www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 7 July 2017, Page No. 22116-22120
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i7.35

A Study on VMM and Resource Allocation Strategies in Cloud Computing Environment

Navdeep kaur[1], Pooja Nagpal[2]

[1]Research Scholar
Rayat Institute Of Engineering and Information Technology, Punjab, India.
[1]navisaini90@gmail.com

[2] Faculty of Computer Science and Information Technology, Rayat Institute Of Engineering and Information Technology, Punjab, India.
[2]rieit.cse.pooja@gmail.com

ABSTRACT: Cloud computing is an internet dependent approach for providing shared resources on demand with the management of storage, networks, servers, services and applications that needs management optimum effort. Virtual machine migration is playing a significant role for improving resources utilization; processing nodes load balancing, application isolation, fault tolerance in virtual machines, to enhance the nodes portability and to maximize the physical server efficiency. To balance the cloud with its resources for better performance with the services to the endusers of the cloud and at the identical time, numbers of users are served by application deployments in the environment of cloud is the main task. The users of cloud may request or rent the resources when they become essential. This paper has provided an overview of the essential components of computing with the discussion of virtual machine migration, resource allocation in cloud computing with its challenges and risks.

Keywords: ***Cloud computing, Virtual Machine Migration, Resource allocation, Resource Utilization***

1. INTRODUCTION

The most emerging research area for the past few years is cloud computing that is offering shared computational power of the resources on demand (pay-as-per use model). It is an internet technology that uses the central remote servers and internet to manage the data and its applications [1]. It has become a highly demanded service due to advantages of cheap cost of services, scalability, availability, high computing power, accessibility. Cloud Service Providers like Google, Amazon offer their services without installation and access their files at any PCs, mobile devices with internet access according to Infrastructure as a Service, Platform as a service and Software as a service. The virtualization is key feature of cloud computing that involves the creation of multiple virtual machines on single physical computer [2]. The multiple OS can execute on the single OS underlying the same hardware platform. In virtualized data centers, multiple clients can share the same hardware resources (workload) and the resources supplied to clients can be scaled dynamically.

2. VIRTUAL MACHINE MIGRATION

For eliminating the supervision of human beings, the virtual machine migration technology is assisted by the various cloud service providers to manage resources [3]. It can achieve multiple goals such as load balancing, green computing, energy efficiency, fault tolerance and real time server maintenance. In this technology the overloaded or under loaded VM is transferred from one server (machine) to another as shown in Fig 1. Unlike non-live migration, live migration does not suspend application service prior to VMM process.

Algorithms used for live migration are as follows:-

Pre-copy: In the pre-copy algorithm, the entire contents of memory resources are copied from source to destination by interchanging the execution states at targeted host. If the migration is failed the VM is still responsive at source and the migration process can be reverted at source. Most popular hypervisors like VMWare, Xenand KVM, etc have adopted this approach [4].

Post-copy: In the postcopy approach, if migration fails the VM is stopped at the source host and switches the execution states at destination host to resume the VM. The VM at destination host start responding immediately.

Hybrid algorithm: Firstly at source, required memory pages migrated in precopy phase, then the execution states will be interchanged and VM resumed at destination host. After that, the remaining memory pages will be processed by postcopy algorithm via network link. Fever number of pages need to be accessed from the source thus this leads to increase the performance. Total migration time is also better than precopy and postcopy approach.

However, the cloud data centers contains the physical resources with higher storage capacity and high network speed but for the better utilization of all the hardware and software resources, data centers need best scheduling and load balancing algorithms. Load balancing is the process of reassigning the total load(CPU, Memory, Network) to the individual nodes (physical machines, servers, network interfaces, hard drivers or other computing resources) of system with eliminating the condition where some of the nodes are highly loaded and some are lowly loaded [5]. These algorithms are beneficial for proper utilization of all the resources and to improve the response time of the job. The goal of load balancing is to fulfill the request at low cost with reliable resources, scalability, stability, to improve the system condition and performance under high load or request rate. Various types of load balancing algorithms are Sender initiated, Receiver initiated, symmetric depend upon the current condition of the system i.e. Static or Dynamic. Load balancing techniques can be compared on basis of some metrics like throughput, SLA violations, overhead, response time, resource utilization, scalability, performance, fault tolerance [6].

