A REVIEW ON AN EFFICIENT TASK ALLOCATION STRATEGY IN CLOUD COMPUTING

Dhrupal Pargi[1], Mohammed H Bohara[2],

[1]PG Student, Computer Engineering, PIET Vadodara, Gujarat, India
[2] Assistant Professor, CSE Department, PIET Vadodara, Gujarat, India

ABSTRACT

Cloud Computing is the use of computing resources like hardware and software that are use as a service over a internet to the end users. In Cloud Computing multiple users can request number of cloud services, so there must be a technique that all resources are made available to user in efficient manner. In Cloud Computing there are many jobs required to be executed by the available resources to meet best performance, minimal total execution time, shortest response time, utilization of resources and etc. so we need an efficient algorithm for task scheduling in cloud environment. In this various task allocation strategy is discussed.

Keyword : - *Cloud Computing, Task Scheduling, Virtual Machines*

1. INTRODUCTION

Cloud Computing refers to application and service that run on distributed network using virtualized resources and accessed by common internet protocol and networking standards. Cloud computing is a specific style of computing where everything from computing power to infrastructure, business apps are provided as a service. The basic principle of cloud computing is that user data is not stored locally but is stored in the data center of internet. The companies which provide cloud computing service could manage and maintain the operation of these data centers. The users can access the stored data at any time by using Application Programming Interface (API) provided by cloud providers through any terminal equipment connected to the internet[7]. The characteristics of cloud is measured service, on demand self service, resource pooling, network access.

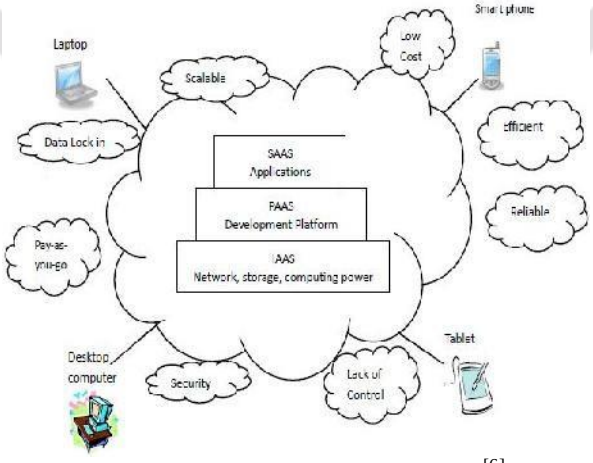

Fig -1: Overview of cloud computing[6]

Cloud Computing Provides service model, service model refers to different type of application or service provided by different vendors across cloud. Services in cloud are of three types, Infrastructure as a service(IaaS), Platform as a Service(PaaS), Software as a service(SaaS). And Deployment models are Public Cloud, Private Cloud, Hybrid Cloud.

1.1 Service Model

Infrastructure-as-a-service (IaaS*):* IaaS Provides virtual machines, virtual storage, virtual infrastructure, and other hardware assets as resources that clients can provision. The IaaS service provider manages all the infrastructure, while the client is responsible for all other aspects of the deployment. Infrastructure as a service is provided by Amazon AWS, Go Grid etc[8].

Platform-as-a-service (PaaS): PaaS provides virtual machines, operating systems, applications, services, development frameworks, transactions, and control structures. The client can deploy its applications on the cloud infrastructure or use applications that were programmed using languages and tools that are supported by the PaaS service provider. The service provider manages the cloud infrastructure, the operating systems, and the enabling software. The client is responsible for installing and managing the application that it is deploying. Google AppEngine is an example of Platform as a Service[8].

Software-as-a-service (SaaS): SaaS is a complete operating environment with applications, management, and the user interface. In the SaaS model, the application is provided to the client through a thin client interface (a browser, usually), and the customer's responsibility begins and ends with entering and managing its data and user interaction. Everything from the application down to the infrastructure is the vendor's responsibility[8].

1.2 Deployment Model

Public Cloud: Public cloud, facilities are providing on the Internet. These services are open for public usage, in which cloud services may be free or offered on a pay as you use manner[9].

Private cloud: In private cloud infrastructure is controlled and used by a single organization. Which is hosted either by externally or internally and managed by internally or by a third-party. The main advantage of private cloud is that the security, compliance and QoS are under the control of the enterprises[9].

Hybrid cloud: Hybrid cloud is an aggregation of more than two clouds i.e. public, private or community which have discrete objects however are bound together and achieving the advantages of multiple models. Main benefit of hybrid cloud is to extend the capabilities of the enterprise to deliver a specific business service through the addition of externally available public cloud services[9].

Community cloud: Community cloud shares resources between different enterprises from a specific community with common purpose (security, compliance, jurisdiction etc.). It is controlled by a third party or internally, and either hosted externally or internally. The cost spread over fewer users than a public cloud (but more than a private cloud)[9].

2. INTRODUCTION ABOUT TASK SCHEDULING

The scheduling of tasks in cloud means choose the best suitable resource available for execution of tasks or to allocate computer machines to tasks in such a manner that the completion time is minimized as possible. Task scheduling plays a key role to improve flexibility and reliability of systems in cloud. The main reason behind scheduling tasks to the resources in accordance with the given time bound, which involves finding out a complete and best sequence in which various tasks can be executed to give the best and satisfactory result to the user[7].

