

A Review On various Resource Scheduling for efficient cloud services

¹Kuldeep Kaur,²Rupali Zakhmi
Department of Computer Science Engineering
^{1,2}SVIET, Banur, Punjab

Abstract: As cloud computing is growing rapidly, efficient task scheduling algorithm plays a vital role to improve the resource utilization and enhance overall performance of the cloud computing environment. However, task scheduling is the severe challenge needed to solve urgently in cloud computing. Therefore, the simulated annealing multi-population genetic algorithm (SAMPGA) is proposed for task scheduling in cloud computing, which is the combination of simulated annealing algorithm (SA) and multi-population genetic algorithm (MPGA) in this paper. In population initialization, SAMPGA adopts max-min algorithm to enhance the search efficiency. SA incorporated into SAMPGA is employed to avoid local optimum and improve the performance of global optimum, while a family evolution strategy based on adaptive mechanism in MPGA is proposed to find better solution and improve convergence speed. Finally, experiments are conducted to evaluate the efficiency of the proposed method in MATLAB. Compared with MPGA, SA and simulated annealing genetic algorithm (SAGA), the results of simulation show that the SAMPGA has more excellent performance in terms of the completion time, completion cost, convergence speed and degree of load imbalance.

Keyword: MAX-MIN, RASA, VMs

I. INTRODUCTION

Cloud computing is based on sharing of resources to achieve cost effective system also the economies of scale. Now a day's cloud computing is becoming popular. Because it provides services to large number of users with cost benefits. There are vast variety of applications which keeps large amount of data also generates large amount of data in every second. so cloud computing provides large scale memory utility for these type of clients. They can store large data on remote repository. Also there will be shared processing at the cloud. For processing purpose client also can request to the cloud virtual machine. Cloud while doing sharing services is vulnerable to various types of utility problems. Like Quality of service, Load Balance and other performance issues. These issues directly effects the user consumption of resources controlled by cloud infrastructure[1]. In cloud scheduling for effective dealing with the requests by the users as well as to the allocation of the resources to the clients there are various scheduling algorithms like Min-Max, Min-Min and RASA etc.

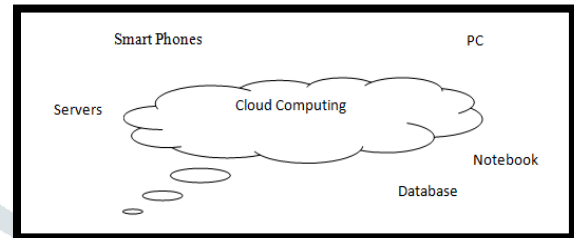


Fig. 1 Cloud architecture

A cloud is highly unpredictable environment. There are large amount of complexities that are involved in accurate information on the state of the system. There are large number of resources which requires highly managed system for better utilization of the resources. There are various factors which effects the performance of the cloud are Time, Cost, functionality etc. There are highly range based max and min utilization of the resources. There highly pre planned way which plan all the issues regarding resource utilization at Max Load. There is a planned spike and other is unplanned spike in workload. For the first the situation can be predicted. But in unplanned there is unpredictable nature. For the first type of situation the prediction is done in advance. For the second there should be a policy of scheduling which make ensure that the resources are allocated on demand and taken back immediately when there utilization will be completed. That means the scheduling algorithms which are viable in current time is of highly dynamic in nature and also there requires better problem solver compared to the older ways of scheduling of the resources. The general policies to be considered in cloud resource management are Admission control: takes decision whether to admit a job/request to be processed in the cloud, Resource allocation: provisions Virtual Machines (VMs) onto Physical Machines (PMs) and jobs onto VMs, Quality of Service (QoS): refers to metrics like response time, operational cost, throughput, maximization of profit and so on, Workload balancing: load balancing of jobs between the resources so as to improve its utilization, Energy Management: refers to optimized use of energy in the datacenter [1]. Resource allocation in cloud can be classified into two types:

First is Mapping of Virtual Machines (VMs) onto Physical Machines: Resources of cloud include the software and hardware required to execute user workloads. Examples of such resources are memory, CPU, bandwidth,

storage and network. Resource allocation is the process of allocating optimal resources to the jobs requested by the user, so these jobs can be processed efficiently. In a cloud environment, resource allocation generally means allocating a Virtual Machine satisfying the configurations specified by the user. The configurations include the operating system, MIPS, network bandwidth, storage and so on. This method of allocation can be referred as mapping of VMs onto Physical Machines[3].