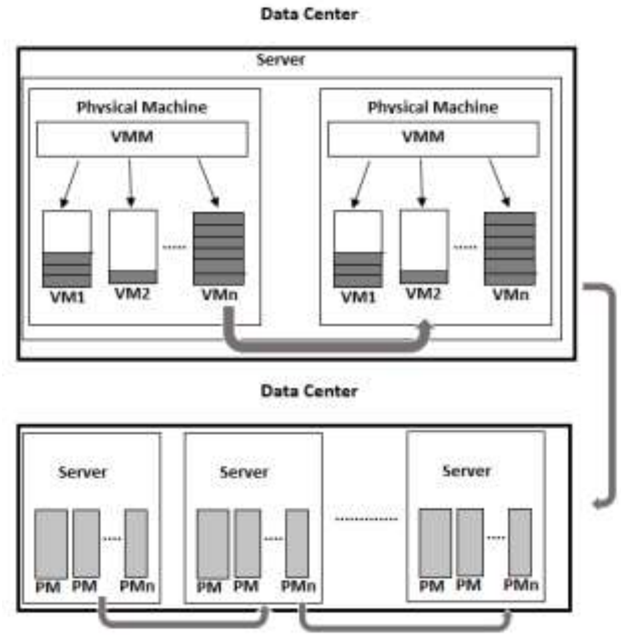

Fig. 1: Load Balancing using Virtual Machine Migration

3. RESOURCE UTILIZATION IN CLOUD

The issues of resources are relevant to the cloud. Especially the elements of resources provisioning, flexibility, and multi-tenure relate to the subject of resources usage. With cloud computing, the client can procure resources consequently; it is regularly not clear ahead of time what resources are required for a specific workload [7]. On account of more than once executed workload, the client may increase a few instinctive comprehension of the conduct of the workload on different sorts and measures of accessible cloud resources, yet for new workloads or variable workloads this is not all that effectively replied. Cloud suppliers normally offer a wide assortment of incidences, differing in the velocity and number of CPUs accessible to the virtual machine, the kind of nearby capacity framework utilized (e.g. single hard circle, plate cluster, SSD storage), whether the virtual machine might impart physical resources to other virtual machines (potentially having a place with different clients), the measure of RAM, system transfer speed, and so on [8]. Likewise, the client must choose what number of examples of every sort to procurement. In the perfect case, more hubs means speedier execution, however issues of heterogeneity, execution unusualness, arrange overhead, and information skew imply that the genuine advantage of using a bigger number of occasions can be not exactly expected, prompting a higher expense per work unit. These issues likewise imply that not all the provisioned resources might be ideally utilized for the span of the application. Provisioning bigger or higher execution occasions is comparatively not generally ready to yield a relative advantage [9].

The main aspects of resource allocation in accordance with security in the cloud are that specific data which need to be safeguarded while in idle state, in transit state, as well as in use state, and access to the data must be controlled, that is to say [10]:

- In order to assure that data does not get corrupted or hijacked, it is very important to have safety techniques in specific place which would protect transfer of the data to and from the databases that be located in the cloud.
- To confirm high confidentiality, it is essential that the out-sourced data in the stored in cloud databases be encrypted at all times

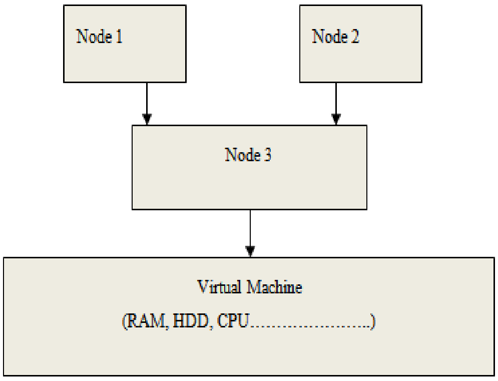

Fig. 2: Load Balancing using Virtual Machine Migration

3.1 Need of Resource Allocation

In cloud computing paradigm, the main challenge is the allocation of several accessible resources between various end-users which are having varying requests of resources dependent upon their patterns of application usage [11]. The random as well as varying requests need to run on data-center resources through Internet. The goal of resource allocation for any specific cloud provider could be either to enhance applications' Quality of Service or increase utilization of resource along with energy proficiency [12]. The key objective is to augment Quality of Service parameters (i.e. response time) which measures the competence of resource-allocation irrespective of the category of ICT resources assigned to any specific end-users.

