A Survey on Dynamic Multiple Request Resource Allocation in Cloud Computing

Nancy Goel[1] Sukhvir Singh[2]
[1,2]NCCE, Israna, Panipat

Abstract— Cloud computing is a vital component of modern computing systems. Cloud computing supply on-demand service. In Cloud computing environment many cloud users can request more than one cloud services in parallel. So there must be an allocation, that all resources which are made available to requesting user in well-planned manner to serve their requirement. In this Review paper, various strategies for dynamic resource allocation in cloud computing are shown based on Linear Scheduling Strategy for Resource Allocation, Topology Aware Resource Allocation (TARA) and Dynamic Resource Allocation for Parallel Data Processing. These algorithms are compared on the basis of these parameters:-Resource Contention, Scarcity of resources, Resource fragmentation, Over provisioning, Under Provisioning. The major limitation of these algorithms is that their resource allocation procedure is not a Dynamic, therefore we are using multiple request resource algorithm to remove this limitation and also to improve the whole performance.

Key words: Topology Aware Resource Allocation, Resource fragmentation

I. INTRODUCTION

Cloud Computing is one of the emerging areas in the various fields related to computer science and can be utilized for on demand effective resource allocation for serving genuine and guaranteed services to the user. Cloud systems are 'on demand' as they offer a way of serving needed resources to consumers as and when they require it. It is pay-as-you-use manner service. In Cloud systems, a pool of services is shared by several consumers. So making these resources free in most optimal way is a dare task. A Multi-Request algorithm is proposed such that it dynamically allocates resources to the users. It helps various clients as they can send their requests at the same time to get the resources as they need. This enables the efficient use of available resources. Then Multiple Request De-allocation algorithm will be used to de-allocate the resources as the client's work will be done. This helps to reduce power consumption. And this makes the whole procedure energy efficient.

A. Cloud Computing Architecture

Cloud Computing: The cloud is a symbol for the Internet and is an idea for the complex groundwork it secret. Cloud computing is a model for enabling worldwide, suitable, on-demand network avail to a shared pool of configurable evaluating services (e.g., web, servers, stowage, applications and capitals) that can be quickly provisioned and liberate with minimal management attempt or service provider interaction. It defines in three models Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS). Fig 1 shows the architecture of the cloud computing. Cloud computing system scales applications by maximizing simultaneousness and using computing resources more efficiently. One must optimize locking time-span, statelessness, sharing pooled services such as task threads network connections bus, cache source data and partition large databases for climbing services to a large number of users. IT companies with fresh ideas for new application services are no longer needed to make large capital expenditure in the hardware and software infrastructures. By using clouds as the application hosting platform, IT companies are freed from the trivial task of setting up basic hardware and software substructures. Thus they can accent more on revolution and creation of business values for their application services. Some of the conventional and appearing Cloud-based application services include social networking, web hosting, content delivery and real time instrumented data processing.

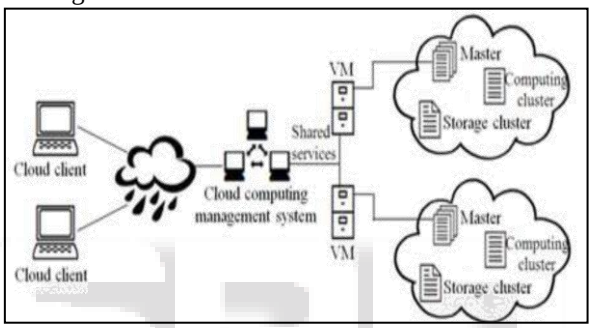

Fig. 1: Cloud Computing Architecture

Cloud computing also describes applications that are extended to be available with the help of Internet. These cloud applications use large data centers and powerful servers that host network applications and network services. Anyone with a suitable Internet connection and a standard browser can access a cloud application.

B. Virtualization

Virtualization, in computing, is the formation of a virtual (rather than actual) version of somewhat, such as a hardware platform, operating system, a storage space device or network resources" .we are receiving many welfare of using virtual machine like rapid scalability, Live migration is possible, consolidation in a Data Center is feasible, Low availability of maintenance, Virtual hardware supports inheritance operating systems well- organized, Security and fault segregation. Virtualization is causing a major change in enterprise data centers and giving rise to a new paradigm: shared virtualized infrastructure. When the data center technicians had problems with traditional data centers (static data centers), the dynamic solution was the solution to their problems. The important fault was very simple, if they had high pressure on their data centres periodically, they could not handle it easily, hence the greatest attainment was the new technology called as server virtualization. They designed the software which creates different isolated parts and shares all of the hardware services between them and give them services as they require. These parts as they called virtual machines (VM), acts and services the clients like real servers with their cons and their pros. The server virtualization

