ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
ICRITCSA
M S Ramaiah Institute of Technology, Bangalore
Vol. 5, Special Issue 2, October 2016

Comparative Study of Load Balancing Algorithms for Best-effort Applications in Cloud

Chethan Venkatesh [1], Shiva Murthy .G[2]

Assistant Professor, Department of Computer Application, M.S. Ramaiah Institute of Technology, Bangalore, India [1]

Associate Professor, Department of MCA, VTU Center for PG studies, Bangalore, India [2]

Abstract: Resource provisioning is the main requirement in a cloud environment, where several resources like memory, CPU, storage, etc.… are allocated to the requesting consumer process in such a way that the resources are utilized in an efficient manner and distributed fairly among the requesting processes. Load balancing happens to be a vital task in resource provisioning to ensure fairness in resource allocation. Best-effort applications do not place any constraints on the amount or the quantity of resources allocated and the timing of scheduling. This paper presents a comparative study of several exiting approaches and certain modified versions of the existing approaches for load balancing algorithms for best-effort applications.

Keywords: Cloud Computing, Resource Provisioning, Load Balancing, Fault Tolerance

I. INTRODUCTION

Cloud is a combination of distributed and parallel system of interconnected network of nodes which supports execution of tasks remotely through virtualization. The advantage of cloud computing is its "pay as you use" model where the user can pay only for the resources used without upfront investment on hardware and software. The usage is facilitated through Service Level Agreement (SLA) which is an agreement between service providers and consumers.

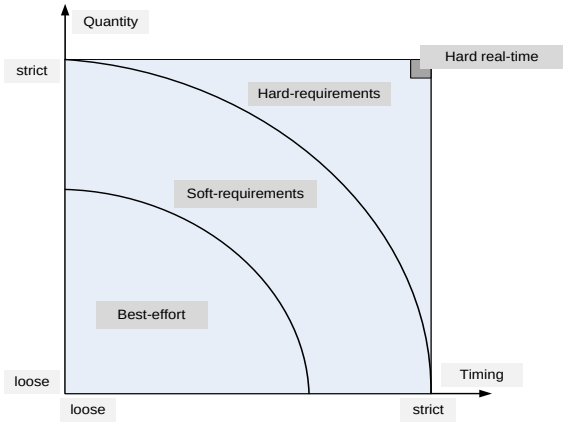

Fig1.1

Resources are the key ingredients of a cloud, resources can include CPU, Memory, Storage, Bandwidth, etc... These resources are virtualized and provided to the consumer on-demand in such a way that every request is satisfied. This is known as Resource Provisioning.

Resource Provisioning is a two stage process which including Scheduling and allocation. Resources can be provisioned statically or dynamically. We focus on dynamic resource provisioning as in cloud requests come at discrete time intervals and not at regular time intervals or known a priori. Hence resource scheduling and allocation becomes rather crucial which need to be managed efficiently in order to utilize resources optimally without resulting in neither over provisioning or under provisioning. The rest of the paper we will discuss various algorithms proposed for the scheduling resources.

Resource management has two dimensions [16] (i) Amount or quantity of resources allocated (ii) timing when access to resource is granted. Figure 1.1 shows different classes of resource allocation requirements in the space defined by two dimensions as best-effort, soft requirements, and hard requirements.

Best-effort: do not impose requirements regarding either the amount of resources allocated to an application, or the timing when an application is scheduled. (Includes batch applications and analytics)
Soft-requirements: require statistically guaranteed amounts and timing constraints.
Hard-requirements: demand strict timing and precise amounts of resources. [16]
There exists several algorithms namely First Come First Serve (FCFS), Shortest Job First (SJF), Shortest Remaining Time Next (SRTN), Round Robin (RR), for "Best-effort Applications", which do not impose requirements on amount of resource allocated and the timing of scheduling. This paper makes an attempt to compare the existing approaches. The primary objective of the algorithms discussed in this paper is to achieve load balancing while scheduling the jobs in cloud.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

II. LOAD BALANCING

Load balancing is one of the key aspects of provisioning the resources in cloud. Load balancing is required when there are multiple instances of resources across the cloud, resource requests must be distributed across all instances uniformly. [6]

There are several measurement parameters for load balancing [7]

- Throughput: It is the output a system can produce within a given time and a set of resources.
- Response time: Time when the request gets first response after being submitted for execution.
- Fault tolerance: It is the ability of the load balancing algorithm to ensure system continues to work in the event of failure of a node.
- Scalability: It is the ability to expand itself according to variations or increase in requests.
- Performance: It is the overall check of the algorithms working. It comprises the completion of the given task against present known standards like accuracy, cost and speed.
- Resource utilization: It is used to keep track of the utilization of various resources.
- Context Switch: number of times we need to remove the process and schedule other process for execution.