Following are the key points to be considered while allocating the resources [13]:

1) Since users do not hold ownership over the resources but only rent resources from remote servers for their

purpose, they do not have control over their resources. Hence users or clients are popularly called tenants and not owners.

2) Migration problem occurs, when the users wants to switch to some other cloud provider for the better storage of their data. It is not easy to transfer enormous amount of data from one provider to the other.
3) In public cloud domain, the clients' data are prone to hacking or phishing attacks. Since the servers on cloud are interlinked, it is easy for malware to spread. Hence, security issues in cloud are the major limitations to resource allocation strategy.
4) Peripheral devices like printers or scanners might not work well with cloud. Many of them require software to be installed locally and require constant internet connection to use and access devices and resources, even in transit.
5) More and deeper enlightenment is required for allocating and properly managing resources in cloud, since all knowledge about the functioning of the cloud mostly relies on the cloud-service-provider (CSP).

3.2 Security in Cloud Computing

Cloud Computing has gained popularity in recent years. Cloud facilitates the storage of various sorts of data [14]. Cloud is highly scalable when it comes to huge data and can provide infinite computing resources on demand. Clients can use cloud services without any installation and the data uploaded on cloud is accessible from any corner of the world, all it needs to be accessed is a computer with active internet connection on it. The users can subscribe high quality services of data and software which resides solely on the remote servers and enjoy the provision of on-demand provision of services. As a customizable computing resources and a huge amount of storage space are provided by internet based online services, the shift to online storage has contributed greatly in eliminating the overhead of local machines in storage and maintenance of data. The cloud provides a number of benefits such as flexibility, disaster recovery, and pay-per-use and easy to access and use model which contribute to the reason of moving into cloud. A large number of clients store their important data in the cloud without keeping a single copy of this data in their local computers. Thus, cloud helps free up the space on the local disk, hence also called as 'A Hard-disc in the sky'. Even though immense advantages are offered by cloud, a lot of security concerns still exist in it. The most worrisome concern is its storage security. Most of the times, the user does not maintain any copy of outsourced data in their local system. The question regarding data security becomes crucial when it comes to confidential data. The integrity of the data has to be looked upon seriously in order to gain user trust and satisfaction. However, maintaining security is a challenging task [15].

Load balancing becomes more complex in heterogeneous data centers, so VM migration is used to overcome the problems of traditional load balancing strategies. In VM migration the overloaded VMs are migrated from one PM to another for proper utilization of resources. In this paper, we presented various VM migration strategies and their performance has been evaluated by using various parameters like SLA Violations, number of migrations, throughput, downtime etc. The results prove that these approaches worked well as compare to previous approaches. In future work, these algorithms are combined with some heuristic algorithms like genetic algorithm or use improved algorithms of these approaches like max-min ACO, ABC etc [16].

4. Challenges and Risks in Cloud Computing

In this section, the challenges and risks in cloud computing are described:

i. Bandwidth, nature of administration and information limits

Cloud computing requires "broadband of extensive pace" Whilst numerous sites are usable on non-broadband associations or moderate broadband associations; cloud-based applications are frequently not usable.

ii. Cost

Cloud computing can have high expenses because of its necessities for both a "dependably on" association, and additionally utilizing a lot of information back in-house

iii. Pricing

Investigation can be made between altered costs and variable costs

iv. Security, Privacy and Trust

Security and protection influence the whole cloud computing stack, subsequent to there is a monstrous utilization of outsider administrations and bases that are utilized to have essential information or to perform basic operations. In this situation, the trust towards suppliers is essential to guarantee the sought level of security for applications facilitated in the Cloud. Legitimate and administrative issues likewise require consideration.

v. Data Recovery

All business applications have Service level understandings that are stringently taken after. Operational groups assume a key part in administration of administration level understandings and runtime administration of utilizations

vi. Management Capabilities

Regardless of there being different cloud suppliers, the administration of Service and framework is still in its earliest Services. Highlights like "Auto-scaling‟ for instance, is a critical necessity for some endeavors.

