that the system utility is maximized at all times [3].

Resource Management is one of the major concern in the cloud computing. Resource management includes resource monitoring and resource discovery these processes manage physical resources such as CPU cores, disk space, and network bandwidth. This resources must be sliced and shared between virtual machines running potentially heterogeneous
Workloads.

III. RESOURCE MANAGEMENT TECHNIQUES

A.STATIC RESOURCE MANAGEMENT TECHNIQUES:

Fetahi Wuhib and Rolf Stadler has described the protocol qualitatively behaves as expected based on its design. For instance, regarding fairness, the protocol performs close to an ideal system for scenarios where the ratio of the total memory capacity to the total memory demand is large. More importantly, the simulations suggest that the protocol is scalable in the sense that all investigated metrics do not change when the system size increases proportional to the external load [4].

Vignesh.V has proposed the Process which gives the available resources and user preferences. The computing resources can be allocated according to the rank of job. The analysis of resource scheduling algorithms. The time parameters of three algorithms, viz. Round Robin, Pre-emptive Priority and Shortest Remaining Time First have

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

been taken into consideration. From this, it has been computed that SRTF has the lowest time parameters in all respects and is the most efficient algorithm for resource scheduling [5]

Kangas.M has analyzed resources dynamically within the cloud of mobile users subject to the longterm average power constraint in order to stabilize the queues and maximize the long-term average throughput of the system. By modeling the problem as a finite horizon Markov Decision Problem. Dynamic control policy that making opportunistic cooperative control decisions adaptively allocates resources over time varying fading channels and maximizes the long-term average throughput of the system. In addition, the concept of inter system networking is introduced to provide performance bound for the resource allocation policies of the cooperative communication network [6].

B. DYNAMIC RESOURCE MANAGEMNT TECHNIQUES:

Jeongseob Ahn evaluated memory-aware cloud scheduling techniques, which do not require any
prior knowledge on the behaviors of Virtual machine. The virtual machine live migration can also be used to mitigate micro-architectural resource contentions, and the cloud-level VM scheduler must consider such hidden contentions. We plan to extend our preliminary design of Non uniform memory access aware scheduling for more efficient Non uniform memory access affinity supports with hot page migrations. Also, we will investigate a systematic approach based on a cost benefit analysis for VM migrations and contention reductions[7].

Chenn jung hung has proposed prediction mechanism by using support vector regressions to estimate the number of resource utilization according to the SLA of each process, and the resources are redistributed based on the current status of all virtual machines installed in physical machines. Notably, a resource dispatch mechanism using genetic algorithms is proposed in this study to determine the reallocation of resources. The experimental results show that the proposed scheme achieves an effective configuration via reaching an agreement between the utilization of resources within physical machines monitored by a physical machine monitor and service level agreements between virtual machines operators and a cloud services provider. In addition, our proposed mechanism can fully utilize hardware resources and maintain desirable performance in the cloud environment [8].

Narendar kumar has proposed priority based dynamic resource management framework in cloud environment and a multi-index transportation problem cloud resource scheduler mechanism with mathematical formulation and an algorithm named MultiIndexed Cloud Resource Scheduling Algorithm with state transitions, deterministic finite automation and we present different test cases to evaluate and validate the results for calculating the total cost of processing requests as well as business process diagram and sequence diagram using unified modeling language . A resiliency analysis is also presented to allocate the resources according to their priority and processing cost is also optimized. Finally it will be able to calculate the processing cost, time, profit of service providers as well as customers benefits and optimum use of cloud resources of various clusters [9].

Alexander Ngenzi has analysed the consequences of improper resource management may result into underutilized and wastage of resources which may also result into poor service delivery in these datacenters. Resources like; CPU, memory, Hard disk and servers need to be well identified and managed. In this Paper, Dynamic Resource Management Algorithm shall limit itself in the management of CPU and memory as the resources in cloud datacenters. The target is to save those resources which may be underutilized at a particular period of time. It can be achieved through Implementation of suitable algorithms. Here, Bin packing algorithm can be used whereby the best fit algorithm is deployed to obtain results and compared to select suitable algorithm for efficient use of resources[10].

