Scheduling Algorithms in Cloud Computing

Damanbeer Kaur
M.Tech Scholar
Amritsar College of Engineering & Technology
Amritsar, India

Tejinder Sharma
Associate Professor
Amritsar College of Engineering & Technology

ABSTRACT

Cloud Computing has come to be perception for large scale of distributed computing and parallel processing. Cloud computing is a form of internet based computing that provides shared computer processing resources and data to computers and other devices on demand. The execution and suitability of cloud computing services always depends upon the completion of the user tasks affirmed to the cloud system. Task scheduling is one of the main types of scheduling performed. Scheduling is the major issue in establishing cloud computing system. The scheduling algorithms should order the jobs in a way where balance between improving the performance and quality of service and at the same time maintaining the efficiency and fairness among the jobs. This paper aims at studying various scheduling methods. A good scheduling technique also helps in proper and efficient utilization of the resources. Many scheduling techniques have been developed by the researchers like GA (Genetic Algorithm), PSO (Particle Swarm Optimization), Min-Min, Max-Min, Priority based Job Scheduling Algorithm

Keywords

Keywords: Cloud computing, heuristic, metaheuristic, FIFO, RR, ACO, PSO, GA, Cloud Scheduling.

1. INTRODUCTION

Cloud computing is an evolving technology. Cloud computing delivers infrastructure, platform, and software that are made available as subscription-based services in a pay-as-you-go model to consumers. In Cloud Computing the term Cloud is used for the service provider, which holds all types of resources for storage, computing etc. The cloud definition provided by the National Institute of Standards and Technology (NIST).The NIST cloud computing definition:"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[3]The reliability and performance of cloud services depends up on various factors like scheduling of tasks. Scheduling can be done at task level or resource level or workflow level. [7]

Scheduling is accomplished on the basis of different parameters so that it expanding the long-term cloud performance. A task may combine entering data, processing, accessing software, or storage functions. The data center divides tasks according to the service-level agreement and requested services. Each task is then authorizing to one of the convenient servers. In turn, the servers implement the requested task, and response, or result, is forward back to the user [20].The main goal of scheduling in clouds computing is to

- Improve the utilization of servers designate to the jobs.
- To process the job higher accept priority.
- Improve the resource utilization.
- Minimizes the completion time
- Minimizing the waiting time[7]

Cloud computing is greatly dynamic; therefore scheduling tasks and resources is limited challenging. Scheduling is the set of policies to authority the order of work to be performed by a computer system. The main benefits of scheduling algorithm are to attain a high performance computing and high system throughput. [3]

Scheduling instruct availability of CPU memory and gives maximum utilization of resource. Scheduling of task is based on the individual parameters. Scheduling problem is NP Hard. The objective of the scheduling algorithms in cloud computing environment is to employ the resources accurately while controlling the load between the resources so that to get the minimum execution time. [3]

Datacenters are essential parts of cloud computing environment. In a single datacenter commonly hundreds and thousands of virtual machines run at any instance of time, hosting many tasks and simultaneously the cloud system keeps accepting the batches of task requests. Throughout this process, few target machines out many powered on machines, can conclude a batch of incoming tasks. So Task scheduling is an essential issue which greatly influences the performance of cloud service provider [6]. One of the incitation of scheduling is for load balancing of the tasks and to ensuring availability of resources.

2. CLOUD COMPUTING MODELS

2.1 Software as a Service (SaaS)

It refers to giving an ability to the user to use the software and its functions on demand remotely through internet. Saas removes the responsibility of organizations such as set-ups, installation, maintenance and daily preservation [5].

2.2 Platform as a Service (PaaS)

This model can be described as application development environments offered by cloud provider as a – Service. It is giving the user capability to arrange his application onto cloud‟s infrastructure

2.3 Infrastructure as a Service (IaaS)

It is providing the infrastructure such as servers, hardware, storage, routers and the other networking modules to the users. According to requirement of user, he can use some or all of these infrastructure components and pay for what he have used only [5].