vii. Performance

Cloud computing usually suffers from extreme execution issues. Load balancer, information replicators, top of the line servers must introduced when required

viii. Regulatory necessities

What administrative, legal, administrative and arrangement situations are cloud-based data subject to? This inquiry is difficult to discover due to the decentralized and worldwide structure of the web, and additionally of cloud computing. The data put away by cloud administrations is liable to the lawful, administrative and approach situations of the nation of habitation of the cloud administration, and the nation in which the server base is based.

ix. Data lock-in and institutionalization

Clients might need to move information and applications out from a supplier that does not meet their prerequisites. Nonetheless, in their present structure, cloud computing Infrastructures and Services don't utilize standard techniques for putting away client information and applications.

Subsequently, user doesn't interoperate and the client information is not convenient.

x. Availability, adaptation to internal failure and catastrophe recuperation

It is normal that clients will have certain assumptions about the administration level to be given once their applications are moved to the Cloud. These desires incorporate accessibility of the administration, its general execution, and in addition what measures are to be taken when something turns out badly in the framework or its parts.

xi. Resource administration and energy proficiency

One imperative test confronted by suppliers of cloud computing administrations is the productive administration of virtualized resources pools. The multi-dimensional nature of virtual machines confuses the action of finding a decent mapping of VMs onto accessible physical hosts while amplifying client utility.

xii. Dynamic versatility

The cloud hubs are scaled all over progressively by the application as indicated by the reaction time of the client's questions. The reservation delays included are genuine concern which prompts the need of powerful and element load administration framework.

5. Related Work

W. E. Walshet.al [7] utilized the feedback algorithm for the management of virtualized machines. Here, VM machines are grouped together into shared pool them as per SLO agreement and the VM allocation takes place. Resource allocation is one of the fundamental technologies of cloud-computing domain, which utilizes the computing resources like bandwidth, energy, and delay and so on in the network to facilitate the execution of cumbersome tasks that require large-scale computation. Shi J.Yet.al [8] proposed 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. Resource allocation is one of the challenges of cloud-computing since end-users could easily access resources from anyplace and at any time. The resources present in a cloud could not be demanded straightly but it could be opened using SOAP/Restful web APIs. B.Weiet.al [9] implemented a framework for live migration and dynamic re-allocation of VMs according to current utilization. While ensuring reliable QoS and minimize power consumption and delay using two algorithms, Modified Best fit decreasing algorithm and genetic algorithm. The idea of Virtual Machines (VMs) is connected to diminish the energy utilization as it essentially decreases the rate of idle power in the general base. S.Esfandiarpoor et.al [11] considered four energy-aware resource management algorithms for virtualized data centers so that, total energy consumption of data centre is minimized. The author has proposed a new algorithm (OBFD) that sorts a list of VMs in decreasing order of their required MIPS instead of current CPU utilizations. Soodeh Farokhi [18] developed a framework for resource allocation in a multi-cloud system from the perspective of the SaaS level, agreed SLA, and service provider conditions. The proposed model utilizes a selection engine, construction engine, and SLA violation detection and monitoring with the use of the service provider's QoS parameters. There are few models that focuses on both cloud provider and consumer perspectives. Parekh et al. [19] addressed the problem of building an effective external controller for automated adaptive scaling of applications deployed in the cloud. They recommended the Proportional Thresholding approach which dynamically adjusts the target range i.e. high and low thresholds based on the number of accumulated virtual machine instances. Thus the relative effect of allocating resources becomes finer as the number of accrued resources increases; eventually resulting in being adaptive and more resource efficient. Jiayin et.al [20] proposed an adaptive resource allocation algorithm for the cloud system with pre-empt able tasks. There is two major contributions of this work, done by the author. First, the author has presented a resource allocation mechanism in cloud systems which enables pre-empt able task scheduling, which is suitable for the autonomic feature within clouds and the diversity feature of VMs. Second, they have proposed two adaptive algorithms for resource allocation and task scheduling in IaaS cloud computing. These algorithms have adjusted the resource allocation adaptively based on the updation of the actual task executions.

6. Conclusion

Cloud computing offers an efficient way for delivering the services in the Internet with varied resources being pooled and configured. This review has studied the concept of Virtual machine migration in cloud computing with the resource utilization and its need for achieving user satisfaction and for enhancing the profit for cloud service providers.