IV. RESOURCE MANAGEMENT STRATEGIES

In the recent years researchers has proposed many strategies for managing resource management. For efficiently make use of the cloud resources resource management techniques to be used. There are many resource management techniques[11].

A.LINEAR SHEDULING STRATEGY:

The resource allocation is taken into consideration commonly the parameters like CPU utilization, memory utilization and throughput etc. The cloud environment has to take into reflection all these things for each of its clients and could provide maximum service to all of its clients. When we are taking the scheduling of resources and tasks separately it imposes large waiting time and response time. In order to overcome this shortcoming this paper introduces a new approach namely Linear Scheduling for Tasks and Resources (LSTR). Here scheduling algorithms primarily focus on the distribution of the resources along with the requestors which will make best use of the selected QoS parameters. The QoS parameter selected in this approach is the cost function. The scheduling algorithm is designed based on the tasks and the available virtual machines together and specified as LSTR scheduling strategy. This is designed so as to maximize the resource utilization.

B.MATCH MAKING AND SHEDULING:

Match making is the first step and scheduling is second in the resource allocation in cloud environment. Matchmaking is the procedure for allocating jobs associated with user requests to resources selected from the available resource pool. Scheduling refers to determining the order in which jobs mapped to a specific resource are to beexecuted . It also points out that there are some uncertainties that are associated with such type of match making and scheduling. They can be like Error Associated with Estimation of Job Execution Times. It is

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

considered that estimating the execution time for a line of work is a very hard labour and faults may happen very often. There is one deviant condition known as the formation of resource idle time . It is bechanced because of certain unwanted statuses like jobs may run for a smaller time in comparison to their estimated execution time. There is one more reason is that abnormal ending of the jobs.

C.PRE COPY APPROACH:

In "Pre-Copy Approach" pages of memory are iteratively copied from the source machine to the destination host and in addition there is a fact that all these things are done without ever stopping the execution of the system. Page level protection hardware is used to make sure that a consistent snapshot is transferred. For controlling the traffic of other running services a rate-adaptive algorithm is used. And during the final phase it pauses the virtual machine and copies any remaining pages to the destination and after that resumes the execution.

D.ERROR ASSOCIATED WITH ESTIMATION OF JOB EXECUTION TIMES

It is considered that estimating the execution time for a job is a terribly laborious task and errors might happen fairly often. There is one abnormal condition referred to as the formation of "resource idle time". It is happened because of certain unwanted conditions like jobs may run for a smaller time compared to their estimated execution time. There is one more reason known as abnormal termination of the jobs. These give rise to a serious degradation in system performance because jobs that could have used the resource during these idle time periods might have been turned away by the matchmaker that expected the resource to be busy executing the job with an over estimated execution time. Other problem raises the

V. RESOURCE MANAGEMENT POLICIES

- **Admission control:** It helps in preventing the system from accepting the resources in the violation of the high level policies.
- **Load balancing:** It helps to Uniformly distribute the resources among the server.
- **Allocation of the capacity:** This process is used for allocating the resources individually.
- **Guaranteeing Quality of service:** As specified by the Service Level Agreements should satisfy both the timings and other conditions. There are some of the mechanism for implementing the resource management policies they are machine learning, utility based computing, etc[12].

VI. SERVICE LEVEL AGREEMENTS (SLA) BASED RESOURCE MANAGEMENT

An SLA represents an agreement between a service user and a provider in the context of a particular service provision. An SLA may exist between two parties, for instance, a single user and a single provider, or between multiple parties, for example, a single user and multiple providers. SLAs contain certain quality-of-service (QoS) properties that must be maintained by a provider during service provision—generally defined as a set of service-level objectives (SLOs). These properties need to be measurable and must be monitored during the provision of the service that has been agreed in the SLA.

