Cost Effective Resource Scheduling for Infrastructure as a Service (IaaS) in Cloud Computing

G Dinesh, C. K Aravind Raj, S. Muthu Saravanan, M Tamil Thendral

Abstract: ***Resource scheduling assigns the precise and accurate task to CPU, network, and storage. The aim behind this is the extreme usage of resources. However, well organized scheduling is needed for both cloud providers and cloud users. We propose a cost effective Resource scheduling algorithm that minimizes the utilization cost of the resources while meeting the deadline in cloud computing environment. we are calculating processing cost based on MIPS (Machine Instruction Per Second), For Memory (RAM) and Bandwidth based on MBPS(Megabytes Per Second) and For storage is based on MBs consumed to store the data. Our proposal considers all the essential characteristics of the cloud as well as VM performance variation and acquisition delay.***

Index Terms: Resource management, IaaS cloud, Resource scheduling, Resource discovery, Resource allocation.

I. INTRODUCTION

Cloud computing is a paradigm that changes the behavior in the consumption and delivery of information technology (IT) services. Users use web service interfaces to demand resources and pay only for the resources that they actually consume [1,2]. In addition to providing on-demand resources, clouds can deploy a custom- made environment for a given application. With the dynamical provision of resources. Cloud computing is also that changes the behavior in the consumption and delivery of information technology (IT) services[3]. Workflow model can describe any application that has jobs and flow of data among jobs. The workflow scheduling problem is NP-complete and it is a problem of assigning jobs to processors in multiprocessor environment[4,5,6] workflows have been frequently used to model large scale scientific problems in areas such as bioinformatics, astronomy, and physics. However, with the emergence of new paradigms such as cloud computing, novel approaches that address the particular challenges and opportunities of these technologies need to be developed [7]. So, that novel approach that uses the opportunities and challenge of cloud and gives a cost effective schedule. Further, the cloud provides computing as a utility service, and its services are categorized as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [9], [10]. Whereas, IaaS cloud provides the hardware resources in the form of a virtual resource such as computation, memory, storage, and networking. PaaS cloud offers an environment for users to develop and deploy their applications, and SaaS, cloud provides web applications/software over the internet, running on cloud infrastructure. PaaS and SaaS are not feasible for large scale scientific workflow applications. Because they provide only an environment to design, develop and test some web based applications [11]. This proposal focuses on IaaS clouds that provide several cost and performance effective benefits, as compared to cluster and grid computing. First, it provides on demand resource provisioning and these resources are controlled by service consumers. Second, a cloud service provider allows the end user to procedure and release computing resource as per their demands [8]. With the increasing complexity of scientific computing systems, such workflow applications have become big data applications, which demand large-scale infrastructures in order to be executed within a reasonable amount of time [12].The size of the workflow may be small or vast as per the type of scientific applications. To run these workflows of varying size, scientists needed a high performance computing environment such as cluster computing, grid computing, and the latest one is cloud computing .Many existing projects are designed to execute large scientific workflow applications. Therefore, the scientific applications can grow or shrink their resource pool as per the need of their applications. The cloud allocates only required computing resources from the resource pool so that overall resource utilization can be increased while reducing the total execution cost [9]. Third, cloud service provider's charges to the service consumers using pay-per-uses price model, in which they have to pay only for the computing resources they have used. Most of the service providers charge to the user for the whole time interval if they used only a fractions of last time interval For example, suppose the user uses computing resources 61 minutes and the time interval is 60 minutes, then the user has to pay for two-time intervals. In this research, we use a similar pricing model that provided by Amazon [13], Google [14] and Cloud Sigma [15] [8].

Although, Cloud computing have various benefits, but still it have some issues that need to be addressed. First, the performance of VMs can vary due to virtualization of hardware resources, the multi-tenancy of cloud infrastructure and the overall CPU performance of the VM can vary up to 24 percentages in Amazon public cloud. However,

Revised Version Manuscript Received on April 03, 2017.

G. Dinesh, Student, Department of Computer Science & Engineering, Kingston Engineering College, Vellore (Tamil Nadu)-632059, India.

C. K. Aravind Raj, Student, Department of Computer Science & Engineering, Kingston Engineering College, Vellore (Tamil Nadu)-632059, India.

S. Muthu saravanan, Student, Department of Computer Science & Engineering, Kingston Engineering College, Vellore (Tamil Nadu)-632059, India.