III. RELATED WORK

Several existing algorithms and its variants are proposed by researchers. The paper compares these approaches.
Aditya Marphatia et al.[1] have proposed optimization of FCFS which follows Greedy approach where in the resources are given in parts for the requests unlike the traditional FCFS algorithm, where the consumer process is made to wait if the complete resource is not available.

Dr. Thomas Yeboahiya et al. [2] have proposed integration of Round Robin with Shortest Job First algorithm. The proposed algorithm follows the traditional round Robin algorithm with a time slice for one iteration. After all the process are given a chance to execute, the process are assigned a priority based on the remaining execution time. Where lowest number corresponds to highest priority.
Mohammed Aboalama et al. [3] have proposed Enhanced Job Scheduling Algorithm using Shortest Remaining Job First approach. This approach uses traditional Preemptive Shortest Job First algorithm to address dynamic request handling. Rakesh Kumar Sanodiya et al [4] proposed Highest Response Ratio Next (HRRN) algorithm for load balancing. The proposed algorithm is classified into two cases cases (i) According to system load, (ii) according to system topology. The algorithm is similar to Shortest Job First, but eliminates the possibility of starvation by increasing the priority if the waiting time is longer.
Stuti Dave et al. [5] proposed fair Round Robin approach which implements dynamic time quantum.
Sanjaya Kumar Panda et al. [9] have proposed A Group based Time Quantum Round Robin Algorithm which uses Min-Max Spread Measure, the proposed approach assigns different time quantum for group of processes. The algorithm aims to reduce number of context switches, turnaround time and waiting time.

Komal Mahajan et al. [12] have proposed a Round Robin approach with server affinity for VM load balancer where in it saves the state of previously allocated VM unlike traditional Round Robin approach. This is achieved through two data structures namely hash map (These stores the status for the last VM allocated to a request from a given Userbase.) and VM state list (this stores the allocation status (i.e., Busy/Available) for each VM. When a request is received from the Userbase).
Gaurav Raj et al. [15] have proposed a modified load balancing algorithm which is a combination of Round Robin and Batch Mode Heuristic Priority algorithms, for better performance.
Supreeth S et al. [17] have proposed weighted round robin algorithm, it allocates all incoming requests to the available virtual machines in round robin fashion based on the weights without taking into account the current load on each virtual machine..

IV. COMPARATIVE REVIEW OF THE RELATED WORK

The proposed algorithm [1] takes into consideration deadline constraint and cost constraint. The algorithm results in minimum turnaround time and increase resource utilization. However in a cloud environment requests are dynamic in nature, which is not handled by the algorithm, the algorithms does not discuss load balancing to optimize resource allocation.

The proposed algorithm [2] results in higher efficiency, reduces context switching, average waiting and turnaround times. It also eliminates starvation. The proposed algorithm does not take into account dynamic request handling. Scheduling in case of multiple VM's is not been discussed. As in case of multiple VM's, there is a need to have a load balancer to increase the efficiency of the system.

The proposed approach [3] does not handle multiple VM's scheduling and balancing the load in case of multiple VM's.
The proposed approach [4] does not consider handling dynamic request and also scheduling and balancing of load in case of multiple VM's.
The proposed method [5] handle requests which arrive a priori and do not address scheduling and load balancing for multiple instances of resources.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

The proposed approach [9] does not address the load balancing issue adequately, it does not consider dynamic requests and multiple instances of resources.

The proposed algorithm [12] does not handle multiple instance of VM's and runtime requests, since the algorithms maintains the information of userbase with a hash map, and if the particular VM is available, executing the Round Robin algorithm can be avoided to save time.

The proposed algorithm [15] does not address the parameters such as dynamic request, load balancing in case of multiple instances of resources adequately.

The proposed approach [17] can increase the waiting time for certain requests if the allocated VM state is busy, the algorithms discussed above are mainly intended for neither best-effort applications, which do not impose neither the amount nor quantity of resources allocated nor the timing of scheduling. Further the above discussed algorithms do not address situations where a fault occurs in the system (failure of VM's or nodes).

V. CONCLUSION

Scheduling and allocation are vital in the context of resource management in cloud. The various algorithms discusses in the paper provides a reasonable solution for best-effort applications, where the amount of resources allocated or the timing of scheduling is not a constraint.

However load balancing and fault tolerance is of primary importance in order to effectively utilize the available resources and ensure system is fault tolerant.

For the comparative study we have found that the algorithms proposed do not handle load balancing adequately, when the requests are dynamic and also do not address multiple instances of resources effectively in terms of load balancing and fault tolerance in the larger context.

