Various Job Scheduling Algorithms in Cloud Computing: A Survey

Yash P. Dave, Avani S. Shelat, Dhara S. Patel
Department of Computer Science and Engineering
Shri Sa'd Vidya Mandal Institute of Technology
Bharuch, India.
E-mail: yashpdave@gmail.com

Rutvij H. Jhaveri
Department of Computer Science and Information Technology
Shri Sa'd Vidya Mandal Institute of Technology
Bharuch, India.
E-mail: rhj_svmit@yahoo.com

***Abstract*—Cloud Computing provides a Computing environment where different resources, infrastructures, development platforms and software are delivered as a service to customers virtually on pay per time basis. Low cost, scalability, reliability, utility-based computing are important aspects of cloud computing. Job scheduling is an essential and most important part in any cloud environment. With increasing number of users, Job scheduling becomes a strenuous task. Ordering the jobs by scheduler while maintaining the balance between quality of services (QoS), efficiency and fairness of jobs is quite challenging. Scheduling algorithms are implemented considering parameters such as throughput, resource utilization, latency, cost, priority, computational time, physical distances, performance, bandwidth, resource availability. Though there are different scheduling algorithms available in cloud computing, a very less comparative study has been done on performance of various scheduling algorithms with respect to above mentioned parameters. This paper aims at a comparative study of various types of job scheduling algorithms that provide efficient cloud services.**

Keywords—Time Optimized Scheduling Algorithms, Cost Optimized Scheduling Algorithms, Cloud Computing.

I. Introduction

Cloud computing is a horizon that is emerging out in virtualization, computing and web technologies. Cloud computing is a distributed computing environment which includes a large set of virtualized computing resources, various development platforms, infrastructures and useful software are delivered as a service to customers on a pay as per use basis usually over the Internet [1]. A major constraint i.e. cutting down the operational expenses can be overcome by the use of cloud computing technology. This will help grow business organizations, government organizations as well as academic institutions. Cost effectiveness, scalability, reliability, fault tolerance, service-orientation, utility based, virtualization and service level agreement (SLA) are some of the salient features of Cloud Computing [2] [3].

Cloud provides on demand computational resources in the form of virtual machines (VMs) deployed in a cloud provider's data center. The computational resources are shared among different cloud consumers who pay for the service accessed as per the usage. Allocation of resources and proper scheduling has a considerable impact on the performance and efficiency of the system. The main goal of cloud computing is to provide efficient access to remote and geographically distributed resources.

An efficient scheduling is a key to manage the access to different resources, load balancing as well as resource allocation [4]. Different types of resource scheduling algorithms are available in cloud computing based on certain parameters like time, cost, performance, priority, utilization of resources, throughput, bandwidth, resource availability and physical distances.

The paper is organized as follows: In section II, we have discussed about the theoretical background of cloud computing. In section III, we have discussed about the literature survey we have done related to scheduling algorithms used in cloud computing.

II. Theoretical Background

Resource scheduling and allocation are important features that affect the performance of cloud computing [5]. Many researchers have developed various algorithms for scheduling and scaling of resources in an efficient manner in a cloud. These algorithms are round robin, first come first serve, min-min, max-min, earliest dead line first, and greedy [6] [7].

Two major types of techniques are used to categorize algorithms. They are Time-based optimization techniques and Cost-based optimization techniques. Time-based algorithms concentrate on minimization of execution time, excluding other factors like cost as well as QoS levels. Cost-based optimization techniques concentrate on satisfying QoS user level [8] [9].

A. Taxonomy of Cloud Computing

Cloud Computing provides various services related to infrastructure, software, and platform. The three basic models of Cloud Computing are Infrastructure as a service (Iaas), Platform as a service (Paas), and Software as a service (Saas) as shown in Fig. 1 Service layers of Cloud Computing [10].

*1) Infrastructure-as-a-Service (Iaas):*Infrastructure-as-a-Service is the foundation layer of cloud computing. Iaas delivers hardware as a service. It includes servers, network, storage, virtualization technology, file systems, and operating systems. Iaas cloud providers supply the above mentioned resources on-demand from their large pools installed in data centers [11].