3. CLOUD DEPLOYMENT MODEL

3.1 Public Cloud:

A cloud is to be entitled as public cloud when the services are being provided over network that are available publically, anyone can access it[6]. In this manner, the foundation of an open cloud is shared between the clients.

3.2 Private Cloud:

The private cloud is more secure and costlier than open cloud. It is devoted to single associations to do its undertakings, it is works inside the association and overseen in the same organization

3.3 Hybrid cloud:

It is a combination of public and private cloud. It is useful when the organization have some critical data/applications which need high security to be stored in private cloud while others does not need high security can be stored in public cloud[6].

4. CLOUD COMPUTING CHARACTERISTICS

4.1 On Request Self Administrations:

A cloud may separately acquire processing liabilities, according to the utilization of various servers, arrange putting away, as when asked for, without speaking with cloud supplier.

4.2 Wide Network Access:

Administrations are dispatched over the Internet inside a standard system and access to the administrations is conceivable through various client apparatuses.

4.3 Asset pooling:

A numerous model is hired to serve distinctive sorts of customers by creating pools of various assets, according to the interest of clients these have diverse assets which can be doled out and reassigned powerfully.

4.4 Quick Versatility:

Abilities may be flexible, procured or expeditiously freed. From clients see, they gave conceivable outcomes turned out to be boundless and must have the capacity to buy in any amount at any time [4].

4.5 Measured Administrations:

The arrangement obtained by various customers is measurable. The utilization of benefit will be coordinated, evaluated, and charged for supporting and asset [5].

5. OVERVIEW OF TASK SCHEDULING

Task scheduling is a significant task in cloud computing environment. It is used to classified specific effort to specific resources in specific time. In the cloud environment, task scheduling is an enormous and challenging concern. The main goal of task scheduling is to upgrade the enforcement and quality of service and in the same manner maintaining the effectiveness and integrity among the tasks. It is also used to reduce the execution cost. An effective task scheduling approach must need to allow low response time so that the execution of submitted tasks acquire among a possible minimum time. There are several scheduling approaches that should get all these concern.

5.1 Scheduling Types

There are dissimilar kinds of scheduling based on different standards, such as static vs. Dynamic, centralized vs. Distributed, offline vs. Online etc.

1. *Static* Scheduling**:** Pre-Schedule jobs, all info recognized about obtainable resources and tasks and a task assigned once to a reserve, so it is calmer to adapt based on scheduler‟s outlook

2. *Dynamic Scheduling***:** Jobs dynamically exist for scheduling over time via the scheduler. It is more elastic than static scheduling, to be able of decisive run time in fee. It is more serious to include load equilibrium as a foremost factor to acquire stable, accurate and effective scheduler algorithm.

3. *Centralized Scheduling:* As stated in dynamic scheduling, it is an accountability of centralized / distributed scheduler to make worldwide choice. The main welfares of centralized scheduling are comfort of employment; efficacy and more control and nursing on resources. On the other hand, such scheduler lacks scalability, liability tolerance and effectual performance. Since of this disadvantage, it is not indorse for large-scale grids.

4. *Distributed /Decentralized Scheduling*: More of realistic for actual cloud not withstanding of its weak competence likened to unify scheduling. There is no vital control entity, so local schedulers‟ requests to achieve and uphold state of jobs queue.

5. *. Preemptive Scheduling:* This type allows all jobs to interject during implementation and a job can travelled to another resource sendoff its originally owed resource, available for other jobs. If restraints such as priority are careful, this type is more helpful.

6. *Non Preemptive Scheduling:* Scheduling process, in which capitals are not being allowed to be re-allocated until the consecutively and scheduled job ended its execution.

7. *Co-operative scheduling:* Here, system have already several schedulers, both one is responsible for execution certain action in scheduling process to common system wide range based on the cooperation of events, given rubrics and present scheme users.

8. *Immediate /Online Mode:* Here, scheduler schedules any freshly arriving job as soon as it reaches with no waiting for next time intermission on available resources at that instant .

9. *Batch / Offline Mode:* The scheduler supplies inward jobs as collection of problems to solve over consecutive time intervals, so that it is well to map a job for fit resources depending on its features.

