An Efficient Scheduling Algorithm for Cloud Computing

Anupama Shukla[1] Bhavesh Patel[2]
[1,2]Department of Information Technology Engineering
[1,2]SVIT, Vasad, Anand, Gujarat, India

Abstract— Cloud computing has been promising as a supple and potent computational structural design to put forward omnipresent services to users. It accommodates inter coupled hardware and software assets in a cohesive way, which is dissimilar to conventional computational environments. A diversity of hardware and software resources are incorporated together as a resource pool, the software is no longer resided in a sole hardware environment, it is performed upon the schedule of the resource pool for optimized resource deployment. The optimization of energy utilization in the cloud computing environment is the issue how to use diverse energy conservation strategies to efficiently allocate resources. In this paper, we study the association among infrastructure components and power utilization of the cloud computing environment, and converse the matching of job types and component power tuning methods, and then we present a resource scheduling algorithm of Cloud Computing based on energy efficient optimization methods.

Key words: Cloud computing, Scheduling, Optimization, Allocation, Energy-efficient

I. INTRODUCTION

Cloud computing is the exercise of computing resources like hardware and software which are delivered as a service over a network (Internet). Load balancing and provisioning in cloud computing systems is truly a dispute at present. Constantly a distributed way out is necessary. Since it is not always practicable or cost efficient to retain one or more inactive services just as to fulfil the essential demands. Jobs cannot be assigned to apposite servers and clients independently for proficient load balancing as cloud is a incredibly multifaceted composition and components are there the whole time in wide spread region [1]. However Cloud Computing performances boast extra on the scheduling algorithm and appropriate load balancing algorithm. Scheduling algorithm will create such a series of procedure that the throughputs are amplified and load balancing algorithm segregate the load accurately among all existing resources. Cloud Computing is a parallel processing representation where these problem is of essential value so they also have substance. While cloud is a pay-go-model, the business performance wishes to be accelerated which is a demanding issue in the realm. So we have preferred this area to discover an enhanced approach to improve the performance in cloud computing atmosphere.

II. RELATED WORK

First of all Naidila et.al [7] have presented an in depth evaluation on the three computing models, cluster, grid and cloud computing. The issues associated to these computing models are highlighted. Cloud computing has few challenges like dynamic scalability, multi-tenancy, querying and access, standardization, reliability & Fault-tolerance, Debugging & profiling, Security, and Power [7].

Parameters	Clusters	Grids	Cloud
SLA	Limited	Yes	Yes
Allocation	Centralized	Decentralized	Both
Resource Handling	Centralized	Distributed	Both
Loose coupling	No	Both	Yes
Reliability	No	Half	Full
Virtualization	Half	Half	Yes
Interoperability	Yes	Yes	Half
Task Size	Single large	Single large	Small & medium

Table 1: Comparison of Cluster, Grid and Cloud Computing

In the experiments and results examination, they find that in a realistic Cloud Computing Environment, using one intact Cloud node to estimate a single task will dissipate a lot of energy, nevertheless when the structure of cloud framework logically support paralleled process. Moreover, they found that it is not the higher the frequency of CPU, the quicker the execution of program, there subsist a turning point which possibly will achieve stability between frequency and energy consumption. Unfortunately, based on their more than a few experiments, this turning point is floating depend on program job, hardware framework or networks. As a result, they have to deploy an automatic process to discover the apt CPU frequency, main memory's mode or disk's mode or speed

