ISBN: 978-93-83758-09-8

Survey of Process Scheduling in Cloud Computing

Umesh Kumar Verma
Computer Science and Engineering Department
Apiit SD India
Panipat
umesh@apiit.edu.in, ukverma.verma@gmail.com

***Abstract*-Scheduling of jobs is a foremost and difficult issue in cloud computing. Utilizing cloud computing resources efficiently is one of the cloud computing service provider's ultimate goals. Today cloud computing is on demand as it offers dynamic flexible resource allocation for trustworthy and definite services in pay-as-you-use manner, to cloud service users. So there must be a provision that all resources should be made available to demanding users in proficient manner to satisfy their needs. Hence researchers are paying attention in developing various scheduling algorithms which helps both consumers as well as providers so that balance must be maintained. In this paper systematic study of various scheduling algorithms and issues related to them in cloud computing is presented.**

Keywords: Scheduling algorithm, Cloud, Survey.

I. INTRODUCTION

Cloud computing has provided new paradigm by providing computing as a utility service rather than a product, whereby shared resources, software and information are provided to users over the network. Cloud computing providers share application via the Internet, which are accessed from web browser, while the business software and data are stored on servers at a distant location. Cloud providers are trying to attain the agreed SLA, by scheduling resources in efficient manner and by deploying application on proper VM as per the SLA objective and at the same time performance of the applications must be optimized. As cloud computing is the most recent buzz, there are many existing issues like Resource Provisioning, Load Balancing, Virtual Machine Migration, Server Consolidation, Energy Management, etc. that are not fully addressed. Process or Job scheduling problem is a foremost and challenging issue in cloud computing. How to use cloud computing resources competently and gain the highest profits with job scheduling system is one of the decisive objectives of cloud computing service providers'.

The objective and motivation of this survey is to give a systematic review of existing job or process scheduling techniques or algorithm in cloud computing and encourage researchers and scholars in this field, so that they can contribute in developing more efficient load balancing algorithm.

II. SCHEDULING

Today Cloud computing is on demand as it offers dynamic flexible resource allocation, for reliable and guaranteed services in pay-Today Cloud computing is on demand as it offers dynamic flexible resource allocation, for reliable and guaranteed services in pay-as-you-use manner, to Cloud service users. So there must be a provision that all resources are made available to requesting users in efficient manner to satisfy their needs. Seminal work of [3] describes scheduling as a process of finding the capable resources that can execute the cloud requests (tasks) at specific times that satisfy specific performance quality measure such as execution time minimization, as specified by cloud users. The main goal of job scheduling is to achieve a high performance computing and the best system throughput.

The main purpose of job scheduling is to achieve a high performance computing and the best system throughput. Clouds are mainly determined by economics—the pay-peruse pricing model like similar to that for basic utilities, such as electricity, water and gas. Schedulers employ a function that takes into account the essential objectives to optimize a specific outcome. The commonly used scheduling reason in a cloud computing environment is related to the tasks completion time and resource utilization. The scheduler uses a particular policy for mapping the tasks to suitable cloud resources in order to satisfy user requirements. However, the bulk of these scheduling strategies are static in nature [18]. They produce a good plan given the current state of Cloud resources and do not take into account changes in resource accessibility. On the other hand, dynamic scheduling [19] considers the current state of the system. It is adaptive in

ISBN: 978-93-83758-09-8

nature and able to fabricate efficient schedules, which ultimately reduces the completion time of tasks as well as improves the overall performance of the system. When a job is allocated to the clouds, it is usually partitioned into several tasks. Following questions are to be considered when applying processing in executing these tasks. (1) How to allocate resources to tasks? (2)What is the execution order of the task? (3) How to reduce schedule expenditure?

III. PROCESS SCHEDULING ALGORITHMS REVIEW

Influential work of [13] had suggested a new task scheduling algorithm RASA. It is composed of two traditional scheduling algorithms; Max-min and Min-min. RASA uses the advantages of Max-min and Min-min algorithms and covers their disadvantages. Though the deadline of each task, arriving rate of the tasks, cost of the task execution on each of the resource, cost of the communication are not considered. The experimental results show that RASA is outperforms the existing scheduling algorithms in large scale distributed systems.

Significant effort of [6] has recommended a new algorithm based on impact of RASA algorithm. Improved Max-min algorithm is based on the expected execution time instead of complete time as a selection basis. Petri nets are used to model the concurrent behavior of distributed systems. Max-min demonstrates achieving schedules with comparable lower make span rather than RASA and original Max-min.

Remarkable work of [4] put forwarded a reliable scheduling algorithm, (RSDC) in cloud computing environment. In this algorithm major job is divided to sub jobs. In order to balance the jobs the request and acknowledge time are calculated separately. The scheduling of each job is done by calculating the request and acknowledges time in the form of a shared job. So that efficiency of the system is increased.