5.2 Phases of Scheduling

In Cloud, Scheduling process can be divided into three stages as described below:

1. *Resource* discovering *and filtering* -Datacenter Broker finds the resources present in the network system and gathers status data related to them.

2. *Resource selection* - Target resource is chosen takinginto account certain parameters of task and resource. This is choosing stage

3. *Task submissions* - Task is submitted to resource chosen. [3]

5.3 Existing Task Scheduling Algorithms

In this paper we are classifying task scheduling into three different categories as aforementioned. They are heuristic, hybrid and energy efficient task scheduling algorithms.Each of this categorization is explained briefly in the coming sub sections with their sub classifications:

5.3.1 Heuristic Task Scheduling Approaches

Heuristics scheduling provides an optimal solution in which it uses the knowledge bases for taking the scheduling decisions. Heuristic approaches can be either static or dynamic. First we will look at the static scheduling algorithms [21], [27],[28].

5.3.1.1 Static Scheduling Methods:

Static scheduling algorithm considers that all tasks arrive at the same instant of time and they are independent of the system resources states and their availability. The static heuristics include the basic simple scheduling strategies like First Come First Serve and Round Robin methods.

➢ FCFS methods collects the tasks and queues them until resources are available and once they become available the are assigned to them based on their arrival time. It is less complex in nature but does not consider any other criteria for scheduling the tasks to machines.

➢ On the other hand RR method uses the same FIFO algo. technique for doing the scheduling of the tasks but it allots a resource for each task for a particular time quantum [25],[11].After that the task is pre-empted and queued until its next chance for execution. Opportunistic load balancing is another heuristic method of scheduling in which it tries to schedule the tasks to the next available machines based on their expected completion time. It will result in poor make span even though it tries to utilize the resources equally making all machines busy at the same time.[20]

➢ Minimum Execution Time and Minimum Completion Time are other two heuristic strategies in which MET assigns tasks on the machines based on which machine it takes less execution time. It selects the best machine for execution but do not consider the availability of resources at the time of scheduling so load imbalance will occur. Minimum Completion Time Algorithm selects machines for scheduling the tasks based on the expected minimum completion time of

tasks among all the machines available. It considers the load of the machine also before scheduling the task on that machine. The task may not have minimum execution time on the same machine. Completion time of a task on a machine can be defined as the sum of the execution time of the task on that machine and the ready time of that particular machine. Execution time is the actual time needed for executing a task. [20]

➢ Minimum Execution Time and Minimum Completion Time[21],[11] are other two heuristic strategies in which MET assigns tasks on the machines based on which machine it takes less execution time. It selects the best machine for execution but do not consider the availability of resources at the time of scheduling so load imbalance will occur. Minimum Completion Time Algorithm selects machines for scheduling the tasks based on the expected minimum completion time of tasks among all the machines available. It considers the load of the machine also before scheduling the task on that machine. The task may not have minimum execution time on the same machine. Completion time of a task on a machine can be defined as the sum of the execution time of the task on that machine and the ready time of that particular machine. Execution time is the actual time needed for executing a task.

➢ Min-Min and Max-Min[25],[21] are two other heuristic methods used for task scheduling. Min-min heuristic selects the smallest task first from all the available tasks and assigns it to a machine which gives the minimum completion time (fastest machine) for that task. It increases the total completion time of all the tasks and hence increases the makes pan. But it does not consider load of the machines before scheduling as simply assigning smaller tasks on faster machines. Here the expected completion time and execution time for a task are considered to be almost same values or close values. The long tasks have to wait for completing the execution of smaller ones. But the method improves the system’s overall throughput.

➢ Max-Min is similar to min-min except that it selects the longest task (with maximum completion time) first to schedule on the best machine available based on the minimum completion time of that particular task on all available machines. Here the smaller tasks have to starve and load balancing is also not considered. Anyway it increases the make span and system throughput than the min-min strategy since the longest task determines the make span of all the available tasks in the system. Hence in max-min the longer tasks can be executed first in faster machines as well as smaller tasks can be executed in parallel on other possible machines which results in better make span and balanced load than the previous method.[20].

