Review of Literature Related to Efficient Multi-Processor Scheduling in Cloud based on the Hybridization Process

[1]Dr. Rajeev Yadav and [2]Shubham Gaur

[1]Professor, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)
[2]Research Scholar, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)
Email- shubhamgr90.2@gmail.com

Abstract: In recent years, the concept of cloud computing has been gaining traction to provide dynamically increasing access to shared computing resources (software and hardware) via the internet. It's not secret that cloud computing's ability to supply mission-critical services has made job scheduling a hot subject in the industry right now. Cloud resources may be wasted, or in-service performance may suffer because of under-utilization or over-utilization, respectively, due to poor scheduling. Various strategies from the literature are examined in this research in order to give procedures for the planning and performance of Job Scheduling techniques (JST) in cloud computing. To begin, we look at and tabulate the existing JST that is linked to cloud and grid computing. The present successes are then thoroughly reviewed, difficulties and flows are recognized, and intelligent solutions are devised to take advantage of the proposed taxonomy. To bridge the gaps between present investigations, this paper also seeks to provide readers with a conceptual framework, where we proposed an effective job scheduling technique in cloud computing. These findings are intended to provide academics and policymakers with information about the advantages of a more efficient cloud computing setup. In cloud computing, fair job scheduling is most important. We proposed a priority-based scheduling technique to ensure fair job scheduling. Finally, the open research questions raised in this article will create a path for the implementation of an effective job scheduling strategy.

Introduction:

A solution of the green cloud is not only to save the consumption of energy but expressively reduce operational costs. The main objective is to make the comprehensive computing influence of a vast collection of resources offered to a single application. High consumption of energy is the main concern in green cloud computing because of high computations happened in tightly connected data centers that need managed resources and smooth operations. The research is implemented on green cloud computing based on the scheduling process for energy conservation which is the main concern. The groups deal with the high efficient arrangement in which every industry is required with high uniformity, scalability, and effective performance over different cloud computing scenarios. This research work deals with the efficient performance to achieve efficient load balancing and the optimization process for high energy efficiency which will be the key feature of our proposed system. The proposed work deals with the hybrid scheduling process i.e. priority-based weighted round-robin and minimum completion time to reduce the energy consumption in green cloud systems. The performance will be evaluated for low error rates and achieve high energy efficiency to balance loads of the requests. Now, Cloud Computing has grown in popularity as a medium for scientific applications. To facilitate scientific study, cloud computing aims to share large-scale resources and equipment in the areas of processing, storage, information, and expertise with other researchers. Cloud computing's job scheduling algorithms are among the most difficult theoretical problems to solve Ghanbari and Othman (2012). Cloud computing uses a scheduler (broker) to figure out how to best allocate a limited number of resources to incoming activities and applications in order to achieve a variety of scheduling goals (e.g., monetary cost, computational cost, makespan, availability, reliability, response time, resource utilization, energy consumption, etc.) Lee, 1996, Allahverdi et al., 2008. One of the most notable uses of contemporary scheduling has been the allocation

of distributed computing systems of limited resources to jobs submitted by Internet users since their establishment in 1980. In the last few years, a new technology called "cluster systems" has emerged, which combines several separate computers into a single unit. Grid systems were developed in response to the weakness of cluster systems, which only utilize local resources, by gathering together all heterogeneous resources accessible in geographically distant places. Weinhardt et al. (2009). Cloud computing is a relatively new technology which makes use of the advantages of both clustered and grid-based systems.

Due to the huge solution space, many scheduling issues that are NP-hard or NP-completely take a long time to implement an optimum or sub-optimal solution in the shortest time. Due to the limited resources in modern computer systems, there is no polynomial time-scheduling technique which could be used to improve the constrained resources scheduling. Using a simple example from Taillard (1990), we can see that just about 0.02 percent of the possible solutions use between 1 and 1.01 times the time required to find the ideal answer. Finding the best answer to a complex problem is quite challenging, as this example illustrates. As a result, most scholars have been motivated to look for a quick but effective solution to these kinds of scheduling challenges. The two most basic forms of scheduling methods are static and dynamic scheduling strategies. However, because cloud settings are inherently dynamic, additional dynamic algorithms must be incorporated into the cloud scheduling process to achieve outstanding results in this field. Static algorithms, on the other hand, are only utilized when the workloads vary just slightly. As a result, adopting deterministic ways to tackle the job scheduling problem is unfeasible in this circumstance. Allahverdi (2015). Nondeterministic meta-heuristic algorithms have been offered as a way to considerably address this challenge in a polynomial amount of time.