Two type of scheduling can be done. 1)Static scheduling schedule tasks in known environment i.e. it already has the information about complete structure of tasks and mapping of resources before execution, estimates of task execution/running time. 2) Dynamic scheduling must depend on not only the submitted tasks to cloud environment but also the current states of system and computer machines to make scheduling decision. Task scheduling is an important issue which greatly influences the performance of cloud computing environment. For providing Task scheduling different types of VM are needed. The cloud service provider and clients have different objectives and

requirements. Task scheduling is again new significant issues which affect the performance of cloud computing environment. The number of tasks in the cloud is huge and the system is dealing with massive tasks all the time, so it is difficult to handle and manage. Because of the high virtualized service provided by the cloud it is necessary to do scheduling. There are many algorithm and techniques available to do an efficient scheduling and improve the result and the performance of the cloud environment.

3. EXISTING SCHEDULING APPROACH

Improved Max-Min heuristic model for task scheduling in cloud[1]
In this paper, a simple modification of Max-min algorithm is proposed. This algorithm is built based on RASA algorithm and the concept of Max-min strategy. An Improved Max-min algorithm is developed to outperform scheduling process of RASA in case of total complete time for all submitted jobs. Proposed Max-min algorithm is based on expected execution time instead of complete time. So the scheduling tasks within cloud environment using Improved Max-min can achieve lower makespan rather than original Max-min[1].

Dynamic Resource Allocation scheme in cloud computing [2] .
This paper focuses on allocation of VM to the user, based on analyzing the characteristics of the job. Main principle of this work is that low priority jobs should not delay the execution of high priority jobs and to dynamically allocate VM resources for a user job within deadline. This method avoids creation of new virtual for the execution of newly arrived job. The method has less overhead in executing all jobs, when compared with creation of new VM[2].

Cloud Task Scheduling based on Load Balancing Ant Colony Optimization [3]
In this paper, a Load Balancing Ant Colony Optimization (LBACO) algorithm is Proposed to find the optimal resource allocation for each task in the dynamic cloud system. Not only does it minimize the makespan of a given tasks set but it also adapts to the dynamic cloud computing system and balance the entire system load [3]

Priority Based Dynamic Resources Allocation in Cloud Computing with Modified Waiting Queue [4]
In this paper propose an algorithm which considered preemptable task execution and multiple SLA parameters such as memory, network bandwidth, and required CPU time. In this paper author focus on dynamic resource provisioning. In order to achieve goal by pre-empting the current executing task having low priority with high priority task and if preemption is not possible due same priority then by creating new VM form globally available resources. If global resource not available task will be placed in waiting queue. When appropriate VM become free that advanced reservation task will be selected from waiting queue and allocated for execution to that Virtual machine[4].

Application of selective Algorithm for effective resource provisioning in cloud computing environment [5]
In this paper author discuss a selective algorithm for allocation of cloud resources to end-users on-demand basis. This algorithm is based on min-min and max-min algorithms. These are two conventional task scheduling algorithm. The selective algorithm uses certain heuristics to select between the two algorithms so that overall makespan of tasks on the machines is minimized. To achieve goal for minimizing the overall makespan of tasks on machines and provide better quality of service author design an algorithm that assigns tasks to best machines in such a way that it provides satisfactory performance to both, cloud users and providers[5].

4 CONCLUSION

In this paper, we analysis the Task Scheduling Approaches for mapping of task and resources in accordance with some certain principles for achieving the desired goal. The number of task in the cloud is huge and every task should deal with proper VM, so it is difficult to manage. Because of high virtualized service provided by cloud it is necessary to do task scheduling. A brief introduction of the task scheduling technique is discussed in this paper. Further the more efficient algorithm can be developed for efficient use of virtual machines.

6. REFERENCES

[1]. S. Devipriya,C. Ramesh. "Improved Max-Min heuristic model for task scheduling in cloud". IEEE, 2013
[2]. Saraswathi AT a, Kalaashri.Y.RA b, Dr.S.Padmavathi. "Dynamic Resource Allocation Scheme in cloud computing". Elsevier. Procedia Computer Science 47(2015)30-36.
[3]. Kun Li, Guangyu Zhao, Yushuang Dong, Dan Wang, "Cloud Task Scheduling Based on Load Balancing Ant Colony Optimization", IEEE, 2011
[4]. Chandrashekhar S Pawar, Rajnikant B wagh, "Priority Based Dynamic Resource Allocation in Cloud computing with Modified waiting Queue", IEEE, 2013
[5]. Mayanka Katyal, Atul Mishra, "Application of Selective Algorithm for Effective resource Provisioning in Cloud Computing Environment", IJCCSA, 2014
[6]. Hamid Shoja, Hossein Nahid, Reza Azizi, "A comparative survey on load balancing algorithm in cloud computing". IEEE, 2014
[7]. Dr. R Sridaran, Dr Padmavathi Ganapathi. "A Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Advance Computer Science and Application(IJACSA), 2015
[8]. Barrie Soinsky, "" Cloud Computing Bible", Wiley Publishing.Inc
[9]. Rajkumar Buyya, James Boberg, Andrej Goscinski, " Cloud Computing Principle and Paradigms"