conceptually designed for data centers that they can actively control and share all accessible resources over their data centers and it could be more or less on-demand. With the help of this technology, we can easily generate different virtual machines (not physically) and share all accessible services between these virtual machines. Each virtual machine can have its own services based on the duty which explained for it. VMs are the virtual plotted onto hosts (physical machines) based on hardware need (processing cores, memory and storage). Processing capabilities of host are measured with the help of MIPS (Million Instruction per second). File size is used as the important unit to compute the size of the problem(cloudlet).

C. Resource Allocation

Resource Allocation (RA) is a process of allotting available services to the required cloud applications over the internet. Resource allocation unfed services if the allocation is not maintained exactly. Resource provisioning solves that fault by allowing the service providers to manage the resources for each individual part. Resource Allocation Strategy (RAS) is all about integrating cloud provider activities for using and allocating scarce services within the limit of cloud environment so as to meet the needs of the cloud application. It requires the type and amount of resources needed by each application in order to fulfil a user job. The order and time span of allocation of resource are also an input for an optimal RAS. An optimal RAS should ignore the following criteria as follows:

a) Resource contention situation arises when two applications try to access the same service at the same time.
b) Scarcity of resources arises when there are limited resources.
c) Resource fragmentation situation arises when the services are limited.
d) Over-provisioning of resources arises when the application gets surplus resources than the demanded one.
e) *Under-provisioning* of resources occurs when the application is assigned with fewer numbers of resources than the demand.

II. LITERATURE REVIEW

In 2012, Dr. Padmavathi Ganapathi [1] have discussed Cloud computing technology is increasingly being used in venture and business areas. In cloud paradigm, an effective resource allocation planning is required for achieving user gratification and maximizing the benefit for cloud resource providers. This paper summarizes the classification of RAS and its effects in cloud system. Some of the strategies discussed above mainly focus on CPU, memory resources but are lacking in some features. Hence this survey paper will hopefully inspire future researchers to come up with smarter and chained optimal resource allocation algorithms and framework to strong the cloud computing paradigm.

In 2012, S.V.Hemanth [2] have discussed the challenges and chances for positive parallel data processing in cloud environments and presented Nephele, the first data processing framework to misuse the dynamic service provisioning offered by today's IaaS clouds. We have described Nephele's basic architecture and presented a performance comparison to the well-maintained data processing framework Hadoop. The performance evaluation gives a first mark on how the capability to allocate specific virtual machine types to specific tasks of a processing job, as well as the feasibility to automatically allocate/de-allocate virtual machines in the course of a job completion, can help to improve the whole resource utilization and, consequently, decrease the processing cost. With a framework like Nephele at hand, there are a variety of open research problems, which we strategize to address for future work. In particular, we are interested in improving Nephele's capability to adapt to service overutilization or underutilization during the job execution automatically. Our present profiling approach builds a valuable base for this, however, at the time the system still needs a reasonable amount of user notations. In general, we think our work presents a vital contribution to the growing field of Cloud computing services and points out exciting new chances in the field of parallel data processing.

In 2012, Venkatesa. Kumar [3] have discussed cloud service providers, Infrastructure as a Service (IaaS) provides a Compute resources for demand in various applications like Parallel Data processing. The computer resources offered in the cloud are extremely dynamic and probably heterogeneous. Now their algorithm, Our Algorithm utilizes the Turnaround time Utility efficiently by differentiating it into a gain function and a loss function for a single task. The algorithm also assigns high priority for task of early completion and less priority for abortions/deadlines issues of real time tasks. The algorithm has been experimented on both preventive and Non-preventive methods. The experimental results show that it out performs the existing utility based scheduling programs and also compare its performance with both preventive and Non-preventive scheduling methods. Hence, a novel time utility planning approach which scars on both more priority and the less priority functions that reaches for happening is proposed.

In 2013,January, Amit Kothari [4] have discussed that After the implementation of proposed strategy, user will not have to worry about the scarcity of resource availability or problems like host failure. The cloud administrator also gets the same benefit. Resource manager can take the advantages of getting help from remote cloud using the "quality based methodology". All the tasks will be done dynamically. User will get full transparency regarding the accessibility of resources without worrying about which server or cloud resources are provided. The purpose of the study is that the user will have their resources in a fastest and most secure manner with lesser power consumption.

