IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 4, April 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

Job Scheduling in Cloud Computing Using Genetic Algorithm

Jasdeep Kaur [1], Kailash Bahl[2]

Jasdeep kaur[1], Research Scholar, Department of Computer Science . P.I.E.T, Patiala, India
Kailash[2], Professor, Department of Computer Science and Engineering, P.I.E.T, Patiala, India

ABSTRACT

Cloud computing is one of device technology trends in the future since it combines the advantages of both device computing and cloud, Recent years have seen the massive migration of enterprise applications to the cloud. Cloud computing used in business organizations and educational institutions. One of the challenges posed by cloud applications is Quality-of-Service (QoS) management, which is the problem of allocating resources to the application to guarantee a service level along dimensions such as performance, availability and reliability. To improve the QoS in a system one must need to reduce the waiting time of the system. Genetic Algorithm (GA) is a heuristic search technique which produces the optimal solution of the tasks. This work produces one scheduling algorithm based on GA to optimize the waiting time of overall system. The cloud environment is divided into two parts mainly, one is Cloud User (CU) and another is Cloud Service Provider (CSP). CU sends service requests to the CSP and all the requests are stored in a Request Queue (RQ) inside CSP which directly communicates with GA Module Queue Sequencer (GAQS). GAQS perform background operation, like daemon, with extreme dedication and selects the best sequence of jobs to be executed which minimize the Waiting time (WT) of the tasks using Round Robin (RR) scheduling Algorithm and store them into Buffer Queue (BQ). Then the jobs must be scheduled by the Job Scheduler (JS) and select the particular resource from resource pool (RP) which it needs for execution.

Keywords: *Genetic Algorithm, Cloud computing, Cloud Service Provider, Request Queue, GA Module Queue Sequencer, Buffer Queue, Waiting Time, Round Robin Scheduling Algorithm, , Resource Pool.*

1. INTRODUCTION

Cloud computing, often referred to as simply "the cloud," is the delivery of on-demand computing resources, everything from applications to data centers, over the Internet on a pay-for-use basis. Cloud computing environment is highly dynamic; the system load and computing resource utilization exhibit a rapidly changing characteristic over time. Therefore Cloud service provider normally over-position computing resources to accommodate the peak load and computing resources are typically left under-utilize in nonpeak time. Cloud environment allows users to use applications without installation and access their personal files at any computer with Internet access. End users access cloud based applications through a web browser or a light weight desktop while the business software and data are stored inside CSP at a remote location. Cloud application providers strive to give the better service and performance than if the software programs were installed locally on end- user machines. Cloud environment is used in lot of fields like in IT industries, educational institute as well as in other industries. In this paper we have proposed Cloud Service Provider, figure 1, which includes mainly three parts- GA Module Queue Sequencer, Job Scheduler (JS) and Resource Pool (RP). All service requests which are coming from Cloud Users domain are stored in RQ which is in GAQS. Now the requested processes must communicate with GAQS processor (GAP) and the processor finds out the appropriate sequence of tasks which reduce the waiting time of the tasks. GAQS processor then communicate directly with JS which schedules the tasks using Round Robin scheduling algorithm and communicate with RP and tries to assign each of these jobs as per their requirement to the resources.but the main problem here is that to find out the best sequence of the tasks from all possible sequence of tasks and JS schedules those tasks and optimize total waiting time of those jobs.The jobs assignment task is done by JS. So JS must need to assign the task such a way that assignments of the jobs to the resources must be fruitful as per as CU requests and the total execution time must be optimal of the whole operations . In next two sections discuss about our proposed model of CSP and one Genetic based

scheduling an algorithm which assigns the task to the resource as per the CU's demand and also to optimize the total waiting time of those tasks.

Scheduling Algorithms

There has been various types of scheduling algorithm exist in distributed computing system. Most of them can be applied in the cloud environment with suitable verifications. The main advantage of job scheduling algorithm is to achieve a high performance computing and the best system throughput. Traditional job scheduling algorithms are not able to provide scheduling in the cloud environments. According to a simple classification, job scheduling algorithms in cloud computing can be categorized into two main groups; Batch mode heuristic scheduling algorithms (BMHA) and online mode heuristic algorithms. In BMHA, Jobs are queued and collected into a set when they arrive in the system. The scheduling algorithm will start after a fixed period of time. The main examples of BMHA based algorithms are; First Come First Served scheduling algorithm (FCFS), Round Robin scheduling algorithm (RR), Min–Min algorithm and Max–Min algorithm.