M. Tamil Thendral, Assistant Professor, Department of Computer Science & Engineering, Kingston Engineering College, Vellore (Tamil Nadu)-632059, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

[16] Reported that a typical cloud environment the performance variation can be up to 30(%) in execution time and up to 65(%) in data transfer time. Further, due to the variation in performance of VMs the overall deadline can be missed or budget can be increased of a scientific workflow. So, the performance variations have a significant impact on the scheduling of workflows on cloud computing environment. Our propose algorithms consider the virtual machines performance variation to meet the user defined deadline constraint. Second, when a VM is leased, it takes the time to proper initialization (acquisition delay) and similarly, whenever computing resources will release, they will take the time to shut-down (termination delay). So, the longer time in Resource acquire will increase the total execution time and longer time in the shutdown will increase the overall cost of the workflow. Thus, these delays will affect the overall performance and cost of the scientific workflows. In our proposed work, we assume an average acquisition delay for each type of VM, which can recognize the overall delay in execution of workflows [8].

II. RELATED WORK

The cloud providers and users requirements in cloud computing is surveyed and explained in detail by (Venter's and Whitley, 2012). There view of the technology has been done by interviews with the cloud providers and users having distinct requirements. It also states about the features of the cloud computing and the users demand. It provides two main dimensions, the technological and the service dimension. The main challenges in research of cloud computing are the efficiency, creativity and simplicity, with considering the knowledge and the trust among the cloud providers and users. Tsai and Rodrigues (2014) review the literatures about meta- heuristic scheduling techniques for cloud computing and present the applications using metaheuristics, some main issues and challenges. Moreover, provides the instructions for researchers to move on meta-heuristics techniques instead of traditional sche- duling technique in cloud computing. Further, Kalra and Singh (2015) deliver an inclusive survey and comparative analysis of various meta-heuristic scheduling techniques for cloud and grid environments including the Ant Colony Optimization (ACO), BAT algorithm, Genetic Algorithm(GA), League Championship Algorithm (LCA) and Particle Swarm Optimization (PSO). Optimization metrics, observations, open issues and challenges for further re- searches in cloud computing are also presented. However, Madni etal.(2016) present an appraisal of various types of resource allocation meta-heuristic algorithms that have been utilized in IaaS cloud computing environment. Furthermore, elaborates the various issues addressed through the resource allocation metaheuristics algorithms, the comparative parameters and also the experimental tools used for validation of the various techniques. There view and classification can serve as a foundation for further researches in resource allocation in IaaS cloud computing. Similarly, Zhang etal.(2016) delivers a comprehensive summary of the modern resource provision algorithms, intending on the state-of-art, existing techniques, and addressed objectives in different deployment stages. Some newly issued features of cloud computing and insufficiency of the existing techniques are highlighted to inspire future research guidelines.

Huang and Ou (2014) explain the PSO is a kind of replicated evolutionary computation algorithm of the congregate for aging behavior in cloud environment by using the PSO for task scheduling and it can get favorable results. Moreover, the ACO has excellent and well manufactured computing process along with the advantages of extension as it consists of a significant part of cloud environment task completion and scheduling issues. Furthermore, the study analyzes the standard and appliance with layered structural design of CloudSim. The architecture includes the resource layer, code layer, service layer, VM service layer, network layer, and structure interface layer. Depending on the work, considering the task scheduling as a research object in the cloud environment. CloudSim is anticipated and stretched by the five task scheduling algorithms. Afterthat, a specific cloudsim utlation prospect is prepared, and simulations are conducted five times for each algorithm. Simulation and the enhancement of task scheduling approach for cloud computing in CloudSim come to be a valuable way for the researchers (He etal.,2013). Resource scheduling indicates the process of organizing there sources among the different cloud users according to certain rules and regulation of resources usage under a specified cloud environment. Resource scheduling in resource management is the basic technology of cloud computing. It reviews the algorithms to increase the performance, involving with dynamic scheduling depend up on the threshold, optimized genetic algorithm with double suitability and enhanced ant colony algorithm for scheduling proposed by Huangetal.(2013). While the following areas are involved in Map reduce scheduling research like, graph methods, utility based optimization, dynamic priority, customization, temporary weight modifications, prediction, adaptive scheduling, equality be tween several users, study of map reduce interdependence and improving the reducing phases. The main task is to improve the overall performance, enhance throughput and response time, pro vides improved locality and fairness. The open area of research for new application, improvement of the make span and enhancing the fairness among multiple users (Elghoneimyetal, 2012).

III. PROBLEM DEFINITION

This section explains the resource scheduling problem in cloud computing environment and also presents a classification of the resource scheduling schemes.