In October 2013, Rakesh Rajani [5] have discussed The author highlights use cases which are presently poorly supported by parallel data processing frameworks and describes how a tighter merge between the processing framework and the underlying cloud system can help to less the financial processing cost for the cloud consumer. With a framework like Nephele at hand, there are a variety of open research problems, which we stratagize to address for future work. We are interested in improving Nephele's capability to adapt to resource overload or underutilization during the job completion automatically. Our present sketching approach builds a valuable basis for this; however, at the moment of the system still needs a reasonable cost of user notations.

In June 2014, M . Dillibabu [6] have discussed the challenges and opportunities for efficient parallel data processing in cloud and presented the new algorithm for dynamic resource allocation in cloud computing environment using "Need Based Resource Allocation". This algorithm allocates the resources for the tasks in dynamic manner with maximum utilization of resources and thus meets the required QoS to the users. They conclude that no resources are over utilized and underutilized using our proposed method. As future work, they plan to add more security with our existing work. And also to integrate the mobile environment where it requires optimum resource utilization.

In 2014,Vinayak Awasre [7] have discussed that Cloud computing technology is increasingly being used in enterprises and business markets. A review shows that dynamic resource allocation is growing need of cloud providers for more number of users and with the less response time. In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the main types of RAS and its impacts in cloud system. Some of the strategies discussed above mainly focus on memory resources but are lacking in other factors. Hence this survey paper will hopefully motivate future researchers to come up with smarter and secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

In August 2014, Priyanka Mod[8] have discussed the proposed method is Ant Colony Optimization Algorithm (ACO). ACO adapt genetic operations to enhance ant movement towards solution state. This paper provides complete description of the resource allocation techniques in cloud, for cloud users a comparative study provide the detail about the different Resource allocation methods. After that, the research problems and challenges are explored to promote the development of cloud computing.

In September 2015, Rajeev Pandey [9] have discussed that Design, implementation, and evaluation of a resource management system for cloud computing services is presented. The system multiplexes virtual to physical resources adaptively based on the changing demand. The skewness metric is used to combine VMs with different resource characteristics appropriately so that the capacities of servers are well utilized. Proposed algorithm achieves both overload avoidance and green computing for systems with multi-resource constraints. By switching of idle machines power can be saved at cloud data centers. Live migration helps in saving power by migration of the virtual machines causing overload and underutilization of the host. Simulation shows around 55 % of Energy saving utilizing our resource allocation approach.

In May 2015, Rutvik Mehta [10] , Cloud computing technology is increasingly being used in enterprises and business markets. Resource management is one of the most important job of cloud computing and its mostly accomplished by resource allocation. In cloud paradigm, an effective resource allocation strategy is required for achieving consumer satisfaction and maximizing the benefit for cloud service providers. In this paper we analyzed resource allocation in detail and its various plannings are reviewed in detail. It is believed that this paper would profit both cloud consumers and researchers in understanding the concepts of resource allocation.

In November 2015, Y. Bharath.Bhushan [11] have discussed that they presented the design, implementation and evaluation of a resource management system for cloud computing services. The performance evaluation gives a first impression on how the ability to assign specific virtual machines to specific tasks of a processing job, as well as the possibility to automatically allocate/de-allocate virtual machines in the course of a job execution can help to improve the overall utilization.

In May 2016, Ashwini E. [12] Cloud Computing is definitely a promising tendency to sought high demanding applications and its related faults. Main objective of the cloud computing environment is to balance load and achieve good performance. Dynamic nature and complexity web make load balancing very complex and vulnerable to faults. To maintain entire load of nodes is very hard due to dynamic nature of resources in a network environment. There are huge number of features, which can effect the server performance like load balancing and resource sharing in the network environment. It focuses on load balancing and presents features due to which load balancing is started, compare load balancing programs and proposes an efficient dispatcher algorithm for network environment.

In May 2016, Paresh patil [13] have discussed that Cloud computing technology is increasingly being used in venture and business global areas. Dynamic resource allocation is growing requirement of cloud providers for many users and with the less response time. In cloud paradigm, an effective service allotation plan is need for achieving user satisfaction and maximizing the benefit for cloud service providers. This paper summarizes the service allocation and its impacts in cloud system.

In November 2016, Dr. P. Srinivasan [14] In this paper, we have discussed the scheduling strategies followed in the cloud environment. Different research works based on service scheduling were analyzed. Scheduling is done between tasks and resources. Allocation of resources to virtual machines called resource provisioning, virtual machines to physical machines consisting of single core and multiple cores were analyzed. After scheduling has been done, monitoring was also done to follow up the resources.