By On-line mode heuristic scheduling algorithm, Jobs are scheduled when they arrive in the system. Since the cloud environment is a heterogeneous system and the speed of each processor varies quickly, the on- line mode heuristic scheduling algorithms are more appropriate for a cloud environment. Most fit task scheduling algorithm (MFTF) is suitable example of On-line mode heuristic scheduling algorithm.

a. First Come First Serve Algorithm:

Job in the queue which come first is served. This algorithm is simple and fast.

b. Round Robin algorithm:

In the round robin scheduling, processes are dispatched in a FIFO manner but are given a limited amount of CPU time called a time-slice or a quantum. If a process does not complete before its CPU-time expires, the CPU is pre-empted and given to the next process waiting in a queue. The preempted process is then placed at the back of the ready list.

c.Min–Min algorithm:

This algorithm chooses small tasks to be executed firstly,which in turn large task delays for long time.

c. Max – Min algorithm:

This algorithm chooses large tasks to be executed firstly, which in turn small task delays for long time.

d. Most fit task scheduling algorithm:

In this algorithm task which fit best in queue are executed first. This algorithm has high failure ratio.

e. Priority scheduling algorithm:

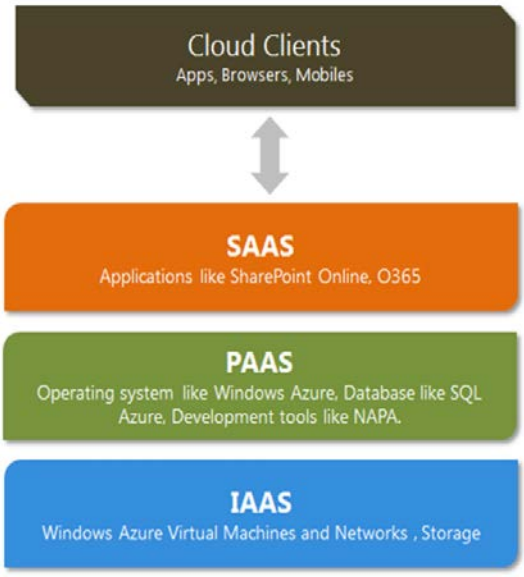
The basic idea is straightforward: each process is assigned a priority, and priority is allowed to run. Equal- Priority processes are scheduled in FCFS order. The shortest-Job-First (SJF) algorithm is a special case of general priority scheduling algorithm. An SJF algorithm is simply a priority algorithm where the priority is the inverse of the (predicted) next CPU burst. That is, the longer the CPU burst, the lower the priority and vice versa. Priority can be defined either internally or externally. Internally defined priorities use some measurable quantities or qualities to compute priority of a process.

Cloud Computing

Cloud computing, also known as 'on-demand computing', is a kind of Internet-based computing, where shared resources, data and information are provided to computers and other devices on-demand Cloud computing promises several attractive benefits for businesses and end users. Three of the main benefits of cloud computing include:

a. Infrastructure As A Service (IaaS)

In the most basic cloudservice model, providers of IaaS offer core IT services, for example, computer(physical or virtual machines) networks, compute, security, operating systems, middleware devices, load balancers and block and file or object storage. IaaS clouds regularly provide extra resources such as a virtualmachine disk image library, IP addresses, firewalls, virtual local area networks (VLANs) and other software bundles. IaaScloud providers provide these resources ondemand from their big pools installed in data centres. For widearea network connectivity, businesses can use either the Internet or carrier clouds (dedicated virtual private networks) that IaaS clouds offer. For the deployment of their applications, users install operatingsystem instances and their application software on the cloud infrastructure. The cloud user also patches and maintains the operating systems and the application software.



where the cost reflects the size of resources allocated and
consumed.

b. Platform As A Service (PaaS)