Consumers and producers of cloud services can benefit from a variety of advantages because to dynamic work scheduling approaches and virtualization technology. Resource (task) scheduling that is effective not only minimize resource consumption (increasing the resource used), but also assures that new jobs are completed as promptly as possible (the minimizes of makespan). Job scheduling has become most important due to the possibility of a scarcity of cloud resources as a result of the continual increase in workloads at cloud datacenters. This has resulted in a significant increase in the importance of task scheduling. As a result, more study into the still-developing topic of cloud job scheduling is required to push for things like more effective mapping of incoming job to available resources and improved criteria for measuring how efficiently a service is provided. Scheduling algorithms can be used to optimize a variety of quality of service (QoS) parameters, for example resource use and utilization, task rejection ratio, energy consumption, and other constraints, by determining the optimal set of resources available to carry out incoming tasks (underutilization and over utilization). The primary objective of a scheduling approach is to find the most efficient use of the available resources (SLA).

Review of Literature

As the purpose of cloud computing is to maximize the usage of virtual machines (VMs) while minimizing data center operational expenses, resource scheduling is crucial. This leads to an increase in the quality of service (QoS) metrics in cloud computing. To accomplish the aims of both cloud service providers and users, resource scheduling manages a huge number of user insistence and distributes them all to the most applicable virtual machines. We searched the literature for scheduling algorithms and determined that just a few well-known surveys have been revealed in the cloud computing. Raghava and Singh, 2014, Milani and Navimipour, 2016, Thakur and Goraya, 2017, Ghomi et al., 2017. These questionnaires show how this suggested algorithm works in its most basic form. Example, Randles et al. (2010) assess several load balancing solutions based on one crucial performance indicator, throughput, while the methodology ignores other factors like flowchart, taxonomy, and other survey characteristics. Despite the fact that Raghava and Singh (2014) presented a brief overview of the existing scheduling techniques based on QoS criteria, all of the surveys described above focus on only a few elements (QoS parameters, year-wise analysis, state-of-art).

Further, none of the existing surveys is complete, and none of the existing surveys considers all of the QoS characteristics at the same time Ghomi et al., 2017, Thakur and Goraya, 2017 improve the survey methodology and considers more QoS metrics, besides taxonomy, a visual representation, and flow chart. In Kumar et al. (2019), the authors conducted an excellent review on cloud computing, in which they included only the work scheduling algorithms and did not mention any obstacles that

they encountered during their research, which was a mistake.Houssein et al. (2021) did another survey in which they present a taxonomy as well as a thorough discussion of the study subject. They did not provide any framework for successful job scheduling in this section, which we added in our research. Research in the field of job scheduling, despite this, its development is still at an early phase. As a result, we seek a comprehensive survey that will assist us in expanding and integrating study findings into resource scheduling on a continuous basis. This work represents a complete and systematic analysis of job scheduling strategies, as well as an assessment of present and future research challenges originating from the employment of cutting-edge scheduling approaches. We have done our best to apply our experience in this work.

