A New Scheme For Maximum Utilization Of Resources In Load Balancing Algorithm's in Cloud Computing Architecture

K Ravindranath, Jigisha D Gadhia, K. Samaikya Reddy

Abstract--- Cloud is one of the most widely used storage facility now in present days. With the advent of cloud services, more number of users are accessing the cloud and at the same time cloud should be ready to assign the tasks to appropriate machines. In order to have a smooth client server interactions and to handle all tasks simultaneously, a better way of load balancing algorithm is to be designed. So we have designed a new algorithm which uses the resources efficiently by dividing the tasks among them (makes use of array for the efficiency of resources) in a better way. This algorithm not only balances the tasks but also makes them execute in a better way by assigning to the resources based up on the execution time of the task.

1. INTRODUCTION:

Cloud Computing now a days have become most commonly used facility for storage in less cost. People doesn't require any hard-disk or some huge memory card to store their files/media and carry them everywhere they go in order to access their data. Cloud have made that easy by storing that information in data centers and making the clients use them to retrieve required data from anywhere around the globe. The storage space required for you depends on Pay-As-You-Use (PAYU) service, this is one of the advantage of cloud. It mainly provides good services like Platform as a service (PAAS), Infrastructure as a service (IAAS) and Software as a service (SAAS). Cloud do have many other advantages like it is Scalable, easy to manage, accuracy, high speed access over the data, cost efficient etc.

In the client server interaction, we can say the services are good only if the requested jobs/tasks are done in a faster way. Due to the increase in the population more number of clients are accessing the cloud. In order to maintain the same efficiency (with less clients) in handling efficient Load-Balancing Algorithms are to be implemented. In simple we can describe Load Balancing as, distributing number of tasks among the multiple servers to execute client requested jobs simultaneously. Most frequently used Algorithms Round Robin Algorithm (RRA), Waited Round Robin Algorithm (WRRA) etc.

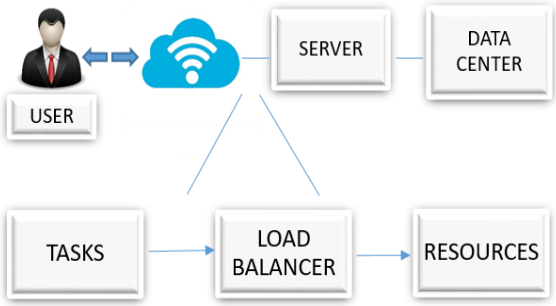

Fig 1: LOAD BALANCING TECHNIQUE

Here we introduce a new way in balancing the tasks requested by the clients. let there be clients/users (CLT's), a pack of servers/resources (RES's) and a data center with user's data. The load balancer takes in the task and based on the task execution time it assigns them by more efficient server/resource with the fast execution time task, moderated efficient server with the slowest execution time task and less efficient server with the moderate execution time task. In order to check the efficiencies of the task every time the Load Balancer makes it easy by storing all the efficiencies in an array and every time as a new task comes in it assigns by looking at the effeminacy resource array. This algorithm could be more efficient because all the small tasks will be executed fast in more efficient resource rather than assigning it in FIFO or random like that of in Round Robin.

This paper contains 1. Introduction, 2. Related Work, 3. Proposed Scheme, 4. Result and Analysis, 5. Conclusion and future work, 6. Acknowledgement, 7. References.

2. Related Work:

Min-min and max-min algorithms had proven they are best in their own way. min-min algorithm works well when there are more minimum completion tasks and max-min performs well when there are more completion time tasks. We cannot perfectly give an outstanding performance of anyone of the tasks; they work well according to the completion time of the tasks. (1) Any client can gain a major advantage with the use of cloud computing elasticity. But it has its own issues in enveloping the resources for the required capacity with no down time. Some of the issues that can be faced by elasticity would be major trafficking and automatic scaling of resources. (2) Another major role

Revised Manuscript Received on March 08, 2019.

Dr K Ravindranath, Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, 522501, Andhra Pradesh, India.