The first time in cloud computing, in which power management has been applied at the data center level, has been done by Pinheiro et al. [10]. In this job the authors have proposed a technique for minimization of power utilization in a diverse cluster of computing nodes serving multiple web-applications. The core technique applied to reduce power consumption is concentrating the workload to the minimum of physical nodes and switching idle nodes off. This approach requires dealing with the power/performance trade-off, as performance of applications can be corrupted due to the workload consolidation. Requirements to the throughput and implementation time of applications are clear in SLAs to guarantee reliable QoS. The proposed algorithm at regular intervals monitors the load of resources (CPU, disk storage) and makes decisions on switching nodes on/off to reduce the overall power utilization, while providing the expected performance. The definite load balancing is not handled by the system and has to be managed by the applications. The algorithm runs on a master node, which creates a Single Point of Failure (SPF) and may become a performance blockage in a large system. In addition, the authors have pointed out that the reconfiguration operations are prolonged, and the algorithm adds or removes only one node at a time, which may also be a reason for slow reaction in major environments. The proposed approach can be applied to multi-application mixed-workload environments with fixed SLAs. Chase et al. [11] have well thought-out the problem of energy-efficient management of uniform resources in Internet hosting centers. The main challenge is to find out the resource demand of each one application at its current request load level and to allocate resources in the most resourceful way. To deal with this problem the authors have applied a cost-effective framework: services ''bid'' for resources in terms of volume and quality. This enables negotiation of the SLAs according to the available funds and current QoS requirements, i.e. balancing the cost of resource usage (energy cost) and the profit gained due to the usage of this resource. The system maintains an active set of servers certain to serve requests for each service. The network switches are dynamically reconfigured to modify the active set of servers when crucial. Energy utilization is reduced by switching inactive servers to power saving modes. The system is targeted at the web workload, which leads to a ''noise'' in the load data. The authors have addressed this difficulty by applying the statistical ''flip-flop'' filter, which diminish the number of idle reallocations and leads to a more constant and efficient control. The proposed approach is appropriate for multi-application environments with variable SLAs and has created a base for numerous studies on power-efficient resource allocation at the data center level. However, in compare to [10], the system deals simply with the management of the CPU, but does not consider other system resources. The latency due to switching nodes on/off as well is not occupied into account. The authors have noted that the management algorithm is quick when the workload is constant, however turns out to be quite costly all through noteworthy changes in the workload. Furthermore, likewise [10], heterogeneous software configurations are not handled, which can be addressed by applying the virtualization technology. The procedure is to estimate the overall CPU frequency necessary to supply the needed response time, determine the optimal number. Raghavendra et al. [12] have investigated the dilemma of power management for a data center environment by combining and coordinating five heterogeneous power management policies. It is claimed that, likewise to [13], the approach is independent of the workload category. Similar to most of the prior works, the system deals simply with the CPU management. On the other hand, the system fails to sustain strict SLAs as well as inconsistent SLAs for diverse applications. This results in the aptness for enterprise environments, but not for Cloud computing providers, where more complete support for SLAs is vital. Cardosa et al. [14] have proposed an approach for the difficulty of power-efficient allocation of VMs in virtualized diverse computing environments. They have leveraged the min, max and shares parameters of VMM, which signify minimum, maximum and fraction of the CPU allocated to VMs sharing the identical resource.

III. PROPOSED WORK

Technologies in cloud computing have a tremendous rise. Due to dynamic changing environment of Cloud, it is desirable to design effective algorithms that can adapt its behaviour to the request in the system and prevents over-provisioning and perform efficiently.

In this proposal we propose an algorithm in which we will create VM with required best configuration to prevent over-provisioning so that resources can be optimized.

A. For implementation in AWS following ateps should be followed

1) Create an Amazon account to utilize AWS.
2) AWS SDK is integrated with ECLIPSE (LUNA).
3) Simulating existing energy efficient algorithms in Amazon EC2.
4) Developing API prerequisite code in EC2.
5) Customize the APIs.
6) The Proposed algorithm is implemented.
7) Live results and analysis of the cloud is done.
8) Compare the issues of existing algorithms and find out which parameters they don't cover.
9) Create a matrix of the entire algorithm existing and proposed as well.
10) Compare the results of simulation.

A. AWS SDK:

AWS stands for Amazon Web Services. The SDK helps seize the complication out of coding by providing JAVA APIs for many AWS services including Amazon S3, Amazon EC2 and more.

In our proposed System, we will have an algorithm which will prevent the concept of over-provisioning. In this algorithm by considering the parameters to create VM, so that it can prevent over-provisioning and the energy in the form of resources can be optimized.

B. Steps of Proposed Strategy:

Condition: Create a new VM

1) Step1: Check average throughput of VM at tx(Time).

– Calculate throughput by formula I = R*T
– Where
– I= Throughput
– R= Rate at which the Task will processed
– T = Processing Time

Parameters included in calculating throughput are:

– CPU utilization
– Network Latency
– Disk Usage

2) Step 2: If average throughput is greater than the required throughput then create a new VM.
3) Step 3: Again go for step 2.
4) Step 4: Create VM

We are considering these parameters to create VM

– Bandwidth requirement.
– Type of storage required.
– CPU utilisation.
– Type of request
– Storage space.
– Budget and deadline.