The particular QoS attributes that are used must be preagreed to between the user and provider(s), before service provision begins, and also they define the obligations of the user/client when the provider meets the quality specified in the SLA. The SLA must also contain a set of penalty clauses when service providers fail to deliver the preagreed to quality. Although significant work exists on how SLOs may be specified and monitored, not much work has focused on actually identifying how SLOs may be impacted by the choice of specific penalty clauses. The participation of a trusted mediator may be necessary in order to resolve conflicts between involved parties. Automating this conflict resolution process clearly provides substantial benefits. Different outcomes from such a process are possible. These include monetary penalties, impact on potential future agreements between the parties and the enforced rerunning of the agreed service. Market mechanisms provide an important basis for attributing the cost of meeting/violating an SLA. SLA goes through various stages within its lifecycle.

Assuming that an SLA is initiated by a client application, these stages include the following[13].

- **Identifying the Provider**: This could either be hardwired or obtained through the use of a discovery service. Provider selection isan activity often outside the scope of the SLA lifecycle, but nevertheless an important stage to be executed.
- **Agreeing on the Terms of the SLA**: This stage involves identifying the constraints that must be met by a provider during service provisioning.
- **Monitoring SLA Violation**: Who does the monitoring and how often is an aspect that needs to be considered at this stage.
- **Destroying SLAs**: Once a service provision has completed, the SLA must be destroyed.
- **Penalties for SLA Violation**: Once a service provision has completed, the Monitoring data may be used to determine whether any penalties need to be imposed on the service provider.

VII. CONCLUSION

In cloud computing resource management is an attractive technology as it is based on pay-as-you-go model and the reservation model for ensuring the guaranteed performance for many applications. These techniques are used to improve SLA. The ultimate goal of resource management is used to maximize the cost from the cloud provider's perspective and the cloud users perspective for reducing the cost.

There are many challenges resource management strategies. Mechanisms have to be proposed to efficiently make of cloud resources so that QoS is met and SLA Violation is minimized.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 9, September 2015

REFERENCES

[1] Juhnyoung Lee, A View Of Cloud Computing, *International Journal of Networked and Distributed Computing, Vol. 1, No. 1 (January 2013), 2-8*
[2] Hong Xu, , Baochun Li, Anchor: A Versatile and Efficient Framework for Resource Management in the Cloud, *IEEE Transactions On Parallel And Distributed Systems, 201x*
[3] Fetahi Zeebenigus Wuhib, Distributed Monitoring and Resource Management for Large Cloud Environments
[4] Fetahi Wuhib and Rolf Stadler Mike Spreitzer Gossip-based Resource Management for Cloud Environments, I*nternational conference on network and service management 2010*.
[5] Vignesh V, Sendhil Kumar KS, Jaisankar N,Resource management and scheduling in cloud environment, *International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 1 ISSN 2250-3153*
[6] kangas.M, Glisic.S, Throughput optimal resource management of cooperative networks with mobile clouds, *IEEE 2011*
[7] Jeongseob Ahn, Changdae Kim, Jaeung Han, Young-ri Choi†, and Jaehyuk Huh, Dynamic Virtual Machine Scheduling in Clouds for Architectural Shared, *IEEE 2013*
[8] chenn jung Hung, chih Tai Guan, Heng min chen, An adaptive resource management scheme in cloud computing,2015 Elsevier
[9] Narendar kumar, Shalini Agarwal nd Vipin Saxena, LP based Adaptive Resource Management Framework in Cloud Environment, *IJCEM International Journal of Computational Engineering & Management, Vol. 17 Issue 1, January 2014 ISSN.*
[10] Nadjia Kara, Mbarka Soualhia, Fatna Belqasmi, Christian Azar and Roch Glitho, Genetic-based algorithms for resource management in virtualized IVR applications, *Kara et al. Journal of Cloud Computing: Advances, Systems and Applications 2014.*
[11] Ajay Gulati, Anne Holler, Cloud-Scale Resource Management: Challenges and Techniques,*2010*
[12] cloud resource management and scheduling[online].
[13] Jordi Guitart, Mario Maci_As, Omer Rana, Philipp Wieder, Ramin Yahyapour, And Wolfgang Ziegler, SLA-Based Resource Management And Allocation, *IEEE 2009*