Second is Mapping of Workloads onto VMs: there is the case when the cloud has fixed resources and the virtual machines. The resources are in the form of memory, CPU and bandwidth. While user submit its request for the processor and resource utilization. This request is of varying time and deadline based. These workloads need to be allocated to the optimal resources such that the workloads are processed efficiently. This type of allocation is referred as mapping of workloads onto VMs. This article presents a discussion on various issues and challenges of resource allocation in cloud computing. Research issues include resource provisioning, job scheduling, resource overbooking, scalability, pricing, load balancing, multi-tier applications, availability, overheads in network I/O workloads and Quality of Service (QoS) constraints. Open challenges in resource management for cloud are also listed.

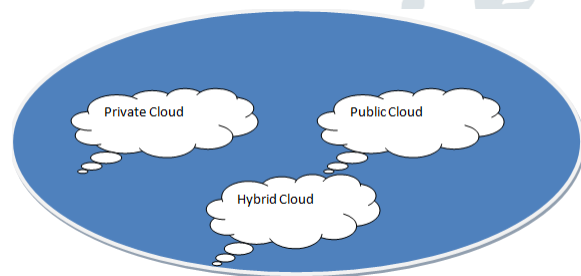


Fig. 2 Types of Clouds

1.2 Features of Cloud

- **Worldwide Access.** Cloud computing increases mobility, as you can access your documents from any device in any part of the world. For businesses, this means that employees can work from home or on business trips, without having to carry around documents. This increases productivity and allows faster exchange of information. Employees can also work on the same document without having to be in the same place.
- **More Storage.** In the past, memory was limited by the particular device in question. If you ran out of memory, you would need a USB drive to backup your current device. Cloud computing provides increased storage, so you won't have to worry about running out of space on your hard drive.
- **Easy Set-Up.** You can set up a cloud computing service in a matter of minutes. Adjusting your individual

settings, such as choosing a password or selecting which devices you want to connect to the network, is similarly simple. After that, you can immediately start using the resources, software, or information in question.

- **Automatic Updates.** The cloud computing provider is responsible for making sure that updates are available – you just have to download them. This saves you time, and furthermore, you don't need to be an expert to update your device; the cloud computing provider will automatically notify you and provide you with instructions.
- **Reduced Cost.** Cloud computing is often inexpensive. The software is already installed online, so you won't need to install it yourself. There are numerous cloud computing applications available for free, such as Drop box, and increasing storage size and memory is affordable. If you need to pay for a cloud computing service, it is paid for incrementally on a monthly or yearly basis. By choosing a plan that has no contract, you can terminate your use of the services at any time; therefore, you only pay for the services when you need them[5].

1.3 Scheduling Process in Cloud

The main advantage of job scheduling algorithm is to achieve a high performance computing and the best system throughput. The available resources should be utilized efficiently without affecting the service parameters of cloud. Scheduling process in cloud can be categorized into three stages they are Resource discovering and filtering, Resource selection, and Task submission. In resource discovery datacenter broker discovers the resources present in the network system and collects status information related to them. During resource selection process target resource is selected based on certain parameters of task and resource. Then during task submission task is submitted to the selected resource.

Cloud computing relies on sharing of resources to achieve coherence and economies of scale similar to a utility over a network. Cloud Computing is getting popular every day. Cloud service providers provide services to large scale cloud environment with cost benefits. Also, there are some popular large scaled applications like social networking and internet commerce. These applications can provide benefit in terms of minimizing the costs using cloud computing. Cloud computing is considered as internet based computing service provided by various infrastructure providers based on their need, so that cloud is subject to Quality of Service(QoS), Load Balance(LB) and other factors which have direct effect on user consumption of resources controlled by cloud infrastructure .In Cloud scheduling process need to achieve several factors. So it needs to use the effective algorithm for allocating proper task to the

proper resources. Various task scheduling algorithms has been proposed, most important task

scheduling algorithms are Min-min, Max-min, RASA[1], etc.