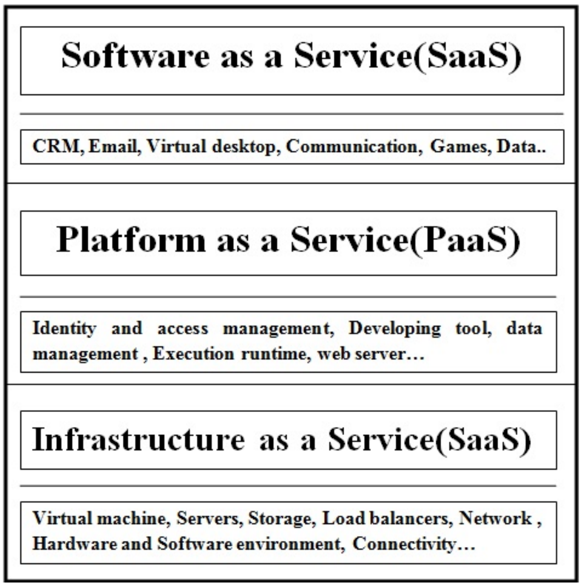

Figure 1. Service layers of Cloud Computing

2) Platform-as-a Service (Paas): PaaS model includes development tools, database, webservers, Execution runtime environment. This model concentrates on providing cost-effective, efficient environment for the development of high-quality applications [11].

3) Software-as-a-Service (SaaS): SaaS also known as "on-demand-software" is generally priced on a pay-per-use basis. In Saas model, the application software are installed on the cloud by the cloud service providers and accessed by the cloud users from the cloud itself. This eliminates the need of installing and running the applications on the cloud user's personal computers [12].

B. Types of Cloud Computing

Cloud computing environment can be categorized into private cloud, public cloud and hybrid cloud. A private cloud is limited to only an organization managed by the organization itself or any third party cloud service provider. A public cloud is available on the network and is open publicly. A hybrid cloud is basically the combination of a public cloud and a private cloud.

C. Scheduling in Cloud Computing

The allocation of system resources to various tasks is known as scheduling. Scheduling is used in cloud computing to achieve high performance and the best system throughput. Speed, efficiency, utilization of resources in an optimized way depends largely on the type of scheduling selected for the cloud computing environment. Various criteria for scheduling are max CPU utilization, Max throughput, Min turnaround time, Min waiting time and Min response time. Throughput is the number of processes that complete their execution per time unit [13]. Turnaround time is the amount of time to execute a particular process, which is the interval of time of submission of a process to the time of completion. Waiting time is the sum of the periods spent waiting in the ready queue. Response time is the time from the submission of a request until the first response is produced [14].

Various issues exist in scheduling algorithms based on different optimization criteria. Turnaround time and throughput are the two required criteria in batch systems, response time and fairness are the two criteria required in interactive system, whereas in real-time system meeting deadlines is an important aspect [15]. Therefore a scheduling algorithm must be selected in such a way that it satisfies the required criteria and provide efficient service and proper allocation of resources.

III. Literature Survey

In this section we have presented comprehensive briefing of various algorithms. Based on the existing Cloud computing related review, various scheduling algorithms are mentioned.

A. Round Robin

Round Robin (RR) algorithm is one of the algorithms employed by process and network scheduler in computing. Fairness is primarily focused in Round robin. RR scheduling is simple, easy to implement, and starvation-free. The scheduler assigns a fixed time unit per process, and cycles through them.

RR uses the ring as its queue to store jobs. Each job has the same execution time and it will be executed in turn. If a job can't complete its work during its turn, it will be stored back to the queue waiting for the next turn. Main feature of RR algorithm is execution of each job in turn and it doesn't have to wait for the previous one to get completed. But if the load is found to be heavy, RR will take a long time to complete all the jobs [16]. The CloudSim toolkit supports RR scheduling strategy for internal scheduling of jobs [17].The drawback of RR is that the largest job takes enough time for completion.

B. Earliest Deadline First

The basic principle of this algorithm is very intuitive and simple to understand. In this scheduling algorithm, at every scheduling point the task having the shortest deadline is taken up for scheduling. Earliest Deadline First (EDF) or Least Time

ISBN No.978-1-4799-3834-6/14/$31.00©2014 IEEE

to go is a dynamic scheduling algorithm used in real-time operating systems that places processes in a priority queue. Whenever a scheduling event occurs (end of task, release of new task.) then the queue will be searched for the process that is closest to its deadline, the found process will be the next that is going to be scheduled for execution [18].

C. Deadline Distribution Algorithm

The deadline distribution algorithm [19] meets the deadline for delivering results to minimize the cost. This algorithm partitions a workflow and distributes the overall deadline into each task based on their workload and dependencies.

Deadline distribution algorithm uses Synchronization Task Scheduling (STS) for synchronization tasks and Branch Task Scheduling (BTS) for branch partition respectively. Each task has its own sub deadline and a local optimal schedule can be generated for every task. If each local schedule guarantees completion of their task execution within their sub-deadlines, the whole workflow execution will be completed within the overall deadline. Similarly, the result of the cost minimization solution for each task leads to an optimized cost solution for the entire workflow [20]. Therefore, an optimized workflow schedule can be constructed from all local optimal schedules. The schedule allocates every workflow task to a selected service such that it can meet its assigned sub-deadline at a low execution cost.