Outstanding work of [7] advised a new scheduling algorithm based on multi – criteria and multi - decision priority driven scheduling algorithm. This scheduling algorithm consist of three level of scheduling: object level, attribute level and alternate level. In this algorithm priority can be set by job resource ratio. Then priority vector can be compared with each queue. This algorithm has higher throughput and less finish time.

Prominent work of [1] has offered a differentiated scheduling algorithm with non-preemptive priority queuing model for activities performed by cloud user in the cloud computing environment. In this approach one web application is created to do some activity like one of the file uploading and downloading then there is need of efficient job scheduling algorithm. The Quality of Service (QOS) requirements of the cloud computing user and the maximum profits of the cloud computing service provider are achieved with this algorithm.

Striking effort of [14] had proposed an improved cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud. Scheduling algorithm divides all user tasks depending on priority of each task into three different lists and measures both resource cost and computation performance, also improves the computation/communication

Seminal work of [17] has presented a gang scheduling algorithm with job migration and starvation handling in which scheduling parallel jobs, already. The model is studied through simulation in order to analyze the performance and overall cost of Gang Scheduling with migrations and starvation handling. Results highlight that this scheduling strategy can be effectively deployed on Clouds, and that cloud platforms can be viable for HPC or high performance enterprise applications.

Incredible work of [2] presents a job combination & dispatching strategy for scheduling jobs to client users which connect to data servers. Based on the job combination and dispatching strategy algorithm (JCDS), the optimization algorithm was proposed in this study; JCDS with dynamic programming (JCDS-D).Authors limelight the job allocating and the communication overheads minimizing in grid system. The experimental results illustrate that the JCDS and JCDS-D provide enhancement in terms of performance and processors' utilization

Comprehensive evaluation provided by [8] compared the performance of a bi-criteria scheduling algorithm for Work Flows with Quality of Service (QoS) support. Suggested work serves as basis to implement a bi-criteria hybrid scheduling algorithm for work flows with QoS support, aiming to optimize the criteria chosen by the users and based on the priority ordering and relaxation specified by them. The proposed model aims to permit scheduling with, reduced the response time to the user, improved use of resources, reducing the make-span and choosing the best resource using the historical data from user applications. Results verify that criteria reliability and runtime are somewhat conflicting and need to be treated independently, but this does not prevent them to be used jointly.

Focus of paper [9] is to provide a scheduler that aims to maximize user satisfaction. Thus the job details submitted by the user will include job prioritization criteria: the allocated budget and the deadline required by the user, enabling the scheduler to maximize CPU utilization while remaining within the constraints imposed by the need to optimize user Quality of Service (QOS).

Remarkable work of[10] addressed the problem of scheduling of consumers' service requests (or applications) on service instances made accessible by providers taking into account costs—incurred by both consumers and providers—as the

ISBN: 978-93-83758-09-8

most vital factor. Author's contributions include the development of a pricing model using processor-sharing for clouds, the application of this pricing model to composite services and the development of two sets of profit-driven scheduling algorithms explicitly exploiting key characteristics of service requests including precedence constraints.

Decisive work of [15] had scrutinized the resource scheduling to achieve SLA-aware profit optimization in cloud services. Individual-job-based SLA was analyzed in the paper. Authors had appraised 2 metrics a) the number of hard-deadline violations and b) cost. Assessment between FCFS, SJF and CBS scheduling policies is done. The results show that CBS every time shows better performance compared to FCFS and SJF for both deadline enforcement and cost.

[16] Proposed an optimized scheduling algorithm to accomplish the optimization or sub-optimization for cloud scheduling problems. Authors investigated the possibility to assign the Virtual Machines (VMs) in a flexible way to permit the maximum usage of physical resources. Authors suggested use of an IGA for the automated scheduling policy. The tests exemplify that the speed of the IGA almost twice the traditional GA scheduling method in Grid environment and the utilization rate of resources always higher than the open-source IaaS cloud systems.

Proficient work of [11] had put forwarded dynamic priority scheduling algorithm, in which average value of processing time and variance of process time was measured and compared with FCFS and SPSA is done. Demonstrate that the DPSA has a improved efficiency and a better fairness than the FCFS, and is more realistic than SPSA.

[12] Explored some of the effects that the paradigm of Cloud Computing has on schedulers when executing scientific applications. Author present premises regarding to provisioning and architectural aspects of a Cloud infrastructure, that are not present in other environments, and which implications they may have on scheduling decisions in presence of relevant policies like improving performance. Authors proposed and test a preliminary workload classification, based on usage modes that may improve early scheduling decisions as Authors research towards automatic deployment of scientific applications.

IV. EVALUATION OF SCHEDULING APPROACHES

Author has appraised the above describe ways for scheduling techniques by different scholars in form of a table by showing the chosen parameters and final findings. Both non preemptive and preemptive algorithms were studied. Dynamic and static algorithms were taken into consideration. Table 1 shows the critical evaluation.