Jigisha D Gadhia, Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, 522501, Andhra Pradesh, India. (jigishagadhia007@gmail.com)

K. Samaikya Reddy, Department of Computer Science Engineering. Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, 522501, Andhra Pradesh, India. (saisamaikya10996@gmail.com)

in cloud computing technology is load balancing, it plays a major role, also a major challenge for cloud developers to manage the responses, allocation of the tasks to machines. A load balancer can maximum utilizes a resource and also minimize the execution time of the task.

Maintaining the system stability is the major problem for load balancer. (3)RAIL is a methodology implemented to dynamically assign weights to different resources according to their usage on a machine, this will help reduce the execution time also avoids future imbalance of allocating tasks to machines, but the only disadvantage in this methodology will be overhead and less efficiency in maintaining a load balancer, this would be less likely can maintain and persists a load balancer.(4)A dynamic load balancing algorithm was introduced to dynamically allot the tasks to resources, but this would not be a best idea, because small tasks may be allotted to high performance machines, which would reduce or waste the capacity of the machine, so while allocating it must also consider the capacities of both tasks and machines(5).

Power usage is another biggest factor to reconsider for the completion times of each tasks, the more the tasks, the more the working time of each machine, when a high performance machine is given no.of small completion time tasks then it would complete quick and outperforms more no.of small tasks. So it is recommended to assign small tasks to high completion machine to save time and power.(6) Allocating the resources with changeable resources and capacities at any time of the execution is a flexible item for a load balancer, assigning a task at any point of time i.e reallocation of resources would be a risky process in terms of data consistency and durability.(7) Usage of resources in cloud and mode of payment is a related concern for a client, for that cloud has majorly provoked the flexibility of pay-as-you-go mode. Accessing the resources at any point of time at anywhere as paying according to our usage would blow out the customers into cloud including no maintenance cost or owing of any hardware. (8) Dynamic evaluation indicators and self-adaptive methods, for resource requests from client, resource computing capacity and resource strength to handle the tasks are important to resolve the state of the tasks (9). New methodology to evaluate the load balancing and workload methods for QoS. In any load balancer the tasks should be assigned to machines in such a way that would effectively utilize the machine.

Through this literature survey, we have come to a conclusion that if a task's capacity is known, based upon this we can allocate it to a resource. Initially all the resources capacity will be known, so that a suitable task can be assigned to a suitable machine.

3. PROPOSED SCHEME

Our new scheme consists an array of values as execution time for efficient resources. Therefore, when a task is given input the load balancer checks the minimum execution time required for the task and then sends it to the appropriate resource for execution w.r.t the resources execution array values.

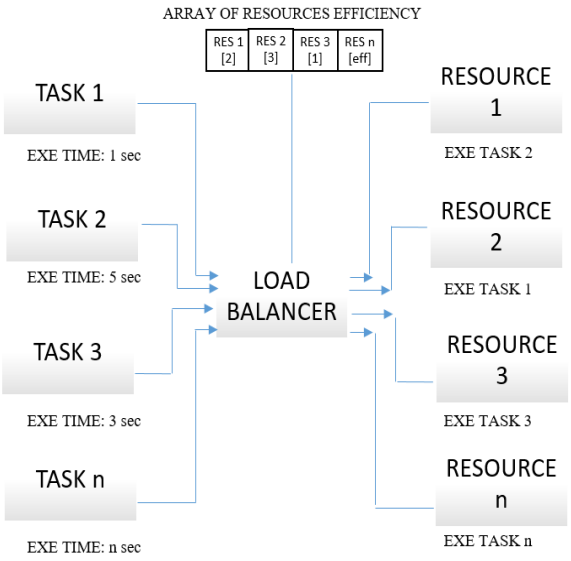

Fig 2: PROPOSED SCHEME REPRESENTATION

The assigning takes place as said in the below algorithm:

1. Upon receiving the client request we will configure the algorithm based on the First Come First Serve (FIFO).

2. We consider a set of weighted resources in which "weighted" determines the capacity of a resource i.e. the number of tasks it can handle at a time.

3. Here we store the capacities of the resources in an array.

4. Now we check the efficiency of the clients requested tasks and assign it to the resources as follows:

> If task execution time is low
If resource with high efficiency is free:
Assign the task for execution.
Else:
Check for next efficient resource.