➢ Genetic Algorithm and Simulated Annealing are two other general methods in heuristic approach which is used to perform near optimal scheduling. In Genetic Algorithm approach [23] we perform four different operations, evaluation, selection, cross over and mutation. The initial population represents the possible mappings of the given task list on the available machines. Each job is represented as a vector in which each position of that vector represents a task in the task list. The value in each position represents the machine to which the task is mapped. Each job represents a chromosome. Every chromosome has a fitness value indicating the overall execution time of all the tasks (makespan) which is formed from the mapping of tasks to resources constituting that chromosome and it is selected such that it reduces makespan. This method uses past results with present results to get better possible mappings and survival of the fittest takes place.

➢ Simulated Annealing is an iterative method which can be represented similar to genetic algorithm in which it starts with a single solution (mapping) selected from a random distribution. The initial version of SA is evaluated to get a better version. After mutation the new makespan is analyzed.

If it is lower (better) than the previous one then replace the old one with the new makespan. Simulated Annealing find poorer solutions than Genetic Algorithm. The features of genetic algorithm and simulated annealing can be combined to get a better scheduling solution. [24],[26]

5.3.1.2 Dynamic Scheduling Methods:

In dynamic scheduling methods [11] [25] tasks are dynamic in nature. Here tasks arrive at different points of time and it is dependent on the system machines state. Dynamic scheduling algorithms are classified into two categories:

(1) online mode and (2) batch mode
In online mode tasks are assigned instantly once they arrive in the system like most-fit task scheduling algorithm where as **In batch mode** tasks are collected as a group and scheduled at predefined times. Min-min, max-min, round robin are some examples for batch mode. MCT, MET, OLB belongs to online mode, and works similar to static algorithms.

➢ Switching algorithm is another algorithm in which it switches between MET and MCT as per the load of the system. K-Percent Best is another heuristic of same kind in which a subset of k computationally higher ranking machines is first selected during the scheduling process. A good value of k shows that it always assigns a task to a machine from this list only. This method leads to a better makespan compared to MCT. It preserves machines which are more suitable for yetto- arrive tasks.

➢ In batch mode along with max-min, min-min methods, and another heuristic is called sufferage heuristic in which the tasks are scheduled based on a sufferage value. It is calculated from the first and second earliest completion times of a task. The sufferage values are compared for different tasks and the task with higher sufferage is selected for scheduling on a same resource [11],[21].

6. METAHEURISTIC SCHEDULING ALGORITHM

A metaheuristic is a high level problem-independent algorithmic framework that provides a set of guidelines or strategies to develop heuristic optimization algorithms. Metaheuristic include genetic algorithms, Particle swarm Optimization (PSO), Ant Colony Optimization.

6.1 Particle Swarm Optimization (PSO) Algorithm:

Particle Swarm Optimization (PSO) as a meta-heuristics method is a self-adaptive global search based optimization technique introduced by Kennedy and Eberhart. The PSO algorithm is alike to other population-based algorithms like Genetic algorithms (GA) but, there is no direct recombination of individuals of the population. The PSO algorithm focuses on minimizing the total cost of computation of an application workflow. PSO balances the load on compute resources by distributing tasks to available resources.

6.2 Ant Colony Optimization (ACO):

An artificial Ant Colony System (ACS) is an agent-based system, which simulates the natural behavior of ants and develops mechanisms of cooperation and learning. ACS was proposed by Dorigo (Dorigo and Gambardella, 1997) as a new heuristic to solve combinatorial optimization problems. This new heuristic, called Ant Colony Optimization (ACO) has been found to be both robust and versatile in handling a wide range of combinatorial optimization problems. The main idea of ACO is to model a problem as the search for a minimum costpath in a graph.

6.3 Artificial Bee Colony (ABC) Algorithm:

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga for optimizing numerical problems in. The algorithm simulates the intelligent foraging behavior of honey bee swarms. It is a very simple, robust AND population based stochastic optimization algorithm. In ABC algorithm, the colony of artificial bees contains three groups of bees; employed bees, onlookers and scouts. A bee waiting on the dance area for making a decision to choose a food source is called onlooker and one going to the food source visited by it before is named employed bee. The other kind of bee is scout be that carries out random search for discovering new sources.