Table 1: Evaluation of Existing Scheduling Approaches

Algorithm	**Approach**	**Parameters**	**Findings**
RASA	[13]	Make span	It is used to reduce Make span
RSDC	[5]	Processing time	1. It is used to reduce processing time and increase load balancing.
Petri Net based Max-min Scheduling	[6]	Load balancing, finish time	1. More efficient load balancing. 2. Used to remove limitation of max-min algorithm.
PJSC	[7]	Three level parameters were used i.e. scheduling, resource and job level	1.Less finish time
PSSP	[4]	Quality of Service, Service request time	1. High QoS 2.High throughput
CBTS	[14]	Cost, Performance	1. Measures both resource cost and computation performance 2. Improves the computation /communication ratio
GSA	[17]	Performance, Cost	1. The application of migrations and starvation handling had a significant effect on the model. 2. It improves performance.

V. CONCLUSION

Job scheduling problem is important and challenging issue in Cloud Computing. Utilizing cloud computing resources proficiently and gaining the highest profits with job scheduling system is one of the Cloud computing service providers' ultimate goals. A lot of research work has been done in this area which mainly focuses on allocating of jobs to machines efficiently but still problem of starvation persists. New algorithm is required to reduce average waiting time, average turnaround time and total finish time of jobs and starvation problem is optimized.

VI. REFERENCES

[1] Ambike, S., Bhansali, D., Kshirsagar, J., & Bansiwal, J. (2012). An Optimistic Differentiated Job Scheduling System for Cloud Computing. International Journal of Engineering Research and Applications (IJERA) ISSN, 2248-9622.

[2] Chen, T. L., Hsu, C. H., & Chen, S. C. (2010). "Scheduling of job combination and dispatching strategy for grid and cloud system". In Advances in Grid and Pervasive Computing (pp. 612-621). Springer Berlin Heidelberg.

[3] Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility". Future Generation computer systems, 25(6), 599-616.

[4] Dakshayini, D. M., & Guruprasad, D. H. (2011). An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment. International Journal of Computer Applications (0975–8887) Volume.

[5] Delavar, A. G., Javanmard, M., Shabestari, M. B., & Talebi, M. K. (2012). RSDC (RELIABLE SCHEDULING DISTRIBUTED IN CLOUD COMPUTING). International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol, 2.

[6] El-kenawy, E. S. T., El-Desoky, A. I., & Al-rahamawy, M. F. (2012)Extended Max-Min Scheduling Using Petri Net and Load Balancing. International Journal of Soft Computing, 2.

[7] Ghanbari, S., & Othman, M. (2012). A Priority based Job Scheduling Algorithm in Cloud Computing. Procedia Engineering, 50, 778-785.

[8] Kloh, H., Schulze, B., Mury, A., & Pinto, R. C. G. (2010, November). "A scheduling model for workflows on grids and clouds". In Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (p. 3). ACM.

[9] Kumar, P., Nitin, N., Sehgal, V., Chauhan, D. S., & Diwakar, M. (2011). "Clouds: Concept to optimize the Quality of Service (QOS) for clusters." Information and Communication Technologies (WICT), .IEEE.

[10] Lee, Y. C., Wang, C., Zomaya, A. Y., & Zhou, B. B. (2010). "Profit-driven service request scheduling in clouds". In Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. IEEE

[11] Lee, Z., Wang, Y., & Zhou, W. (2011). "A dynamic priority scheduling algorithm on service request scheduling in cloud computing". In Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on (Vol. 9, pp. 4665-4669). IEEE.

[12] Mc Evoy, G., & Schulze, B. (2011, December). "Understanding scheduling implications for scientific applications in clouds". In Proceedings of the 9th International Workshop on Middleware for Grids, Clouds and e-Science (p. 3). ACM.

[13] Parsa, S., & Entezari-Maleki, R. (2009). RASA: A new task scheduling algorithm in grid environment. World Applied sciences journal, 7, 152-160.

[14] Selvarani, S., & Sadhasivam, G. S. (2010, December). Improved cost-based algorithm for task scheduling in cloud computing. In Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on (pp. 1-5). IEEE.

[15] Yun C., & Hacigumus, H. (2010, July). SLA-aware profit optimization in cloud services via resource scheduling. In *Services (SERVICES-1), 2010 6th World Congress on* (pp. 152-153). IEEE.

[16] Zhong, H., Tao, K., & Zhang, X. (2010, July). "An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems". In ChinaGrid Conference (ChinaGrid), 2010 Fifth Annual (pp. 124-129). IEEE.

[17] Zoschakis, I. A., & Karatza, H. D. (2012). Evaluation of gang scheduling performance and cost in a cloud computing system. The Journal of Supercomputing, 59(2), 975-992.

[18] Zsai, W. T., Sun, X., & Balasooriya, J. (2010). "Service-oriented cloud computing architecture". In Information Technology: New Generations (ITNG), 2010 Seventh International Conference on (pp. 684-689). IEEE.

[19] Zhong, H., Tao, K., & Zhang, X. (2010). An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems. In ChinaGrid Conference (ChinaGrid), 2010 Fifth Annual (pp. 124-129). IEEE.