In the PaaS models, cloud providers provide a computing platform, usually including operating systems, programming language running environment,database and web servers.
Application developers can develop and run their software solutions on a cloud platform minus the cost and complexity of purchasing and managing the underlying hardware and software layers.As per some PAAS offers like Microsoft Azure and Google App Engine.the core computer and storage resources scale by design to meet application demand so that the cloud user does not need to allot resources manually.This has also been planned by an architecture targeting to facilitate realtime in cloud environment.Media encoding as an application can also be provided via Paas.
Paas platforms are not very common because service providers often cannot offer customers the control and variety that they need for their applications.Furthermore,vendor lockin is another concern.Once the application starts to use any proprietary tools or interfaces that a Paas provider makes accessible,migration to another provider may become challenging.

c**. Software As A Service (SaaS)**

In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud service providers manage the infrastructure and platforms that run the applications. SaaS is on occasions cited to as "ondemand software" and is typically rated on payperuse or on a subscription fee basis.

In the SaaS model, cloud providers manage the application software in the cloud and cloud users access the software via cloud clients. The cloud infrastructure and platform where the application runs is transparent to the cloud users. The need to install and run the application on the cloud user's own computers is not needed, which simplifies maintenance and support. Cloudbased applications are unlike other applications in their scalability, which is achieved by duplicating tasks onto several virtual machines at runtime to meet varying demand with load balancers distributing the work. The user sees only a single access point. Cloud applications can be multitenanted to accommodate a huge number of cloud users having any machine serving more than one cloud user organisation.The pricing model for SaaS applications is usually a monthly or yearly flat fee per user, consequently price is scalable and flexible if users are added or removed at any point.Supporters claim SaaS lets a business the potential to lessen IT operational costs by subcontracting hardware and software maintenance and support to the cloud provider. This facilitates the business to reapportion IT operations expenses away from hardware/software and staff costs. Besides when applications are hosted centrally, updates can be done without the need for users' intervention.One shortcoming of SaaS is that the users' data are stored on the cloud provider's server which could be subject to unauthorised access to the data. Hence, users are progressively implementing intelligent thirdparty key management systems to protect their data.

2. LITERATURE REVIEW

Various modifications to task scheduling in cloud computing and genetic algorithm have been proposed by several authors. These modifications can be classified as follows

In 2003,H.xiaoshan et al [1] suggested a QoS Guided Min-Min heuristic [Batch mode heuristic algorithm] was introduced in that some task require higher network bandwidth to exchange a large amount of data among processors, whereas some can be satisfied with the lower network bandwidth. In this algorithm the matching of the QoS request and services between the tasks and hosts based on conventional Min-Min. Firstly each task with the high QoS request in the Meta task, the algorithm finds the earliest completion time and the host that obtains it, in the entire QoS Qualified host. Secondly find the task with the minimum earliest completion time and assigns the task to the host that give the earliest completion time to task. In this algorithm they have addressed only one-dimension QoS issue, because they worked only bandwidth constraint

In 2006 F.Dong et al.. [2] proposed a QoS priority grouping algorithm which considers deadline and acceptation rate of the task and the makespan as main factor of task scheduling in whole system.
It achieves better acceptance rate and completion

time for the submitted task then Min- Min and QoS Guided Min-Min.

In 2008, C.Hsu et al [3] carried out two optimization schemes MOR (Makespan Optimization Rescheduling) and ROR (Resource Optimization Rescheduling). MOR focus on improving the makespan to pull off the better performance and in ROR focus on the redispatch tasks from the machine with the minimum number of tasks to other machine, which is helpful to reduce the resource need. Both this technique achieves low complexity, high effectiveness, good performance than QoS Guided scheduling algorithm and Min-Min algorithm.

In 2008, M.Singh et al [4] proposed a QoS based predictive Max-Min, Min-Min switcher algorithm. In this algorithm, scheduling of the next job is based on appropriate selection among QoS based min-min or QoS max-min algorithm. The effect on the execution time grid jobs has been reduced due to non-dedicated resources. It normally uses the history information about the execution jobs to predict the performance of non-dedicated resources. This algorithm merges the efficiency of max-min along with min-min and also considers both QoS and non-dedicated property of grid resources.

In 2009, S.Parsa et al [5] introduced a new task scheduling algorithm called RASA which has the advantage of both Min-Min and Max-Min algorithm. In this first estimate the completion time of the tasks on each resource and then applied both the algorithm. RASA use the Min- Min strategy to execute the small task first then long task and then applied Max-Min to avoid the delays in the execution of large task and support concurrency in the execution of the large and small tasks. It achieves the lower Makespan with good QoS

In 2011, C.Zhao et al [6] proposed a Berger Model in Cloud computing in that algorithm scheduling process establish dual fairness constraint. First constraint is to classify user task by QoS preferences, and establish the general expectation function in accordance with the classification of tasks to restrain the fairness of the resources in the selection process. Second constraint is to define resource fairness justice function to judge the fairness of the resources allocation. According to constraint, the algorithm always assigns tasks on the optimal resources in order to satisfy the QoS requirement of user and it avoid to consider a long task for execution. Experiment result of this algorithm shows effective execution of the user tasks and manifest better performance.

In 2013, X.Wu et al [7] introduce a task scheduling algorithm based on QoS-driven in cloud computing (TS-QoS). In this TS-QoS algorithm compute the priority of the task according to the special attributes of the tasks, and then sort tasks based on priority. Then the algorithm calculate the completion time of each task on different services.
But in this process priority can change dynamically an increase continuously this can help to solve the ―starvation‖ problem and follow FCFS principle. Experimental result achieves well performance and load balancing by QoS driving form both priority and completion time.

3. PROPOSED MODEL

Before starting to discuss about Cloud queueing model first we discuss about the Genetic algorithm and then site our proposed scheduling algorithm based on Genetic algorithm.

Genetic algorithms (GA) were first proposed by the John Holland in the 1960s The GA is a heuristic search technique that simulates the processes of natural selection and evolution. Genetic algorithm (GA) is a promising global optimization technique.

It works by emulating the natural process of evolution as a means of progressing towards the optimal solution. A genetic algorithm has the capability to find out the optimal job sequence which is to be allocated to the processor.
General Algorithm perform its general operations using the following steps-

- Select the fixed size chromosomes from the from the population set.
- Perform any one type encoding operation on the chromosomes of the chromosome sets.
- Select the best two chromosomes from the chromosome set using their fitness value.
- Perform the crossover between two chromosomes and get two different offspring.
- Perform the mutation operation on those offspring just interchanging the bit positions.
- Continue the steps A to B until get the best solution of the population.
- Finally perform the elitism operation of the chromosomes means store the best chromosomes in to the system for future use.

REFERENCES

[1] XiaoShan He,Xianhe Sun and Gergor von Laszewski. QoS guided Min-Min heuristic for grid task scheduling. Journal of Computer Science and Technology, 2003, 18(4), p.442-451.

[2] Dong. F, Luo. J, Gao. L and Ge. L, "A Grid Task Scheduling Algorithm Based on QoS Priority Grouping," In the Proceedings of the Fifth International Conference on Grid and Cooperative Computing (GCC'06), IEEE, 2006.

[3] Ching-Hsien Hsu,Zhan. J.,Wai-Chi Fang, et al. ―Towards improving QoS-guided scheduling in grid Third ChinaGrid Annual Conference(CHINAGRID). Dunhuang, Gansu,China, 2008, p.89-95.

[4] M.Singh and P.K.Suri; ―QPSMax-Min<>Min-Min : A QoS Based Predictive Max-Min, Min-Min Switcher Algorithm for Job Scheduling in a Grid, ―International Technology Journal7(8) :p.1176-1181,2008

[5] Saeed Parsa and Reza Entezari-Maleki, RASA: A New Task Scheduling Algorithm in Grid Environment in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009. [8] Mrs. S.Selvarani1, Dr.G.Sudha Sadhasivam, improved cost-based algorithm for task scheduling in Cloud computing ―in IEEE 2010.

[6] Baomin Xu,Chunyan Zhao,Enzhao Hua,Bin Hu. ―Job scheduling algorithm based on Berger Model in Cloud Environment Advance in Engineering Software, 2011, 42(7), p.419-425.

[7] Xiaonian Wu, Mengqing Deng, Runlian Zhang, Bing Zeng, Shengyuan Zhou. ―A task scheduling algorithm based onQoS-driven in Cloud Computing Information Technology and Quantitative Management(ITQM2013).2013,p.1162-1169 [11] J. Kennedy and R. Eberhart, (1995), Particle swarms optimization In IEEE International Conference on Neural Networks, volume 4, pages 1942–1948.