As per the definition provided by the National Institute for Standards and Technology (NIST) (Badger et al., 2011), "cloud computing is a model for enabling 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". It represents a paradigm shift in information technology many of us are likely to see in our lifetime. While the customers are excited by the opportunities to reduce the capital costs, and the chance to divest themselves of infrastructure management and focus on core competencies, and above all the agility offered by the on-demand provisioning of computing, there are issues and challenges which need to be addressed before a ubiquitous adoption may happen. Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. There are four basic cloud delivery models, as outlined by NIST (Badger et al., 2011), based on who provides the cloud services. The agencies may employ one model or a combination of different models for efficient and optimized delivery of applications and business services. These four delivery models are: (i) Private cloud in which cloud services are provided solely for an organization and are managed by the organization or a third party. These services may exist off-site. (ii) Public cloud in which cloud services are available to the public and owned by an organization selling the cloud services, for example, Amazon cloud service. (iii) Community cloud in which cloud services are shared by several organizations for supporting a specific community

that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). These services may be managed by the organizations or a third party and may exist offsite. A special case of community cloud is the Government or G-Cloud. This type of cloud computing is provided by one or more agencies (service provider role), for use by all, or most, government agencies (user role). (iv) Hybrid cloud which is a composition of different cloud computing infrastructure (public, private or community). An example for hybrid cloud is the data stored in private cloud of a travel agency that is manipulated by a program running in the public cloud.

Sung Ho Jang, et al. (2012) Discovered a task scheduling approach which is based upon the genetic algorithm (GA), which focuses on improving the profit for cloud provider. The set of task schedules are produces by genetic algorithm scheduling function which calculates the population with the help of fitness function that considers both user satisfaction and availability of virtual machine. To produce the best task schedule as a result the function iterations reproduce. The task scheduling using genetic algorithm shows the effectiveness and efficiency in results as compared to another task scheduling models like ABC based model of task scheduling and round robin model of task scheduling.

Xiaonian Wu, et al. (2013) described an algorithm that is based upon quality of service driven. The algorithm finds out the priority level of different type of tasks which is based upon different types of attributes of the tasks and also applies the sorting on tasks onto a service which can further finished the task. First of all the proposed algorithm find outs the priority of different tasks in accordance with special type of attributes of tasks in order to reflect the tasks precedence relation after that it sort out tasks according to their priority. Secondly, the algorithm calculates the completion time of each task on different type of services and schedules the each task onto that service which can easily complete the task as soon as possible according to the sorted task queue. The results are based upon Cloud Sim and shows that good performance and load balancing by QoS which is drived from both priority and completion time achieves by the algorithm.

Mrs.S.Selvarani, et al. (2008) has proposed a improved cost based algorithm. This algorithm employs the efficient allocation to the resources that are available in cloud. The main goal of this paper is to schedule the groups of tasks in cloud computing environment. Resources have different

costs of resource and performance of computation at this cloud environment. Resources optimization is based on ratio of computation per communication. An algorithm depends on both types of costs is produced. The proposed scheduling algorithm describes both types costs and performance of computation which improves the ratio of computation per communication. The results which are produced with the help of improved cost algorithm are better as compared to ABC algorithm. Finally the conclusion of this whole approach is that the scheduling algorithm which is described is better as compared Activity based costing algorithm.

Jung, Lim, et al. (2010) proposed the scheduling scheme for workflow that reduces waiting time of tasks when an instance occurs the out-of-bid situation. It executes the jobs of user' within the selected instances and expands the suggested user budget. The results produced are compared with various types of instances. The method used in this paper improves performance. And, the cost is higher when instance with low performance on other hand it is lower when an instance with high performance.

Naghibzadeh, et al. (2017) proposed workflow which is based upon quality of service scheduling it helps to decrease the cost of workload. According to their expected Quality of Service the users request for the available services. These services are charged on the basis of pay-per-use basis. The most difficult problem is scheduling of workflow in cloud computing, which means satisfying the users QoS and lower the cost of executing workflow. This paper discovered a new workflow scheduling algorithm based upon quality of service. This algorithm tries to decrease the cost of execution of workflow when meeting deadlines. The results describe very promising performance of this algorithm.

Ram Kumar Sharma, et al. (2016) Proposed a new algorithm which is based upon 3-tier cloud architecture includes consumer, Service Provider and the Resource Provider. This algorithm is beneficial from both the user and the provider of the service because it has an effective schedule reallocation which shows better resource utilization. The main goal of this paper is to show maximum utilization on both client and server side which is accessed in the cloud environment. This algorithm also improves the rate of efficiency. The requirements of user and provider of service fully satisfied in this algorithm because it follows the efficient scheduling approach and also based upon the priority reassignment.