5) Step 5: Check step 2 and remove the extra VM if not required.

The consolidation is done in this system. First the load will be checked if the load is high on any VM it will be balanced hence load balancing concept is applied over here. While, when there won't be much load on VM the tasks will be consolidated i.e all the task will b combined at one VM and the others will b turned off hence it will save energy. There is assured vital feature like:

– Job Scheduling
– Load Balancing

– Resource Allocation

So we have preferred this area to discover an enhanced approach to improve the performance in cloud computing atmosphere.

IV. CONCLUSION

Cloud computing has been promising as a supple and potent computational structural design to put forward omnipresent services to users. It accommodates inter coupled hardware and software assets in a cohesive way, which is dissimilar to conventional computational environments. A diversity of hardware and software resources are incorporated together as a resource pool, the software is no longer resided in a sole hardware environment, it is performed upon the schedule of the resource pool for optimized resource deployment. The optimization of energy utilization in the cloud computing environment is the issue how to use diverse energy conservation strategies to efficiently allocate resources. In this report, we study the association among infrastructure components and power utilization of the cloud computing environment, and converse the matching of job types and component power tuning methods, and then we present a resource scheduling algorithm of Cloud Computing based on energy efficient optimization methods.

REFERENCES

[1] Chunye Gong, Jie Liu, Qiang Zhang, Haitao Chen and Zhenghu Gong :The Characteristics of Cloud Computing, 39th International Conference on Parallel Processing Workshops 2010.
[2] Yogita Chawla and Mansi Bhonsle: A Study on Scheduling Methods in Cloud Computing, IJETTCS Sept 2012
[3] Ashish Kumar Singh, Sandeep Sahu Kamalendra Kumar Gautam, Mangal Nath Tiwari; Private Cloud Scheduling with SJF, Bound Waiting, Priority and Load Balancing ,IJarcsse January 2014.
[4] V.Vinothina, Dr.R.Sridaran, Dr.Padmavathi Ganapathi: A Survey on Resource Allocation Strategies in Cloud Computing, IJACSA-2012.
[5] Chunye Gong, Jie Liu, Qiang Zhang, Haitao Chen and Zhenghu: The Characteristics of Cloud Computing. 2010.
[6] Naidila Sadashiv and S. M Dilip Kumar Cluster, Grid and Cloud Computing: A Detailed Comparison ICCSE 2011.
[7] Liang Luo, Wenjun Wu and Dichen Di,Fei Zhang,Yizhou Yan,Yaokuan Mao:A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods. IEEE 2012.
[8] Young Choon Lee and Albert Y. Zomaya: Energy efficient utilization of resources in cloud computing systems,Springer 2012.
[9] Mayank Mishra, Anwesha Das, Purushottam Kulkarni, and Anirudha Sahoo: Dynamic Resource Management Using Virtual Machine Migrations, IEEE 2012
[10] E. Pinheiro, R. Bianchini, E.V. Carrera, T. Heath, Load balancing and unbalancing for power and performancee in cluster-based systems, in: Proceedings of the Workshop on Compilers and Operating Systems for Low Power, 2001.
[11] J.S. Chase, D.C. Anderson, P.N. Thakar, A.M. Vahdat, R.P. Doyle, Managing energy and server resources in hosting centers, in: Proceedings of the 18th ACM Symposium on Operating Systems Principles, ACM, New York, NY, USA,2001.
[12] R. Nathuji, K. Schwan, Virtualpower: coordinated power management in virtualized enterprise systems, ACM SIGOPS Operating Systems Review 41 (6) (2007).
[13] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, No ‘‘power’’struggles: coordinated multi-level power management for the data center, SIGARCH Computer Architecture News 36 (1) (2008).
[14] M. Cardosa, M. Korupolu, A. Singh, Shares and utilities based power consolidation in virtualized server environments, in: Proceedings of the 11th IFIP/IEEE Integrated Network Management, IM 2009, Long Island, NY, USA, 2009.
[15] www.hyperstratus.com
[16] http://aws.amazon.com