6.4 Firefly Optimization (FA):

Nature-inspired meta-heuristic algorithms, especially those based on swarm intelligence, have attracted much attention in the last ten years. Firefly algorithm appeared in about five-six years ago, its literature has expanded dramatically with diverse applications. Firefly algorithm was developed by in late Xin She Yang 2007 and 2008 at Cambridge University, Which is based on the flashing patterns and behavior of Firefly.

Table 1. Comparison of different scheduling algorithms

Scheduling Algorithms	Parameters Considered	Objectives	Waiting Time
FCFS	Simplest Scheduling Algorithms	CPU is allocated in the order to which the process arrive	More
RR Algorithm	Performance heavy depends upon the size of time quantum	The preemption take place after a fixed interval of time	More than all
Min-Min, Max-Min Algorithm	Makespan	Promised the guarantee regarded the provided resources	lesser
PSJN	Cost and Time	Effective and Fast execution	Lesser
Optimized algorithm	Arrival time, process time, deadline and IO requirement s	Effective resource allocation under defined parameters	lesser
Cost based Algorithm	Cost and task grouping	Minimizing the cost and completion time	lesser
Genetic Algorithm	Complexity depends on the task to be scheduled	Minimizing the cost and time	lesser
ACO	Cost and time	Improve the efficiency and reliabilities in all conditions	more

7. CONCLUSION

Scheduling mechanism is an important issue in case of cloud computing. It is necessary to check server and resource utilization to increase the performance of the system. A study of existing task scheduling algorithms is done in this paper. It considers some heuristic and metaheuristic methods for study. A better scheduling algorithm can be developed from the existing methods by adding more number of metrics which can result in good performance and outputs that can be deployed in a cloud environment in future.

8. REFERENCES

[1] Peter Mel, Timothy Grace, "The NIST definition of Cloud Computing (September, 2011)", Accessed on May, 2014.

[2] Shahdan Sudin, Sophan Wahyudi Nawawi, Amar Faiz, Zainal Abidin, Muhammad Arif Abdul Rahim, Kamal Khalil, Zuwairie Ibrahim, Zulkifli Md Yusof, " A modified gravitational search algorithm for discrete optimization problem", A Modified Gravitational Search Algorithm for Discrete Optimization Problem, IJSSST, 15:51-55, 2014.

[3] Anagha Yadav, S.B.Rathod "Study of scheduling Techniques in Cloud Computing Environment"International Journal of Computer Trends and Technology [IJCTT], Volume 29 Number 2, Nov 2015

[4] Sonia Sidhu "Task Scheduling in Cloud Computing "International Journal of Advanced Research in Computer Engineering& Technology (IJARCET), Volume 4 Issue 6, June 2015

[5] P.Nagendra Babu, M. Chaitanya Kumari, S.Venkant Mohan "A Literature Survey on Cloud Computing" International Journal of Engineering Trends and Technology (IJETT), Volume 21 Number 6, March 2015

[6] Padmaja.K, Dr.R Seshadri, P.Anusha "Different Scheduling Algorithms in Types of Clouds" International Journal of Computer Science Trends and Technology [IJCST] Volume 4 Issue 5, Sept-Oct 2016

[7] Nidhi Arora "Review on Task Scheduling Algorithms in Cloud Computing Environment" International Journal of Advanced Research in Computer Science Volume 8 number 4, May 2017

[8] Partibha Rani, Akhilesh K. Bhardwaj "A Review: Metaheuristic Technique in Cloud Computing" International Research Journal of Engineering and Technology (IRJET) Volume 04 Issue 08, Aug-2017

[9] Rajveer Kaur, Supriya Kinger, "Analysis of Job Scheduling Algorithms in Cloud Computing" International Journal of Computer Trends and Technology (IJCTT) Volume 9 Number 7, Mar 2014

[10] S.Rekha and R.Santhosh Kumar "Priority Based Job Scheduling For Heterogeneous Cloud Environment" IJCSI International Journal of Computer Science Issues, Volume 11 Issue 3 Number 2, May 2014

[11] S.Nagadevi, K.Satyapriya, Dr.D.Malathy, "A Survey On Economic Cloud Schedulers For Optimized Task Scheduling", International Journal of Advanced Engineering Technology, 2013.

