IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

A Review on Techniques used for Resource Allocation in Cloud Computing

Mast. Neil Narse[1], Mast. Pravin Chaudhari[2], Ms. Khushi Nikumbh[3], Ms. Mansi Shardul[4]
Department of Computer Engineering, Sandip Institute of Technology, Nashik, Maharashtra, India[1]
Department of Computer Engineering, P. V. G College of Engineering, Nashik, Maharashtra, India[2]
Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, India[3]
Department of Computer Engineering, K. K. Wagh College of Engineering, Nashik, Maharashtra, India[4]
eon.neil@gmail.com[1], pravinchaudhari2002@gmail.com[2], khushinikumbh33@gmail.com[3],mansishardul@gmail.com[4]

Abstract: *In the cloud computing, RA Allocation (RA) is the process of providing resources available to the required cloud systems via the Internet. The process of redistributing resources is key to embracing unexpected demands and improving return on investment from cloud- based infrastructure. The cloud compound consists of multiple clients requesting resources in a dynamic environment with potential obstacles. In cloud computing the distribution of resources for every highly profitable and profitable client is the main task of the cloud controller. Cloud Administrator has trouble sharing the same resources with multiple users. So when allocating resources to prioritize activities is very important. The paper deals with review on techniques or editing algorithms and six distribution methods and parameters for scheduling success. An analysis of the methodology of this study helps researchers discover the value of a planning algorithm or techniques for a particular method of distribution in order to maximize resource allocation in the cloud.*

Keywords: Resource allocation, information technology, task scheduling, cloud applications, cloud computing.

I. INTRODUCTION

Cloud Computing is a model that provides computational resources through the internet and central secluded servers. It acts as a service rather than a merchandise. Refining cloud computing and making it more efficient is the key to the next generation of computation. Companies and even normal people can rent resources from the cloud for Computational purposes. This not only reduces the cost required by infrastructures but in the long run even helps reduce E-waste because the person wanting to upgrade his computation need not throw his old hardware but instead have cheap access to rent and have the hardware resource on the cloud allotted to him when needed. Resource allocation system in cloud computing is the mechanism to guarantee that all the user's requirements are met and are more importantly, met efficiently. An important problem that must be addressed effectively in the cloud is to manage the quality of service and maintain service level. Agreement for cloud users that share the resources. Cloud Computing is of namely three types

1.1 Infrastructure as a Service (IaaS)

In the early days of the cloud, most providers offered Infrastructure as a Service (IaaS), where consumers could buy access to virtual machines and storage. This allows businesses to be agile, respond quickly to market changes, and save on IT costs.

1.2 Cloud Platform as a Service (PaaS)

In this service model, instead of using locally run operating systems and applications, the cloud consumer uses applications that are hosted by the provider in the cloud. Cloud consumer can build and deploy their own applications on the platform without having to build and maintain their own development platforms. Examples of PaaS include Salesforce.com and Microsoft Azure.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

1.3. Software as a Service (SaaS)

In this service model, cloud providers offer a set of software applications and services to consumers over the internet. Cloud consumer can access the software and services through the internet without actually having to install and maintain the software on their own servers. Examples of SaaS include email services such as Gmail, word processing software such as Microsoft Office 365, and online storage such as Google Drive. These services are hosted by the provider, who updates and maintains the software on their servers. The consumer only pays for what they use, typically a monthly subscription fee.

The main advantage of the cloud is that it allows companies to focus on their core business rather than on managing large data centers. Instead, they can buy computing resources and storage from the cloud when they need them, paying only for what they use.

Allocation of resources is an important component of cloud computing as its efficiency will directly have an impact on the performance of the overall cloud environment.

Regardless of the services clouds can be classified as

A. Private Cloud

A private cloud is a cloud computing platform that is only accessible to an organization and is managed by the organization itself or by a third-party vendor.

B. Public or Hosted Cloud

The cloud services are available to the general public without any fee. Unlike the private cloud, the public cloud is not managed by the organization. Instead, the organization's services are provided to the general public. Cloud services offered by the organization are free to the general public.

C. Community Cloud

The cloud is available to specific group of people. All the cloud services are shared by all these community people.

D. Hybrid Cloud

It is a combination of two or more cloud models.

II. Significance of Resource Allocation in Cloud Computing

Resource Allocation (RA) refers to the process of assigning the available resources to which are required by the customer over the internet. Resources are not allocated if the Resource Allocation Strategy (RAS) is not managed adequately

Resource Allocation Strategy (RAS) is about combining cloud provider activities for utilizing and allocating scanty resources which are in the limit of cloud environment so as to meet the needs of the cloud application. It requires the desired type and sufficient amount of resources needed by each application in order to complete a user job.

The order and time of allocation of resources is considered as an input for an optimal RAS. Under-provisioning of resources occurs when the application is assigned with fewer numbers of resources than the demand. The cloud user or the customer roughly estimates the resources required to complete a job. On the other hand, allocating the resources before the roughly estimated time may lead to over provisioning of resources.

To avoid the over provisioning and under provisioning of resources inputs are requires from the customer (cloud user) as well as the cloud provider for an efficient Resource Allocation Strategy (RAS).

Allocation the Resources over Cloud consist of physical and virtual resources. The physical resources are shared across multiple compute requests through virtualization and provisioning while, the request for virtualized resources is described through a set of parameters detailing the processing, memory and disk. Provisioning satisfies the request by mapping virtualized resources to physical ones. The hardware and software resources are allocated to the cloud applications on-demand basis.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

III. Methods for Resource Allocation in Cloud Computing

3.1 Tenant Based Resource Allocation

This technique is used to allocate recourses in an efficient manner over the SaaS provider or applications taking into consideration multitenancy. Multitenancy is when each tenant can interact with the application as if it were the exclusive owner. Here tenant is the customer using the applications over the SaaS provider. For deployment the following strategies are used- The resource request is distributed according to the tenant information using the tenant- based load balancing The encapsulation of each tenant execution is done using the tenant-based isolation The Virtual machine (VM) instance is used to determine the number of VM required based on workload.

3.2 Economy Based

As the name suggests, it takes into consideration the incentives of both the parties that is, the provider as well as the consumer. It uses the economy-based algorithm. Economy based method aims for the concept of optimizations of resources as wellas the costs.

3.3 Rule Based

This technique is used to access the resources one of the challenges of allocating resources dynamically virtual machines such that the quality of service constraints are maintained and operating costs are minimized. The arrangement between the above mentioned goals can be expressed by a utility function.in Utility based resource allocation , a two-tier resource management approach based on sufficient utility functions is presented, on IaaS layer as and when demanded. Rule based resource manager is developed to enhance the system capabilities in case of private cloud. It also aims for decreasing the cost for hybrid environment and in turn giving time to private and public cloud for providing the services efficiently.

3.4 Agent Based

In this technique the resources are allocated mainly based on two techniques –

- The work volume at centers
- The distance between the provider and the center
- Agent based resource allocation allows decentralization of resources. Eventually there can be many improvements in agent based resource allocation such as
- Computational load distribution and reduced computation time.

3.5. Preference Based

In order to provide services to multiple users at the same point of time with differing resource requirements, a cloud provider needs to find and put into use an efficient resource allocation technique which meets the requirements of the user as well as the provider. To address the issue, Offline and online auctions are considered one of the best ways of resource allocation where a user is selected based on his/her payment capacity and resource requirements. The bidding price reflects the needs and capacity of the user.

3.6 Utility Based

This technique uses Master-Slave framework which is a role based access control considering the trust of the node and meets the requirements using the services. This technique provides assessment according to the real-time condition of the system and then allocates resources based on their assessment generated. This kind of dynamic feedback mechanism guarantees the firmness of the system and reliability of the services effectively.

Consisting of local controllers that dynamically allocates CPU shares to the virtual machines to maximize a local node utility function and a global controller that initiates live migrations of virtual machines to other physical nodes to maximize a global system utility function. The Table summarizes the work done by various researchers and future work and/or gaps in their existing work.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Year	Author	Techniques/Algorithms	Tools used	Future work and/or gaps in existing technologies
2009	Qiang Li, Qinfen Hao, Limin Xiao and Zhoujun Li [11]	Adaptive Management of Virtualized Resources using Feedback Control	KVM	Only KVM model, network I/O performance, Still better modelling can be done for resource sharing.
2012	Mayank Mishra, Anwesha Das, Purushottam Kulkarni and Anirudha Sahoo [12]	Live Virtual Machine Migration	Not Mentioned	Only load on the virtual machine for migration is considered. Consumer requirements and priority of job is not considered.
2011	T. R. Gopalkrishnan Nair and Vaidehi M [13]	Rule Based Resource Allocation Model (RBRAM), M-P-S (Memory-Processor-Storage) Model	Not Mentioned	Only Cloud Efficiency Factor is considered to evaluate the performance level of cloud system.
2010	Rosy Aoun, Elias A. Doumith and Maurice Gagnaire in [14]	Mixed Integer Linear Program for resource provisioning	Globus Toolkit and Condor -dispatcher (Sebastian 2)	Dynamism is considered based on pre-planned traffic. Resource allocation algorithm is executed offline. Observed simulation runtimes exceed the considered scheduling period.
2011	Justin Y. Shi, Moussa Taifi and Abdallah Khreishah [15]	Timing Model with Amazon EC2	Amazon EC2	Authors mainly focused on cost effectiveness parallel processing
2013	Chenn-Jung Huang, Chih-TaiGuan, Heng-MingChen, Yu-WuWang, Shun-ChihChang, Ching-Yu Li and Chuan-HsiangWeng [16]	Resource Allocation based on Support Vector Regression (SVR) and Genetic Algorithm (GA)	CloudSim	Authors plan to modify the algorithms to decrease the calculation time in terms of the prediction process to improve the GA's convergence speed.
2012	Zhen Xiao, Weijia Song and Qi Chen [17]	Resource Allocation based on "Skewness" Metric	Xen Virtual Machine	The migration of job has to be there but which job to migrate that is not specified. The effect of partial machine migration is not discussed.
2011	Amit Nathani, Sanjay Chaudhary and Gaurav Somani [18]	Policy based resource allocation, Haizea scheduler	Haizea	Backfilling algorithm is not implemented yet. Response time of best effort service can be enhanced.
2011	Weiwei Lina, James Z. Wangb, Chen Liangc and Deyu Qia [19]	Threshold based dynamic resource allocation	CloudSim	Experiments were done only for Internet Applications. Overhead of virtual resources are considered not for physical resources.
2010	Yichao Yang, Yanbo Zhou, Lei Liang, Dan He and Zhili Sun [20]	Cloud Infrastructure Service Framework for QoS requirements with Service-oriented Resource Broker.	Not Mentioned	Optimization of combined resources is a challenge.
2011	Kejiang Ye, Xiaohong Jiang, Dawei Huang, Jianhai Chen and Bei Wang [21]	Live Migration of Virtual Machines	Xen and VMWare	Intelligent live migration machine can be future work.
2011	Congfeng Jiang, Xianghua Xu, Jilin Zhang, Yunfa Li and Jian Wan [22]	Resource allocation and scheduling heuristics algorithms based on real time knowledge of workload and performance feedback	Xen	Cache allocation and contention not considered which may improve prediction accuracy and allocation efficiency.
2010	Guiyi Wei, Athanasios V. Vasilakos, Yao Zheng and Naixue Xiong [23]	Game theoretic method for fair resource allocation to solve sophisticated parallel computing problem	Not Mentioned	Applications are not clearly defined.
2011	Baomin Xu, Chunyan Zhao, Enzhao Hu and Bin Hu [24]	Job scheduling algorithm based on Berger Model based on dual fairness constraints	CloudSim	More accurate vector value of the general expectation vector can obtained.
2011	Linlin Wu, Saurabh Kumar Garg and Rajkumar Buyya [25]	Resource allocation algorithm for maximizing total profit and customer satisfaction.	CloudSim	Some more services can be included for improving the performance such as spot pricing.
2009	Borja Sotomayor, Ruben Santiago Montero Ignacio Martin Llorente and Ian Foster [26]	Prediction of various run time overheads for advanced reservations, Haizea based	Xen Virtual Machine	Experiments are done on only Xen VM, not on KVM VM or any other.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Figure 1: Names of authors and researchers.

Why strategies? ->Weakness/problems to be avoided:
Resource unavailability conditions occurs when two or more users attempt to access same content same time.
Lack of resources/VM on cloud from cloud service provider.
Over-provisioning of resources happens when the said tasks get spare resources than the actual demand.
Under-provisioning of resources befalls when the user's tasks are allotted with fewer quantities of resources than the actual demand.
Resource fragmentation problem need to be Consider as resources are in accessible sometimes.

IV. Resource Allocation Strategies

- Dynamic resource allocation.
- Linear scheduling strategy.
- Particle swarm optimization.
- Ant colony optimization (ACO) algorithm.
- Reliability Aware Scheduling
- Cost based resource allocation.

4.1 Dynamic Resource Allocation.

Due to lack of resource availability and under provisioning dynamic resource allocation come into picture, which solves problems by emphasizing on-demand services and resource scalability. This is one of best optimization strategies for cloud computing by distributing loads to various virtual machines (VM). As in e-commerce's websites and banking application changes their resource dynamically in order to satisfy end users need. Utilizing cloud dynamic resource allocation method reduces low response time and serve user with resources dynamically.

4.2 Linear Scheduling Strategy

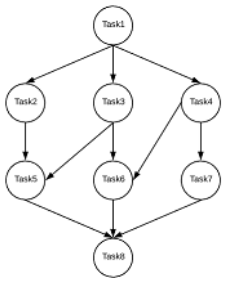
Linear scheduling strategy can be also called as service level agreement as it focus on response time, throughput, and maximize quality of service QOs. Software as a Service (SaaS), the main consideration here is the quality of service parameters of the cloud service provider's side, like a load on the current environment and cost associated with the data center. Resource utilization in cloud environments is based on memory usage, CPU usage, and throughput. Thus, the scheduling algorithm uses a server node to determine the most applicable virtualization method.

4.3 Ant Colony Optimization (ACO) Algorithm

ACO algorithm technique derived from ants as release their hormone known as pheromone on ground in order to make a path for other colony ants to follow. ACO is significant way to solve load balancing problem over network and improve computation time. ACO manages the usage of virtual machines, memory, and the number of clouds.

4.4 Reliability Aware Scheduling

With the increase in number of users on cloud, the network failures and other performance related issues have become inevitable.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

Figure 2: Graph of task scheduling designing of cloud architecture.

To achieve reliability we need to focus on designing of cloud architecture. This directed acyclic graph (DAG) focuses to achieve best task scheduling. The reliability analysis block will figure out the processor's reliability with respect to different frequency in order to maintain the reliability of the system. Finally, the scheduler will schedule the user's task based upon the DAG.

4.5 Cost Based Resource Allocation

In cloud computing, typically using a "pay- as-you-go" model which is nothing but to charge user on utilizing resources shared by cloud. In cloud computing consists of computing resources such as RAM, CPU, memory, bandwidth etc. which will be shared by multiple users. The goal of the cloud service provider is to generate the profits with its user's gratification. The billing will be generated based on the size and time of the resources used by the end users. This strategy used to focus on completion time of particular task and execution time in order to make more profits with high user satisfaction.

V. LIMITATIONS

There are two players in a cloud computing environment, the service providers and the users and they both have different end goals. The service providers want maximum resource usage to earn more revenue and the users want to minimize their expenses while meeting their performance requirements. Therefore it is difficult to allocate resources in a mutually beneficial way. The increase in variability and uncertainty in resources everyday cannot be solved by traditional resource allocation which poses a difficult challenge for both parties. Many resource allocation schemes have been shared in the development of cloud computing however the major difficulty is optimization of the resources being allocated.

CONCLUSION

Various techniques are used for ensuring optimal resource allocation in cloud computing. environments have been surveyed and investigated both at the advanced level as well as the due to increase in the use of cloud users efficient resource allocation has been a problem which needs and effective mechanism that will guarantee the customer satisfaction and cost optimization for the provider. Though some substantial results have been obtained for enhancing the dynamic resource allocation, there is scope for further betterment. Hopefully this paper will motivate future researchers to come up with smarter and optimal ideas for allocation of resources in cloud computing.

REFERENCES

[1]. 2013 Nirma University International Conference on Engineering (NUiCONE) Swapnil M Parikh Department of Computer Science and Engineering,BITS edu campus, Varnama.

[2]. Resource Allocation Techniques in Cloud Computing: A Review and Future Directions by Muhammad Faraz Manzoor Adnan Abid, Muhammad Shoaib Farooq, Naeem A. Nawaz.

[3]. A survery on resource allocation stargeies in Cloud computing by

[4]. A comparative study on Resource Allocation techniques in Cloud Computing by Pandaba Pradhan, Prafulla Ku. Behera, B N B Ray in International Journal of Advanced Computer Engineering and Communication Technology (IJACECT). Resource Allocation Techniques in Cloud computing-Research Challenges for Applications by N.R.RamMohan Research Scholar, Computer and Information Tech., Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India. And E. Baburaj Professor, Dept of Computer Science and Engg, Sun College of Engineering and Technology, Nagercoil, Tamilnadu, India. In 2012 fourth International Conference on Computational Intelligence and communication networks. A Survey on Resource Allocation Strategies in Cloud Computing by V.Vinothina, Sr.Lecturer, Dept. of Computer Science, Garden City College, Bangalore, Karnataka, India. Dr.R.Sridaran, Dean Faculty of Computer Applications, Marwadi Education Foundation's Group of Institutions, Rajkot, Gujarat, India.

[5]. Dr.Padmavathi Ganapathi Professor and Head, Dept. of Computer Science, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, Tamil Nadu. India in (IJACSA) International Journal

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

of Advanced Computer Science and Applications, Vol. 3, No.6, 2012