D. Compromised-time-cost scheduling Algorithm

S Pandey et al. presented a compromised-time-cost scheduling algorithm in which they considered cost constrained workflows by compromising execution time and cost with user input [21]. This algorithm considers the characteristics of cloud computing to accommodate instance-intensive cost-constrained workflows by compromising execution time and cost with user input enabled on the fly [22].

E. Back-tracking Algorithm

Backtracking algorithm finds solutions to computational problem that incrementally builds candidates to the solutions, and abandons each partial candidate c as soon as it determines that c cannot possibly be completed to a valid solution [23]. The back-tracking algorithm [24] assigns available tasks to the least expensive computing resources. The unscheduled task's parent tasks are scheduled.

If more than one task is available, the algorithm assigns the task with the largest computational demand to the fastest resources in its available resource list. This process is repeated until all tasks have been scheduled. After each step, the execution time of the current assignment is computed. If the execution time exceeds the time constraint, the back tracks the previous step, removes the resource with minimum expense from its resource list and reassigns tasks with the reduced resource set. Backtracking keeps the previous steps where the resource list is empty and reduces the relative resource list and reassigns the tasks.

F. Genetic Algorithm

Genetic Algorithm (GA) is a heuristic that mimics the process of natural selection. It starts with a set of initial solution and then with the help of genetic operators, will generate the new solution [25] [26]. This algorithm handles a large search space, applicable to the complex objective function with the help of a local optimum solution, avoiding the trap. Authors have developed a genetic algorithm, which provide a cost based multi QoS scheduling in cloud [27] [28]. Genetic algorithms belong to the larger class of Evolutionary Algorithm (EA).

G. Modified ant colony optimization scheduling algorithm

Modified ant colony optimization is an approach for diversified service allocation and scheduling mechanism in cloud paradigm. The main aim of this optimization method is to minimize the scheduling throughput to service all the diversified requests according to the different resource allocator available under cloud computing environment [29].

H. Improved activity based cost based algorithm

This algorithm is used for efficient scheduling of tasks to available resources in cloud. In this algorithm the cost and computation performance are scheduled. It also improves the computation or communication ratio by grouping the user tasks according to a particular cloud resource's processing capability and sends the grouped jobs to the resource [30].

I. Particle Swarm Optimization algorithm

Particle swarm optimization is used to minimize the total cost of execution of application workflows on Cloud computing environments. It calculates the total cost of execution by varying the communication cost between resources and the execution cost of computing resources [31].

J. Market Oriented Hierarchical Scheduling

Market-oriented hierarchical scheduling strategy consists of service-level scheduling and task-level scheduling. Service-level scheduling deals the Task-to-Service assignment and the task-level scheduling deals with the optimization of the Task-to-VM assignment in local cloud data centers. It can be used to optimize the time and cost simultaneously [32].

K. Profit-driven Service Request Scheduling

Two sets of profit-driven service request scheduling algorithms are presented. These algorithms are devised incorporating a pricing model using process sharing (PS) and two allowable delay metrics based on service and application. Authors have demonstrated the efficiency of consumer applications with interdependent services. The evaluation of the algorithm was done on the basis of maximum profit and utilization [33].

A comparative study has been done to analyse the working of different scheduling algorithms.

ISBN No.978-1-4799-3834-6/14/$31.00©2014 IEEE

IV. Result And Discussions

The algorithm with the best efficiency for Cloud environment is determined by various parameters. Cloud service providers also offer extra resources to users on demand with some extra charges. Therefore, Scheduling policies must work within the deadline and time constraints.

Scheduling algorithms are generally classified as Time based and Cost based. Cost based algorithms include Deadline distribution algorithm, Compromised time-cost algorithm, and Genetic algorithm. Time based algorithms include Backtracking algorithm, Extended Min-Min algorithm, Round robin algorithm, and Earliest deadline algorithm.

Cost optimization techniques minimize the cost for running the applications whereas Time optimization techniques minimize the completion time for applications. Cloud service providers use several scheduling algorithms as per the cloud environment to provide efficient services to the cloud users. Some of the leading cloud providers have been mentioned in the TABLE I. along with the algorithms they use for allocation of resources [34] [35].

