A Survey And Taxonomy on Task Scheduling In Cloud Computing Platforms

Ms. Bhawana Lakhani[1], Prof. Amit Agrawal[2]
[1]Dept of Computer Science
[2]Professor, Dept of Computer Science
[1, 2] Medicaps University, Indore (M.P.)

Abstract- *Cloud computing can be thought of as a virtualization based approach for rendering various types of services to remote computers on demand. The services include software, hardware or even rendering platform for clients. Task Scheduling in Cloud based platforms is a critical task that is pivotal for load balancing for cloud computing applications. The task scheduling approaches are framed with an aim to optimize the latency and throughput of the system. The paper starts with the basics of cloud computing, the need for task scheduling and the salient parameters which are needed to evaluate the performance of the system. A survey of previous existing systems focusing on task scheduling for optimal load balancing have been analyzed with their salient characteristics.*

Keywords- Cloud Computing, Load Balancing, Task Scheduling, Latency, Throughput.

I. INTRODUCTION

The heart of cloud computing lies in the fact that services can be accessed by clients from diverse and distributed servers. The remote machines or clients can access the services through the concept of virtualization in which the remote servers can be accessed by the machines using a virtual platform called the virtual machine (VM). [1] The virtual machine mimics the environment of the actual remote server on the local machine of the client so as to render the requisite services. The major categories of services rendered by the cloud based platform are:[2],

IaaS: It stands for infrastructure as a service. This is the kind of cloud service that enables the IT firms to rent the cloud resources from the cloud provider. The provider gives services of only the basic types.

The virtual machine could be employed for the proper delivery of the IaaS. The resources pertaining to a proper infrastructure is of immense importance.[3] This cloud service helps with the correct utilization of the cloud tools and infrastructure to various business firms.

PaaS: It stands for Platform as a Service. It is kind of a developer platform for developers. It is the form of cloud service that is provided as an entire platform for the development of any product or to support any service on cloud platform. PaaS not only has the provision of making a suitable environment for running and developing application but it also gives certain standard executable environment for any kind of development.

SaaS: It refers to software as a service. Any kind of application or software service is directly made available to the intended customers and users. It provides software solutions and services in a direct manner through the use of cloud storage. It is a very optimal method that ensures that even when there is any problem in traditional storage space or someone is in a moving environment.[4] Then SaaS is a great and add greater amount of flexibility.

Hybrid and Multi-Cloud Services: They are also now an important part of cloud landscape. Hybrid as the name suggests, is a mix of a multiple kind of cloud workloads. The infrastructure within the workloads may vary. Put together, they give rise to a complete cloud solution.[5],[6] On the other hand, a multi cloud service can refer to a hybrid service or also a single solution of cloud where the information is processed on multiple numbers of clouds.

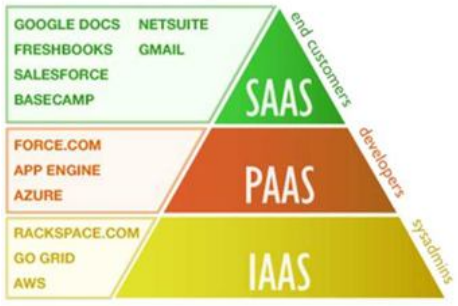

Fig.1 Types of Cloud Services with examples

The above figure depicts the cloud service services with certain examples.

II. SCHELULING AND LOAD BALANCING

The major challenge pertaining to the cloud resource management is the task scheduling among the cloud servers so as to optimize the latency and throughput. The empirical approaches along with their applications are given below:[7]

2.1 Ant Colony Optimization

This technique derives its name from practical ants which travel from their food sources to their colonization area. In this case, as the ant travels, its leaves behind pheromones which other ants can follow for later movements form food sources to the colony. The technique is typically useful for circumventing the limitations of centralized and distributed VM Migration using the agents for local monitoring.

2.2 Particle Swarm Optimization (PSO)

In this case, the model tries to emulate the behavior of particles in a group or swarm. Each particle has a characteristic position and location which keep updating with each passing iteration. The approach generally uses virtual machine based cloud load balancing.

2.3 Genetic Algorithm

In this approach, every chromosome is a representative of a plausible outcome which is again a function of the composite genes. The initial population is chosen arbitrarily. The applicability of the chromosome for modeling a suitable outcome is based on a mathematical parameter called the fitness function. On turning out to be successful in the fitness function test, subsequent mutation and cross over is designed to generate or produce new off springs of the chromosome.