Using scheduling techniques, performance, bandwidth utilization and quality of service could be increased with the decrease in make-span time and cost. Virtualization is the technology in cloud computing which makes the resources spread across the world available to the end-users on a marginal cost. The various available scheduling algorithms in cloud computing environment and the performance which are improved or affected were listed out. More research work needs to be done on virtualization, where provisioning of resources need attention to decrease the execution time and to improve availability of resources.

III. CONCLUSION

Cloud computing technology is increasingly being used in venture and business areas. A review shows that dynamic service allocation is growing need of cloud providers for more number of users and with the less response time. This paper summarizes the algorithms and their impact on cloud system. Some of the algorithms mainly focused on the

memory resources allocation due to which they are facing problems in other parameters. So with this review paper hopefully, the future researchers could get more ideas and can give more secure and optimized results.

REFERENCES

[1] V.Vinothina, Dr.R.Sridaran, and Dr.Padmavathi Ganapathi, "A Survey on Resource Allocation strategies in Cloud Computing," International Journal of Advanced Computer Science and Applications(IJACSA),vol. 3 no. 6, pp. 97-104, 2012
[2] D. Prasad, Fathima Zehra, S.V.Hemanth "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud," International Journal of Computer Science And Technology(IJCST),vol. , issue 2, pp. 610-622, June 2012.
[3] Venkatesa Kumar, V. and S. Palaniswami, "A Dynamic Resource Allocation Method for Parallel Data Processing in Cloud Computing,"Journal of Computer Science, pp. 780-788, 2012.
[4] Amit Kothari, Nootan Padia, and Maulik Parekh, "Real Time Resource Allocation Methodology and Quality based Framework in Cloud Computing," International Journal of Computer Application(IJCA),vol. 62 no. 11, pp. 31-34, January 2013.
[5] Rupali Shelke, and Rakesh Rajani, "Dynamic Resource Allocation in Cloud Computing," International Journal of Engineering Research and Technology(IJERT),vol. 2 issue 10, pp. 843-846, October 2013.
[6] B. Karthikeyan, T. Sasikala , and M. DilliBabu, "Dynamic Resource Allocation for Cloud Computing including the feasibility study on Green Computing," International Journal of Science and Research (IJSR),vol. 3 issue 6, pp. 1230-1232, June 2014.
[7] Vinayak Awasare, and Sudarshan Deshmukh, "Survey and Comparative Study on Resource Allocation Strategies in Cloud Computing Environment," IOSR Journal of Computer Engineering(IOSR-JCE),vol. 16 issue 2, pp. 94- 101, April 2014.
[8] Priyanka Mod, and Mayaank Bhatt, "A Survey on Resource Allocation Technique in Cloud Computing," International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE),vol. 3, issue 8, pp. 7815-7818, August 2014.
[9] Saima Israil, Rajeev Pandey, and Uday Chorasia, "Dynamic Resource Allocation for Overload Avoidance and Green Cloud Computing," International Journal of Computer Applications (IJCA), vol. 126, no. 12, pp. 6-13, September 2015.
[10] Abhijits inh Parmar, and Rutvik Mehta, "An Extensive Analysis on Resource Allocation in Cloud Computing Environment," International Journal of Advance Engineering and Research Development(IJAERD), vol. 2 issue 5, pp. 303-308, May 2015.
[11] Y. Bharath. Bhushan, and V. Bhavani, "Dynamic Resource Allocation using Virtual Machines and Parallel Data Processing in Cloud," An International Journal of Advanced Computer Technology, vol. 4 issue 11, pp. 2001-2005, November 2015.
[12] Ashwini. E, "Dynamic Allocation of Virtual Machines in Cloud Computing," Global Journal for Research Analysis(GJRA),vol. 5 issue 5, pp. 266-268,May 2016.
[13] Vinayak Sontakke, Paresh Patil, Sumedh Waghamare, Rajani Kulkarni, N. S. Patil, and M. Saravanapriya, "Dynamic Resource Allocation Strategy for Cloud Computing using Virtual Machine Environment," International Journal of Engineering Science and Computing(IJESC),vol. 6 issue 5, pp. 4804 – 4806,May 2016.
[14] C. Vijaya, DR. P. Srinivasan, "A Survey on Resource Scheduling in Cloud Computing," International Journal of Pharmacy and Technology(IJPT), vol. 8 issue 4, pp. 26142 26162, November 2016.