A. Resource Scheduling Problem

In simple words, scheduling is a way of determining which activity should be performed. It is a demand of resources to the applicable consignment of tasks to the resources accessible for processing, network and storage, such that there is an extreme usage of resources. Well organized resource scheduling is compulsory for cloud providers besides cloud users, which is also the key technology of cloud computing (Acharetal.2012).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Scheduling algorithms are generally used to decrease the execution cost and time. Scheduling tackles the problem of which resources needed to be allocated to the received task or cloudlet. Efficient scheduling algorithms should consider the balancing the load of the system, the total execution time of the available resources, quick recovery, fault tolerance and migration with no interruption of service (Javanmardietal.,2014). Similar to the general scheduling problem, resource scheduling problem incloud computing is also to find the optimal planning, it can explain with the help of an expression.

$$\sum_{x=0}^{m,n}(Rx + Sx) * Tx \rightarrow \bigcup_{x} z$$

that assigns *m* required numbers of cloudlets/task (or virtual resources) $T=(T_1, T_2, T_3, \ldots\ldots T_m)$ on to *n* available physical resources in cloud data centers $R = (R_1, R_2, R_3, \ldots\ldots R_n)$ and $S= (S_1, S_2, S_3, \ldots\ldots S_n)$ to the cloud user $U=(U_1, U_2, U_3, \ldots\ldots U_n)$ such that the fitness of *z* particular objectives $F=(F_1, F_2, F_3, \ldots\ldots F_z)$ are maximized.

The problem of resource scheduling in cloud computing is illustrated in Fig.1. Furthermore, the scheduling of resource distribution is based on various aspects for example, how many quantity of resources is required in overall scenario, the number of cloud providers providing the services, how much availability of resources are available in their data center and many other conditions. The following objectives are often considered for the optimal resource scheduling in cloud computing those are cost, time, makespan, QoS, energy, load balancing, availability, reliability, failure rate, etc.

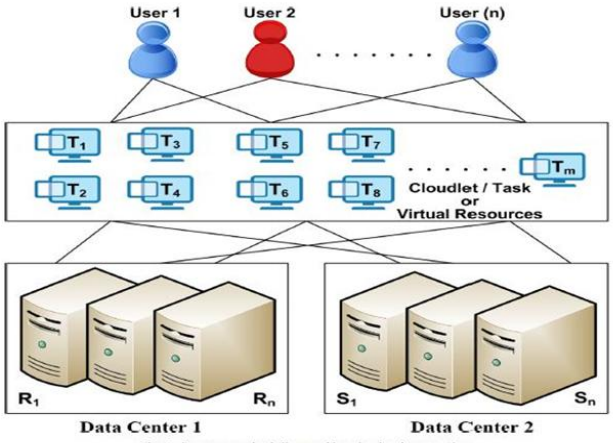

Fig. 1. Resources scheduling problem in cloud computing.

B. Classification of Resource Scheduling

The resource scheduling schemes are classify into six hybrid categories including cost aware resource scheduling, efficiency aware resource scheduling, energy aware resource scheduling, load balancing aware resource scheduling, QoS aware resource scheduling and utilization aware resource scheduling as shown in Fig. 2. The purpose of these classifications is to build the basis for future researchers in cloud computing environment. This classification is based on the parameter used in the evaluation of the performance in the various studies for the resource scheduling. The ovals shapes in Fig. 2.

The first classification focuses on cost aware resource scheduling, which includes cloud providers revenue and profit, users expenditure, cost of resources and total cost (we will discuss briefly on upcoming session). While the second category focuses on efficiency aware resource scheduling, that enhances the performance by including priority, reducing the execution time, execution cost and make span, also, increasing the bandwidth and speed. The third type presents the energy aware resource scheduling that elaborates on minimizing the power and energy consumptions in the datacenters. Fourth group represents the load balancing aware resource scheduling by efficiently managing the workload of multiple users among different data centers. Fifth class presents QoS aware resource scheduling that deals with the deals with reliability, availability, SLA, fault tolerance, throughput and recovery time. Finally, the sixth category deals with utilization aware resource scheduling. It focuses on maximizing the usage in an efficient way.