2.4 Artificial Bee Colony Optimization.

The approach resembles the behavior of bees in a colony. The bee colony generally consist of three groups of bees i.e. Scouts, onlooker & employee. The employee is generally connective to a unity food source rather than an amalgamation of food sources. The bees belonging to the employee group generally visit the food source only to reach the destination to mark a symbolic finding. The onlooker is responsible for inferring the food source. The scouts go for search of new food sources.

2.5 Bat

The bat algorithm tries to emulate the behavior of the bat in searching for the gap between the bat and the prey. The bat has the capability of modifying the position, location and velocity as per the conditions of the prey

Table 2.1 Comparative Performance evaluation of different approaches [16-25]

Approach / Idea	Approach	Limitations
MOACO (Multi-Objective Optimization Based on Ant Colony Optimization)	Optimize Utilization of Resource utilization and reduce makes-pan	Weak Energy saving
ACO-VMM - VM migration Based Ant Colony Optimization	Optimize resource utilization and reduce migration rate and SLA	Weak Security aspect
ACO-LB (Load balancing optimization algorithm based on ant colony)	Optimize utilization of Resource Utilization and Load balancing	Weak Security
ACOPS(Ant colony optimization with particle swarm)	Optimize utilization of Resource Utilization with reduction in make-span and response time	No consideration of SLA
Modified Ant colony optimization	Optimize distribution of Load and Resources in Cloud and Grid	Poor Throughput
Load Balancing Based on BAT Algorithm	Reduction in response time	Poor Power Efficiency
VM Migration Based PSO and BAT	Optimize energy consumption	Poor security
PSO	Reduce Make-span	Do not consider other QoS parameter
PSO based Heuristic	Reduce task execution time	No focus on cost minimization
PSO with Cuckoo search	Reducing make-span	Poor time efficiency
PSO	Minimization of Task transfer time, execution time , energy consumption,	Heterogeneous s/m is not considered
Improved HBB-LB	Minimize make-span	.Priority parameter not consider while migration.
Efficient Bee Colony Algorithm	Reduce response time	Weak Security
Enhanced Bee Colony Algorithm	Reduce response time	Weak Security
Hybrid ACO & BEE (HACOBEE) Based Load Balancing	Reduce response time	Poor throughput
Genetic Algorithm (GA) based Load Balancing	Improve Quality of Service parameters (especially response time)	Weak Security
Job spanning time and load balancing genetic algorithm(JLGA)	Minimize Job Spanning	Low throughput
GA and Gravitational Emulation Based Hybrid Load Balancing	Reduce make-span	Weak Security
Multi-QoS human-centric Load Balancing (MQLB- RAM)	Minimize cost and time	Low throughput
Context and Load-Aware Family Genetic Algorithm	Minimize make-span	Does not consider other QoS parameters

III. EVALUATION PARAMETERS

It can be seen from table 2.1 that different approaches have their own pros and cons. However, to evaluate the performance of any approach, the following parameters are considered:

1) Response Time
2) Migration Rate
3) Energy Saving
4) Security
5) Resource Utilization
6) Throughput

Often a tradeoff between parameters is seen in terms of the evaluation parameters leading to heuristic and meta-heuristic approaches to be employed for optimized task scheduling and load balancing.

IV. CONCLUSION

It can concluded from previous discussions that that task scheduling is a critical parameter for the optimization of

resource utilization and load balancing. Often there is no clear choice for the approach to be used to attain the best values of the evaluation parameters. Hence a heuristic or meta-heuristic approach may be needed to solve the optimization problem in terms of the scheduling. The paper introduces the basics of the common approaches used along with their empirical derivatives. A comprehensive tabulation of the same helps in rendering insight into the approaches.

REFERENCES

[1] F Juarez, J Ejarque, RM Badia, "Dynamic energy-aware scheduling for parallel task-based application in cloud computing", 2018 ELSEVIER.

[2] Lan Wang and Erol Gelenbe, "Adaptive Dispatching of Tasks in the Cloud", 2018 IEEE.

[3] Martin Duggan, Karl Mason, Jim Duggan, Enda Howley, Enda Barrett, "Predicting Host CPU Utilization in Cloud Computing using Recurrent Neural Networks", 2017 IEEE.

[4] Ning Liu, Zhe Li, Jielong Xu, Zhiyuan Xu, Sheng Lin, Qinru Qiu, Jian Tang, Yanzhi Wang, "A Hierarchical Framework of Cloud Resource Allocation and Power Management Using Deep Reinforcement Learning", 2017 IEEE.