VI. FUTURE WORK

There is a need to develop new techniques which address all the parameters of load balancing discussed in section 2 of this paper. Even though best-effort applications do not have strict timing constraint as far as scheduling is concerned or the quantity of resource allocated, it will be beneficial if better algorithm is proposed which can handle multiple instances of resources and load balancing in the event of a system failure.

ACKNOWLEDGMENT

I thank **Prof. Dr. G Shivamurthy** for his support and help in preparing this paper. I also thank my HOD and colleagues for their constant encouragement and support.

REFERENCES

[1] Aditya Marphatia, Aditi Muhnot, Tanveer Sachdeva, Esha Shukla, Prof. Lakshmi Kurup, " Optimization of FCFS Based Resource Provisioning Algorithm for Cloud Computing", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 10, Issue 5 (Mar. - Apr. 2013), PP 01-05.

[2] Dr. Thomas Yeboah, Prof. Odabi I. Odabi, Kamal Kant Hiran, "An Integration of Round Robin with Shortest Job First Algorithm for Cloud Computing Environment", International Conference On Management, Communication and Technology (ICMCT), April 2015, Vol III Issue – 1 ISSN: 2026 – 6839.

[3] Mohammed Aboalama and Adil Yousif, " Enhanced Job Scheduling Algorithm for Cloud Computing Using Shortest Remaining Job First (SRJF)", IJCSMS (International Journal of Computer Science & Management Studies) Vol. 15, Issue 06 Publishing Month: June 2015 An Indexed and Referred Journal with ISSN (Online): 2231–5268.

[4] Rakesh Kumar Sanodiya, Dr. Sanjeev Sharma, Dr. Varsha Sharma, "A Highest Response Ratio Next (HRRN) Algorithm Based Load Balancing Policy For Cloud Computing", International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 1, Jan-Feb 2015.

[5] Stuti Dave,Prashant Maheta, "Utilizing Round Robin Concept for Load Balancing Algorithm at Virtual Machine Level in Cloud Environment", International Journal of Computer Applications (0975 – 8887) Volume 94 – No 4, May 2014.

[6] Chethan Venkatesh, Shiva Murthy G "A Survey on Resource Provisioning Schemes in Cloud", International Journal of Engineering Research & Technology (IJERT), NCSE'14 Conference Proceedings.

[7] Abhinav Hans, Sheetal Kalra, "A Comprehensive Study of Various Load Balancing Techniques used in Cloud Based Biomedical Services", International Journal of Grid Distribution Computing Vol.8, No.2 (2015), pp.127-132.

[8] Rakesh Kumar Mishra, Sandeep Kumar, Sreenu Naik B "Priority Based Round-Robin Service Broker Algorithm for Cloud-Analyst", 2014 IEEE.

[9] Sanjaya Kumar Panda, Debasis Dash, Jitendra Kumar Rout" A Group based Time Quantum Round Robin Algorithm using Min-Max Spread Measure" International Journal of Computer Applications (0975 – 8887) Volume 64– No.10, February 2013.

[10] H.S.Behera, R.Mohanty, Debashree Nayak, "A New Proposed Dynamic Quantum with Re-Adjusted Round Robin Scheduling Algorithm and Its Performance Analysis" International Journal of Computer Applications (0975 – 8887), Volume 5– No.5, August 2010.

[11] Pooja Samal, Pranati Mishra," Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing" , International Journal of Computer Science and Information Technologies, Vol. 4 (3) , 2013, 416-419.

[12] Komal Mahajan, Ansuyia Makroo, Deepak Dahiya, "Round Robin with Server Affinity: A VM Load Balancing Algorithm for Cloud Based Infrastructure", http://dx.doi.org/10.3745/JIPS.2013.9.3.379, J Inf Process Syst, Vol.9, No.3, September 2013.

[13] Dr. Hemant S. Mahalle, Prof. Parag R. Kaveri and Dr.Vinay Chavan, 2013 Load Balancing in Cloud Data Centers.

[14] Prabhjot Kaur and Dr. Pankaj Deep Kaur, "Efficient and Enhanced Load Balancing Algorithms in Cloud Computing", International Journal of Grid Distribution Computing Vol.8, No.2 (2015), pp.9-14.

[15] Gaurav Raj, Navreet Singh, Dr. Dheerendra Singh, "Resource Provisioning Using Batch Mode Heuristic Priority with Round Robin Scheduling", International Journal of Engineering and Technology (IJET), ISSN: 0975-4024 Vol 5 No 3 Jun-Jul 2013.

[16] Dan C.Marinescu, "Cloud Computing – Theory and Practice", Morgan Kaufmann, Elsevier Inc, 2013.

[17] Supreeth S, Shobha Biradar, "Scheduling Virtual Machines for Load balancing in Cloud Computing Platform", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, Volume 2 Issue 6, June 2013.