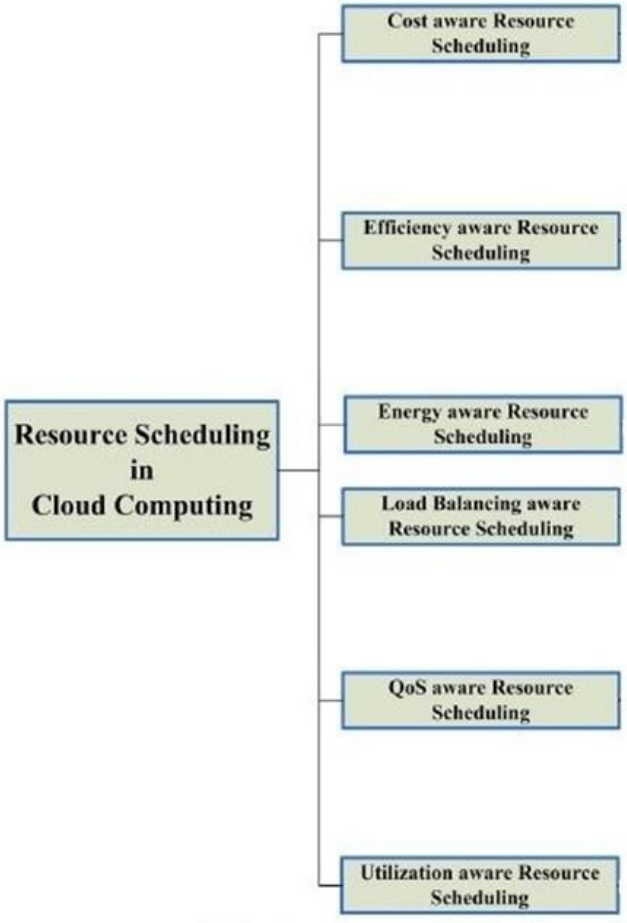

Fig. 2. Classification of resource scheduling in cloud computing.

IV. ANALYSIS

This section reviews the resource scheduling schemes and algorithms which have been used in the existing research work. However, analyzing existing techniques and understanding their focus is necessary for developing some additional applicable techniques and schemes. It can be an enhancement to the existing techniques or take the advantages from the previous studies.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Cost aware resource scheduling plays an important role in cloud computing, because as the definition of cloud declares that it delivers the services in cheapest amounts (Qian etal.,2009). A cloud provider is responsible for delivering the users demands as a service, which results in provider revenue, profit and user expenditures. The growth of revenue and profit with maximum utilization of resources is a desire of every cloud provider. On the otherside, the cloud users desire is to get services with high performance and minimum expenses(Zhang etal.,2010). An efficient resource scheduling improves the overall system perfor- mance and helps cloud provider to deliver resources as the desire or requirement of users, with maximum utilization and economically (Li andGuo,2010). The main objective is to enhance the internal resource utilization and reduce the cost of outsource tasks for the cloud users. The Cuckoo Search driven Particle Swarm Optimization (CS_PSO) approach is formulated as an integer programming model to resolve this problematic issue. It execute the local search extra profession- ally and avoids the local optima problem of PSO. The simulation out comes show that the proposed approach obtained high average profit as compared to the Standard PSO and Self Adaptive Learning PSO for the problem of non-trivial size (Rajuetal.,2016). Bansal etal.(2015) examine that the QoS scheduling algorithm works efficiently with make span, latency, load balancing and cost factor. This algorithm efficient with first three parameters except the cost is introduced by Wuetal.(2013). This paper adds the additional parameter of cost factor to complete this algorithm with the optimal results. In the paper, a new modified cost effective algorithm is proposed which minimizes the overall resource cost along with work load balancing. Simulations results demonstrate that the proposed algorithm performs better than the greedy algorithm, to reduce the overall resource cost. Furthermore, the technique deals with the energy efficiency, and also contribute to the green computing (Kapur, 2015).

V. CERS ALGORITHM

Our propose algorithms consider the virtual machines performance variation to meet the user defined deadline constraint. Second, when a VM is leased, it takes the time to proper initialization (acquisition delay) and similarly, whenever computing resources will release, they will take the time to shut-down (termination delay). So, the longer time in resource acquire will increase the total execution time and longer time in the shutdown will increase the overall cost of the workflow. Thus, these delays will affect the overall performance and cost of the scientific workflows. In our proposed work, we assume an average acquisition delay for each type of VM, which can recognize the overall delay in execution of workflows.