Shaminder kaur, et al. (2016) have proposed a new algorithm which decreases the time of execution and cost of processing. This algorithm is based upon modification of Genetic Algorithm in task scheduling called modified or improved GA. An improved genetic algorithm is produced by combination of two existing algorithms. The results produced in this paper describe that algorithm shows very good performance in the presence of heavy workload. The initial population with Shortest Cloudlet to Fastest Processor and longest Cloudlet to Fastest Processor is also modifies in this algorithm.

T. Sandholm, et al. (2016) proposed parallel task scheduler for Hadoop. This type of task scheduler is based on dynamic priority which allows the user to adjust their spending time to control the allocation capacity. With the help of this technique the task scheduler improves the efficiency and makes more efficient decisions. It prioritizes the jobs. It helps the users to scale back their jobs in case of high demand because the cost of running is very high. This scheduler works with in budget.

J. Dean, et al. (2016) proposed that map reduce is also a programming model. The computation of user is based upon both map and reduces function. The computation of machines having large cluster automatically parallelized the runtime system. User also handles failure of machine so that network and disks are used efficiently. The whole system is easy to use and easily find by programmers. From last four years Google implements more than ten thousand different map and reduce programs.

K. Kc, K. Anyanwu (2016) described that deadlines which satisfied the requirements of user are very important. Existing cloud computing techniques such as Hadoop not considered such type deadlines. Jobs are scheduled with the help of FIFO algorithm in current implementation. In this paper some type of technique is developed to schedule the jobs which are based on user specified deadline. Other similar works has been presented in.

CONCLUSION

The environment of cloud computing has two types of players such as provider of the cloud and another one is the user of cloud. Both type of player works on different goals. The providers of the cloud want to increase the utilization of resources on the other hands user of the cloud want to decrease the cost when meeting their requirements. But there is lack of sharing of information between cloud provider and cloud user so it is difficult to allocate the resources. This paper

provided the work done and challenging problems existing in the cloud computing.

References:
1. Arora, Swinky, et al. "Novel stress calculation in parallel processor systems using buddy approach with enhanced short term CPU scheduling." Communication and Computing Systems. CRC Press, 2016. 663-672.
2. Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., and Warfield, A. (2003). Xen and the Art of Virtualization. Technical Report, University of Cambridge. Available online at: www.cl.cam.ac.uk/research/srg/netos/papers/2003-xensosp.pdf (Accessed on: January 23, 2013).
3. Bhattacherjee, B., Abe, N., Goldman, K., Zadrozny, B., Chillakuru, V. R., Del Caprio, M., and Apte, C. (2006). Using Secure Coprocessors for Privacy Preserving Collaborative Data Mining and Analysis. In 41 Proceedings of the 2nd International Workshop on Data Management on New Hardware (DaMoN'06), Chicago, Illinois, USA, June, 2006 Article No 1, New York: ACM Press.
4. Bhullar, Rohit K., et al. "Intelligent stress calculation and scheduling in segmented processor systems using buddy approach." Journal of Intelligent & Fuzzy Systems 32.4 (2017): 3129-3142.
5. Bhullar, Rohit K., Lokesh Pawar, and Vijay Kumar. "A novel prime numbers based hashing technique for minimizing collisions." Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on. IEEE, 2016.
6. Boneh, D., and Waters, B. (2007). Conjunctive, Subset, and Range Queries on Encrypted Data. In Proceedings of the 4th Conference on Theory of Cryptography (TCC'07), pp. 53-534.
7. Brandic, I., Music, D., Leitner, P., Dustdar, S. (2009). VieSLAF Framework: Enabling Adaptive and Versatile SLA-Management. In Proceedings of the 6th International Workshop on Grid Economics and Business Models (GECON'09), pp. 60-73, August 25-28, 2009, Delft, The Netherlands.
8. Cavoukian, A. (2008). Privacy in the Clouds: A White Paper on Privacy and Digital Identity: Implications for the Internet. Available online at: http://www.ipc.on.ca/images/resources/privacyintheclouds.pdf.
9. Chappel, D. (2008). Introducing the Azure Services Platform. Available online at: http://download.microsoft.com.(Accessed on: January 23, 2013).
10. Chong, F., Carraro, G., and Wolter, R. (2006). Multi-Tenant Data Architecture. Available online at: http://msdn.microsoft.com/en-us/library/aa479086.aspx. (Accessed on: January 23, 2013). Cloud Computing Security: Making Virtual Machines Cloud Ready. Available online at: http://www.techrepublic.com/whitepapers/cloud-computing-security-making-virtual-machines-cloudready/1728295. (Accessed on: January 23, 2013).
11. Creeger, M. (2009). Cloud Computing: An Overview. ACM Queue- Distributed Computing, Vol 7, Issue 5, p. 2, June 2009. New York: ACM Press.
12. DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., and Vogels, W. (2007). Dynamo: Amazon's Highly Available Key-Value Store. In Proceedings of the 21st ACM SIGOPS Symposium on Operating Systems Principles (SOSP'07), pp. 205- 220, Stevenson, WA, USA, October 2007.
13. Desisto, R. P., Plummer, D. C., and Smith, D. M. (2008). Tutorial for Understanding the Relationship between Cloud Computing and SaaS. Stamford, CT: Gartner, April 2008.
14. Emig, C., Brandt, F., Kreuzer, S., and Abeck, S. (2007). Identity as a Service- Towards a ServiceOriented Identity Management Architecture. In Proceedings of the 13th Open European Summer School and IFIP TC6.6 Conference on Dependable and Adaptable Network and Services (EUNICE'07), pp. 1-8, July 2007, Twente, The Netherlands.
15. Everett, C. (2009). Cloud Computing- A Question of Trust. Computer Fraud & Security, Vol 2009, Issue 6, pp. 5-7 June 2010.
16. Gellman, R. (2009). Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing. World Privacy Forum (WPF) REPORT, February 23, 2009. Available online at: http://www.worldprivacyforum.org/cloudprivacy.html (Accessed on: January 23, 2013).
17. Golden, B. (2009). Capex vs. Opex: Most People Miss the Point about Cloud Economics. URL:

http://www.cio.com/article/484429/Capex_vs._Opex_Most_People_Miss_the_point_About_Cloud_Econ omic.

18. Heritage, T. (2009). Hosted Informatics: Bringing Cloud Computing Down to Earth with Bottom-Line Benefits for Pharma. Next Generation Pharmaceutical, Issue 17, October 2009.

19. J. Dean, S. Ghemawat. "MapReduce: simplified data processing on large clusters" Magazine Communication of ACM, Volume 51, Issue 1, pp- 107-113, January 2008. Kamal Kc, Kemafor Anyanwu. "Scheduling Hadoop Jobs to meet deadlines", IEEE Second International Conference, pp- 388-398, 2010.

20. L. Pawar, R. Kumar, S. Arora, A. K. Manocha, "Optimized Route Selection On The Basis Of Discontinuity And Energy Consumption In Delay Tolerant Networks", Springer: Advances in Intelligent Systems and Computing, ISBN: 978-981-10-3769-6, 2016.

21. Rohit Kumar "Pragmatic Implementation of Power Optimization in Wireless Sensor Networks" MATRIX Academic International Online Journal of Engineering and Technology 1.2 (2016): 1-6.

22. 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, December, 2012.

23. Verma, Shiv Kumar, Manpreet Kaur, and Rohit Kumar. "Hybrid Image Fusion Algorithm Using Laplacian Pyramid and PCA Method." Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. ACM, 2016.

24. Xiaonian Wu, Mengqing Deng , Runlian Zhang , Bing Zeng, Shengyuan Zhou "A task Scheduling algorithm based on QOS- driven" International conference on Information Technology and quantitative Management(ITQM) pp. 1162- 1169 , ELSEVIER(2013).

9/21/2024