> Else If task execution time is moderate
If resource with low efficiency is free:
Assign the task for execution.
Else:
Check for next efficient resource.

> Else If task execution time is high
If resource with moderate efficiency is free:
Assign the task for execution.
Else:
Put in a waiting queue and Check for next fastly available efficient resource.

Proposed Algorithm

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

4. RESULTS AND ANALYSIS:

Result:

From the algorithm build, all the tasks will be assigned to resources based upon their pre-requisite. When a heavy task is queued, it will have assigned to a machine which can handle a single heavy task. Smaller tasks will be assigned to machines which can handle number of resources together, that means the resources will be distinguished initially and the tasks will be allotted based upon its requirement.

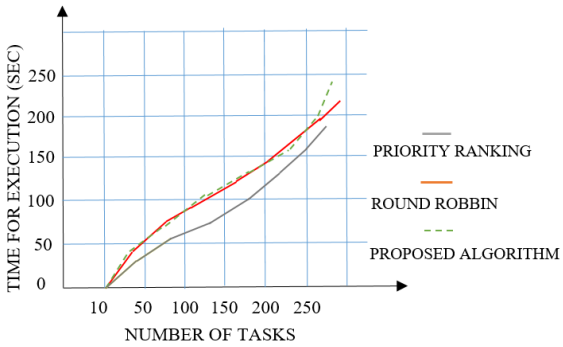

Fig 3: PROPOSED vs OTHER ALGORITHMS EXECUTION TIME GRAPH

Analysis:

This algorithm will differentiate different resources based upon its capacity which will further help in assigning tasks to resources with no resources kept unused, during runtime of the tasks. When this part is successfully done, there will be quick and adequate results, this algorithm takes the strategy of round robin and implements the allocation of tasks to resources.

5. CONCLUTION AND FUTURE WORK

We can say that our proposed scheme works efficiently when compared to that of other load balancing algorithms. If the more efficient resource is completing all the smaller tasks fast, then the newly added smaller tasks can start executing in a faster way so that there would be less waiting time for remaining tasks. In future we would like to enhance this algorithm with adding queues in the client side to suppress the loss of instructions, when there are set of tasks given by the users are more due to increase in clients/users.

6. ACKNOWLEGMENT

We Jigisha D Gadhia, K Samikya Would like to thank Dr K Ravindranath for his consistent support and valuable guidance. We also like to thank Koppada Gowtham for his support in writing the paper.

REFERENCES

1. An In-depth Analysis and Study of Load Balancing Techniques in the Cloud Computing Environment. 2015,P.P. GeethuGopinath,Shriram K.Vasudevan.
2. Elasticity in cloud computing: a survey. August 2015, Paulo Antonio Leal RegoDanielo Gonçalves Gomes
3. Load balancing in cloud computing: Challenges & issues. Ms. Shalini Joshi ,Dr. Uma Kumari.
4. RIAL: Resource Intensity Aware Load Balancing in Clouds. August 2017, Haiying Shen.
5. Load balancing in cloud computing using dynamic load management algorithm. 14 January 2016,Reena Panwar, Bhawna Mallick.
6. Energy-Aware Load Balancing and Application Scaling for the Cloud Ecosystem. Ashkan Paya and Dan C. Marinescu 2017.
7. MPSO: A Novel Meta-Heuristics for Load Balancing in Cloud Computing. 2017, Subhadarshini Mohanty, Prashanta Kumar Patra, Subasish Mohapatra, Mitrabinda Ray.
8. furnish users access to IT resources anytime and anywhere using pay-per-use fashion. Alireza SadeghiMilaniNima JafariNavimipour,2016.
9. Dynamically Weighted Load Evaluation Method Based on Self-adaptive Threshold in Cloud Computing. Liyun ZuoEmail authorLei ShuShoubin DongChunsheng ZhuZhangbing Zhou,2017.
10. Multi-Class load balancing scheme for QoS and energy conservation in cloud computing. 2017, Olasupo O. Ajayi, Florence A. Oladeji, Charles O. Uwadia.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication