International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 5, May - 2014

Priority Based Resource Allocation in Cloud Computing

Savani Nirav M.
Student at M.E in Computer Engineering
Gujarat Technological University
Ahmedabad, India

Prof. Amar Buchade
Computer Engineering
Pune Institute of Technology
Pune, India

***Abstract*—Due to increase in the usage of cloud computing there is a need for a efficient and effective resource allocation algorithm which can be used for proper usage of the resources and also check that the resource is not wastage. In this we propose a priority based resource allocation algorithm which can be used for proper allocation of resources and also the resources are allocated efficiently and effectively.**

Keywords—cloud computing, resource allocaation, priority based allgorithm

I. INTRODUCTION

Today, cloud computing is emerging at a rapid rate. Many of the organizations has already started moving towards the cloud computing due to their "on-demand service and pay for what you use" services. In general we can define cloud computing is style of computing in which IT-related capabilities are provided-as a service, allowing users to access technology-enabled services from the Internet without the knowledge of expertise with or control over the technology infrastructure that supports them. Email applications were probably the first service on the cloud. As now a day's many computing industry is shifting toward providing Platform as a Service, Infrastructure as a Service and Software as a Service for consumers and enterprises to access on demand regardless of time and location, which helps in avoiding the over-supplying of the resources when used with utility pricing, which meets the demands of the millions of users.

Types of services in cloud computing are as follows:-

A. Software as a Service

In the business model using Software as a Service, user's are provided access to the application software and databases. Cloud Provider manages the infrastructure and platform that run the applications. SaaS is sometimes also known as "on-demand software" and is usually priced on pay-per use basis.

B. Platform as a Service

In the PaaS model, cloud provider provides a computing platform, which includes operating system, programming language executing environment, database and a web server. Developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the hardware and software layers.

C. Infrastructure as a Service

Infrastructure as a Service is a provision model in which an organization outsources the equipment used to support operations, including storage, servers and networking components.

II. TYPES OF CLOUD

There are normally three types of clouds in cloud computing:-

A. Private Cloud

A private cloud is one in which the services and infrastructure are maintained on a private network which generally a datacenter within an organization. These clouds offer the greatest level of security and control, but they still require the company to purchase and maintain all the software and infrastructure, which can significantly reduce cost savings.

B. Public Cloud

A public Cloud is one in which the services and infrastructure are provided off-site over the internet. At its essence, "Cloud Computing" refers to the public cloud. These clouds offer the greatest level of efficiency in shared resources as well as efficiency in cutting spending. However, they are also more vulnerable than private cloud.

C. Hybrid Cloud

A hybrid cloud offers a variety of public and private options with multiple of providers. By using a hybrid approach, you are able to spread things out over a number of providers to keep each aspect of your business in the most efficient possible environment. The major downside here is of having to keep track of multiple security platforms and make sure all aspects of business can communicate with each other.

III. RELATED WORK

As per the method proposed in [1] for resource allocation method, the author says that the resources which are demanded by the customer for servicing their request of the job should be identified. After identifying the resources, the amount of the resources which are needed by the customer and the center which is having that particular resource in the largest proportation should be kept reservered for future use.

In [2], the author Soumya Ray and Ajanta De Sarkar, has proposed a novel approach for resource allocation method. In this method a matrix for allocating the resources is to be created by which the provider will be able to identify the amount of resources which are available with the cloud provider and the amount of resources which are currently being accessed by the customer. The load capacity of each and every resource can be identified with the help of the resource occupancy matrix.

There are many different methods for allocating resources into the cloud computing. From all the methods, due to the less wastage of the resources and proper allocation of the resources into the cloud computing, priority based resource allocation method is been used.

In priority based resource allocation, there are many different methods in which we can use this method into the cloud and grid computing.

Algorithm			Complexity		Focus On			Scalability		
Name	Method	Factor	Cost	Time	Makespan	Resource utilization	Speed	Dynamic	Static	Recommended in
Prioritizes Round Robin Algorithm	Batch Mode	Meta tasks	x	√	x	√	x	x	√	Grid
Priority based Weighted Fair Scheduling Algorithm	Batch Mode	Meta tasks	x	√	x	√	√	x	√	Cloud
Priority based Job Scheduling Algorithm	Batch Mode	Meta tasks	√	√	√	√	x	x	√	Cloud
Agent based Priority Heuristic	Batch Mode	Meta tasks	√	√	√	x	x	√	x	Grid
Deadline based Priority Heuristic	Batch Mode	Meta tasks	x	√	√	√	x	√	x	Grid
Two stage priority rule based algorithm	Batch Mode	Meta tasks	x	√	x	√	x	x	√	Grid

The above methods gives us the idea about how the priority based resource allocation method is to be used in different ways for allocating of the resources into the cloud computing.

IV Different Scheduling based on Different Criteria

There are different types of scheduling techniques based on different criteria's, such as Static Scheduling, Dynamic Scheduling, Dynamic Scheduling, Centralized Scheduling, Decentralized Scheduling, offline or online scheduling:

1. **Static Scheduling:** This scheduling method is used for pre-schedule of the jobs. In this all the information are known about the available resources and tasks and a task which are assigned to a resource. So its easier to adapt based on schedulers policy.
2. **Dynamic Scheduling:** In this method, jobs are dynamically allocated over time by the scheduler. It is more likely to use in the real time scenario than static scheduling. It is more critical to include load balance as a main factor to get stable, correct and efficient algorithm.
3. **Centralized Scheduling:** The main advantage of centralized scheduling are ease of implementation; efficiency and more control and monitoring on resources.
4. **Decentralized Scheduling:** This type of scheduling is more realistic for real cloud environment despite of its weak efficiency compared to centralized scheduling.

V WORKING OF ALGORITHM

The whole algorithm will be working accordingly:-

Once the user's request will be received at the cloud end, after that according to the user's requirement, the resources will be checked for assigning to the user. Batches of the user's requirement will be created according to the type of task, the amount of processor required by the user, and time for the execution of the user.

If the resources are not available then the user needs to wait for the resources to be available. The user's waiting request will be compared with all the waiting resources and priority will be assigned accordingly. The throughput value is calculated according to the usage of the processor and ram.

If the request of two same requirements having the same priority then at that point of time the resources will be allocated on the basis of FCFS(First Come First Served).

The whole algorithm is simulated into the CloudSim 3.0 where there are 12 number of users in which two is the Datacenter's which is the resource provider to each and every 10 customers. If the resource VM is not available into the Datacenter 1 then the request will be checked into the Datacenter2 and VM will be allocated accordingly for execution of the task of the user.

According to the allocated VM into the Datacenters the cost of each and every customer will be defined and the SLA will be prepared accordingly so that there would not be any violation of the SLA from any of the side.

How Priority will be assigned to the resources?

As the resource request in the cloud computing is not fixed, the user's demand for many of the resources based on what type of resources are required by the user. So the priority is decided with the help of Analytical Hierarchy Process(AHP). AHP is a multi-criteria decision making and multi-attribute decision making model. The main use of the AHP is the comparison matrix which can be shown as:-

$$A = \begin{cases} Atf = \frac{1}{aij} & i \neq j \\ 1 & i = j \end{cases}$$

Each value in the matrix A is positive (a_{ij}). Here, A is a square matrix ($A_{n\times n}$). The most important step in AHP is to calculate the vector of weights(ώ). Vector of weights can be computed through $A\acute{\omega} = \lambda_{max}\acute{\omega}$.

If A is absolutely consistent then $\lambda_{max} = \eta$. In this case A will be consistent. For consistency a consistency ration is also generated which can be defined as

CR = $\frac{CI}{RI}$ where RI is the random index randomly based on rank of comparison matrix.

CI can be calculated as $CI = (\lambda max - \eta)/\eta$

VI RESULTS

The whole algorithm is implemented in the cloudsim 3.0.3 and the obtained results of the algorithm are as follows:

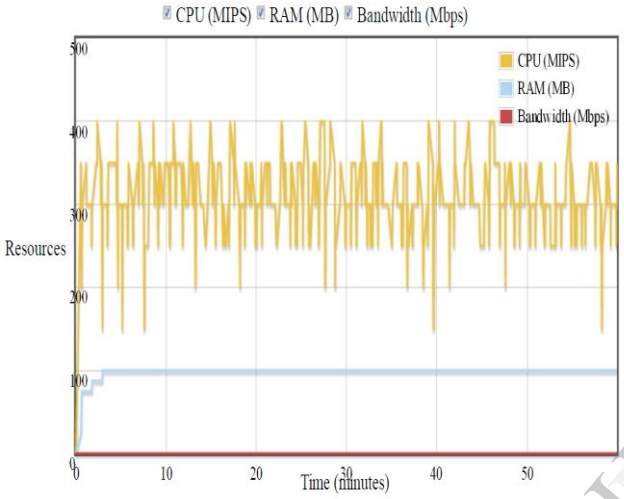

Figure 1. Overall Resource Utilization

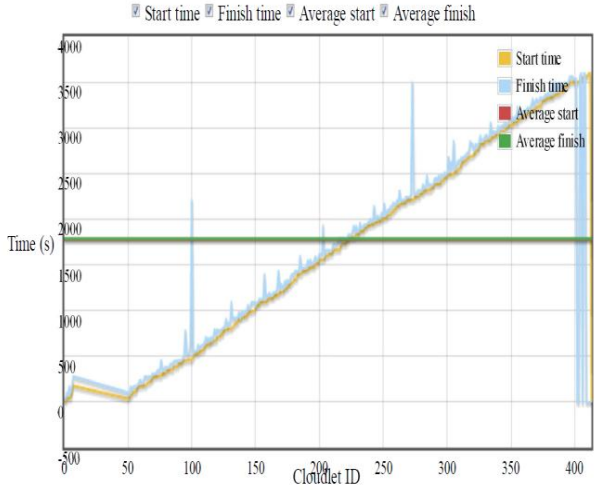

Figure 2. Time Required for execution of resources

VI CONCLUSION AND FUTURE WORK

This algorithm is tested and according to the results which are generated shows that the resources in the cloud are allocated according to the priority which is assigned to each and every customer and also the resources are allocated efficiently and effectively. As this algorithm is tested on the simulator, the future work of this algorithm is to be tested on a real cloud environment and needs to be check for the resource allocation to be done efficiently and effectively. Load Balancing on each and every datacenter for allocating of the resources will be the future work of this algorithm

VII REFERENCES

1. Tsumura, S.; Kuribayashi, S.-i., "Simultaneous allocation of multiple resources for computer communications networks," Communications, 2006. APCC '06. Asia-Pacific Conference on , vol., no., pp.1,5, Aug. 31 2006-Sept. 1 2006
2. Hatakeyama, K.; Tsumura, S.; Kuribayashi, S.-i., "Fair joint multiple resource allocation method in all-IP networks," Communications, 2008. APCC 2008. 14th Asia-Pacific Conference on , vol., no., pp.1,4, 14-16 Oct. 2008
3. Shin-ichi Kuribayashi, "Optimal Joint Resource Allocation Method For Cloud Computing Environments", International Journal of Research and Reviews in Computer Science (IJRRCS), Vol. 2, No. 2, March 2011
4. Yuuki Awano and Shin-ichi Kuribayashi, "Proposed Joint Multiple Resource Allocation Method for Cloud Computing Services for Cloud Computing Services with Heterogeneous QoS", 3rd International Conference on Cloud Computing, 2012 pp. 1-6 July 2012.
5. Awano, Y.; Kuribayashi, S.-I., "A joint multiple resource allocation method for cloud computing environments with different QoS to users at multiple locations," Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on , vol., no., pp.1,5, 27-29 Aug. 2013.
6. Soumya Ray, Ajanta De Sarkar, ―Resource Allocation Scheme in Cloud Infrastructure‖, 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies.
7. Bernardetta Addis, Danilo Ardagna, Barbara Panicucci, Mark S. Squillante, Li Zhang, ―A Hierarchical Approach for the Resource Management of Very Large Cloud Platforms‖, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 10, NO. 5, SEPTEMBER/OCTOBER 2013.
8. Chrysa Papagianni, Aris Leivadeas, Symeon Papavassiliou, Vasilis Maglaris, Cristina Cervello-Pastor, Alvaro Monje, ―On Optical Allocation of Virtual Resources in Cloud Computing Networks‖, IEEE TRANSACTIONS ON COMPUTERS, VOL. 62, NO. 6, JUNE 2013.
9. Aceto, Giuseppe; Botta, Alessio; de Donato, Walter; Pescape, Antonio, "Cloud monitoring: Definitions, issues and future directions," *Cloud Networking (CLOUDNET), 2012 IEEE 1st International Conference on*, vol., no., pp.63,67, 28-30 Nov. 2012 doi: 10.1109/CloudNet.2012.6483656
10. Dhingra, M.; Lakshmi, J.; Nandy, S. K., "Resource Usage Monitoring in Clouds," *Grid Computing (GRID), 2012 ACM/IEEE 13th International Conference on* , vol., no., pp.184,191, 20-23 Sept. 2012 doi: 10.1109/Grid.2012.10
11. N. Krishnaveni, G. Shivakumar, ―Survey on Dynamic Resource Allocation Strategy in Cloud Computing Environment‖, International Journal of Computer Applications Technology and Research, Volume 2 – Issue 6, 731-737, 2013.
12. Pawar, C.S.; Wagh, R.B., "Priority based dynamic resource allocation in Cloud computing with modified waiting queue," *Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on* , vol., no., pp.311,316, 1-2 March 2013 doi: 10.1109/ISSP.2013.6526925
13. Shamsollah ghanbari, Mohamed Othman. "A Priority based Job Scheduling in Cloud Computing," International Conference on Advances Science and Contemporary Engineering 2012, Procedia Engineering 50 (2012) 778 – 785.
14. Vinothina, V., Dr R. Sridaran, and Dr PadmavathiGanapathi. "A survey on resource allocation strategies in cloud computing." International Journal of Advanced Computer Science and Applications 3.6 (2012).