1.4 Scheduling Process In Cloud

The main advantage of job scheduling algorithm is to achieve a high performance computing and the best system throughput. The available resources should be utilized efficiently without affecting the service parameters of cloud. Scheduling process in cloud can be categorized into three stages they are Resource discovering and filtering, Resource selection, and Task submission [1]. In resource discovery datacenter broker discovers the resources present in the network system and collects status information related to them. During resource selection process target resource is selected based on certain parameters of task and resource. Then during task submission task is submitted to the selected resource.

II. LITERATURE SURVEY

[1] **Xing Jia Wei (2017) et. al:** Cloud computing is growing at fast rate. Where number of users which need the resources from the cloud is growing. That means there is a pressure on the cloud resources. To overcome this better utilization, scheduling technique is required. In current research paper the author has put SAMPGA based technique. It is based on two types of issues like simulating annealing and multi population Genetic algorithm. This paper has used Max-Min type of technique for increase in the search mechanism.

[2] **Shuibing He (2016) et. al:** In this paper, they have considered to improve scientific workflows in cloud environments where data transfers between tasks are performed via provisioned in-memory caching as a service, instead of relying entirely on slower disk-based file systems. However, this improvement is not free since services in the cloud are usually charged in a “pay-as-you-go” model. To further show the values of this concept, we also implement these algorithms and apply them, through a simulation study, to improve deadlock resolutions in workflow-based workloads when memory resources are constrained.

[3] **S.Devipriya(2013) et al:** Cloud computing is the use of computing resources that are delivered as a service over a network. It supplies a high performance computing based on protocols which allows shared computation and storage over long distances. In cloud computing, many tasks need to execute at a time by the available resources in order to achieve better performance, minimum completion time, shortest response time, resource utilization etc.

[4] **Hitoshi Matsumoto (2011) et al:** Two mechanisms of cooperative PSO and CPSO are analyzed and the load-balance requirement of equipotent CPSO mechanism was

discussed. Then the CPSO load-balance architecture was set up, control parameters were chosen, and the criterion of PSO convergence degree was established. Finally, the control strategy for CPSO’ load-balance was proposed. Two tests show that the proposed technique improved the CPSO in precision and efficiency.

[5] **Shaminder Kaur (2012) et al.** Discussed cloud computing is the use of computing resources that are delivered as a service over a network. In cloud computing, many tasks need to execute at a time by the available resources to achieve better performance, minimum completion time, shortest response time, resource utilization etc . Because of these different factors, an Improved Max-min algorithm is designed to outperform scheduling process of RASA in case of total complete time for all submitted jobs. So the scheduling tasks within cloud environment using Improved Max-min can achieve lower make span rather than original Max-min.

[6] **Rajesh Piplode (2012) et al:** An optimal power flow model was established for Available Transfer Capability (ATC) under the static security constraints. The maximum active power of all load nodes in receiving area was taken as objective function. To aim at the low accuracy and premature convergent in ATC optimization algorithms, the chaos cloud particle swarm algorithm based on golden section evaluation criteria (CCGPSO) was proposed. This method divided the particle swarm into standard particle, chaos cloud particle and cloud particle, which used the golden section judge principle according to fitness level. Every sub-swarm particle had respective different algorithm operations.

[7] **Vignesh V (2013) et al.** Discussed resource management is the primary issue as the demand grows for provisioning resources and computation in cloud systems. It presents various research issues pertaining to the management of cloud resources while a comparison is made between existing resource allocation systems. The issues and challenges discussed are resource provisioning, job scheduling, load balancing, scalability, pricing, energy management and availability.