TABLE I. SCHEDULING ALGORITHMS IN DIFFERENT CLOUD COMPUTING ENVIRONMENT

Cloud Provider	**Open Source**	**Scheduling algorithm used**
Eucalyptus	Yes	Greedy algorithm , First fit algorithm, Round Robin algorithm
Rackspace	Yes	Round Robin algorithm , Weighted round robin algorithm, Least connections algorithm, weighted least connections algorithm
Open Nebula	Yes	Rank matchmaker scheduling algorithm, Preemption scheduling algorithm
Nimbus	Yes	Virtual machine schedulers PBS and SGE
Amazon	No	Xen ,swam, Genetic algorithm
RedHat	Yes	Breath first algorithm, Depth first algorithm
Lunacloud	Yes	Round Robin algorithm

Based on the above study few parameters have been identified to classify the scheduling algorithms. Some algorithms can optimize the time span whereas some can minimize the cost as shown in TABLE II.

Thus, based on customers requirement service providers can use various algorithms for enhancing the efficiency and obtain optimized resource allocation based on certain parameters.

TABLE II. FACTOR BASED CLASSIFICATION OF SCHEDULING ALGORITHMS

Time	**Cost**
Round Robin	Dead line distribution
Earliest Dead Line first	Compromised time-cost
Back Tracking	Genetic algorithm
Modified ant colony optimization	Improved activity based cost
Extended Min-Min	A particle swarm optimization
Compromised time cost	Profit driven service oriented algorithm

Conclusion

In cloud computing environment, resources with diverse characteristics are served virtually. To manage this type of heterogeneous resources in optimized way efficient scheduling is imperative. Efforts have been put to study various scheduling algorithms in cloud environment. After a brief study, various scheduling algorithms have been classified on the basis of factors such as cost and time and have been presented in this paper. This comparative study shall be helpful in selection of appropriate scheduling algorithms for using different types of services as per the requirements of cloud consumers as well as cloud service providers. This study may be useful to cloud research enthusiasts for development of efficient algorithms to enhance the overall efficiency of cloud computing environments in future.

References

[1] Saeed Parsa and Reza Entezari-Maleki."RASA: A New Task Scheduling Algorithm in Grid Environment" .World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009 ISSN 1818.4952© IDOSI Publications, 2009J.

[2] Lu Huang, Hai-shan Chen, Ting-ting Hu,"Survey onResource Allocation Policy and Job Scheduling Algorithms of Cloud Computing"Journal of software, vol. 8, NO. 2, feb2013.

[3] O. M. Elzeki, M. Z. Reshad and M. A. Elsoud, "Improved Max-Min Algorithm in Cloud Computing",International Journal of Computer Applications (0975 – 8887).

[4] Braun, T.D., Siegel, H.J., Beck, N., Boloni, L.L.,Maheswaran, M., Reuther, A.I., Robertson, J.P., et al. "A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems", Journal of Parallel and Distributed Computing, Vol. 61, No. 6, pp.810–837, 2001.

[5] Greg Boss, Padma Malladi, Dennis Quan, Linda Legregni, Harold Hall. "Cloud Computing", High Performance On Demand Solutions(HiPODS).

[6] Chunmei Chi, Feng Gao." The Trend of Cloud Computing " in China[J]. Journal of Software, VOL.6, NO.7, 2011,7:1230~1234.

[7] IsamAzawiMohialdeen. " Comparitive Study of Scheduling Algorithm in cloud computing environment", in 2013 ISSN 1549-3636.

[8] Rodrigo N. Calheiros1,2, Rajiv Ranjan1, César A. F. De Rose2, and RajkumarBuyya, "CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services",1Grid Computing and Distributed Systems (GRIDS) Laboratory Department of Computer Science and Software Engineering The University of Melbourne, Australia, Pontifical Catholic University of Rio Grande do SulPortoAlegre, Brazil.

[9] Jagbeer Singh, BichitranandaPatra, Satyendra Prasad Singh, "An Algorithm to Reduce the Time Complexity of Earliest Deadline First Scheduling Algorithm in Real-TimeSystem",(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.2, February 2011.

[10] Omar Serfaoui, Mohammed Aissaoui, MohsineEleuldj , "OpenStack: Toward an Open-Source Solution forCloud Computing", International Journal of Computer Applications (0975 - 8887)Volume 55 - No. 03, October 2012.

[11] K. Keahey and T. Freeman, ―Science Clouds: Early Experiences in Cloud Computing for Scientific Applications,□ in proceedings of Cloud Computing and Its Applications 2008, Chicago, IL. 2008.

[12] B. Furht, and A. Escalante, "Handbook of cloud computing," Cloud computing fundamentals chapter by B. Furht, Springer, 2010.

[13] Salim Bitam,"Bees Life algorithms for job scheduling in cloud computing", International Conference on computing and Information Technology, 2012.