[12] Huankai Chen, Frank Wang, Na Helian, Gbola Akanmu,"User-Priority Guided Min-Min Scheduling Algorithm for Load Balancing in Cloud Computing", IEEE, 2013

[13] Lipsa Tripathy, Rasmi Ranjan Patra "Schedulinin Cloud Computing" International Journal on Cloud Computing: Services and Architecture (IJCCSA),Volume 4, Number 5,October 2014

[14] Dr Ajay jangra, Tushar Saini, "Scheduling Optimization in Cloud computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume Number.3, Issue 4, April 2013.

[15] Er. Shimpy, Mr. Jagandeep Sidhu, "Different scheduling algorithms in different cloud environment", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 9, Sept 2014.

[16] Poonam Devi, Trilok Gaba ,"Implementation of Cloud Computing By Using Short Job Scheduling" International Journal of Advanced Research in Computer Science and Software Engineering, Vol. no.3, Issue 7, pp 178-183, July 2013.

[17] Nitish Chopra, Sarabjeet Singh, "Deadline and Cost based Workflow Scheduling in Hybrid Cloud", International Conference on Advances in Computing, communications and Informatics (ICACCI), IEEE, 2013.

[18] Wang Zong jiang, Zheng Qiu sheng, "A New Task Scheduling Algorithm in Hybrid Cloud Environment" International Conference on Cloud Computing and Service Computing, 2012.

[19] E.Huedo and et al., ―A modular meta-scheduling architecture forefacing with pre-WS and WS Grid resource management services,Future Generation Computing Systems,Volume.23Number.2,PP.252–261,2000

[20] Teena Mathew, K.Chandra Sekaran and John Jose"Study and Analysis of various Task Scheduling Algorithms in the cloud Computing Environment" International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014

[21] Tracy D. Braun, Howard Jay Siegel, Noah Beck," A Comparison of Eleven Static Heuristics or Mapping a Class of independent Tasks onto heterogeneous Distributed Computing System" ,Journal of Parallel and Distributed Computing 61,pp.810-837,2001

[22] Weicheng Huai, Zhuzhong Qian, Xin Li, Gangyi luo and Sanglu Lu ,"Energy Aware Task Scheduling in Data Centers, Journal of Wireless Mobile Networks. Ubiquitous Computing and Dependable Applications," Volume 4, Number 2, PP- 18-38, 2013

[23] Sung Ho Jang, Tae Young Kim, Jong Sik Lee School," The Study of Genetic Algorithm- based Task Scheduling for Cloud Computing", International Journal of Control and Automation Volume 5,Number 4,December 2012.

[24] Shenai Sudhir,"Survey on Scheduling Issues in Cloud Computing" Procedia Engineering, Elsevier, 2012.

[25] Archana mantri, Suman Nandi, Gaurav Kumar, Sandeep Kumar,"High Performance Architecture and Grid Computing", Internationals Conference HPAGC 2011, Chandigarh, India ,July 2011

[26] M. Coli, P. Palazzari, “Real Time Pipelined System Design through Simulated Annealing,” Journal of Systems Architecture, Volume 42, Number. 6-7, PP. 465-475, 1996.

[27] J.Cao, D.P. Spooner, S.A.Jarvis, G.R. Nudd,” Grid Load Balancing Using intelligent Agents,”Future Generation Computer Systems. Vol.21, Number 1, 2005, PP.135-149

[28] H.Izakian, A.Abraham, V.Snasel, “Comparison of Heuristics for Scheduling Independent Tasks on Heterogeneous environments,” in Proc. of the International Joint Conference on Computational Sciences and Optimization, IEEE, Volume 1,2009,PP,8-12.