A. Cost Effective Resource Scheduling Algorithm

1. BEGIN
2. Initialize data center.
3. Verify the available resources in Datacenter
4. Get User Virtual Machine request and VM Configuration.
   i. The Processing Element based on Machine Instructions per Second
   ii. Ram, Storage, Bandwidth based on Megabytes per Second
5. Initialize Requested Virtual Machine and Start Virtual Machine.
6. Start the entities and submit the Cloud resources List to user.
7. Virtual Machine is allocated to the Host's and Calculate Latest start time Latest Start Time = VM Start time – Acquisition delay
8. Cloudlets (or) Tasks sent to the Virtual Machines
9. After completing the user tasks Virtual Machine will be terminated by user (or) if any risk Occurs then Go to step 10.
10. Save the User data and Terminate Virtual Machine. Latest Finish Time = VM Termination Time – Latest Start Time
11. Shutdown Host and Datacenter.
12. Calculate Total Execution cost and Total Execution time.
13. END In our CERS (cost effective resource scheduling) algorithm, we calculate cost based on users usage. if users terminated the VM or risk occurs in the sense users suppose to pay only how much resources they utilized.

VI. CONCLUSION AND FUTURE WORK

The resource scheduling techniques should be as simple as possible that require less execution time, makespan and computation power so that they consume less energy and produce less heat to work as part of green computing. Furthermore, resource scheduling techniques should provide guarantee of reliability, availability and fault tolerance. There should be a procedure to keep track of ongoing operation of resource scheduling. In future memory, ram and bandwidth cost can be calculate based on kilobytes per second.

REFERENCES

1. Amazon Elastic Compute Cloud (Amazon EC2), http://aws.amazon.com/ec2, 2008.
2. Amazon Simple Storage Service (Amazon S3), http://aws.amazon.com/s3, 2008.
3. Chia-Ming Wu∗, Ruay-Shiung Chang, Hsin-Yu Chan,"A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters" in Future Generation Computer Systems 37 (2014) 141–147
4. Yu J, Kirley M , Buyya R. Multi-Objective Planning for Workflow Execution on Grids, 8th IEEE/ ACM International Conference on Grid Computing, IEEE Computer Society, USA, 2007, p.10−17.
5. Chen WN, Zhang J. An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem with Various QoS Requirements, IEEE Transaction System Man Cybern. 2009; 39(1):29−43
6. Zhang F, Cao J, Wu HK. Ordinal Optimized Scheduling of Scientific Workflows in Elastic Compute Clouds, 3rd IEEE International Conference on Cloud Computing Technology and Science, China, IEEE, 2011, p. 1−9.
7. Maria Alejandra Rodriguez and Rajkumar Buyya, Deadline Based

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds in IEEE Transactions On Cloud computing, vol. 2, no. 2, april-june 2014
8. Jasraj Meena, Malay Kumar, And Manu Vardhan, Cost Effective Genetic Algorithm for Workflow Scheduling in Cloud Under Deadline Constraint published in ieee access on cloud computing,vol.,4 ,September 28, 2016.
9. P. Mell, ``The NIST definition of cloud Computing," Commun. ACM, vol. 3, no. 6, p. 50, 2010.
10. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, ``Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," Future Generat. Comput. Syst., vol. 25,no. 6, pp. 599_616, 2009
11. J. Meena, M. Kumar, and M. Vardhan, ``Efficient utilization of commodity computers in academic institutes: A cloud Computing approach," Int. J. Comput., Elect., Autom., Control Inf. Eng., vol. 9, no. 2, pp. 498_503,2015.
12. Kashlev, S.Y. Lu, A system architecture for running big data workflows in the cloud, in: 2014 IEEE International Conference on Services Computing, SCC 2014, pp. 51–58.
13. Amazon EC2 Pricing, accessed on Feb. 19, 2016. [Online]. Available: https://aws.amazon.com/ec2/pricing/
14. Google Cloud Platform, accessed on Feb. 24, 2016. [Online]. Available: https://cloud.google.com/compute
15. CloudSigma, accessed on Feb. 25, 2016. [Online]. Available: https://www.cloudsigma.com/pricing
16. Cloudsim simulation tool. Available: http://opensourceforu.com/2014/03/cloudsim-framework-modelling-simulating-cloud-environment
17. Resources chedulingforinfras tructureasa service (IaaS) incloud computing: Challenge sand opportunities on 22 April 2016. Syed Hamid Hussain Madni n, Muhammad Shafie AbdL atiff, Yahaya Coulibaly, Shafi'i Muhammad Abdulhamid.

(Authors)

G. Dinesh, pursuing B.E. Computer Science and Engineering in Kingston Engineering College, Vellore.

C.K. Aravind Raj, pursuing B.E. Computer Science and Engineering in Kingston Engineering College, Vellore.

S. Muthu Saravanan, pursuing B.E. Computer Science and Engineering in Kingston Engineering College, Vellore.

M. Tamil Thendral, B.Tech, M.E., Assistant Professor in Kingston Engineering College, Vellore.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.