[14] He. X, X-He Sun, and Laszewski. G.V, "QoS Guided Min-min Heuristic for Grid Task Scheduling," Journal ofComputer Science and Technology, Vol. 18, pp. 442-451, 2003.

[15] Mei, L., Chan, W.K., Tse, T.H., ―A Tale of Clouds: Paradigm Comparisons and Some Thoughts on Research Issues□, In: APSCC 2008, pp. 464-469.

[16] S. Sadhasivam ,N.Nagaveni ,R. Jayarani,and R. Vasanth Ram , "Design and Implementation of an efficient Twolevel Scheduler for Cloud Computing Environment', International Conference on Advances in Recent Technologies in Communication and Computing,2009.

[17] Shin-ichiKuribayash, "Optimal Joint Multiple Resource Allocation Method for Cloud Computing Environments", International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 2, No. 1, March 2011.

[18] J. Singh, B. Patra, S. P. Singh ,"An Algorithm to Reduce the Time Complexity of Earliest Deadline First Scheduling Algorithm in Real-Time System", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.2, February 2011.

[19] J. Yu, R. Buyya and C. K. Tham, "A Cost-based Scheduling of Scientific Workflow Applications on Utility Grids", Proc. of the 1st IEEE International Conference on e-Science and Grid Computing, Melbourne, Australia, December 2005 , pp140-147.

[20] Kaur, P.D., Chana, I. ―Unfolding the distributed computing paradigm□ ,In: International Conference on Advances in Computer Engineering, pp. 339-342 (2010).

[21] "Overview of cloud computing and Scheduling policies 2.0", http://archives.opennebula.org/documentation:archives:rel2.0:schg.

[22] Ke Liu, Jinjun Chen, Hai Jin, Xiao Liu, Dong Yuan, "A Compromised Time Cost Scheduling Algorithm in SwinDew C for Instance Intensive Cost Constrained Workflows on a Cloud Computing Platform",July 4,2012.

[23] Gilles Brassard, Paul Bratley(1995). *Fundamentals of Algorithmics*.Prentice-Hall.

[24] D. A. Menasc and E.Casalicchio, "A Framework for Resource Allocation in Grid Computing", Proc. of the 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, Volendam, The Netherlands, October, 2004,pp 259-267.

[25] Sung Ho Jang, Tae Young Kim, Jae Kwon Kim and Jong Sik Lee"The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing".International Journal of Control and Automation Vol. 5, No. 4, 2012.

[26] ShaminderKaur ,AmandeepVerma, **"**An Efficient Approach to Genetic Algorithm for Task Scheduling in Cloud Computing Environment" I.J. Information Technology and Computer Science, 2012, 10, 74-79.

[27] J. Yu and R. Buyya, "Scheduling Scientific Workflow Applications with Deadline and Budget Constraints using Genetic Algorithms", Scientific Programming Journal, 14(3-4), 217-230, IOS Press, 2006.

[28] D Dutta,R C Joshi ,"A Genetic –Algorithm Approach to Cost-Based Multi-QoS Job Scheduling in Cloud Computing Environment", International Conference and Workshop on Emerging Trends in Technology (ICWET 2011) – TCET, Mumbai, India,2011.

[29] S Banerjee, I Mukherjee, and P.K. Mahanti "Cloud Computing Initiative using Modified Ant Colony Framework" , World Academy of Science, Engineering and Technology, , 56 2009, pp 221-224.

[30] Selvarani, S.; Sadhasivam ,G.S.,"Improvedcostbased algorithm for task scheduling in cloud computing", Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International conference on 2010 , pp 1–5.

[31] K. Liu, Y. Yang, J. Chen, X. Liu, D. Yuan , H. Jin, "A Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-intensive Cost-Constrained Workflows on Cloud Computing Platform", Int. Journal of High Performance Computing Applications, Volume 24 Issue 4, 2010.

[32] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing", in 2009 IEEE International Symposium on Parallel and Distributed Processing.

[33] Young Choon Lee, Chen Wang, Albert Y. Zomaya, Bing BingZhou,"Profit-driven Service Request Scheduling in Clouds", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing,2010.

[34] Sonali Yadav, "Comparative Study on Open Source Software for Cloud Computing Platform: " ,International Journal Of Engineering And Science Vol.3, Issue 10 (October 2013), PP 51-54.

[35] Anita S. Pillai , prof. L. S. . Swastimathi, "A Study on open source cloud computing platforms ",EXCEL International Journal of Multidisciplinary Management Studies Vol.2 Issue 7, July 2012, ISSN 2249 8834.