[8] **Hongsheng Su (2013) et al.** Literature meaning of cloud computing is distributed computing, storing, sharing and accessing data over the Internet. It provides a pool of shared resources to the users available on the basis of pay as you go service, means users pay only for those services which are used by him according to their access times. The data processing and storage amount is increasing quickly day by day in cloud environment. This leads to an uneven distribution of overall work on cloud resources. So a proper balance of overall load over the available resources is a major issue in cloud computing paradigm. Load balancing ensures that no single node will be overloaded and used to distribute

workload among multiple nodes. It helps to improve system performance and proper utilization of resources. It also minimizes the time and cost involved in such big computing models. Load balancing and better resource utilization is provided by many existing algorithms.

III. CONCLUSION

As number of the users of the cloud environment is growing the resources are becoming scarce. Multiple requests need to be addressed with less amount of resources. Under such condition a efficient technique will be required which can schedule the requests and the resources. So that their efficient utilization can be ensured. cloud efficiency will depends upon this issue that how well cloud schedules the resources amongst different processes. MAX-MIN and MIN-MIN individually are not so efficient because there may be various longer or even shorter tasks. For optimization of the selection process in current research we have used PSO on Improved Max-Min. This technique identifies the best possible resources amongst the multiple available resources.

IV. FUTURE WORK

In current study of various research papers MAX-MIN and MIN-MIN scheduling algorithms are being used for resource efficiency. In future some optimization based technique can be used for efficient utilization of the resources.

REFERENCES

- [1] Xing Jia Wei, Wang Bei, Li Jun," SAMPGA Task Scheduling Algorithm in Cloud Computing",IEEE,Vol. 4,PP:5633-5640, 2017.
- [2] Shuibing He, Yang Wang, Xian-He Sun IEEE Fellow, and Chengzhong Xu," Using MinMax-Memory Claims to Improve In-Memory Workflow Computations in the Cloud", 1045-9219 (c) 2016 IEEE.
- [3] S.devipriya, c.ramesh," improved max-min heuristic model for task scheduling in cloud", 978-1-4673-6126-2/13 [1] hitoshi Matsumoto, Gavriil Tzortzakis, and Alex Delis, "dynamic Resource Management in cloud environment". (Elsevier Press). pp. 41–85. 2010.
- [4] Hitoshi Matsumoto,Mala Kalra,,"Cloud Computing Security Issues in Infrastructure as a Service", IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS), (Guilin), pp. 673-675, 2012.
- [5] Shaminder Kaur, Elina Pacini& Carlos Garc Garino, An Efficient Approach to Genetic Algorithm for Task Scheduling in Cloud Computing Environment, (2012).
- [6] Rajesh Piplode, and M. Naghibzadeh, " An Overview and Study of Security Issues & Challenges in Cloud Computing" The Third IEEE/IFIP International Conference on Internet, (Uzbekistan), 2012.
- [7] Vignesh V, resource management and scheduling in cloud environment, Cloud Computing and Distributed Systems Laboratory, Department of Computer ,2013
- [8] Hongsheng Su "The Available Transfer Capability Based on a Chaos Cloud Particle Swarm Algorithm ", IEEE Sixth International Conference on Grid and Cooperative Computing, 2007. GCC 2007, (Los Alamitos), CA, pp. 221-227, 2007
- [9] Srushti Patel and Pranati Mishra, " A Survey of Resource Allocation Policies in Cloud Computing", International Journal of Computer Science and Information Technologies, pp. 416-419, Vol. 4(3), 2013.
- [10] S.DEVIPRIYA, Han Ruilian. Study on cloud computing task schedule strategy based on MACO algorithm. Computer Measurement & Control, Vol.19 (5): 1203~1204, 1211, 2013.
- [11] Shamsollah Ghanbaria& Mohamed Othmana, A Priority based Job Scheduling Algorithm in Cloud Computing, Procedia Engineering 50, and pp. 778 – 785, 2012.
- [12] S.Thamarai Selviand R. Eberhart, Resource Allocation Issues and Challenges in Cloud Computing, vol. 4, pp 1942–1948, 2014.
- [13] yousef fahim, Sarabjit Singh, and Meenakshi Sharma," load balancing improvement of a data center by a hybrid algorithm in cloud computing", World Academy of Science, Engineering and Technology, 2014.