[5] Liyun Zuo, Shoubin Dong, Lei Shu, Senior Member, IEEE, Chunsheng Zhu, Student Member, IEEE, and Guangjie Han, Member, IEEE, "A Multiqueue Interlacing Peak Scheduling Method Based on Tasks' Classification in Cloud Computing", 2016 IEEE.

[6] Yazhou Hu, Bo Deng, Fuyang Peng and Dongxia Wang, "Workload Prediction for Cloud Computing Elasticity Mechanism", 2016 IEEE.

[7] Ji Xue, Feng Yan, Robert Birke, Lydia Y. Chen, Thomas Scherer, and Evgenia Smirni, "PRACTISE: Robust Prediction of Data Center Time Series", 2015 IEEE.

[8] Mehmet Demirci, "A Survey of Machine Learning Applications for Energy-Efficient Resource Management in Cloud Computing Environments", 2015 IEEE.

[9] Sherif Abdelwahab, Member, IEEE, Bechir Hamdaoui, Senior Member, IEEE, Mohsen Guizani, Fellow, IEEE, and Ammar Rayes, "Enabling Smart Cloud Services Through Remote Sensing: An Internet of Everything Enabler", 2014 IEEE.

[10] Chin-Feng Lai, Member, IEEE, Min Chen, Senior Member, IEEE, Jeng-Shyang Pan, Chan-Hyun Youn, Member, IEEE, and Han-Chieh Chao, Senior Member, IEEE, "A Collaborative Computing Framework of Cloud Network and WBSN Applied to Fall Detection and 3-D Motion Reconstruction", 2014 IEEE.

[11] Ian Davis, Hadi Hemmati, Ric Holt, Mike Godfrey, Douglas Neuse, Serge Mankovskii, "Storm Prediction in a Cloud", 2013 IEEE.

[12] Abul Bashar, "Autonomic Scaling of Cloud Computing Resources using BN-based Prediction Models", 2013 IEEE.

[13] Sadeka Islam , Jacky Keunga, Kevin Lee, Anna Liu, "Autonomic Scaling of Cloud Computing Resources using BN-based Prediction Models", 2012 ELSEVIER.

[14] Erol Gelenbe, Ricardo Lent and Markos Douratsos, "Choosing a Local or Remote Cloud", 2012 IEEE.

[15] Mohammad Moein Taheri and Kamran Zamanifar, "2-Phase Optimization Method for Energy Aware Scheduling of Virtual Machines in Cloud Data Centers", 2011 IEEE.

[16] Saurabh Kumar Garg, Srinivasa K. Gopalaiyengar, and Rajkumar Buyya, "SLA-Based Resource Provisioning for Heterogeneous Workloads in a Virtualized Cloud Datacenter", 2011 IEEE.

[17] Md. Toukir Imamt, Sheikh Faisal Miskhatt, Rashedur M Rahmant, M. Ashraful Amin, "Neural Network and Regression Based Processor Load Prediction for Efficient Scaling of Grid and Cloud Resources", 2011 IEEE.

[18] Christopher Stewart, Kai Shen, Arun Iyengar, Jian Yin, "EntomoModel: Understanding and Avoiding Performance Anomaly Manifestations", 2010 IEEE.

[19] Maria Couceiro, Paolo Romano, Lu´ıs Rodrigues, "A Machine Learning Approach to Performance Prediction of Total Order Broadcast Protocols", 2010 IEEE.

[20] Haibo Mi, Huaimin Wang, Gang Yin, Yangfan Zhou, Dianxi Shi, Lin Yuan, "Online Self-reconfiguration with Performance Guarantee for Energy-efficient Large-scale Cloud Computing Data Centers", 2010 IEEE.

[21] H Zhang, G Jiang, K Yoshihira, H Chen, "Intelligent workload factoring for a hybrid cloud computing model", IEEE 2009

[22] H Zhang, G Jiang, K Yoshihira, H Chen, "Resilient workload manager: taming bursty workload of scaling internet applications", IEEE 2009

[23] A Quiroz, H Kim, M Parashar, "Towards autonomic workload provisioning for enterprise grids and clouds", IEEE 2009

[24] A Stage, T Setzer, "Network-aware migration control and scheduling of differentiated virtual machine workloads", ACM 2009

[25] C Moretti, K Steinhaeuser, D Thain, "Scaling up classifiers to cloud computers", IEEE 2008