Therefore, this analysis will hopefully encourage future researchers to come up with secured and smarter algorithms for optimal resource allocation with the structure for strengthening the cloud computing paradigm.

References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M., "A view of cloud computing , Communications of the ACM, Vol. 53, Issue 4, pp.50-58,2010.
2. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J. and Brandic, I., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation computer systems, Vol. 25, Issue 6, pp.599-616.
3. Xiao, Zhen, Weijia Song, and Qi Chen, "Dynamic resource allocation using virtual machines for cloud computing environment", IEEE transactions on parallel and distributed systems Vol. 24, Issue 6,pp. 1107-1117,2016.
4. Fang, Yiqiu, Fei Wang, and Junwei Ge, "A task scheduling algorithm based on load balancing in cloud computing", In International Conference on Web Information Systems and Mining, pp. 271-277. Springer Berlin Heidelberg, 2010.
5. Beloglazov, A., Abawajy, J. and Buyya, R., "Energy-aware resource allocation heuristics for

efficient management of data centers for cloud computing", Future generation computer systems, 28(5), pp.755-768, 2012.

6. Mishra, Mayank, Anwesha Das, Purushottam Kulkarni, and Anirudha Sahoo, "Dynamic resource management using virtual machine migrations", IEEE Communications Magazine Vol. 50, Issue 9, 2012.
7. W. E. Walsh, G. Tesauro, J. O. Kephart, and R. Das, "Utility Functions in Autonomic Systems," in ICAC '04: Proceedings of the First International Conference on Autonomic Computing, IEEE Computer Society, pp. 70–77, 2004.
8. Shi J.Y., Taifi M., Khreishah A., "Resource Planning for Parallel Processing in the Cloud," in IEEE 13th International Conference on High Performance and Computing, pp. 828-833, 2011.
9. B.Wei, C.Lin, X.Kong, "Energy Optimized Modeling for Live Migration in Virtual Data Center", International Conference on Computer Science and Network Techno C.Lin,
10. P.Liu, J. Wu, "Energy-Aware Virtual Machine Dynamic Provision and Scheduling for Cloud Computing", IEEE 4th International Conference on Cloud Computing, pp.736-737, 2011.
11. S.Esfandiarpoor, A.Pahlavan, M.Goudarzi, "Virtual Machine Consolidation for Data center Energy Improvement", Computer Engineering Department, Sharif University of Technology, Tehran, Iran, 2013.
12. Sudeepa, R., and H. S. Guruprasad, "Resource allocation in cloud computing", International Journal of Modern Communication Technologies & Research, Vol. 2, Issue 4, 2014.
13. Mangla, Neeraj, and Jaspreet Kaur, "Resource Allocation in Cloud Computing".
14. Xiao Z, Song W, Chen Q, "Dynamic resource allocation using virtual machines for cloud computing environment", IEEE transactions on parallel and distributed systems, Vol.24, Issue 6,pp.1107-17,2013.
15. Gritzalis, Stefanos, Chris Mitchell, Bhavani Thuraisingham, and Jianying Zhou, "Security in cloud computing", 2013.
16. Feng, Deng-Guo, Min Zhang, Yan Zhang, and Zhen Xu, "Study on cloud computing security", Journal of software, Vol. 22,Issue 1, pp.71-83,2011.
17. Takabi, Hassan, James BD Joshi, and Gail-Joon Ahn, "Security and privacy challenges in cloud computing environments", IEEE Security & Privacy, Vol. 8, Issue 6,pp. 24-31,2010.
18. Farokhi, "Towards An Sla-Based Service Allocation In Multi-Cloud Environments", Cluster, Cloud And Grid Computing (Ccgrid), 14th IEEE/Acm International Symposium On, Chicago, pp. 591-594, 2014.
19. Harold C. Lim, Shivnath Babu, Jeffrey S. Chase, Sujay S. Parekh, "Automated Control in Cloud Computing: Challenges and Opportunities".
20. Jiayin Li, Meikang Qiu, Yu Chen., "Adaptive Resource Allocation for Preemptable Jobs in Cloud Systems", IEEE 10thInternational Conference on Intelligent Systems Design and Applications, 2010.